



HAL
open science

Approche physique de la représentation des erreurs de modélisation en prévision d'ensemble atmosphérique

Axelle Fleury

► **To cite this version:**

Axelle Fleury. Approche physique de la représentation des erreurs de modélisation en prévision d'ensemble atmosphérique. Physique [physics]. Université Toulouse 3 - Paul Sabatier, 2022. Français. NNT: . tel-03990894v1

HAL Id: tel-03990894

<https://hal.science/tel-03990894v1>

Submitted on 15 Feb 2023 (v1), last revised 10 Jul 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
Axelle FLEURY**

Le 16 décembre 2022

**Approche physique de la représentation des erreurs de
modélisation en prévision d'ensemble atmosphérique**

Ecole doctorale : **SDU2E - Sciences de l'Univers, de l'Environnement et de
l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :
CNRM - Centre National de Recherches Météorologiques

Thèse dirigée par
François BOUTTIER et Thierry BERGOT

Jury

Martin CHARRON, Rapporteur
Martin LEUTBECHER, Rapporteur
Frédérique CHERUY, Examinatrice
Fleur COUVREUX, Examinatrice
Alan HALLY, Examinateur
Laure RAYNAUD, Examinatrice
François BOUTTIER, Directeur de thèse
Jean-Luc ATTIE, Président

Remerciements

Faire une thèse en sciences de l'atmosphère à Météo France est quelque chose qui n'était pas vraiment prévu avec mon parcours post-bac. Si j'ai pu commencer à travailler dans cet établissement et découvrir le monde de la météo, c'est grâce à François Bouttier et Olivier Caumont qui m'ont permis d'arriver au CNRM en 2019 pour travailler sur l'utilisation de la prévision d'ensemble dans le cadre du projet PICS sur les crues soudaines. Je voudrais donc commencer par leur adresser mes remerciements pour m'avoir donné l'occasion de bifurquer et de découvrir un nouveau domaine scientifique, des questions de recherche stimulantes et des gens passionnés.

Je remercie particulièrement François pour m'avoir proposé ce sujet de thèse, pour son encadrement bienveillant et son soutien depuis que je suis arrivée dans l'équipe. J'ai beaucoup apprécié la grande liberté qu'il m'a donnée pour approfondir les directions que je choisissais, ses encouragements à prendre des initiatives, à interagir avec d'autres collègues, et sa grande disponibilité et réactivité à chacune de mes sollicitations. Malgré ma tendance à travailler un peu (trop) dans mon coin avant de présenter mes résultats, François m'a toujours impressionnée par sa capacité à comprendre très rapidement les problèmes auxquels j'étais confrontée et à me donner des conseils pertinents.

Je remercie chaleureusement Thierry Bergot, mon co-directeur de thèse, avec qui ce fut également un plaisir de travailler. Sa grande connaissance de la physique du brouillard m'a beaucoup aidée à aborder les enjeux concernant ce phénomène et à réfléchir aux perturbations du schéma microphysique et des propriétés optiques des nuages.

Un grand merci à Fleur Couvreur dont l'aide a été précieuse pour tout le travail concernant les perturbations de la turbulence et de la convection peu profonde. Merci de m'avoir aidée dans l'exploitation des données LES, pour tous les conseils avisés et la relecture de l'article.

Un grand merci également à Rachel Honnert, d'abord pour ses cours sur la couche limite atmosphérique qui m'ont permis de me mettre "dans le bain", de commencer à appréhender un peu plus mon sujet d'étude et auxquels je me suis beaucoup référée pour tout le travail sur la turbulence. Ensuite pour sa participation à mon comité de suivi de thèse, sa disponibilité pour répondre à mes questions, ses encouragements et sa relecture de certaines parties de ce manuscrit. Je remercie aussi Alan Hally, également membre de mon comité de suivi et dont la bonne humeur est communicative.

Pour avoir accepté de relire et critiquer ma thèse, je remercie sincèrement Martin Leutbecher et Martin Charron. Merci à Martin Leutbecher pour son accueil durant le workshop à Reading en mai 2022, qui m'a aidée à être un peu moins intimidée, et pour avoir accepté de relire ce manuscrit en français. Merci à Martin Charron pour les questionnements intéressants de son rapport et pour avoir accepté de se connecter à 8h du matin à Montréal pour commencer la

journée par une soutenance de thèse.

Merci également à tous les membres du jury que je n'ai pas déjà cités, présents à Toulouse ou en visio, pour leurs retours bienveillants, et leurs questions intéressantes : Jean-Luc Attié, Frédérique Chéruy, Laure Raynaud.

Je remercie également beaucoup Laurent Descamps et Carole Labadie pour leur aide concernant les simulations 3D, leur intérêt pour mon travail et leur entrain lors de tous nos échanges. Merci à Meryl Wimmer pour m'avoir expliqué son travail de thèse dont les résultats ont alimenté mes réflexions. Merci à Sébastien Riette pour son aide précieuse concernant EPyGrAM et le code AROME, et merci à Laurent Labatut pour son aide informatique. Merci enfin au personnel administratif et particulièrement à PAR que j'ai trop souvent sollicité pour des oublis de badgeage...

Tout le travail présenté dans ce manuscrit n'aurait pu être réalisé sans l'apport scientifique, technique ou humain de toutes les personnes précédemment citées. Mais, sans qu'elles soient en lien direct avec ma thèse, les interactions avec de nombreuses autres personnes à Météo France m'ont permis de passer quatre années réjouissantes au CNRM, ce qui m'a forcément beaucoup aidée. Je suis ainsi très reconnaissante d'avoir pu échanger tant de moments sympathiques avec les gens du couloir PRECIP/PHY-NH au gré des discussions de pause café ou de midi à Eurest. J'ai énormément appris sur Météo France et la prévision météo en général lors de ces moments. Pour cela, un merci tout particulier à Marc, qui a été mon réflexe n°1 en cas de questions sur la météo et que je suis donc souvent allée déranger dans son bureau... Un grand merci à Hugo également pour m'avoir aidée à ne pas me perdre complètement dans les tréfonds de la nomenclature des fichiers GRIB et avoir été un compagnon "prévision d'ensemble" dans l'équipe PRECIP! Énorme merci à Adèle, Mareva et Pauline pour les moments passés ensemble, les discussions, les sorties. J'ai vraiment beaucoup apprécié le temps passé avec vous. Si un jour on crée notre centre de recherche à la campagne-élevage de chèvres-cinéma associatif, ce sera avec plaisir! Merci à Damien et Clément pour leur gentillesse quand je suis arrivée au CNRM pour la première fois. Et merci à tous les autres dont je ne vais pas citer les noms par peur d'en oublier.

Merci à Marie de m'avoir fait découvrir le club poterie et à Iris de m'avoir fait découvrir le tissu aérien, qui a été une petite révolution pour moi. Un grand merci à Yannick pour son accompagnement dans la découverte de cette activité et sa gentillesse, et merci à tous les membres du club, Laetitia, Sébastien, Alice, et celles et ceux que je croisais moins souvent, pour les encouragements mutuels.

Je finirai par dépasser le cadre du bureau pour remercier la "maison bleue" et ses habitant.e.s, qui ont concrétisé l'idée que je me faisais de la vie en commun. Enfin, ma reconnaissance la plus profonde va à ma famille et à Lucas, qui j'espère le savent déjà.

Résumé

La prévision météorologique repose sur des modèles numériques qui ne représentent qu'imparfaitement l'atmosphère et son évolution. Les erreurs commises changent d'un jour à l'autre, selon le lieu et les phénomènes en jeu. Elles sont dues aux diverses approximations utilisées dans les modèles et à la nature-même de leurs équations, qui les rendent très sensibles aux conditions initiales que l'on ne connaît jamais avec exactitude. Savoir quelle confiance accorder dans les résultats d'un modèle est un enjeu essentiel pour la prévision météorologique, en particulier dans le contexte de la prévision d'évènements extrêmes et plus largement pour tous les secteurs de la société fortement dépendants de la météo.

Estimer le degré de confiance, ou, de façon équivalente, le degré d'incertitude associé à une prévision est l'objectif de la « prévision d'ensemble », une approche initiée dans les années 1970 qui consiste à combiner plusieurs scénarios alternatifs et équiprobables d'un même évènement, de façon à produire une prévision non plus déterministe mais probabiliste. Les systèmes de prévision d'ensemble sont construits le plus souvent en introduisant des perturbations, d'une part dans les données d'entrée des modèles (conditions initiales et de bord) et d'autre part à l'intérieur des modèles eux-mêmes. Cette dernière catégorie regroupe un grand nombre de méthodes différentes, dont la majorité se concentre sur la perturbation des paramétrisations des phénomènes physiques de petite échelle non résolus (turbulence, convection, phénomènes microphysiques, etc), qui sont reconnues être une source d'erreur importante. Ces méthodes sont en grande partie empiriques : les perturbations sont réglées de façon à produire en moyenne des scénarios suffisamment différents les uns des autres, mais elles ne sont pas forcément toujours adaptées à la situation météorologique et peuvent ponctuellement produire des prévisions irréalistes. D'autres méthodes, cherchant à relier davantage les perturbations aux phénomènes physiques dont la paramétrisation est jugée incertaine, se sont néanmoins développées ces dernières années et font l'objet d'une recherche active.

Dans cette thèse, on cherche à savoir si ces dernières méthodes, intéressantes d'un point de vue théorique, ont un impact suffisant sur les résultats d'un modèle pour rendre possible leur utilisation opérationnelle dans un système de prévision d'ensemble. Pour cela, certaines d'entre elles sont testées dans le modèle à aire limitée AROME de Météo-France, d'abord dans un cadre simplifié uni-colonne pour des situations de couche limite idéalisées, puis dans une configuration plus opérationnelle.

Dans un premier temps, deux méthodes de perturbation visant à mieux représenter la variabilité des mouvements turbulents et convectifs sous-maille sont étudiées et réglées grâce à la comparaison avec des simulations à haute résolution (LES). Leurs performances sont différentes selon les variables et le cas d'étude, et la comparaison avec les LES montre que leurs objectifs sont partiellement atteints. On étudie par la suite une méthode de perturbation de deux paramètres de réglage de la distribution de taille des gouttelettes nuageuses, intervenant dans les schémas microphysique et radiatif. Cette méthode produit une dispersion importante mais également un biais par rapport à la référence, qui peut être en partie corrigé grâce à la résolution d'un problème inverse. Les premiers tests réalisés avec une version 3D du modèle AROME montrent que certaines perturbations testées dans le cadre uni-colonne ont un impact sur la simulation de brouillard et de précipitations convectives dans un contexte de faible forçage synoptique. En revanche, les premiers scores statistiques sur une période d'un mois ne montrent pas d'impact des méthodes de perturbation de la turbulence et de la convection peu profonde. Il sera nécessaire d'effectuer des simulations sur de plus longues périodes et à d'autres moments de l'année pour préciser ces résultats.

Abstract

Weather forecasting relies on numerical models that do not perfectly represent the real atmosphere and its evolution, because of various approximations and of the chaotic nature of their equations. Meteorological model predictions are thus subject to errors that are time-, location-, and situation-dependent. This raises the question of how much confidence one can have in a forecast, which can be critical when predicting extreme events, and more generally for all weather-sensitive human activities.

Predicting forecast uncertainty is the main goal of "ensemble prediction", an approach developed since the 1970's, in which different forecasts of the same event are combined to produce a probabilistic prediction instead of a deterministic one. Ensembles are usually produced by introducing perturbations into the forecasting system, with the aim of targeting known sources of uncertainty. Two major sources of uncertainty are nowadays acknowledged to influence models results : uncertainties in the initial and boundary conditions on one hand, and model uncertainty on the other hand. Regarding the representation of model uncertainty, a large number of different methods can be employed. Most of them focus on the representation of fine-scale processes (e.g. turbulence, convection, microphysical processes, etc) by the so-called "parameterization schemes" of the models, which are known to be a substantial source of model error. These methods often use perturbations that are empirically tuned to give good long-term statistical scores, but that sometimes lack physical consistency and may not be well adapted to all meteorological situations. Recently, there has been active research on alternative methods that try to link more precisely perturbations to physical processes.

This thesis focuses on this last category of perturbation methods. The objective is to evaluate their impact on a convection-permitting model – the AROME model of Météo-France –, with the long-term goal of seeing whether they could be used in the ensemble prediction system PEARO. This manuscript reports on the implementation of several perturbation methods, first in an idealized framework of single-column simulations, and then in a 3D configuration of the AROME model.

In a first part, process-oriented perturbations of turbulence and shallow convection are studied and evaluated against large-eddy simulations (LES) of several boundary-layer cases. Their performance is found to be case- and variable-dependent. In a second part, a parameter perturbation method is tested on two parameters governing the cloud droplet size distribution inside the microphysics and the radiation schemes. This method is shown to produce significant dispersion, but also bias. The bias, however, can be reduced thanks to an inverse problem solving. First tests indicate that the perturbations evaluated in the single-column framework also affect the AROME model outputs in its 3D configuration, for two real case studies of fog formation and weakly-forced convective precipitations. Statistical scores computed on a one-month period show however no impact of turbulence and shallow-convection perturbations. Further testing on longer periods should be conducted to clarify these results.

Table des matières

Introduction

1	Prévision d'ensemble : comment représenter l'incertitude d'une prévision météorologique ?	4
1.1	Qu'est-ce qu'une prévision numérique du temps?	4
1.2	La nécessité de représenter l'incertitude	5
1.2.1	Modèles atmosphériques et théorie du chaos	5
1.2.2	Émergence de la prévision d'ensemble	6
1.2.3	Différentes sources d'erreur	7
1.3	Perturber les données d'"entrée" du modèle	8
1.3.1	Perturbations de l'état initial	8
1.3.2	Perturbations des conditions latérales	9
1.3.3	Perturbations de la surface	10
1.4	Diverses approches pour la représentation des erreurs modèle	11
1.4.1	Les méthodes multimodèle et multiphysique	11
1.4.2	Les méthodes de perturbation stochastiques "globales"	13
1.4.3	Les méthodes de perturbation de paramètres	17
1.4.4	Les méthodes de perturbation des processus physiques	21
1.5	Représentation de l'incertitude dans les modèles de Météo France	25
1.5.1	Les ensembles d'assimilation AEARO et AEARP	26
1.5.2	Les systèmes de prévision d'ensemble PEARP, PEARO	26
2	Paramétrisations physiques et stochasticité	28
2.1	Paramétrisation de la turbulence locale	28
2.1.1	Généralités sur la turbulence	28
2.1.2	Approche EDMF pour la paramétrisation de la turbulence	30
2.1.3	Paramétrisation de la turbulence locale dans AROME	31
2.1.4	Les sources d'incertitude liées à la paramétrisation de la turbulence	34
2.2	Paramétrisation des phénomènes convectifs	37
2.2.1	Principes généraux des schémas de convection	37
2.2.2	Fluctuations autour de l'équilibre	39
2.2.3	Remise en question des hypothèses de séparation d'échelles dans les modèles kilométriques	40
2.2.4	La paramétrisation de la convection peu profonde dans AROME	44
2.3	Paramétrisation des nuages	46
2.3.1	Principes de l'ajustement à la saturation	47
2.3.2	Le schéma microphysique ICE3	49
2.3.3	Processus microphysiques et incertitude	52

2.4	Paramétrisation des effets radiatifs	54
2.4.1	Quelques éléments sur le schéma radiatif d'AROME	54
2.4.2	L'incertitude liée à la paramétrisation des effets radiatifs	56
3	Description du modèle et des simulations numériques	58
3.1	Description générale du modèle AROME	58
3.2	MUSC : une version 1D du modèle	59
3.3	Simulations 1D de cas d'étude idéalisés	60
3.3.1	ARMCu	61
3.3.2	BOMEX	62
3.3.3	FIRE	63
3.3.4	LANFEX	65
3.4	Le modèle Méso-NH et les simulations LES	67
3.5	Méthodologie pour la production d'ensembles de simulations 1D	68
4	Perturbation des processus physiques d'AROME à l'aide de simulations LES	69
4.1	Perturbation de la turbulence et de la convection peu profonde	69
4.1.1	Introduction et résumé de l'article	69
4.1.2	Article	72
4.2	Recherches sur la perturbation de l'ajustement à la saturation	106
4.2.1	Motivations	106
4.2.2	Distributions dans les LES	106
4.2.3	Perturbations de la relation de CB02	109
5	Combinaison avec d'autres méthodes de perturbation	112
5.1	Perturbation du schéma microphysique ICE3 à l'aide d'une méthode de perturbation de paramètres	112
5.1.1	Motivations	112
5.1.2	Analyse de sensibilité du cas LANFEX aux paramètres N_0 et ν	113
5.1.3	Premiers tests de perturbation des paramètres N_0 et ν	116
5.1.4	Réglage des distributions de N_0 et ν par la résolution d'un problème inverse	118
5.2	Couplage entre perturbation microphysique et perturbation du schéma radiatif .	128
5.2.1	Principe du couplage	128
5.2.2	Étude de sensibilité	130
5.2.3	Test sur le cas FIRE	134
5.3	Intercomparaison entre les différentes méthodes : SPPT, iSPPT, RPP, POP . .	135
5.3.1	Combinaison des méthodes proposées dans les chapitres 4 et 5	135
5.3.2	Comparaison avec iSPPT, RPP	138
6	Évaluation des méthodes de perturbation dans des simulations 3D du modèle AROME	141
6.1	Tests sur quelques journées en 2020 et 2021	141
6.1.1	Sensibilité aux perturbations des paramètres de distribution de taille des gouttelettes nuageuses N_0 et ν	141
6.1.2	Premiers tests des méthodes de perturbation de la turbulence et de la convection peu profonde	151
6.2	Évaluation de trois ensembles sur le mois de septembre 2021	157
	Conclusion	158

Annexes	166
A Calcul de la vitesse terminale de chute des gouttelettes nuageuses	167
B Calcul du rayon effectif à partir de la distribution de taille des gouttelettes nuageuses	170
C Comparaison de trois approches de représentation de l'erreur modèle dans des simulations uni-colonne de couche limite	172

Introduction

Malgré les modèles dont on dispose pour simuler l’atmosphère, il est difficile de prévoir de façon fiable son état futur au delà d’une certaine échéance de temps (quelques jours à quelques mois selon les phénomènes et le degré de précision). Ces limites de la prévisibilité météorologique sont reconnues notamment depuis les travaux d’Edward Lorenz et son constat, à partir de simulations numériques de systèmes météorologiques simplifiés, que certains systèmes dynamiques présentent une forte sensibilité aux conditions initiales, ce qui limite en pratique leur capacité à prédire le futur quand bien même il s’agit de systèmes déterministes (Lorenz 1963). Ces résultats sont connus à présent sous le terme de théorie du chaos et les systèmes d’équations servant à décrire l’atmosphère peuvent être qualifiés de systèmes chaotiques. Étant donné qu’il existe une incertitude inévitable dans l’initialisation du système de prévision – due à la précision finie des instruments de mesure et la rareté des données pour décrire l’atmosphère à un instant donné –, il est impossible de prévoir le temps dans un futur lointain car une petite erreur initiale mène à des prévisions complètement différentes après un certain temps d’intégration des équations.

L’évolution de la qualité d’une prévision est difficile à prédire dans un système chaotique où l’accroissement des erreurs n’est pas linéaire. Pourtant, pour pouvoir réaliser des prévisions météorologiques utiles, il est nécessaire de savoir quelle confiance accorder aux résultats d’un modèle. La figure 1 illustre ce problème dans un cas extrême, celui d’un orage ayant eu lieu en 2021 dans le sud de la France : les deux premières cartes de précipitations sont issues d’un même modèle météorologique lancé à 12 heures d’intervalle : la première simulation prévoit de façon irréaliste des précipitations faibles tandis que la deuxième est beaucoup plus proche des observations, avec des cumuls prévus dépassant localement les 300 mm.

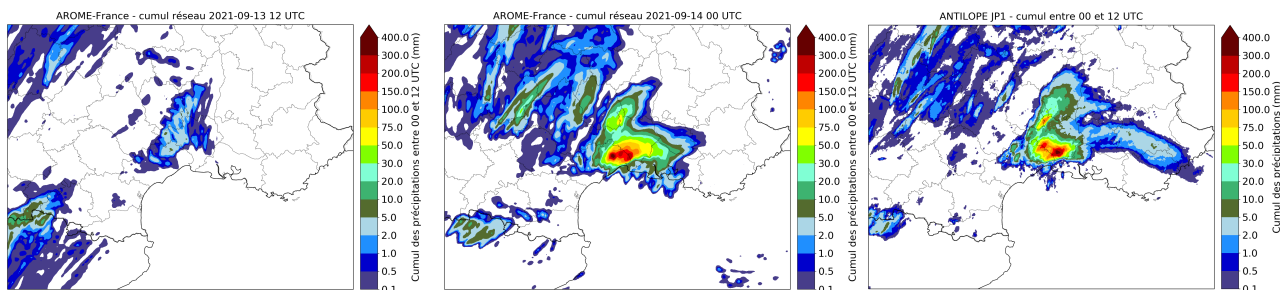


FIGURE 1 – Cumuls de précipitations simulés et observés entre 00 UTC et 12 UTC le 14/09/2021 dans le sud de la France. De gauche à droite : prévisions AROME-France initialisées avec les analyses du 13/09/2021 12 UTC, du 14/09/2021 00 UTC et observations des pluies (analyse de lame d’eau Antilope).

Pour estimer la fiabilité d'une prévision météorologique, une méthode, envisagée dès les années 1970 (Leith 1974), est de produire plusieurs scénarios d'un même évènement, de façon à échantillonner la distribution des futurs possibles de l'atmosphère. Cette technique est désignée en météorologie sous le terme de *prévision d'ensemble*. Elle propose d'envisager une prévision sous un angle non plus déterministe mais probabiliste.

La prévision d'ensemble a fait l'objet de nombreuses recherches depuis les années 1980. Les premiers systèmes de prévision d'ensemble opérationnels ont débuté en 1992 au Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT) et au National Centers for Environmental Prediction (NCEP) aux Etats-Unis (Palmer et al. 1993; Toth et Kalnay 1993), et existent de nos jours dans la majorité des services météorologiques nationaux. Les premières versions de ces systèmes étaient axées sur la représentation de l'incertitude liée à l'initialisation des modèles (Houtekamer 1995; Palmer et al. 1997; Toth et Kalnay 1997; Buizza et al. 2008; Wei et al. 2008) : les ensembles étaient construits en faisant démarrer les modèles avec plusieurs états initiaux perturbés. Néanmoins, il est apparu assez tôt que perturber seulement l'état initial ne suffit pas pour représenter de façon satisfaisante toutes les erreurs de prévision (Houtekamer et al. 1996). En effet, une partie des erreurs vient des modèles eux-mêmes, du fait qu'ils représentent de façon approximative la réalité. Dans le cas des modèles atmosphériques, les équations de la dynamique des fluides décrivant les mouvements des masses d'air sont résolues numériquement, de façon approchée, à l'aide d'une discrétisation spatiale et temporelle. De plus, de nombreux processus d'échelle plus fine que celle de la discrétisation choisie ne sont pas représentés par ces équations et doivent être modélisés par des schémas spécifiques, appelés schémas de paramétrisation, qui reposent eux aussi sur un ensemble d'hypothèses et d'approximations.

Les incertitudes provenant des modèles ont été prises en compte dès 1996 au Canada (Houtekamer et al. 1996) et un peu plus tard au CEPMMT (Buizza et al. 1999) et diverses stratégies ont été adoptées. Certaines reposent sur le fait de combiner des prévisions issues de modèles différents (multimodèle) ou bien d'alterner divers schémas de paramétrisation existants (multiphysique). D'autres consistent à utiliser un seul modèle dans lequel on injecte un bruit, censé représenter les erreurs commises dans diverses étapes du modèle (perturbations stochastiques). Cette dernière catégorie est celle qui nous intéresse dans cette thèse et comprend de nombreuses méthodes qui diffèrent par leur façon d'introduire le bruit dans le modèle : il peut s'agir par exemple de perturber les tendances des variables d'état, ou bien la valeur de certains paramètres ou encore directement les équations modélisant des processus physiques particuliers. Parmi toutes ces méthodes, certaines traitent l'incertitude de façon globale, en essayant de représenter toutes les erreurs modèle à la fois : c'est le cas par exemple de l'approche multimodèle et de l'approche de perturbation stochastique "Stochastically Perturbed Parametrization Tendencies" (SPPT) développée par Buizza et al. (1999). D'autres au contraire tentent de cibler davantage les sources d'incertitude dans le modèle, comme l'approche multiphysique ou les méthodes de perturbation des paramètres. La plupart ont par ailleurs une approche pragmatique de la représentation des erreurs modèle, par exemple en combinant les prévisions de modèles déjà existants ou bien en introduisant un bruit réglé de façon à maximiser la performance des ensembles vis-à-vis d'observations. Certaines approches sont néanmoins fondées sur une justification physique plus poussée : c'est le cas des méthodes de perturbation des processus physiques, que l'on désigne aussi parfois sous le nom de schémas de paramétrisation stochastiques, qui construisent une représentation stochastique de certains processus sous-maille à partir d'arguments théoriques sur ces processus et la façon dont on peut les modéliser.

Malgré les différentes méthodes de perturbation testées depuis plusieurs années, les systèmes de prévision d'ensemble restent souvent sous-dispersifs : la diversité des scénarios qu'ils proposent ne suffit pas à contenir la réalité. Il existe donc des sources d'incertitude qui ne sont toujours pas représentées de façon satisfaisante dans les ensembles. Concernant la perturbation de l'initialisation des modèles, les méthodes employées par les différents centres météorologiques sont aujourd'hui relativement similaires. En revanche, il existe une grande variété de techniques pour représenter les erreurs modèle et leurs performances peuvent varier en fonction du modèle ou même de la zone géographique. Les études récentes semblent tendre à privilégier les méthodes de perturbation ciblant plus précisément les sources d'incertitude, de façon à assurer une meilleure cohérence physique des ensembles, mais également une meilleure compréhension de leur comportement. Néanmoins, leur efficacité pratique n'est pas toujours satisfaisante, de sorte qu'il n'existe pas de consensus clair sur la meilleure façon de tenir compte des erreurs modèle dans les ensembles.

Ces travaux de thèse s'attachent essentiellement à étudier et tester des méthodes de perturbation qui ciblent précisément les sources d'erreurs modèle, comme alternative à la méthode SPPT, largement répandue dans les systèmes de prévision d'ensemble. L'objectif de la thèse est d'évaluer l'impact de ces méthodes de perturbation sur le modèle de fine échelle Application of Research to Operations at MEso-scale (AROME, Seity et al. 2011 ; Brousseau et al. 2016) de Météo France et de comparer leurs performances en terme de production d'ensembles à celles de la méthode SPPT. On s'intéresse en particulier aux approches privilégiant la justification physique des perturbations, mais d'autres méthodes, notamment la perturbation des paramètres, sont également explorées.

Ce manuscrit est composé de six chapitres. Le **chapitre 1** présente de façon générale les méthodes employées pour construire une prévision d'ensemble, en abordant successivement les différentes sources d'incertitude pouvant affecter une prévision météorologique et leur représentation. Cette thèse s'intéresse en particulier aux incertitudes provenant des schémas de paramétrisation physiques utilisés pour représenter tous les processus non résolus explicitement dans le modèle. Ces schémas sont décrits dans le **chapitre 2**, en tentant de mettre en lumière l'incertitude associée à chacun d'eux. La méthodologie utilisée pour traiter du problème est présentée dans le **chapitre 3** : le modèle AROME y est décrit brièvement ainsi que les différents cas d'étude. Les chapitres suivants présentent les principaux résultats obtenus, d'abord dans un cadre de travail simplifié : une première partie est consacrée à deux méthodes de perturbation ciblant les schémas de turbulence et de convection peu profonde et l'apport de simulations à haute résolution (LES : Large-Eddy Simulations) pour le réglage et l'évaluation de ces perturbations (**chapitre 4**) ; d'autres méthodes de perturbation adaptées aux autres schémas de paramétrisation du modèle sont abordées dans le **chapitre 5**. Enfin, le **chapitre 6** explore l'impact des diverses méthodes de perturbation précédemment étudiées dans un cadre plus proche des conditions opérationnelles. Une dernière partie est consacrée aux conclusions.

Chapitre 1

Prévision d'ensemble : comment représenter l'incertitude d'une prévision météorologique ?

1.1 Qu'est-ce qu'une prévision numérique du temps ?

Depuis la deuxième moitié du 20^{ième} siècle, les météorologues utilisent des modèles de prévision numérique du temps (PNT) pour prédire l'évolution de l'atmosphère. Ces modèles reposent en premier lieu sur des équations de la dynamique des fluides permettant de décrire les mouvements des masses d'air dans l'atmosphère : l'équation de Navier-Stokes, qui correspond à un bilan de quantité de mouvement (1.2), l'équation de conservation de la masse (ou équation de continuité, 1.1). On y ajoute généralement une équation thermodynamique exprimant la conservation de l'énergie (1.3), et l'équation d'état (1.4).

$$\frac{D\rho}{Dt} + \rho \vec{\nabla} \cdot \vec{V} = 0 \quad (1.1)$$

$$\frac{D\vec{V}}{Dt} = \vec{g} - \frac{1}{\rho} \vec{\nabla} P + \nu \Delta \vec{V} \quad (1.2)$$

$$\frac{Dc_v T}{Dt} = -\frac{P}{\rho} \vec{\nabla} \cdot \vec{V} + \delta Q \quad (1.3)$$

$$P = \rho R T \quad (1.4)$$

Les équations ci-dessus utilisent la formulation lagrangienne. ρ est la masse volumique du fluide, \vec{V} son vecteur vitesse, P la pression, T la température, \vec{g} le champ de pesanteur terrestre, ν la viscosité cinématique du fluide, c_v la chaleur massique à volume constant, δQ l'échange de chaleur avec l'extérieur par unité de masse et de temps et R la constante spécifique de l'air.

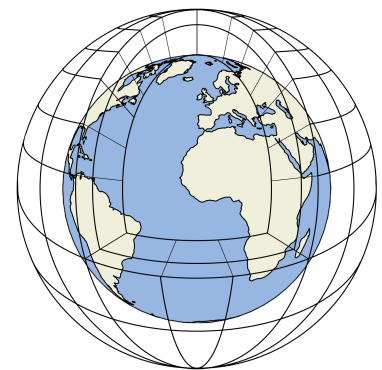


FIGURE 1.1 – Illustration de la discrétisation de l'atmosphère terrestre.

Dans le cadre d'un modèle numérique, ce système d'équations aux dérivées partielles est résolu de façon approchée, grâce à une discrétisation des variables d'espace et de temps. Le domaine de résolution est ainsi partagé en petits sous-domaines élémentaires, appelés *mailles*,

et on calcule l'état du système pour chaque maille à différents instants séparés par un pas de temps prédéfini. Les échelles de discrétisation spatiale et temporelle varient selon les modèles et leurs applications : pour un modèle à aire limitée elles peuvent être de l'ordre de quelques kilomètres et de la minute, tandis que pour un modèle global (c'est-à-dire dont le domaine est le globe terrestre), elles sont de l'ordre de 10 à 100 km et de quelques minutes. Ces échelles évoluent également au fur et à mesure de l'augmentation de la puissance de calcul des ordinateurs : ainsi, la première implémentation opérationnelle du modèle global Action de Recherche Petite Echelle Grande Echelle (ARPEGE ; Courtier et al. 1991) à Météo France en 1992 utilisait une résolution spatiale de l'ordre de 300 km (Pailleux et al. 2000) tandis qu'elle est actuellement d'environ 5 km sur la France (la résolution est variable selon l'endroit du globe et son minimum est de 24 km aux antipodes de la France, vers la Nouvelle-Zélande).

Les modèles atmosphériques sont également utilisés pour prévoir l'évolution de l'atmosphère à très long terme sur le globe. On parle alors de prévisions climatiques. Les échéances étant beaucoup plus lointaines, les modèles de climat utilisent en général une résolution spatio-temporelle plus faible que les modèles globaux de PNT – de l'ordre de la dizaine de minutes et de plus de 100 kilomètres.

Du fait de la discrétisation spatiale employée pour résoudre les équations de Navier-Stokes, de nombreux phénomènes physiques ayant lieu dans l'atmosphère ne peuvent pas être explicitement représentés. Il s'agit de phénomènes *sous-maille*, c'est-à-dire dont l'échelle est typiquement inférieure à celle de la maille du modèle. Bien qu'étant des processus de fine échelle, leur impact sur l'état d'une maille peut être très important et ils ne peuvent donc être négligés. Ils sont représentés dans les modèles de PNT à l'aide de "sous-modèles", appelés *schémas de paramétrisation* (ou simplement *paramétrisations*) et concernent des processus aussi divers que la formation des hydrométéores, l'effet des ascendances convectives ou les mouvements turbulents. Ces paramétrisations sont en général incohérentes spatialement (horizontalement), c'est-à-dire qu'elles sont appliquées indépendamment dans chaque colonne du modèle et leur action est limitée à la colonne (on parle de paramétrisations unidimensionnelles ou 1D). Souvent, elles font l'hypothèse d'une séparation d'échelle entre les processus sous-maille et la taille de la maille, de sorte que les processus sous-maille peuvent être considérés de façon statistique : on s'intéresse à leur effet moyen et on considère que celui-ci est essentiellement déterminé par l'état résolu de la maille. Les paramétrisations appliquent donc un lien déterministe entre état résolu et non résolu dans chaque maille. Comme on le verra par la suite, ces hypothèses peuvent être partiellement remises en cause et pourraient être à l'origine d'une part substantielle des erreurs commises par les modèles.

Par ailleurs, l'atmosphère interagit avec les autres compartiments géophysiques de la planète : les océans et les surfaces continentales. Des modèles spécifiques sont développés pour les décrire et peuvent être couplés aux modèles atmosphériques.

1.2 La nécessité de représenter l'incertitude

1.2.1 Modèles atmosphériques et théorie du chaos

Dans son article de 1963, Lorenz montre que certaines solutions de systèmes d'équations différentielles ordinaires non linéaires sont extrêmement sensibles aux conditions initiales. En modifiant légèrement l'initialisation, le modèle peut produire des résultats très différents (Lo-

renz 1963). Plusieurs systèmes dynamiques modélisant l'atmosphère de façon plus ou moins simplifiée sont étudiés par Lorenz (Lorenz 1963, 1965, 1969, 1982) et présentent tous une instabilité vis-à-vis des conditions initiales. Le terme de chaos est employé plus tard pour désigner ce comportement.

Les conséquences du caractère chaotique des modèles atmosphériques pour la prévision météorologique sont importantes. En effet, du fait de la quantité et de la qualité limitées des observations disponibles, il n'est pas possible de déterminer l'état "véritable" de l'atmosphère à un instant donné pour initialiser un modèle. Les prédictions de ce modèle finiront donc forcément par s'écarter de la réalité, au point d'être inutilisables. Lorenz (1965) montre également que la vitesse à laquelle les prévisions d'un système chaotique s'écartent de la réalité (ou s'écartent les unes des autres) n'est pas toujours la même, elle dépend notamment de la situation météorologique. Ainsi, la capacité des modèles à prédire le futur est limitée mais aussi variable selon la situation de départ.

1.2.2 Émergence de la prévision d'ensemble

Face à ce constat, plusieurs auteurs argumentent que les prévisions des modèles météorologiques devraient être envisagées sous un angle probabiliste (e.g. Epstein 1969), de façon à donner une information sur leur degré d'incertitude. Leith (1974) décrit le problème de la façon suivante. Considérons l'espace des phases, c'est-à-dire l'espace dont le nombre de dimensions équivaut au nombre de variables pronostiques du modèle, et dont chaque point représente un état possible du système. On dispose à l'instant t_0 d'un état initial x imparfaitement connu. On peut considérer qu'il existe une distribution de l'état initial véritable x_0 , autour de x , représentative du fait que l'on ne connaît pas l'état véritable, mais que l'observation imparfaite x nous donne un certain degré d'information sur ce qu'il pourrait être. A cet état initial probabiliste on pourrait associer une prévision également probabiliste, en faisant évoluer tous les points de la distribution initiale selon le modèle météorologique déterministe. On aurait, à l'instant $t_0 + \Delta t$ un ensemble de points dans l'espace des phases qui représenterait la distribution de probabilité de l'état véritable du système à cet instant (figure 1.2).

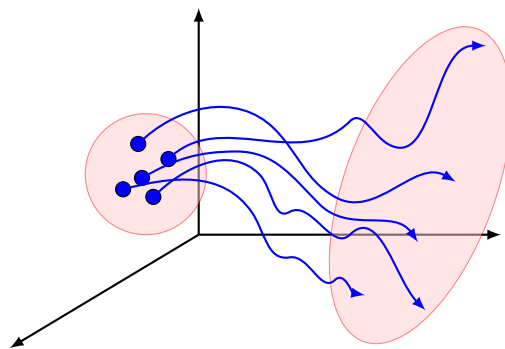


FIGURE 1.2 – Illustration de trajectoires dans l'espace des phases. Les cercles bleus représentent plusieurs états initiaux distincts. Les trajectoires correspondent aux différents états par lesquels passe le système. Les flèches bleues illustrent les états atteints à un instant donné $t > 0$.

Dans les années 1970, plusieurs études se sont intéressées à la façon d'estimer l'évolution de cette distribution des états possibles d'un modèle atmosphérique. L'équation de Liouville, qui est analogue à une équation de continuité appliquée aux probabilités, est étudiée notamment

par Gleeson (1970) et Ehrendorfer (1994) car elle permet en théorie, connaissant la densité de probabilité de l'état du système à l'instant t_0 , de calculer sa densité de probabilité aux instants ultérieurs. En pratique, l'évaluation numérique directe de cette densité de probabilité n'est pas envisageable car elle nécessite de parcourir tout l'espace des phases, ce qui représente un nombre de calculs très élevé en météorologie. Epstein (1969) propose avec l'approche *stochastic dynamic prediction* de réduire ce coût en n'évaluant que les deux premiers moments de la distribution et en utilisant une formulation approchée de leurs équations d'évolution. Cette méthode est utilisée dans plusieurs études (Fleming 1971 ; Pitcher 1977 ; Thompson 1985), mais toujours avec des systèmes dynamiques simplifiés ayant un faible nombre de degrés de liberté. Pour un modèle météorologique, le coût numérique reste prohibitif.

Une autre approche est proposée par Leith (1974), qui consiste à approcher la distribution par un tirage aléatoire. Leith propose de réaliser plusieurs intégrations d'un modèle à partir d'états initiaux échantillonnés par une méthode de Monte Carlo, et à utiliser l'ensemble des simulations pour obtenir des informations sur la distribution des états possibles de l'atmosphère. Cette proposition peut être considérée comme le début de la prévision d'ensemble.

Un ensemble est donc un nombre fini de simulations numériques (appelées *membres*), considéré comme étant un échantillon de la distribution des états possibles de l'atmosphère d'après nos connaissances. L'écart entre les membres (on parle aussi de *dispersion* de l'ensemble) est représentatif de l'incertitude de la prévision. Toutes les techniques de prévision d'ensemble cherchent à échantillonner au mieux la distribution des états possibles. Si les membres sont trop semblables les uns aux autres par rapport aux erreurs de prévision constatées à posteriori, ou que la réalité se trouve trop souvent en dehors des intervalles de valeurs prévues, alors l'ensemble est dit *sous-dispersif*. Dans ce cas, l'ensemble surestime la confiance que l'on peut avoir dans la prévision. A l'inverse, un ensemble peut être *sur-dispersif* quand il sous-estime la confiance dans la prévision.

1.2.3 Différentes sources d'erreur

A la suite des travaux de Lorenz, une attention particulière a d'abord été apportée à l'incertitude des états initiaux et leur impact sur la capacité de prédiction des modèles météorologiques. Les premières études faisaient donc l'hypothèse implicite d'un modèle parfait. C'est le cas par exemple dans l'étude de Leith (1974), qui précise néanmoins que les modèles eux-mêmes sont une source d'erreur de prévision : "One should bear in mind, however, that there is an additional external error generated by the discrepancy between the dynamics of the model and that of the real atmosphere". Ainsi, les différentes trajectoires dans l'espace des phases de la figure 1.2 ne devraient pas seulement être obtenues en ayant des états initiaux différents, mais également en perturbant certains éléments du modèle déterministe pour représenter ses erreurs. Plusieurs études ont en effet montré qu'un ensemble construit pour représenter uniquement l'incertitude des conditions initiales est sous-dispersif et que ses qualités sont améliorées lorsque l'incertitude du modèle est prise en compte (e.g. Stensrud et al. 2000 ; Fujita et al. 2007).

Dans toute la suite de cette thèse, on utilisera indistinctement les termes "erreurs modèle" et "incertitude (du) modèle" pour désigner les erreurs provenant du fait que les modèles ne reproduisent qu'imparfaitement la réalité. Il est par ailleurs possible de faire une distinction entre des erreurs *systématiques*, qui ont une tendance claire sur des statistiques de long terme (par exemple un manque d'humidité) et qui peuvent dans une certaine mesure être attribuées

à certaines déficiences identifiées du modèle, et les erreurs *aléatoires*, qui n'ont pas de tendance nette à long terme et qui peuvent provenir de tout un ensemble d'interactions dans le modèle. Le plus souvent, c'est à cette dernière catégorie d'erreurs que l'on fera référence en parlant de l'incertitude modèle.

Pour construire des prévisions d'ensemble à partir de modèles météorologiques déterministes et représenter les différentes sources d'erreur précédemment citées, un grand nombre de techniques ont été peu à peu mises en place dont on donne un aperçu dans les sections suivantes.

1.3 Perturber les données d'"entrée" du modèle

1.3.1 Perturbations de l'état initial

Réaliser une simulation numérique à l'aide d'un modèle de PNT nécessite d'initialiser ce modèle avec un état de départ qui correspond à l'état de l'atmosphère et éventuellement des autres compartiments – surfaces continentales, océans – à un instant donné. Bien que les observations disponibles soient toujours plus nombreuses, notamment depuis l'essor des données satellitaires débuté dans les années 1990, il n'est pas possible de connaître précisément l'état de l'atmosphère en chaque point du globe terrestre : les instruments de mesure ont une précision limitée et les données ne couvrent pas tout l'espace. Pour estimer un état initial, les observations doivent être traitées et il est notamment nécessaire de procéder à des interpolations en chaque point du modèle. Cette procédure est nommée *assimilation de données* et l'état initial du modèle qui en résulte est l'*analyse*.

L'analyse est un état approché de l'atmosphère à un instant donné. Ne connaissant pas l'état "véritable", on peut considérer qu'il existe une distribution d'états possibles dont l'analyse utilisée pour initialiser un modèle déterministe fait partie. Dans le cadre de la prévision d'ensemble, on cherche à échantillonner cette distribution des états initiaux possibles, de façon à représenter l'erreur d'initialisation. La principale difficulté est de savoir comment générer un échantillon d'états initiaux représentatif, sachant que sa taille est limitée. En effet, pour des raisons de coût numérique, le nombre de membres que l'on peut produire dans le cadre d'un système de prévision d'ensemble reste faible – typiquement quelques dizaines.

L'échantillonnage de l'état initial par tirage aléatoire (méthode de Monte Carlo), étudié d'un point de vue théorique par Leith (1974), est difficilement utilisable en pratique. En effet, l'espace des phases d'un modèle de PNT a un nombre de dimensions très important, beaucoup plus grand que la taille de l'échantillon à générer. Par ailleurs, toutes les perturbations de l'état initial n'ont pas les mêmes capacités à s'amplifier, ce qui peut mener à des prévisions très semblables les unes aux autres (Hollingsworth 1980). Les méthodes purement aléatoires de type Monte Carlo ont donc une efficacité limitée pour échantillonner l'ensemble des trajectoires possibles. Par ailleurs, elles nécessitent de connaître la distribution de probabilité des états initiaux, qui en pratique est rarement connue avec précision. Tout ceci a motivé l'émergence de méthodes d'échantillonnage s'appuyant sur la dynamique des modèles pour sélectionner des états initiaux susceptibles de mener à des trajectoires suffisamment différentes les unes des autres (Mureau et al. 1991 ; Palmer et al. 1993). La technique des vecteurs singuliers, développée au CEPMMT, a ainsi pour principe d'identifier les directions privilégiées dans l'espace des phases du modèle où l'instabilité dynamique est grande, c'est-à-dire où une perturbation sera susceptible de croître de façon importante (Molteni et al. 1996). Elle s'appuie pour cela sur le calcul des vecteurs

singuliers d'un opérateur linéaire modélisant la propagation de perturbations atmosphériques, et ceux-ci sont ensuite utilisés pour perturber l'analyse. De façon similaire, une méthode de "bred vectors" a été développée et utilisée pendant plusieurs années au NCEP pour repérer les erreurs qui s'amplifient au fur et à mesure des cycles d'assimilation et les utiliser pour générer des perturbations de l'analyse (Toth et Kalnay 1993, 1997).

Des méthodes dérivées du filtre de Kalman ont également été utilisées pour générer des perturbations de l'état initial. Le filtre de Kalman (Kalman 1960) est une méthode d'assimilation de données séquentielle consistant à calculer une succession de prévisions/analyses en tenant compte des erreurs contenues dans la prévision et dans les observations utilisées pour construire l'analyse. Une approximation de cette méthode, le filtre de Kalman d'ensemble (EnKF), proposée par Evensen (1994), utilise un ensemble de simulations pour évaluer et faire évoluer la matrice de covariance des erreurs de prévision nécessaire à l'algorithme. Un ensemble d'analyses est produit qui peut être utilisé pour déterminer les perturbations à appliquer à une analyse de référence. Cette méthode est employée actuellement par le Global Ensemble Forecast System (GEFS) du NCEP (Zhou et al. 2022). Une variante (ensemble *transform* Kalman filter) a également été utilisée dans le système MOGREPS-G du service météorologique britannique (Met Office) jusqu'en 2019 (Bowler et al. 2008 ; Tennant et Jardak 2020).

Au lieu de perturber une analyse issue d'un processus d'assimilation de données, il est aussi possible de faire tourner en parallèle plusieurs cycles d'assimilation de données, chacun ayant ses propres perturbations, afin d'obtenir un ensemble d'analyses permettant d'initialiser les différents membres d'un ensemble. On parle alors d'*ensembles d'assimilation* (AE, ou EDA en anglais pour Ensemble Data Assimilation). Houtekamer et Mitchell (2005) ont introduit en premier un ensemble d'assimilation pour initialiser le système de prévision d'ensemble canadien, en utilisant la méthode du filtre de Kalman d'ensemble. Des ensembles d'assimilation ont également été développés au CEPMMT et à Météo-France (Berre et al. 2007 ; Buizza et al. 2008 ; Desroziers et al. 2009) en faisant tourner en parallèle plusieurs instances d'une méthode d'assimilation de données variationnelle (Isaksen et al. 2010 ; Bouttier et al. 2012 ; Descamps et al. 2015). Le Met Office utilise également depuis peu un ensemble d'assimilation formé par plusieurs instances d'une méthode d'assimilation "hybride 4DEnVar" pour initialiser l'ensemble MOGREPS-G (Tennant et Jardak 2020).

1.3.2 Perturbations des conditions latérales

Si les premiers systèmes de prévision d'ensemble ont été construits pour des modèles globaux, il existe aujourd'hui de nombreux ensembles basés sur des modèles de plus fine échelle. Ces modèles à résolution spatiale et temporelle élevée sont plus coûteux en calculs et ne sont donc utilisés que sur des domaines géographiques restreints (on parle de modèles à aire limitée ou LAM, pour Limited Area Model). Utiliser un LAM pour réaliser une prévision nécessite de connaître les conditions aux bords de son domaine tout au long de son intégration. Pour cela, on utilise en général les prévisions d'un modèle ayant un domaine géographique plus étendu, qui contient le domaine du LAM – par exemple un modèle global.

De la même façon qu'une analyse est un état approché de l'atmosphère, les prévisions d'un modèle de PNT quel qu'il soit sont une approximation de l'état véritable de l'atmosphère. Utiliser ces prévisions comme conditions sur les bords latéraux du domaine d'un LAM est donc une source additionnelle d'erreurs (Warner et al. 1997). Dans le cadre d'un système de prévision

d'ensemble basé sur un LAM (que l'on abrègera par LEPS pour Limited area Ensemble Prediction System), plusieurs études démontrent l'importance de représenter cette source d'incertitude (Hou et al. 2001 ; Saito et al. 2010 ; Romine et al. 2014). En effet, utiliser des conditions de bord déterministes pour un LEPS peut aboutir à un ensemble fortement sous-dispersif (Nutter et al. 2004) car la croissance des perturbations introduites dans le modèle de fine échelle – que ce soit au niveau des conditions initiales ou bien du modèle lui-même – est fortement limitée par les conditions de bords qui n'interagissent pas avec le flux simulé par le LAM (Errico et Baumhefner 1987 ; Vukicevic et Paegle 1989 ; Vukicevic et Errico 1990). De nombreux LEPS utilisent donc des conditions de bords perturbées, et il est démontré que celles-ci ont une importance prédominante sur les perturbations des conditions initiales après quelques heures d'intégration (Hohenegger et al. 2008 ; Vié et al. 2011).

Plusieurs stratégies existent pour perturber les conditions de bord. Il est par exemple possible d'appliquer des perturbations aléatoires à une condition de bord déterministe (Torn et al. 2006). Néanmoins, la méthode la plus courante est d'utiliser les membres d'un ensemble de plus grande échelle. Il peut s'agir d'un ensemble formé par des prévisions provenant de différents centres météorologiques (Grimm et Mass 2002 ; García-Moya et al. 2011), ou bien d'un système de prévision d'ensemble basé sur un modèle de grande échelle (e.g. Frogner et Iversen 2002 ; Marsigli et al. 2005 ; Bowler et al. 2008 ; Hacker et al. 2011b ; Wang et al. 2011). Dans ce dernier cas, il est souvent nécessaire d'avoir recours à des méthodes de sélection d'un sous-ensemble de membres car la taille d'un ensemble de grande échelle est souvent supérieure à celle d'un LEPS, du fait du coût numérique moindre des modèles de grande échelle. Pour cela, des méthodes de clustering peuvent être appliquées (Molteni et al. 2001 ; Nuissier et al. 2012 ; Bouttier et Raynaud 2018).

1.3.3 Perturbations de la surface

L'état de la surface est également une condition de bord qui doit être fournie à un modèle de PNT – qu'il soit global ou à aire limitée. Les conditions de surface sont particulièrement importantes car ce sont elles qui modélisent les interactions entre continents, océans et atmosphère.

Les conditions de surface peuvent être obtenues à l'aide de schémas de paramétrisation directement intégrés dans un modèle atmosphérique (e.g. le schéma TERRA ML du modèle COSMO, Doms et al. (2011)), ou bien grâce à des modèles plus indépendants (e.g. SURFEX, Le Moigne 2012 ; Masson et al. 2013 ; JULES, Walters et al. 2019 ; CLM, Lawrence et al. 2011 ; NEMO, Madec et al. 2022). Ces derniers peuvent interagir avec le modèle atmosphérique par un processus dit de *couplage*, grâce auquel ils fournissent des conditions et des flux de surface au modèle atmosphérique qui en retour leur fournit diverses données météorologiques susceptibles de les affecter. Le modèle SURFEX est par exemple utilisé opérationnellement en mode couplage pour fournir les conditions de surface des modèles de PNT et de climat de Météo France (Seity et al. 2011 ; Roehrig et al. 2020).

Contrairement à un modèle atmosphérique, un modèle de surface est rarement de nature chaotique : les erreurs d'initialisation ne s'amplifient pas au cours de l'intégration, au contraire, après une période de transition (spin-up) le modèle atteint un état de quasi-équilibre déterminé par les forçages atmosphériques (Cosgrove et al. 2003). Néanmoins, un modèle de surface peut être une source d'incertitude du fait de ses erreurs internes. Certains phénomènes ne sont pas représentés ou sont simplifiés et, selon la résolution spatiale, les types de surface et leur

hétérogénéité sont plus ou moins bien représentés. Ceci a une incidence sur les prévisions des modèles de PNT car de nombreux phénomènes atmosphériques sont fortement influencés par les conditions de surface, notamment les phénomènes convectifs (Sud et al. 1988 ; Fu et al. 1994 ; Barthlott et Kalthoff 2011). Dans le cadre de la prévision d'ensemble, il est donc nécessaire de représenter aussi les erreurs des conditions de surface, sans quoi l'atmosphère simulée proche de la surface manque de variabilité (Draper 2021).

Plusieurs études ont ainsi montré une amélioration des performances de systèmes de prévision d'ensemble en incluant des perturbations des conditions de surface (Lavaysse et al. 2013 ; Tennant et Beare 2014 ; Bouttier et al. 2015 ; MacLeod et al. 2016 ; Orth et al. 2016). Parmi les méthodes de perturbations testées, on trouve des perturbations appliquées aux tendances des variables du modèle de surface (MacLeod et al. 2016 ; Wang et al. 2019), ou bien directement à certaines variables (McLay et al. 2012 ; Bouttier et al. 2015 ; Draper 2021), ou encore à plusieurs paramètres du modèle (Bouttier et al. 2015 ; MacLeod et al. 2016 ; Draper 2021).

1.4 Diverses approches pour la représentation des erreurs modèle

Les résultats d'un modèle de PNT ne divergent pas seulement de la réalité à cause d'erreurs introduites dans leur initialisation ou dans la prescription des conditions de bord. Les modèles eux-mêmes utilisent de nombreuses approximations et une partie des incertitudes des prévisions météorologiques proviennent de cet écart avec la réalité. Dès le début des expériences de prévision d'ensemble, la nécessité de représenter les erreurs de modélisation, et donc d'écarter l'hypothèse d'un modèle "parfait", est soulignée (Houtekamer et al. 1996). Plusieurs méthodes ont été proposées et testées depuis les années 1990, que l'on tente de résumer dans la figure 1.3. Aujourd'hui encore, les systèmes de prévision d'ensemble des centres météorologiques nationaux utilisent une ou plusieurs de ces méthodes, et il n'existe pas encore de véritable consensus sur la façon la plus efficace de représenter les erreurs modèle.

1.4.1 Les méthodes multimodèle et multiphysique

Une des premières approches envisagées pour représenter les erreurs de modélisation a été de tirer parti de la diversité des modèles météorologiques existants et de leurs différentes configurations possibles (Houtekamer et al. 1996 ; Harrison et al. 1999 ; Stensrud et al. 1999). L'idée générale est que chaque modèle ou chaque version d'un modèle possède des hypothèses et parfois des approches différentes pour modéliser les phénomènes atmosphériques, et donc qu'utiliser les résultats de plusieurs modèles différents peut être une manière d'estimer les erreurs de modélisation.

Comparer les résultats de modèles différents est une pratique ancienne dans certains centres météorologiques pour évaluer la fiabilité des prévisions (Wobus et Kalnay 1995), et des études ont par ailleurs montré que le consensus obtenu à partir de plusieurs prévisions (sous la forme d'une moyenne pondérée par exemple) pouvait avoir de meilleures performances que ces mêmes prévisions prises isolément (Vislocky et Fritsch 1995 ; Brown et Murphy 1996 ; Fritsch et al. 2000). Avec l'émergence de la prévision d'ensemble, cette approche, qualifiée de *multimodèle*, a été utilisée pour produire et évaluer divers ensembles : dans le cadre du projet européen DEMETER (Palmer et al. 2004) un ensemble est construit à partir de 7 modèles différents et produit

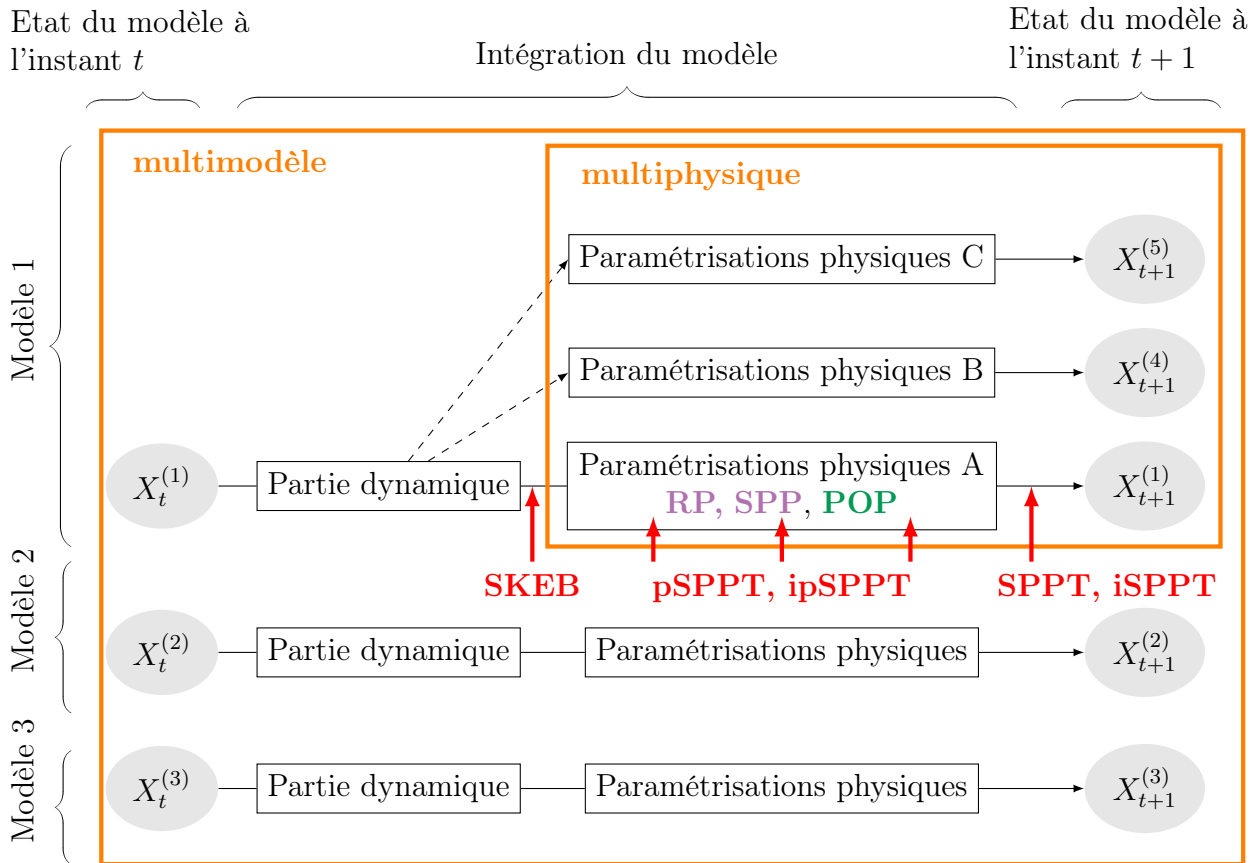


FIGURE 1.3 – Illustration des différentes méthodes de perturbations utilisées pour représenter les erreurs modèle en prévision d'ensemble. En orange sont indiquées les méthodes multimodèle/multiphysique décrites dans la section 1.4.1, en rouge sont représentées les méthodes de perturbation stochastique "globales", décrites dans la section 1.4.2, en violet les méthodes de perturbation des paramètres, décrites dans la section 1.4.3 et en vert les méthodes de perturbation des processus, décrites dans la section 1.4.4.

de meilleurs reprévisions que des ensembles basés sur chacun des modèles isolément (Hagedorn et al. 2005) ; Arribas et al. (2005) construisent un ensemble à partir de 14 modèles provenant de 9 centres météorologiques et rapportent des performances comparables au système de prévision d'ensemble du CEPMMT de l'époque ; Zhou et Du (2010) montrent qu'un ensemble combinant deux modèles régionaux dérivés du modèle Weather Research and Forecasting (WRF) améliore les prévisions de brouillard ; un ensemble multimodèle est constitué par le service météorologique espagnol à partir de cinq LAMs (García-Moya et al. 2011).

Au lieu de construire un ensemble à partir de plusieurs modèles déterministes, il est également possible de combiner directement les résultats de systèmes de prévision d'ensemble déjà existants – on peut alors parler de *multiensembles*. Evans et al. (2000) montrent une amélioration des performances des prévisions lorsque les ensembles du CEPMMT et du Met Office sont réunis, qui n'est pas simplement due à l'augmentation du nombre de membres. Dans le cadre du projet TIGGE, Park et al. (2008) rapportent également pour certaines zones du globe l'intérêt d'un grand ensemble formé par la réunion des ensembles de huit centres de prévision. Candille (2009) rapporte de meilleures performances en utilisant une combinaison des ensembles du Centre Météorologique Canadien (CMC) et du NCEP, qu'avec chaque ensemble pris séparément, à nombre de membres égal.

Lorsque ce ne sont pas des modèles différents mais plutôt des configurations différentes d'un même modèle qui sont utilisées, on parle de *multiphysique*. C'est l'approche initialement utilisée par Houtekamer et al. (1996) pour l'ensemble du CMC : des combinaisons de différents schémas de convection, de radiation et de traînées d'ondes de gravité sont utilisées pour chaque membre de l'ensemble afin de tenir compte des erreurs modèle. De façon analogue, Stensrud et al. (2000) construisent un ensemble en faisant varier les schémas de paramétrisation de la convection et de la couche limite dans un modèle de méso-échelle; Hacker et al. (2011b) déterminent 10 configurations physiques optimales pour leur ensemble régional en sélectionnant différents schémas microphysiques, radiatifs, de couche limite, de surface et de paramétrisation des cumulus; Duda et al. (2014) utilisent quant à eux différents schémas microphysiques pour étudier l'incertitude provenant de la microphysique dans le modèle WRF.

Malgré le succès des approches multimodèle et multiphysique pour produire des ensembles (e.g. Mylne et al. 2002; Arribas et al. 2005; Eckel et Mass 2005; García-Moya et al. 2011; Iversen et al. 2011), plusieurs désavantages sont soulignés dans la littérature (voir par exemple Jankov et al. 2017; Gallus et al. 2019). Tout d'abord, ces méthodes peuvent demander un lourd travail de maintenance, car les évolutions et la compatibilité doivent être assurées pour tous les modèles et les schémas de paramétrisation utilisés. Par ailleurs, certaines hypothèses communément admises en prévision d'ensemble ne sont pas forcément vérifiées par les approches multimodèle et multiphysique. Par exemple, on suppose généralement que les membres d'un ensemble sont indépendants et équiprobables : ce n'est pas le cas si les modèles ou les paramétrisations utilisés pour construire l'ensemble sont apparentés, ou si leurs performances sont différentes. Separovic et al. (2016) montrent ainsi une claire séparation des membres de l'ensemble canadien Global Ensemble Prediction System (GEPS) concernant la prévision de la température à 250 hPa dans les tropiques : les membres utilisant le schéma de convection profonde de Kuo (Kuo 1974) ont tous un biais plus froid que les membres utilisant un schéma de Kain-Fritsch (Kain et Fritsch 1990). Par ailleurs, si l'on utilise systématiquement pour chacun des membres le même modèle ou bien la même combinaison de schémas de paramétrisation physique, alors chaque membre d'un ensemble multimodèle ou multiphysique a des biais et des statistiques de long terme (c'est-à-dire une climatologie) différentes. Avec un tel ensemble, on n'échantillonne pas au hasard dans la distribution de probabilité des états du système, on regarde seulement certains types de trajectoires liées à un nombre fini et restreint de modèles. Même si cela peut donner des informations sur l'erreur de modélisation, cela ne permet pas d'explorer toutes les trajectoires possibles, et, comme cela est souligné dans Bowler et al. (2008), le risque est que la distribution des membres de l'ensemble soit davantage influencée par les différences systématiques entre les modèles que par l'incertitude liée à la situation météorologique. Ces différentes limites expliquent que les approches multimodèle et multiphysique soient actuellement de moins en moins utilisées, au profit des méthodes de *perturbations stochastiques* (Bowler et al. 2008; Charron et al. 2010; Hacker et al. 2011b; Jankov et al. 2017; Xu et al. 2020a; Kalina et al. 2021; McTaggart-Cowan et al. 2022b).

1.4.2 Les méthodes de perturbation stochastiques "globales"

Les méthodes de perturbation stochastique proposent de représenter les incertitudes d'un modèle en particulier en lui appliquant des perturbations aléatoires. Concrètement, il s'agit d'introduire un "bruit" à certaines étapes de l'intégration du modèle – le plus souvent celles impliquant les schémas de paramétrisation physique qui sont considérés comme une source ma-

jeu d'incertitude (Palmer et al. 2009).

Bien que les méthodes multimodèle ou multiphysique et les méthodes de perturbation stochastique aient pour but commun de représenter les erreurs de modélisation, Palmer et al. (2005) considèrent que leur philosophie est différente. Les premières partent du constat que l'on ne connaît pas la meilleure façon de modéliser les phénomènes atmosphériques, en particulier les processus sous-maille, et proposent donc de représenter les erreurs qui découlent de ces imperfections en utilisant différentes versions des modèles et des schémas. Les deuxièmes reconnaissent le fait que les erreurs ne viennent pas seulement des détails sur la façon de modéliser tel ou tel processus sous-maille, mais de l'existence-même de la discrétisation spatiale et de l'utilisation de schémas de paramétrisation (Palmer 2001). Ainsi, même en ayant des paramétrisations très performantes, il subsisterait une incertitude.

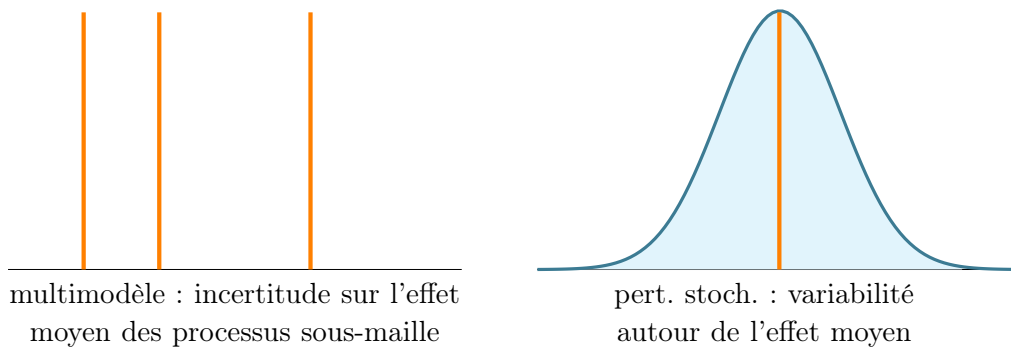


FIGURE 1.4 – Illustration de la différence entre les approches multimodèle et perturbations stochastiques. Voir Palmer (1999).

En effet, les paramétrisations classiques reposent en général sur des hypothèses de séparation d'échelle, parfois d'équilibre local au niveau de la maille entre phénomènes résolus et non résolus, grâce auxquelles elles établissent un lien déterministe entre l'état résolu d'une maille et l'effet moyen des processus sous-maille sur cet état résolu. Or ceci peut représenter une approximation grossière et cacher une importante variabilité, comme le montrent par exemple Shutts et Palmer (2004) et Shutts et Palmer (2007). Dans ces études, les auteurs utilisent deux simulations de convection profonde dans les tropiques, une à haute résolution, permettant de représenter explicitement le phénomène, et l'autre avec un modèle global, qui nécessite un schéma de paramétrisation. Le spectre des tendances de température issues du modèle à haute résolution, moyennées à l'échelle du modèle global, est beaucoup plus large que celui des tendances paramétrées dans ce dernier. Ces constats ont motivé le développement de la méthode SPPT (Stochastically Perturbed Parametrization Tendencies) décrite plus loin dans cette section.

Le fait d'avoir une discrétisation spatiale et des schémas de paramétrisation pourrait également être responsable d'un manque de variabilité dans les modèles de PNT en empêchant une partie des interactions entre mouvements d'échelles différentes d'avoir lieu. En effet, dans l'atmosphère, l'énergie cinétique est transmise des plus grandes échelles aux plus petites (on parle de *cascade d'énergie*), où elle est en partie dissipée, mais les petits mouvements peuvent également s'associer et en former de plus larges, et dans ce cas l'énergie est transmise de façon rétrograde, des petites aux grandes échelles (Gage 1979 ; Lilly 1983). L'utilisation d'une discrétisation spatiale introduit une rupture et nécessite l'utilisation d'un terme de diffusion pour représenter artificiellement la dissipation de l'énergie aux plus petites échelles. Le fait que

cette diffusion puisse être excessive, et que par ailleurs les schémas de paramétrisation ne représentent pas forcément le transport rétrograde d'énergie est avancé comme une cause possible du manque de variabilité des modèles et donc une source d'erreurs modèle (Shutts 2005). La méthode de perturbation SKEB (Stochastic Kinetic Energy Backscatter), décrite ci-après, a été construite en réponse à ce problème.

Dans cette section, on présente donc deux méthodes de perturbation stochastique qui ont été et sont encore largement utilisées : SPPT et SKEB . On les qualifie de méthodes de perturbation "globales", dans le sens où chacune représente avec un unique terme aléatoire l'incertitude provenant de multiples processus, par opposition à des méthodes plus ciblées qui seront décrites dans les sections suivantes.

SPPT : Stochastically Perturbed Parametrization Tendencies

En 1999, une méthode de perturbation stochastique pour représenter l'incertitude modèle est introduite dans le système de prévision d'ensemble du CEPMMT (Buizza et al. 1999 ; Palmer 2019). Désignée ultérieurement par l'acronyme SPPT (Palmer et al. 2009), son principe général est d'appliquer une perturbation multiplicative aux tendances calculées par les schémas de paramétrisation physique du modèle.

Considérons un modèle de PNT que l'on résume par :

$$\frac{dX}{dt} = G(X) = D + P,$$

avec X représentant les variables d'état du modèle, et G le modèle. Ce modèle peut être séparé en une partie dynamique D correspondant à l'intégration numérique des équations du mouvement, et une partie P correspondant à l'effet des processus sous-maille calculés grâce aux schémas de paramétrisation. La méthode SPPT consiste à perturber P de la façon suivante :

$$\frac{dX}{dt} = D + (1 + \epsilon)P,$$

où ϵ est un processus aléatoire. Dans la version initiale de Buizza et al. (1999), ϵ correspond à un champ 2D de valeurs aléatoires issues d'un tirage uniforme. Une corrélation spatiale et temporelle est assurée en ayant des valeurs constantes sur des sous-domaines plus ou moins grands et en les renouvelant avec un pas de temps plus ou moins long.

Par la suite, Palmer et al. (2009) proposent une méthode plus élaborée basée sur un générateur de champs spectraux. Le champ aléatoire 2D est exprimé à l'aide d'un développement en harmoniques sphériques :

$$\epsilon(\lambda, \Phi, t) = \sum_{n=0}^N \sum_{m=-n}^n \phi_{n,m}(t) Y_m^n(\lambda, \Phi)$$

où λ est la longitude, Φ la latitude, $\phi_{n,m}$ les coefficients spectraux et Y_m^n les fonctions de Legendre associées. Chaque coefficient spectral est défini comme étant un processus autorégressif d'ordre 1 :

$$\phi_{m,n}(t + \Delta t) = \beta \phi_{m,n}(t) + a_n \eta$$

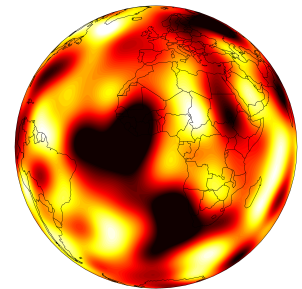


FIGURE 1.5 – Exemple de champ aléatoire ϵ gaussien utilisé dans l'ensemble global PEARP de Météo France.

où η est un bruit blanc gaussien $\eta \sim N(0, 1)$.

La corrélation temporelle du processus AR1 est réglée par le paramètre β , et sa variance dépend de β et du paramètre d'amplitude a_n . Ce dernier est réglé de façon à ce que la variance du processus ϵ soit la même en tout point de la grille du modèle, et permet également de contrôler la corrélation horizontale du processus. Le type de champ aléatoire obtenu avec cette méthode est illustré sur la figure 1.5. Le processus aléatoire ϵ produit donc à chaque pas de temps un champ aléatoire 2D spatialement corrélé. En revanche il n'existe pas de structure verticale du bruit dans la méthode SPPT : en chaque point du modèle, la perturbation $\epsilon(\lambda, \Phi, t)$ est appliquée uniformément sur toute la colonne. Certains ajustements sont parfois réalisés : Palmer et al. (2009) introduisent par exemple une atténuation des perturbations dans les niveaux du modèle proches de la surface ainsi que dans la stratosphère. Un coefficient α dépendant de la verticale est appliqué au processus aléatoire selon :

$$\frac{dX}{dt} = D + (1 + \alpha\epsilon)P.$$

La raison principale de l'atténuation à la surface est d'éviter l'apparition d'instabilités numériques. Quant à l'atténuation dans la stratosphère, elle est due au fait que la contribution des schémas de paramétrisation à cette altitude est essentiellement assurée par le schéma radiatif qui est considéré comme une source d'incertitude de moindre importance dans cette zone de l'atmosphère. Plusieurs systèmes de prévision d'ensemble utilisant la méthode SPPT appliquent cette atténuation verticale (Bouttier et al. 2012; Leutbecher et al. 2017; Frogner et al. 2022; Xu et al. 2022) mais certains systèmes semblent pouvoir se passer de l'atténuation près de la surface (e.g. Klasa et al. 2018).

La méthode SPPT a été appliquée avec succès dans de nombreux systèmes de prévision d'ensemble (e.g. Charron et al. 2010; Yonehara et Ujiie 2011; Bouttier et al. 2012; Romine et al. 2014; Weisheimer et al. 2014; Batté et Doblas-Reyes 2015; Lupo et al. 2020). Néanmoins, plusieurs limites sont soulevées dans la littérature. L'une des plus fréquentes est le manque de cohérence physique de la méthode. En effet, les perturbations sont appliquées seulement aux tendances des variables et n'affectent pas les flux calculés au sommet de l'atmosphère et à la surface, il y a donc un déséquilibre non maîtrisé dans le bilan d'énergie et de quantité d'eau au sein de chaque colonne du modèle. Une autre limite de la méthode SPPT est liée au fait que la perturbation est multiplicative et appliquée à la somme de toutes les tendances des schémas de paramétrisation physique. Ceci permet de conserver la contribution relative de chaque schéma dans la tendance totale, mais suppose que leurs erreurs sont parfaitement corrélées, ce qui n'est pas forcément vérifié (Shutts et Pallarès 2014; Christensen 2020). D'autre part, cela implique que la perturbation est nulle si la somme des tendances est nulle; or la somme des tendances peut être nulle alors que les tendances individuelles de chaque schéma sont grandes mais se compensent. L'absence de perturbation dans ce cas de figure est discutable (voir par exemple la figure 7 de Shutts et Palmer 2007). Enfin, le fait de devoir atténuer les perturbations près de la surface, du fait de problèmes d'instabilités numériques, n'est pas réaliste car les paramétrisations physiques ont un impact important dans cette zone de l'atmosphère, et peuvent être une source d'incertitude non négligeable. Cette atténuation a pu être désignée comme responsable du manque de dispersion de variables de surface dans certains cas (Wastl et al. 2019b).

Plusieurs variantes de la méthode SPPT ont été proposées qui répondent partiellement à ces limites. Arnold (2013) puis Christensen et al. (2017) proposent d'utiliser plusieurs champs aléatoires différents pour pouvoir perturber indépendamment la contribution de chaque schéma

de paramétrisation aux tendances des variables d'état (méthode iSPPT). Dans le même esprit, Wastl et al. (2019b) proposent de perturber indépendamment les tendances après l'appel de chaque schéma de paramétrisation, permettant ainsi aux perturbations d'être propagées successivement à travers les différents schémas (méthode pSPPT). Utiliser des perturbations indépendantes pour chaque variable (température, humidité et vent) est également exploré dans la même étude (méthode ipSPPT).

SKEB : Stochastic Kinetic Energy Backscatter

Une autre méthode de perturbation stochastique très répandue est la méthode SKEB. Initialement proposée par Shutts (2005), en s'inspirant du schéma "backscatter" de Mason et Thomson (1992) destiné à un modèle de turbulence à haute résolution (LES), elle vise à réinjecter sous la forme d'une perturbation stochastique la dissipation excessive d'énergie aux plus petites échelles. Shutts (2005) identifie différentes sources de dissipation d'énergie dans le modèle global du CEPMMT : la diffusion numérique, la paramétrisation des traînées d'ondes de gravité et la paramétrisation de la convection. Pour réinjecter une partie de cette énergie dans le modèle, des perturbations du champ de vorticit   sont appliqu  es    travers une fonction de courant, du type :

$$\psi(\lambda, \Phi, z, t) = \sqrt{bD(\lambda, \Phi, z, t)}\phi(\lambda, \Phi, t).$$

Dans la version de Shutts (2005), ϕ est un champ al  atoire 2D cr  e    l'aide d'un automate cellulaire (CA pour Cellular Automaton en anglais). Dans une version ult  rieure, propos  e par Berner et al. (2009), le CA est remplac   par un g  n  rateur de champs spectraux semblable    celui d  fini pr  c  demment pour la m  thode SPPT. Il est multipli   par un champ d  pendant de D , l'estimation de la dissipation instantan  e d'  nergie cin  tique aux petites   chelles. La proportion d'  nergie dissip  e qui est r  inject  e dans le mod  le est contr  l  e par le param  tre b .

Avec SKEB, les perturbations sont appliqu  es    la partie dynamique du mod  le, contrairement    la m  thode SPPT o   ce sont les tendances issues des sch  mas de param  trisation physique qui sont perturb  es.

La m  thode SKEB est ou a   t   utilis  e dans de nombreux ensembles (e.g. Bowler et al. 2008 ; Charron et al. 2010 ; Berner et al. 2011 ; Shutts et al. 2011 ; Str  mmen et al. 2018 ; Xu et al. 2020b). Selon certains auteurs, son impact serait int  ressant principalement pour les mod  les    basse r  solution tandis que l'emploi de cette m  thode pour des mod  les    fine r  solution serait moins justifi   du fait d'interactions entre   chelles plus compl  tes dans ces derniers (Leutbecher et al. 2017).

1.4.3 Les m  thodes de perturbation de param  tres

Les mod  les m  t  orologiques poss  dent un grand nombre de param  tres r  glables. Ce sont des valeurs fixes intervenant dans les   quations et qui repr  sentent par exemple des coefficients de m  lange, de transfert, des seuils, des longueurs caract  ristiques, des valeurs maximales ou minimales, etc. Les param  tres ne sont pas toujours des grandeurs observables et mesurables. Dans la majeure partie des cas, le r  glage de leur valeur est fait manuellement en s'appuyant sur l'avis d'experts et les r  sultats de diff  rents cas d'  tude. Trouver des valeurs acceptables qui permettent d'obtenir de bonnes performances g  n  re un travail important, d'autant plus que

les schémas de paramétrisation interagissent entre eux et sont de plus en plus complexes. Ainsi, de façon générale, il n'est pas garanti que les valeurs choisies pour les paramètres d'un modèle soit les plus appropriées, ce qui en fait une source potentiellement importante d'incertitude. Des méthodes ont été développées pour tenter de régler de façon plus automatique la valeur des paramètres (Järvinen et al. 2010, 2012 ; Couvreur et al. 2021), mais elles ne sont pas à ce jour utilisées de façon opérationnelle.

Perturber les paramètres des modèles météorologiques pour estimer les erreurs modèle a été envisagé dès le début de la prévision d'ensemble. Dans les expériences menées par Houtekamer et al. (1996) pour construire un ensemble basé sur le modèle canadien, la modification du réglage de la diffusion horizontale fait partie des méthodes utilisées pour tenir compte des erreurs modèle. Les perturbations de paramètres ont été utilisées ensuite dans de nombreuses expériences impliquant des modèles globaux (Murphy et al. 2004 ; Stainforth et al. 2005 ; Yokohata et al. 2010) et des LAMs (Bright et Mullen 2002). Les premières études ont en général utilisé des perturbations statiques (Fresnay et al. 2012 ; Yussouf et Stensrud 2012), c'est-à-dire que chacun des membres a une valeur différente mais constante (en temps, en espace et pour chaque initialisation du système) pour un paramètre donné. Cette approche, qualifiée parfois de *multiparamètre* (Palmer et al. 2005 ; Gebhardt et al. 2011 ; Hacker et al. 2011a ; Yussouf et Stensrud 2012 ; Shiogama et al. 2014), n'est pas à proprement parler une méthode de perturbation stochastique : chaque membre possède sa propre combinaison de valeurs de paramètres fixe, et donc une climatologie distincte. Plusieurs études ont en revanche appliqué des perturbations stochastiques des paramètres, en réactualisant leurs valeurs à chaque initialisation du système et en les rendant variables en temps et/ou en espace. Certains auteurs argumentent en effet que comme les paramètres peuvent servir à résumer de façon simplifiée plusieurs processus, il est probable qu'il n'existe pas une unique valeur optimale pour chacun d'eux, mais qu'au contraire elle varie selon la situation météorologique. Ainsi, par exemple, Bright et Mullen (2002) perturbent le nombre de Richardson critique d'un schéma de couche limite grâce à un processus autorégressif de premier ordre (AR1). Les valeurs de ce paramètre sont donc variables et corrélées temporellement au sein d'une simulation. Le système de prévision d'ensemble MOGREPS du Met Office intègre quant à lui dès ses débuts une méthode de perturbation de paramètres (Bowler et al. 2008) désignée par RP (pour Random Parameter), qui perturbe 8 paramètres autour de leur valeur par défaut grâce à un processus AR1. Des évolutions sont apportées dans Baker et al. (2014), McCabe et al. (2016) et Tennant (2016), qui introduit notamment des variations spatiales. Christensen et al. (2015) étudient l'impact de perturbations fixes ou variables temporellement et spatialement, appliquées à 4 paramètres du schéma de convection du CEPMMT. Ollinaho et al. (2017) introduisent des perturbations de plusieurs paramètres du modèle IFS, appartenant aux schémas radiatif, de convection, de nuage, de turbulence et d'orographie. Leur approche, désignée par l'acronyme SPP (pour Stochastically Perturbed Parametrizations), est caractérisée par des perturbations produites sous la forme de champs 2D aléatoires, indépendants pour chaque paramètre et possédant une corrélation temporelle et spatiale. De façon similaire, Jankov et al. (2017) testent une méthode de perturbation de plusieurs paramètres d'un schéma de convection et d'un schéma de couche limite dans l'ensemble basé sur le modèle Rapid Refresh (RAP).

Concernant la performance des méthodes de perturbation de paramètre dans le cadre de la prévision d'ensemble, plusieurs études rapportent un bénéfice moindre (Hacker et al. 2011b ; Berner et al. 2015 ; Christensen et al. 2015 ; Jankov et al. 2017, 2019) ou mitigé (Ollinaho et al. 2017) par rapport aux méthodes de perturbations stochastiques globales ou à la méthode mul-

tiphysique. Selon certains auteurs, la perturbation des paramètres ne suffit pas à représenter l'ensemble des erreurs modèle et devrait donc être combinée à d'autres méthodes de perturbation. Berner et al. (2015) rapportent par exemple que les meilleures performances dans leurs expériences avec le modèle WRF sont obtenues en combinant les approches multiphysique, SKEB et multiparamètre. Jankov et al. (2017, 2019) préconisent d'utiliser SPP en combinaison avec les méthodes globales SKEB et SPPT. Néanmoins, de récentes études montrent aussi que les méthodes de perturbation de paramètres sont à présent compétitives par rapport aux autres méthodes de perturbations. Ainsi, des améliorations apportées à la méthode SPP permettent d'obtenir des performances similaires à SPPT dans l'ensemble du CEPMMT (Lang et al. 2021). Par ailleurs, l'implémentation de SPP dans l'ensemble HarmonEPS du programme High Resolution Limited Area Model (HIRLAM) donne de meilleurs résultats que SPPT (Frogner et al. 2022). SPP a également été récemment introduit dans le système de prévision d'ensemble global canadien (McTaggart-Cowan et al. 2022a) et les résultats favorables obtenus (McTaggart-Cowan et al. 2022b) ont conduit le CMC à remplacer les méthodes multiphysique et SPPT, jusqu'alors utilisées, par SPP. Kalina et al. (2021) montrent que SPP appliqué au système High-Resolution Rapid Refresh Ensemble (HRRRE) permet d'obtenir des performances similaires ou supérieures à l'ensemble opérationnel High-Resolution Ensemble Forecast system, version 2 (HREFv2) du NCEP, qui utilise une approche multimodèle/multiphysique. Wang et al. (2020) appliquent des perturbations de type RP à 8 paramètres sélectionnés par une méthode d'optimisation dans une version de mésoéchelle du modèle Global/Regional Assimilation and Prediction System (GRAPES) et obtiennent de meilleurs scores qu'avec SPPT. Enfin, Wimmer (2021) rapporte également des résultats positifs avec une méthode de perturbation de type multiparamètre appliquée au modèle AROME et comparée à la méthode opérationnelle SPPT.

Les travaux d'intercomparaison précédemment cités ont des conclusions diverses concernant la supériorité de telle ou telle représentation des erreurs modèle qui doivent parfois être considérées avec précaution étant donnée la diversité des conditions de tests (modèles choisis, cas d'étude, paramètres perturbés, etc). Néanmoins, les études récentes semblent indiquer que les méthodes de perturbation de paramètres (notamment la méthode SPP) ont des performances intéressantes, et sont en tout cas le reflet d'une recherche active à leur sujet, qui va de pair avec la volonté de plusieurs centres de les utiliser de façon opérationnelle (Leutbecher et al. 2017; Kalina et al. 2021; McTaggart-Cowan et al. 2022b). Les méthodes de perturbation de paramètres possèdent en effet plusieurs avantages. Par rapport aux méthodes de perturbation stochastique globales, elles permettent de cibler plus précisément l'incertitude : au lieu de représenter toutes les erreurs modèle avec un seul terme stochastique, elles permettent de perturber indépendamment des processus particuliers grâce aux paramètres intervenant dans leur modélisation. Ainsi, on peut, en sélectionnant tel ou tel paramètre, perturber les processus physiques que l'on estime être une source d'incertitude importante. Par ailleurs, les méthodes de perturbation des paramètres ne présentent pas le problème de déséquilibre des budgets d'énergie ou de quantité d'eau comme la méthode SPPT, car les perturbations sont introduites au sein des schémas de paramétrisation, qui eux sont construits de façon à respecter ces équilibres.

Néanmoins, quelques limites sont aussi identifiées dans la littérature. Palmer (2019) mentionne par exemple le problème de perturber de façon indépendante les paramètres, qui peut mener à détruire l'équilibre entre les paramétrisations et donc entre les processus participant à modéliser un événement météorologique donné. Selon les valeurs aléatoires prises par les paramètres, cela pourrait donner lieu à des simulations non réalistes. Dans certaines études, les perturbations de différents paramètres sont ainsi corrélées ou anticorrélées suivant la nature

des processus en jeu (e.g. Kalina et al. 2021). Par ailleurs, un des désavantages le plus couramment cité à propos des méthodes de perturbation de paramètres est qu'elles sont coûteuses en terme de réglages. De même que la valeur des paramètres est spécifiée manuellement dans les modèles déterministes, le réglage des perturbations appliquées aux paramètres dans le cadre d'un ensemble est aussi la plupart du temps manuel : la sélection des paramètres à perturber ainsi que les intervalles de valeurs possibles pour ces paramètres sont fixés par des experts. Or le choix des paramètres ainsi que la distribution de probabilité dans laquelle sont tirées leurs valeurs semblent influencer grandement sur les résultats (voir par exemple la différence de performance de SPP entre les versions de Ollinaho et al. (2017) et Lang et al. (2021)). De plus cela implique que si les schémas de paramétrisation évoluent, ou sont remplacés par de nouveaux schémas, le travail de sélection de paramètres et le réglage de la distribution est à refaire, ce qui peut générer un travail humain important. Face à cela, certaines études explorent la possibilité d'utiliser des méthodes plus automatiques pour le réglage des perturbations. Ainsi, Christensen et al. (2015) perturbent quatre paramètres d'un schéma de convection grâce à une estimation préalable de leur incertitude par la méthode Ensemble Prediction and Parameter Estimation System (EPPEs), développée par Järvinen et al. (2012) et Laine et al. (2012) et appliquée au système de prévision d'ensemble du CEPMMT par Ollinaho et al. (2013).

Premiers tests de perturbation des paramètres à Météo France

A Météo France, Wimmer (2021) et Wimmer et al. (2022) ont étudié l'impact des perturbations de 21 paramètres du modèle AROME sur le système de prévision d'ensemble associé, PEARO. Les 21 paramètres ainsi que leur intervalle de valeurs acceptables ont été définis avec l'aide d'experts des différentes paramétrisations physiques. Une analyse de sensibilité a ensuite été menée avec deux méthodes (méthode de Morris et méthode des indices de Sobol') pour étudier l'impact de ces différents paramètres sur plusieurs variables (scores de performance) du modèle AROME, pendant trois périodes de 31 jours de l'année 2018 et du début de l'année 2019. Les conclusions de cette étude font état de 8 paramètres ayant un impact prépondérant dans AROME, listés dans le tableau 1.1.

	Schéma	Valeur par défaut	Intervalle
RSWINHF	radiatif	1	0.6 - 1
VSIGQSAT	condensation sous-maille	0.02	0 - 0.1
XCTP	turbulence	4.65	1.035 - 22.22
XCEP	turbulence	2.11	0.225 - 4
XCED	turbulence	0.85	0.4 - 2
SLHDEPSH	diffusion	0.06	0.01 - 0.09
XFRACZ0	surface	5	2 - 10
XCMF	convection peu profonde	0.065	0 - 0.1

TABLE 1.1 – Liste des 8 paramètres d'AROME les plus influents identifiés par Wimmer et al. (2022).

Grâce aux données de l'analyse de sensibilité, plusieurs combinaisons de valeurs de paramètres, performantes du point de vue d'un score ensembliste, ont été déterminées et testées (à

la manière d'une méthode multiparamètre) dans AROME. Leur comparaison avec la méthode actuellement opérationnelle SPPT montre de meilleurs résultats pour la majorité des variables étudiées. Ces résultats constituent une piste intéressante pour l'amélioration de la représentation de l'erreur modèle dans le système de prévision d'ensemble PEARO de Météo France.

Dans le chapitre 5 de cette thèse, des méthodes de perturbation de paramètres sont employées pour construire différents ensembles de simulations avec le modèle AROME. Des perturbations des paramètres étudiés par Wimmer (2021) sont testées dans des cas d'étude simplifiés (section 5.3), et on étudie également la perturbation de deux autres paramètres du schéma microphysique ICE3 d'AROME (section 5.1).

1.4.4 Les méthodes de perturbation des processus physiques

Les méthodes de perturbations stochastiques présentées dans cette dernière partie forment un ensemble moins homogène dans la littérature que les méthodes décrites précédemment. On fait ici le choix de regrouper sous le terme de "méthodes de perturbation des processus" ou "perturbations orientées processus" (POP) les méthodes ayant les caractéristiques suivantes :

- de même que pour les méthodes de perturbation des paramètres, les perturbations ciblent des processus physiques particuliers,
- les perturbations sont appliquées à des variables (pronostiques ou diagnostiques) du modèle et non à des paramètres fixes,
- dans une certaine mesure, les caractéristiques de la perturbation appliquée s'adaptent à la situation météorologique.

Parmi les études précédemment citées utilisant des méthodes de perturbation des paramètres, certaines n'appliquent pas seulement les perturbations à des paramètres fixes, mais également directement à des variables diagnostiques voire pronostiques de leur modèle. Par exemple, des variables diagnostiques comme le taux d'entraînement, l'humidité spécifique de l'environnement dans le schéma de couche limite, ou des longueurs de rugosité sont perturbées dans Kalina et al. (2021). Dans McTaggart-Cowan et al. (2022a), le rayon effectif des gouttelettes nuageuses ou la taille du thermique à la base du nuage dans le schéma de convection profonde sont d'autres exemples de variables diagnostiques perturbées. Ces perturbations ne seront pas qualifiées de "perturbations orientées processus" dans cette thèse car même si les perturbations sont appliquées directement à des variables, leurs caractéristiques sont spécifiées de manière externe au modèle et ne s'activent pas selon des critères physiques.

Une différence importante entre les méthodes de perturbation de paramètre et les méthodes POP, telles que définies ici, est que ces dernières tentent de rendre compte de défaillances structurelles dans les schémas de paramétrisation. Elles repartent souvent des considérations développées par Palmer (2001) et Palmer et al. (2005) sur le problème du lien déterministe entre l'état résolu d'un modèle et son état sous-maille, dont les schémas de paramétrisation classiques font l'hypothèse, mais le traite à un niveau plus précis que la méthode SPPT en examinant individuellement les schémas de paramétrisation. De ce fait, les méthodes POP sont propres à chaque schéma, et ne se ressemblent pas forcément (bien que leur logique soit similaire). Ceci est également une différence importante avec la méthode de perturbation des paramètres, dont le principe peut être étendu à tous les schémas de paramétrisation physique en trouvant les bons paramètres à perturber.

Un premier exemple de méthode POP est le schéma de convection stochastique de Plant et Craig (2008) issu des travaux de Cohen et Craig (2006) et Craig et Cohen (2006). Une description détaillée est donnée dans la partie 2.2.2, mais on présente ici succinctement la méthode. Un schéma de convection a pour but de représenter l'effet des mouvements ascendants convectifs (appelés *thermiques* ou *plumes convectives*) sous-mailles, et des nuages associés, sur l'état moyen de chaque maille du modèle. Les schémas de convection déterministes classiques reposent sur l'hypothèse d'un équilibre entre des conditions de grande échelle (résolues) et la réponse en terme de convection sous-maille (Arakawa et Schubert 1974). Craig et Cohen (2006) soulignent que de tels schémas représentent l'effet *moyen* d'une population de nuages sous-maille, or, ces nuages étant présents en nombre limité et de façon apparemment aléatoire, il existe nécessairement une fluctuation statistique autour de l'effet moyen. Cette fluctuation est d'autant plus importante que leur nombre est limité, ce qui est le cas lorsque la résolution du modèle augmente (et donc que la taille d'une maille diminue), comme cela est montré sur la figure 1 de Plant et Craig (2008). Pour représenter cette variabilité, Plant et Craig (2008) proposent de perturber la condition de fermeture en flux de masse du schéma de convection de Kain-Fritsch (Kain et Fritsch 1990). Leur perturbation est construite en générant un nombre aléatoire de nuages dans chaque maille de modèle, dont la distribution dépend des conditions de grande échelle. Le flux de masse total sur une maille est alors calculé en sommant la contribution de tous les nuages sous-maille générés. Un schéma similaire est proposé pour la paramétrisation de la convection peu profonde par Sakradzija et al. (2015), lui aussi décrit plus en détails dans la partie 2.2.3.

Le schéma de convection stochastique de Plant et Craig (2008), comme celui de Sakradzija et al. (2015), consiste à perturber la fermeture en flux de masse d'un schéma de convection déterministe. In fine, ceci permet, comme avec toutes les méthodes de perturbations stochastiques, de perturber l'évolution du système. Néanmoins, par rapport aux autres méthodes stochastiques, il existe une différence fondamentale dans le dimensionnement des perturbations, et plus précisément dans le réglage de leur amplitude et de leur corrélation spatiale et temporelle. Pour illustrer cette différence, comparons par exemple le schéma de convection stochastique de Sakradzija et al. (2015) et la perturbation du paramètre XCMF d'AROME selon une méthode RP ou SPP. XCMF fait partie des 8 paramètres les plus influents identifiés par Wimmer et al. (2022) et est un coefficient intervenant dans le calcul du flux de masse en surface (qui est la condition de fermeture du schéma de convection peu profonde d'AROME). Avec les deux méthodes on peut donc perturber le schéma de paramétrisation au même niveau, mais les caractéristiques des perturbations sont différentes :

- dans Sakradzija et al. (2015), la condition de fermeture en flux de masse correspond à la somme des contributions d'un nombre aléatoire de nuages, générés à chaque pas de temps dans chaque maille du modèle. Le taux de génération de nouveaux nuages, ainsi que leurs caractéristiques individuelles en terme de flux de masse sont réglés à l'aide de conditions de grande échelle et de réglages ajustés d'après les résultats de simulations à haute résolution. L'amplitude des perturbations est donc indirectement liée à l'état résolu du modèle à travers les conditions de grande échelle. La corrélation horizontale des perturbations est liée à l'utilisation d'un domaine plus grand qu'une maille pour déterminer les conditions de grande échelle. La corrélation temporelle provient de la durée de vie des nuages qui sont gardés en mémoire d'un pas de temps à l'autre du modèle ;
- si l'on perturbe XCMF avec une approche RP, alors les perturbations sont appliquées de façon uniforme sur tout le domaine (la longueur de corrélation spatiale des perturbations est infinie). Avec une méthode SPP, le champ ne sera pas uniforme et une corrélation sera

prescrite, par exemple 200 km dans Frogner et al. (2022). De même, dans les deux cas une corrélation temporelle est prescrite, par exemple 12h dans Frogner et al. (2022). L'amplitude des perturbations est quant à elle réglée par l'intervalle des valeurs "physiquement acceptables" défini sur la base de recommandations d'experts (voir le tableau 1.1).

On voit ainsi que dans le cas des méthodes de perturbation de paramètres, les caractéristiques des perturbations sont réglées de façon arbitraire, souvent pour optimiser leur impact (Ollinaho et al. 2017). Avec une méthode POP au contraire, les caractéristiques des perturbations dépendent du flux résolu, ainsi que d'arguments physiques : par exemple dans Sakradzija et al. (2015), la corrélation horizontale provient de la détermination de conditions de grande échelle sur un domaine plus large que la maille, suffisamment large pour que l'hypothèse d'équilibre d'Arakawa et Schubert (1974) soit considérée comme valide par les auteurs.

Les schémas de convection stochastique de Plant et Craig (2008) et Sakradzija et al. (2015) ne sont pas les seuls exemples de méthodes POP dans la littérature. Dans le tableau 1.2 sont listées de façon non exhaustive plusieurs méthodes de perturbation que l'on pourrait classer parmi les méthodes POP, et que l'on décrit brièvement.

Nom de la méthode	Processus affecté	Référence(s)
Physically-based Stochastic Perturbations (PSP)	turbulence	Kober et Craig 2016 ; Hirt et al. 2019
Conditional Markov chain parametrization	shallow convection	Dorrestijn et al. 2013
Cellular Automata (CA) scheme	deep convection	Bengtsson et al. 2013, 2021
Stochastic triggering	deep convection	Rochetin et al. 2014
Stochastic Eddy-Diffusivity/Mass-Flux (EDMF) parametrization	turbulence & shallow convection	Sušelj et al. 2013

TABLE 1.2 – Liste de méthodes POP affectant la paramétrisation de différents processus sous-maille. Les méthodes indiquées en gris ont été développées pour être utilisées dans un modèle déterministe, mais leur principe pourrait être utilisé pour générer un ensemble.

Physically-based Stochastic Perturbations (PSP)

Le schéma de Kober et Craig (2016) vise à représenter la variabilité sous-maille des tourbillons turbulents. De la même façon que pour les thermiques, les tourbillons peuvent être considérés comme des événements aléatoires dans une maille, présents en nombre fini. La paramétrisation de leur effet moyen ne représente donc pas suffisamment la variabilité de leur impact sur le flux résolu. Kober et Craig (2016) proposent donc de perturber les tendances de température, d'humidité et de vent d'un schéma de paramétrisation de la turbulence en ajoutant un terme aléatoire, dont l'amplitude dépend de la variance sous-maille de ces variables. Cette variance sous-maille est diagnostiquée par le schéma de turbulence. Ainsi, en étant réglée d'après la variabilité sous-maille présumée par le schéma de turbulence, la perturbation s'adapte à la situation météorologique. Ce schéma est décrit de façon plus détaillée dans la partie 2.1.4.

Conditional Markov chain parametrization

Contrairement à la méthode PSP, la méthode proposée par Dorrestijn et al. (2013) n'utilise pas une information déjà présente dans le schéma de paramétrisation (la variance sous-maille pour PSP), mais s'appuie sur des simulations à haute résolution (LES). L'idée est que les LES résolvent explicitement la convection peu profonde et peuvent donc permettre de construire (en appliquant une méthode de coarse-graining) une paramétrisation adéquate. Le schéma de Dorrestijn et al. (2013) consiste ainsi à sélectionner de façon aléatoire, à l'aide de chaînes de Markov, un état sous-maille (des profils de flux turbulents de température et d'humidité) parmi un ensemble d'états sous-maille possibles, dérivé des LES.

Cellular Automata (CA) scheme

Bengtsson et al. (2013) proposent de représenter la variabilité sous-maille de la convection profonde à l'aide d'un automate cellulaire. Ce dernier consiste en une grille, plus fine que celle du modèle météorologique, dont chaque cellule peut, dans la version la plus simple, avoir deux états – 0 ou 1 – qui évoluent selon l'état des cellules voisines avec des règles simples. L'initialisation de nouvelles cellules "actives" (état 1) dépend de l'état résolu du modèle (à travers une condition sur la CAPE) et les règles d'évolution peuvent aussi être modifiées pour prendre en compte certains champs résolus (le vent par exemple). L'automate cellulaire permet ainsi d'obtenir un champ 2D auto-organisé de cellules dans différents états. Ce champ peut alors être couplé à un schéma de paramétrisation de la convection. Dans Bengtsson et al. (2013), le champ est utilisé pour perturber la taille de la plume convective paramétrée par le modèle météorologique dans chacune de ses mailles. Une propriété intéressante de l'automate cellulaire est sa capacité d'auto-organisation, et le fait de modéliser des interactions horizontales, entre les mailles du modèle (ce que ne prévoient pas les schémas de paramétrisation classiques qui agissent indépendamment dans chaque colonne).

Stochastic triggering

Pour accroître la variabilité de la convection profonde dans les simulations climatiques, Rochetin et al. (2014) proposent un critère stochastique de déclenchement de la convection profonde. Dans le critère de déclenchement déterministe, l'énergie disponible pour l'élévation d'une masse d'air (ALE pour Available Lifting Energy) doit dépasser l'énergie d'inhibition de la convection (CIN pour Convective Inhibition). La ALE est déterminée par la vitesse d'ascendance maximale des plumes convectives du schéma de paramétrisation de la convection peu profonde. La formulation de Rochetin et al. (2014) consiste à reformuler le calcul de la ALE, en s'appuyant sur des simulations LES, et à ajouter un critère stochastique : si la ALE devient supérieure à la CIN, la convection profonde est déclenchée à condition qu'un tirage aléatoire entre 0 et 1 dépasse une certaine probabilité, définie à partir de la distribution de taille des plumes convectives peu profondes.

Stochastic EDMF parametrization

Sušelj et al. (2013) proposent de rendre stochastique deux aspects de la paramétrisation de la convection peu profonde dans le schéma de paramétrisation EDMF (Eddy Diffusivity/Mass Flux, Siebesma et al. 2007) : les propriétés thermodynamiques et de vitesse verticale à la base des nuages sont tirées aléatoirement selon une loi normale multidimensionnelle et l'entraînement d'air extérieur dans les nuages est également généré aléatoirement avec une distribution

de Poisson.

Parce qu'elles répondent à certaines limites conceptuelles des schémas de paramétrisation, les méthodes POP peuvent être considérées comme un premier pas vers des schémas de paramétrisation stochastiques, c'est-à-dire conçus dès le départ en intégrant des effets aléatoires pour modéliser les processus physiques de fine échelle. Dans le rapport de Plant et al. (2022), l'intérêt de telles méthodes est souligné, notamment pour leur justification physique, mais des réserves sont émises quant à leur efficacité à produire en pratique des ensembles suffisamment dispersifs. En effet, plusieurs aspects pourraient limiter l'impact des méthodes POP. Tout d'abord, les sources d'incertitude qu'elles représentent sont en général assez précises : par exemple, les schémas de convection stochastique cités précédemment ne perturbent qu'une condition de fermeture ; ils ne considèrent pas les autres incertitudes possibles dans la formulation de la paramétrisation (représentation des thermiques, entraînement et déentraînement, etc). Les méthodes POP ne sauraient donc être utilisées seules pour représenter toutes les erreurs modèle. Par ailleurs, un autre problème des méthodes POP est que les perturbations sont souvent appliquées à l'échelle des processus paramétrés, donc à des échelles spatiales et temporelles restreintes (par exemple quelques kilomètres et 10 minutes pour la méthode PSP). Ces perturbations de petite échelle peuvent croître et affecter les plus larges échelles mais cela prend un certain temps : on peut donc se demander par exemple si la durée des prévisions d'un LAM est suffisamment longue pour que ce type de perturbations ait le temps de se développer. De même, dans un LAM, leur capacité à se développer et affecter significativement le flux résolu peut dépendre du temps que passe une masse d'air donnée dans le domaine : les méthodes POP pourraient donc avoir un impact réduit en cas de forçage synoptique important.

Ainsi, les méthodes POP possèdent un certain avantage conceptuel, mais leur usage en pratique dans les systèmes de prévision d'ensemble est pour le moment encore un défi (Ahlgriim et al. 2022 ; Bengtsson 2022). L'enjeu principal de cette thèse est d'étudier l'apport potentiel des méthodes POP dans le cadre de la prévision d'ensemble basée sur un LAM (le modèle AROME de Météo France). Pour cela, on s'intéresse à deux méthodes POP parmi celles précédemment citées : PSP et le schéma de convection stochastique de Sakradzija et al. (2015). On étudie notamment leur impact sur le modèle AROME et leur capacité à produire de la dispersion entre les membres d'un ensemble, en comparaison avec d'autres méthodes de perturbations stochastiques (SPPT, perturbation des paramètres).

1.5 Représentation de l'incertitude dans les modèles de Météo France

Météo France utilise actuellement deux systèmes de prévision d'ensemble dans sa production opérationnelle. L'ensemble PEARP est construit à partir du modèle global ARPEGE (Courtier et al. 1991) et l'ensemble PEARO à partir du modèle à aire limitée AROME (Seity et al. 2011 ; Brousseau et al. 2016). Ces deux ensembles sont construits pour représenter de multiples sources d'incertitudes : incertitudes des conditions initiales, de surface, incertitudes modèle, ainsi que l'incertitude des conditions latérales pour l'ensemble PEARO.

Le modèle AROME et l'ensemble PEARO ont été utilisés pour l'ensemble des travaux de thèse décrits dans les chapitres 3 à 6 de ce manuscrit.

1.5.1 Les ensembles d'assimilation AEARO et AEARP

L'incertitude des conditions initiales dans les ensembles PEARP et PEARO sont représentées à l'aide de deux ensembles d'assimilation, AEARP (pour Assimilation d'Ensemble ARPège) et AEARO (pour Assimilation d'Ensemble AROme).

AEARP est un système de 50 membres, lancé opérationnellement toutes les 6 heures. L'assimilation de données est réalisée avec une méthode 4DVar, identique à celle utilisée pour le modèle déterministe ARPEGE. Les 50 membres sont obtenus à partir de 50 ébauches provenant du cycle AEARP précédent, et en perturbant aléatoirement les observations assimilées.

AEARO est un système comprenant 25 membres, et est lancé toutes les heures. De même que pour AEARP, les analyses d'un cycle servent d'ébauches au cycle suivant, et les observations à assimiler sont perturbées. De plus, les conditions latérales des membres de l'AEARO sont fournies en sélectionnant des membres de l'AEARP. AEARO intègre également une représentation de l'erreur modèle : les membres sont perturbés lors de leur intégration avec la méthode SPPT.

Pour compenser le surcoût de calcul lié au nombre important de membres dans ces deux systèmes, des versions plus économiques des deux modèles ARPEGE et AROME sont utilisées : dans l'AEARP, une version à basse résolution d'ARPEGE est utilisée, et dans AEARO, une version hydrostatique d'AROME à 3.25 km de résolution horizontale est utilisée.

1.5.2 Les systèmes de prévision d'ensemble PEARP, PEARO

On décrit ici les systèmes PEARP et PEARO dans leur version opérationnelle suite à la mise à jour de la chaîne de PNT de Météo France du 29/06/2022.

PEARP est un système de prévision d'ensemble à 35 membres (dont 1 membre contrôle non perturbé), basé sur le modèle global ARPEGE. Il est lancé 4 fois par jour avec les réseaux suivants : 00, 06, 12 et 18 UTC (temps universel coordonné), et pour chaque réseau 102 heures de prévisions sont produites. La résolution du système est la même que pour le modèle déterministe : une grille horizontale étirée ayant une résolution d'environ 5 km sur la France et 24 km aux antipodes, avec 105 niveaux verticaux. L'initialisation des 35 membres de la PEARP est réalisée en tirant aléatoirement 35 analyses parmi les 50 produites par l'AEARP et en les combinant à des perturbations issues du calcul de vecteurs singuliers sur 7 zones géographiques. Avant la mise à jour de juin 2022, les erreurs modèles dans la PEARP étaient représentées par une approche multiphysique : une sélection aléatoire parmi 10 configurations différentes de schémas de paramétrisation physique était effectuée pour chaque membre. Depuis juin 2022, cette approche a été remplacée par une méthode de perturbation de paramètres, tout en conservant néanmoins la possibilité d'alterner entre deux schémas de paramétrisation de la convection profonde (le schéma "Tiedke-Bechtold", Tiedtke 1989 ; Bechtold et al. 2014, et le schéma PCMT, Piriou et al. 2007).

PEARO est un système de prévision d'ensemble à 17 membres (dont 1 membre contrôle non perturbé), basé sur le modèle à aire limitée AROME. Il est également lancé 4 fois par jour, avec les réseaux suivants : 03, 09, 15, 21 UTC. La résolution du système est identique au modèle déterministe : 1.3 km de résolution horizontale et 90 niveaux verticaux. L'initialisation des membres de la PEARO est issue d'un tirage aléatoire parmi les 25 membres de l'AEARO.

L'incertitude des conditions de bord du modèle est représentée en sélectionnant des membres de la PEARP par une méthode d'échantillonnage automatique (Bouttier et Raynaud 2018) pour les conditions latérales, et en perturbant des paramètres et des variables de l'analyse de surface issue du modèle déterministe (Bouttier et al. 2015). Concernant la représentation des erreurs modèle, la méthode actuellement opérationnelle est la méthode SPPT.

La figure 1.6 résume de façon simplifiée les interactions entre les différents systèmes.

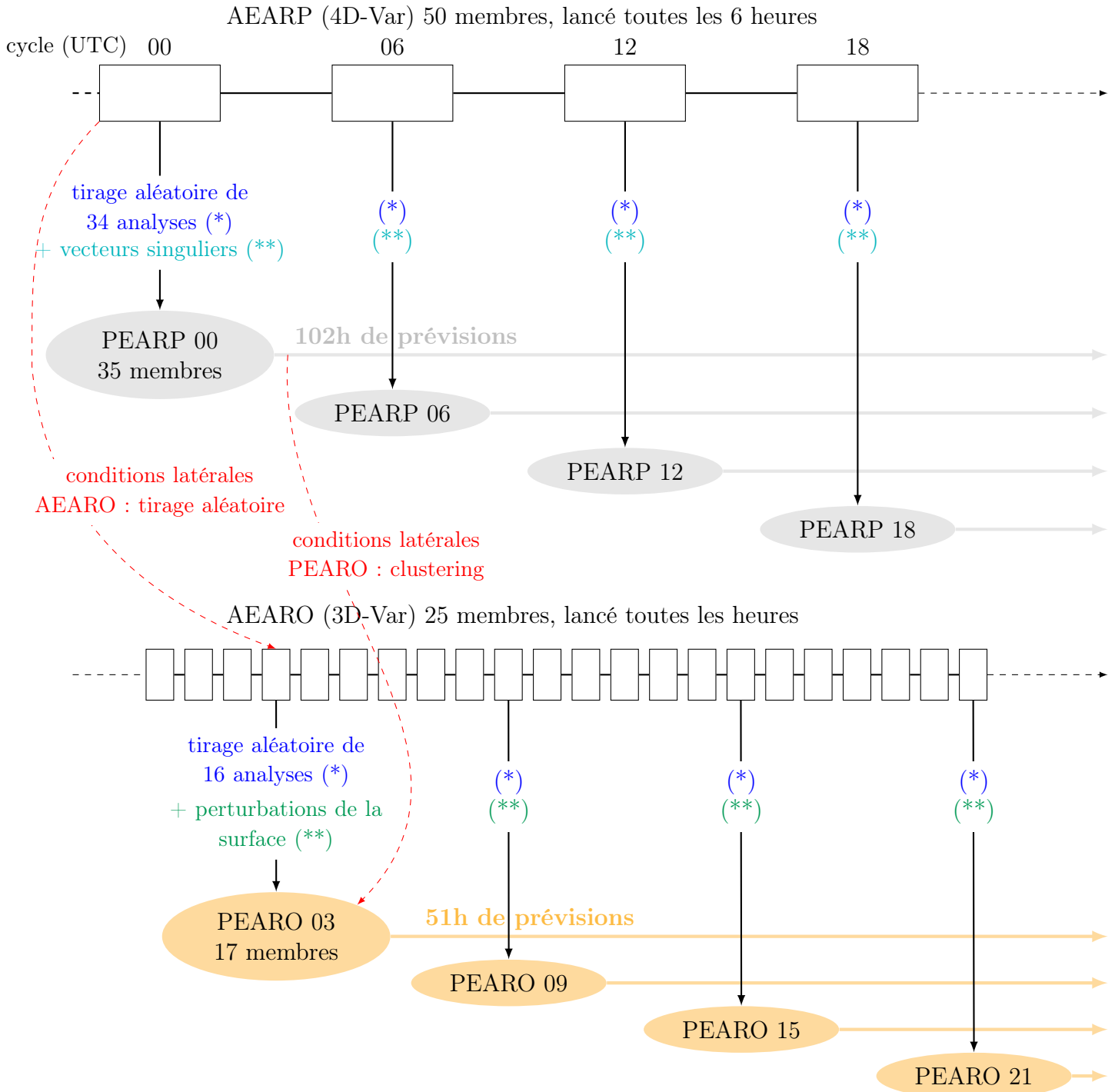


FIGURE 1.6 – Schéma simplifié des interactions entre les systèmes d'assimilation AEARP, AEARO et les systèmes de production PEARP et PEARO.

Chapitre 2

Paramétrisations physiques et stochasticité

Dans ce chapitre, les principaux schémas de paramétrisation du modèle AROME sont présentés : le schéma de turbulence et de convection peu profonde, le schéma de condensation sous-maille, le schéma microphysique et le schéma radiatif. AROME peut également faire appel à un schéma de surface et un schéma de chimie atmosphérique, mais ces derniers n'ont pas été étudiés dans cette thèse et ne sont donc pas décrits ici. Dans les sections suivantes, pour chaque schéma de paramétrisation, on essaiera de souligner certaines sources d'incertitude qui peuvent être identifiées en lien avec leur formulation, et, si elles existent, les méthodes de perturbation qui ont été proposées dans la littérature pour les représenter.

2.1 Paramétrisation de la turbulence locale

2.1.1 Généralités sur la turbulence

Les équations de la dynamique des fluides utilisées pour décrire l'évolution des masses d'air dans l'atmosphère (voir le chapitre 1) sont valables à l'échelle du continuum, c'est-à-dire l'échelle minimale à partir de laquelle un fluide (l'air) peut être considéré comme un milieu continu. Or, la résolution numérique de ces équations dans les modèles atmosphériques nécessite de les discrétiser : on utilise une grille spatio-temporelle, bien supérieure à l'échelle du continuum, et chaque variable du modèle en un point de la grille est censée représenter la moyenne sur un petit domaine spatial (la maille du modèle) et temporel (le pas de temps du modèle). Une conséquence directe de cette discrétisation est donc qu'une partie des mouvements n'est pas résolue.

Une façon de mettre en évidence la part des mouvements non résolus est d'appliquer la décomposition de Reynolds aux équations du mouvement (Reynolds 1895). Pour cela, on considère une variable x comme la somme de sa moyenne statistique \bar{x} et de l'écart à cette moyenne x' (que l'on peut appeler aussi *fluctuation* autour de la moyenne) :

$$x = \bar{x} + x'$$

Dans le cadre de la décomposition de Reynolds, l'opérateur moyenne doit respecter un certain nombre de critères, dont la conséquence est que la moyenne des fluctuations est nulle :

$$\overline{x'} = \overline{x - \bar{x}} = \bar{x} - \bar{\bar{x}} = 0$$

Un exemple de moyenne respectant les critères de Reynolds est la moyenne d'ensemble, c'est-à-dire la moyenne d'un nombre infini d'expériences indépendantes. En pratique néanmoins, on considère le plus souvent des moyennes spatiales et/ou temporelles.

Considérons le vecteur vitesse \vec{V} du fluide et l'équation de bilan de la quantité de mouvement appliquée à une particule de fluide atmosphérique dans le référentiel local :

$$\frac{\partial \vec{V}}{\partial t} + (\vec{V} \cdot \nabla) \vec{V} = -\frac{1}{\rho} \nabla P + \vec{g} + \nu \Delta \vec{V} - 2\vec{\Omega} \wedge \vec{V}, \quad (2.1)$$

avec $\vec{\Omega}$ le vecteur rotation de la Terre. Pour une composante u_i ($i \in \{1, 2, 3\}$) de \vec{V} , dans un système de coordonnées cartésiennes, cette équation peut s'écrire :

$$\frac{\partial u_i}{\partial t} + u_\alpha \frac{\partial u_i}{\partial x_\alpha} = -\frac{1}{\rho} \frac{\partial P}{\partial x_i} - \delta_{i3} g + \nu \frac{\partial^2 u_i}{\partial x_\alpha \partial x_\alpha} - 2\epsilon_{i,\alpha,\beta} \Omega_\alpha u_\beta, \quad (2.2)$$

où $\alpha \in \{1, 2, 3\}$, $\beta \in \{1, 2, 3\}$, δ_{ij} est le symbole de Kronecker, $\epsilon_{i,\alpha,\beta}$ est le symbole de Levi-Civita, et où l'on applique la convention de sommation d'Einstein pour les sommes, c'est-à-dire que le symbole \sum est sous-entendu dès lors qu'un indice est répété dans un même terme, par exemple :

$$u_\alpha \frac{\partial u_i}{\partial x_\alpha} \quad \text{signifie} \quad \sum_{\alpha=1}^3 u_\alpha \frac{\partial u_i}{\partial x_\alpha}$$

Si l'on considère le fluide incompressible, alors le terme $u_\alpha \partial u_i / \partial x_\alpha$ dans l'équation 2.2 peut être remplacé par $\partial(u_i u_\alpha) / \partial x_\alpha$. Si, de plus, on ne tient pas compte des fluctuations de la masse volumique ($\rho = \bar{\rho}$) – ce qui peut se justifier dans la couche limite atmosphérique à laquelle on s'intéresse essentiellement dans cette thèse –, et que l'on applique ensuite la moyenne de Reynolds à l'équation 2.2, on obtient la même équation avec des variables moyennes, excepté pour les termes de transport $\partial(u_i u_\alpha) / \partial x_\alpha$ qui sont non linéaires et qui se développent ainsi :

$$\frac{\overline{\partial u_i u_\alpha}}{\partial x_\alpha} = \frac{\partial (\overline{u_i u_\alpha})}{\partial x_\alpha} = \frac{\partial \overline{u_i} \overline{u_\alpha}}{\partial x_\alpha} + \frac{\partial \overline{u'_i u'_\alpha}}{\partial x_\alpha},$$

étant donné que

$$\overline{u_i u_\alpha} = \overline{(u_i + u'_i)(u_\alpha + u'_\alpha)} = \overline{u_i u_\alpha} + \overline{u_i u'_\alpha} + \overline{u'_i u_\alpha} + \overline{u'_i u'_\alpha} = \overline{u_i} \overline{u_\alpha} + \overline{u'_i u'_\alpha}$$

car d'après les propriétés de la moyenne de Reynolds : $\overline{u_i u'_\alpha} = \overline{u'_i u_\alpha} = 0$.

L'équation du mouvement moyennée avec les hypothèses précédentes est donc :

$$\frac{\partial \overline{u_i}}{\partial t} + \frac{\partial \overline{u_i} \overline{u_\alpha}}{\partial x_\alpha} = -\frac{1}{\bar{\rho}} \frac{\partial \overline{P}}{\partial x_i} - \delta_{i3} g + \nu \frac{\partial^2 \overline{u_i}}{\partial x_\alpha^2} - 2\epsilon_{i,\alpha,\beta} \Omega_\alpha \overline{u_\beta} - \frac{\partial \overline{u'_i u'_\alpha}}{\partial x_\alpha} \quad (2.3)$$

Avec un modèle météorologique, on résout numériquement l'équation du mouvement 2.1 avec des variables moyennes (les variables discrétisées sur la grille du modèle). En faisant cela, une partie des mouvements n'est donc pas prise en compte : c'est le cas des termes bleus de l'équation 2.3. Ces mouvements, dus à l'existence de fluctuations autour de la moyenne et de non linéarités dans l'équation du mouvement, sont appelés *mouvements turbulents*. Les termes $\partial \overline{u'_i u'_\alpha} / \partial x_\alpha$ sont des termes de transport turbulent et $\overline{u'_i u'_\alpha}$ sont appelés les *flux turbulents*. La décomposition de Reynolds s'applique également à l'équation thermodynamique et l'équation

d'évolution de la quantité de vapeur d'eau. Ces équations présentent elles aussi des termes d'advection non linéaires, ce qui fait apparaître les termes de transport turbulent de température potentielle et d'humidité suivants : $\overline{\partial\theta'u'_\alpha/\partial x_\alpha}$ et $\overline{\partial q'u'_\alpha/\partial x_\alpha}$, et les flux turbulents associés $\overline{\theta'u'_\alpha}$ et $\overline{q'u'_\alpha}$.

L'ensemble des équations de type 2.3 forme le système de Reynolds. Les flux turbulents $\overline{X'u'_\alpha}$ ($X = u_1, u_2, u_3, \theta, q$) sont des inconnues, et le système contient donc plus d'inconnues que d'équations. Il est possible d'écrire les équations d'évolution des flux turbulents, mais celles-ci font alors apparaître des termes d'ordre 3, du type $\overline{X'u'_\alpha u'_\beta}$, qui sont eux aussi inconnus. De façon générale, du fait des non-linéarités du système, les équations d'évolution des termes turbulents font toujours intervenir des termes d'ordre supérieur, de sorte qu'il n'est pas possible d'obtenir un système d'équations fermé. Il est donc nécessaire d'avoir recours à des hypothèses dites de fermeture, qui permettent de formuler les termes turbulents en fonction de variables connues. Ceci est le rôle principal des schémas de paramétrisation de la turbulence.

2.1.2 Approche EDMF pour la paramétrisation de la turbulence

Le transport turbulent représenté par les termes $\overline{\partial X'u'_\alpha/\partial x_\alpha}$ correspond à une large catégorie de mouvements d'échelle inférieure à la maille, qui ne peuvent pas tous être représentés par un même schéma de paramétrisation, étant données leurs caractéristiques différentes. On peut en effet distinguer deux grands types de mouvements turbulents (Siebesma et Cuijpers 1995; Soares et al. 2004) : la turbulence "non locale" (aussi appelée turbulence organisée) qui correspond aux ascendances convectives transportant des parcelles d'air verticalement, parfois jusqu'au sommet de la couche limite, et la turbulence dite "locale" qui correspond à tous les autres mouvements tourbillonnaires de tailles très variées et en apparence plus désordonnés.

Une paramétrisation classique de la turbulence locale utilise l'analogie avec la diffusion moléculaire (Lilly 1967; Deardorff 1972b; Louis 1979) : les flux turbulents sont considérés proportionnels aux gradients des variables moyennes, selon

$$\overline{X'u'_\alpha} = -K_X \frac{\partial \overline{X}}{\partial x_\alpha}, \quad (2.4)$$

où K_X est un coefficient d'échange turbulent. De même que pour la diffusion moléculaire, l'équation 2.4 prévoit un transport de X des zones de fortes valeurs de X vers les zones où X est plus faible, c'est-à-dire que le flux est positif lorsque le gradient est négatif (K_X est un terme positif). L'équation 2.4 est en général qualifiée de *fermeture en K-gradient*. Dans le cas d'un modèle à haute résolution, les flux turbulents dans les trois directions de l'espace peuvent être estimés. Dans les modèles où la résolution horizontale est plus basse, on considère en général seulement le terme de transport vertical :

$$\frac{\partial \overline{X'u'_\alpha}}{\partial x_\alpha} \approx \frac{\partial \overline{X'u'_3}}{\partial z}, \quad \text{avec} \quad \overline{X'u'_3} = -K_X \frac{\partial \overline{X}}{\partial z},$$

où $w = u_3$ est la composante verticale du vecteur vitesse \vec{V} .

L'équation 2.4 est qualifiée de fermeture d'ordre 1 : l'hypothèse de fermeture est appliquée directement aux flux turbulents $\overline{X'u'_\alpha}$. Il existe néanmoins d'autres types de fermetures selon les termes auxquels l'hypothèse est appliquée. Par exemple, si les termes $\overline{X'u'_\alpha}$ sont calculés avec leurs équations d'évolution et que l'hypothèse de fermeture s'applique aux termes $\overline{X'u'_\alpha u'_\beta}$,

alors on parlera de fermeture d'ordre 2, etc.

L'approche diffusive que l'on vient de présenter ne permet pas de représenter de façon satisfaisante tous les mouvements turbulents sous-maille. Par exemple, la proportionnalité avec le gradient dans l'équation 2.4 implique que le flux est dirigé selon un gradient négatif. Or, dans l'atmosphère, certains mouvements sont à "contre-gradient" : il est par exemple possible d'avoir un transport de chaleur vers le haut (un flux positif) même si le gradient de température est positif, grâce à des mouvements ascendants convectifs ayant une vitesse verticale suffisante (Deardorff 1972a). De façon générale, l'approche diffusive n'est pas adaptée aux mouvements d'origine convective, qui sont des mouvements turbulents de large échelle, pour lesquels les flux turbulents sont supérieurs aux gradients verticaux (Hourdin et al. 2002). Or ceux-ci peuvent représenter dans certaines situations la majeure partie du transport turbulent (Siebesma et Cuijpers 1995). Pour ces types de mouvements, qui sont liés au phénomène de convection, des paramétrisations spécifiques ont été développées (Ooyama 1971 ; Arakawa et Schubert 1974). La plupart consistent à séparer la maille entre une certaine proportion de zones ascendantes, appelées *updrafts* (ou plumes convectives, ou encore thermiques en français), où les parcelles d'air s'élèvent, et un environnement alentour. Les ascendances sont décrites par une équation d'évolution d'un *flux de masse* M_u sur la verticale et les flux turbulents sont paramétrés par (Arakawa et Schubert 1974) :

$$\overline{X'w'} = \frac{M_u}{\rho}(X_u - X_e),$$

où X_u représente la valeur moyenne de X dans les parties ascendantes de la maille et X_e la valeur moyenne de l'environnement alentour, pour laquelle on utilise souvent l'approximation $X_e \approx \bar{X}$. Le flux de masse est défini par $M_u = \rho a_u w_u$, où a_u est la proportion de la maille occupée par les zones d'ascendance et w_u la vitesse verticale moyenne sur ces zones.

Il existe donc deux grands types de fermeture des flux turbulents sous-maille, l'une utilisant une analogie avec la diffusion, l'autre construite pour représenter les mouvements convectifs sous-maille. L'approche Eddy-Diffusivity/Mass Flux (EDMF) consiste à les unifier, en considérant les flux turbulents comme la somme de ces deux contributions :

$$\overline{X'w'} = -K_X \frac{\partial \bar{X}}{\partial z} + \frac{M_u}{\rho} (X_u - \bar{X}) \quad (2.5)$$

Cette approche a été implémentée dans le modèle de recherche Mésos-NH (Lafore et al. 1998 ; Lac et al. 2018) par Soares et al. (2004), ainsi que dans d'autres modèles (e.g. Sušelj et al. 2014 ; Han et al. 2016 ; Chen et al. 2022), et est actuellement utilisée dans AROME.

Dans la suite de cette section, on décrit la paramétrisation de la turbulence locale (la partie "Eddy-Diffusivity" d'EDMF) dans AROME, avec le schéma de Cuxart et al. (2000). La paramétrisation de la turbulence non locale d'AROME, aussi appelée *convection peu profonde*, est présentée dans la section 2.2.

2.1.3 Paramétrisation de la turbulence locale dans AROME

Le schéma de paramétrisation de la turbulence locale utilisé dans AROME est décrit par Cuxart et al. (2000) et s'appuie entre autres sur les développements de Redelsperger et Sommeria (1981). Il s'agit d'un schéma utilisant une fermeture d'ordre 1.5, c'est-à-dire un système

d'équations partant des équations d'évolution des moments d'ordre 2, mais pour lequel un certain nombre d'hypothèses simplificatrices sont faites. Concrètement, cela permet d'obtenir un ensemble d'équations diagnostiques pour les flux turbulents, mais qui font intervenir une variable pronostique : l'énergie cinétique turbulente (TKE pour Turbulent Kinetic Energy).

La TKE, notée \bar{e} , est définie par

$$\bar{e} = \frac{1}{2} \left(\overline{u'^2 + v'^2 + w'^2} \right),$$

où on note u, v, w les trois composantes du vent \vec{V} . L'évolution de \bar{e} peut être obtenue à partir des équations d'évolution des moments d'ordre 2 puisque

$$\frac{\partial \bar{e}}{\partial t} = \frac{1}{2} \left(\frac{\partial \overline{u'u'}}{\partial t} + \frac{\partial \overline{v'v'}}{\partial t} + \frac{\partial \overline{w'w'}}{\partial t} \right).$$

Les équations d'évolution des moments d'ordre 2 sont elles-même déterminées à partir des équations du mouvement, des équations du mouvement moyennées et des propriétés de la moyenne de Reynolds (voir par exemple Stull 1988). En utilisant l'approximation de Boussinesq, l'équation d'évolution de la TKE peut être écrite (Stull 1988) :

$$\begin{aligned} \frac{\partial \bar{e}}{\partial t} = & - \left(\overline{u_\alpha} \frac{\partial \bar{e}}{\partial x_\alpha} + \frac{\partial \overline{u'_\alpha e}}{\partial x_\alpha} + \frac{1}{\rho_0} \overline{u'_\alpha \frac{\partial p'}{\partial x_\alpha}} - \nu \frac{\partial^2 \bar{e}}{\partial x_\alpha^2} \right) && \text{(transport)} \\ & - \overline{u'_\alpha u'_\beta} \frac{\partial \overline{u_\beta}}{\partial x_\alpha} + \frac{g}{\theta_{v,0}} \delta_{\alpha 3} \overline{u'_\alpha \theta'_v} && \text{(production/destruction)} \\ & - \nu \left(\frac{\partial u'_\alpha}{\partial x_\beta} \right)^2 && \text{(dissipation)} \end{aligned}$$

Les termes de cette équation peuvent être classés en trois catégories : transport, création ou destruction et dissipation visqueuse. Les termes de transport regroupent, dans l'ordre :

- une advection par le vent moyen $\overline{u_\alpha} \partial \bar{e} / \partial x_\alpha$,
- un transport par les fluctuations sous-maille du vent (transport turbulent) $\partial \overline{u'_\alpha e} / \partial x_\alpha$,
- un transport par les fluctuations de pression (termes de presso-corrélations) $\frac{1}{\rho_0} \overline{u'_\alpha \partial p' / \partial x_\alpha}$, où ρ_0 est la densité d'un état de référence et p' les fluctuations de pression,
- un transport diffusif $\nu \partial^2 \bar{e} / \partial x_\alpha^2$, qui est le plus souvent négligé.

Les termes de création/destruction de TKE font intervenir deux phénomènes. D'une part le cisaillement, représenté par le terme $\overline{u'_\alpha u'_\beta} \partial \overline{u_\beta} / \partial x_\alpha$, qui est une source de TKE. D'autre part la flottabilité, représentée par le terme $\frac{g}{\theta_{v,0}} \delta_{\alpha 3} \overline{u'_\alpha \theta'_v}$, où $\theta_{v,0}$ est la température potentielle virtuelle d'un état de référence, θ'_v les fluctuations de température potentielle virtuelle, $\delta_{\alpha 3}$ le symbole de Kronecker, et qui se réduit donc à $\frac{g}{\theta_{v,0}} \overline{w' \theta'_v}$. Ce terme peut être une source ou un puits de TKE, selon le sens du déplacement vertical et celui du gradient de température. Enfin, le dernier terme est un terme de destruction par les effets visqueux : on parle de dissipation visqueuse.

L'équation pronostique de la TKE décrite ci-dessus fait intervenir des variables moyennes, des flux turbulents, mais également plusieurs termes inconnus : le transport turbulent, les presso-corrélations et la dissipation visqueuse. Dans Cuxart et al. (2000), les dérivées spatiales

des fluctuations de pression sont négligées, donc le terme de presso-corrélation de l'équation d'évolution de la TKE est négligé. Le transport turbulent et la dissipation sont en revanche paramétrés. Concernant le transport turbulent, l'approximation d'un transport diffusif (voir l'équation 2.4) peut être utilisée (e.g. Moeng 1984). Dans Redelsperger et Sommeria (1981) et Cuxart et al. (2000), la paramétrisation retenue est

$$\overline{u'_\alpha e} = -C_e L \sqrt{\bar{e}} \frac{\partial \bar{e}}{\partial x_\alpha},$$

où C_e est une constante et L est une longueur, appelée *longueur de mélange*, qui intervient dans de nombreuses équations des schémas de turbulence. Concernant la dissipation visqueuse, souvent désignée par le symbole ϵ , une formulation classique, obtenue par analyse dimensionnelle ou bien d'après les hypothèses de Kolmogorov (Kolmogorov 1941, voir par exemple Lilly 1967) est :

$$\epsilon = C_\epsilon \frac{e^{3/2}}{L_\epsilon},$$

où C_ϵ est une constante et L_ϵ est appelée *longueur de dissipation*.

Ces différentes hypothèses permettent d'aboutir à l'équation de la TKE suivante, utilisée par Cuxart et al. (2000) :

$$\frac{\partial \bar{e}}{\partial t} = - \left(\overline{u_\alpha} \frac{\partial \bar{e}}{\partial x_\alpha} + \frac{\partial}{\partial x_\alpha} \left(-C_e L \sqrt{\bar{e}} \frac{\partial \bar{e}}{\partial x_\alpha} \right) \right) - \overline{u'_\alpha u'_\beta} \frac{\partial \bar{u}_\beta}{\partial x_\alpha} + \frac{g}{\theta_{v,0}} \overline{w' \theta'_v} - C_\epsilon \frac{\bar{e}^{3/2}}{L_\epsilon}$$

Dans le schéma d'ordre 1.5 de Cuxart et al. (2000), les flux turbulents $\overline{X' u'_\alpha}$ sont calculés de façon diagnostique en partant des équations d'évolution des moments d'ordre 2 auxquelles on applique l'hypothèse que la turbulence sous-maille est en équilibre stationnaire avec les conditions au bord de la maille, ce qui permet de négliger la dérivée temporelle et le terme d'advection (cette hypothèse n'est pas faite pour l'équation de la TKE, qui reste une variable pronostique). De cette façon Cuxart et al. (2000) dérivent les équations diagnostiques des flux turbulents $\overline{u'_\alpha u'_\beta}$, $\overline{u'_\alpha \theta'}$, $\overline{u'_\alpha q'}$, ainsi que des covariances et variances suivantes : $\overline{\theta' q'}$, $\overline{u'^2_\alpha}$, $\overline{\theta'^2}$, $\overline{q'^2}$. On rappelle ici seulement les équations des variances, qui seront utilisées par la suite.

Équation diagnostique des co-variances sous-maille des composantes du vent u_i et u_j

$$\overline{u'_i u'_j} = \frac{2}{3} \delta_{ij} \bar{e} - \frac{4}{15} \frac{L}{C_m} \sqrt{\bar{e}} \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} - \frac{2}{3} \delta_{ij} \frac{\partial \bar{u}_\alpha}{\partial x_\alpha} \right), \quad (2.6)$$

et en particulier pour les variances sous-maille des composantes du vent :

$$\overline{u'^2_i} = \frac{2}{3} \bar{e} - \frac{4}{15} \frac{L}{C_m} \sqrt{\bar{e}} \left(2 \frac{\partial \bar{u}_i}{\partial x_i} - \frac{2}{3} \frac{\partial \bar{u}_\alpha}{\partial x_\alpha} \right) \quad (2.7)$$

Équation diagnostique de la variance sous-maille de la température potentielle θ

$$\overline{\theta'^2} = \frac{2}{3} \frac{L^2}{C_s C_\theta} \left(\frac{\partial \bar{\theta}}{\partial x_\alpha} \frac{\partial \bar{\theta}}{\partial x_\alpha} \right) \Phi_\alpha \quad (2.8)$$

Équation diagnostique de la variance sous-maille de l'humidité spécifique q

$$\overline{q'^2} = \frac{2}{3} \frac{L^2}{C_h C_q} \left(\frac{\partial \bar{q}}{\partial x_\alpha} \frac{\partial \bar{q}}{\partial x_\alpha} \right) \Psi_\alpha \quad (2.9)$$

Ces équations font apparaître un certain nombre de constantes ($C_m, C_s, C_\theta, C_h, C_q$), dont les valeurs sont données dans Cuxart et al. (2000), ainsi que des fonctions de stabilité Φ_α et Ψ_α qui dépendent de nombres sans dimension définis dans Redelsperger et Sommeria (1981) et qui sont différentes de 1 seulement sur la verticale.

Le schéma de Cuxart et al. (2000) est un schéma de turbulence à trois dimensions pouvant être utilisé dans les modèles LES, mais également dans des modèles de méso-échelle. Dans ce dernier cas, on considère en général uniquement les gradients verticaux dans les équations précédentes (on parle de turbulence 1D).

2.1.4 Les sources d'incertitude liées à la paramétrisation de la turbulence

Les hypothèses et approximations présentées dans les sections 2.1.1 et 2.1.3 sont en théorie autant de sources d'incertitude dans la paramétrisation des effets turbulents. Parmi elles, nous nous intéressons ici à une hypothèse fondamentale, mentionnée dans la section 2.1.1, qui est d'utiliser des moyennes spatio-temporelles dans le système d'équations de Reynolds. En effet, dans toutes les équations qui en découlent et qui sont utilisées dans les modèles météorologiques, les quantités moyennes \bar{X} sont les valeurs des variables du système résolues en chaque point de grille du modèle, qui représentent théoriquement une moyenne de X sur un petit domaine spatial (la maille du modèle) et temporel (le pas de temps du modèle). Les axiomes de Reynolds, qui servent à dériver les équations de paramétrisation de la turbulence, sont typiquement valides pour des moyennes appliquées à un grand nombre d'évènements indépendants. Pour un processus donné ayant par exemple un temps τ et une longueur ξ d'autocorrélation, les réalisations de ce processus à des instants séparés par un temps supérieur ou égal à τ et des localisations séparées par une longueur supérieure ou égale à ξ peuvent être considérées comme indépendantes. Un moyennage spatio-temporel bien choisi peut donc en théorie permettre de s'approcher d'une moyenne d'ensemble (il s'agit de l'hypothèse d'ergodicité). Néanmoins, comme ce moyennage doit être effectué sur un temps et un domaine plus grands que τ et ξ de façon à mimer une moyenne sur des évènements indépendants, selon les dimensions caractéristiques du phénomène sous-maille considéré, la taille de la maille et le pas de temps du modèle peuvent ne pas suffire. Les variables du modèle résolues en chaque point de grille ne peuvent alors pas être considérées comme des moyennes d'ensemble. Par exemple, pour des tourbillons ayant une taille de l'ordre de 1000 m, la maille du modèle AROME de 1.3 km de côté ne semble pas suffisante pour contenir un grand nombre d'évènements indépendants.

Ce problème est souligné par exemple par Leoncini et al. (2010, 2013), qui mentionnent que les moyennes spatio-temporelles ne respectent pas nécessairement les axiomes de Reynolds. Leoncini et al. (2010) argumentent notamment que pour des modèles à haute résolution, où la taille des mailles est réduite, la moyenne spatio-temporelle s'écarte d'une moyenne de Reynolds et présente des fluctuations dues à l'échantillonnage limité de phénomènes sous-maille indépendants. Ces fluctuations n'étant pas prévues par les schémas de paramétrisation, ils proposent de

les représenter à l'aide de perturbations stochastiques. Les perturbations proposées par Leoncini et al. (2010, 2013) concernent le champ de température potentielle et leurs caractéristiques sont choisies selon la taille typique des phénomènes sous-maille considérés (dans leur cas, des ascendances convectives). Ce raisonnement est prolongé et développé par Clark et al. (2021).

Le schéma stochastique de Clark et al. (2021)

Clark et al. (2021), noté dans la suite CH21, construisent un schéma de paramétrisation stochastique de la couche limite, censé représenter la variabilité due au nombre limité et aléatoire de tourbillons sous-maille dans un domaine donné (par exemple la maille d'un modèle). Ils partent des hypothèses suivantes :

- les tourbillons turbulents sont des évènements aléatoires et indépendants, qui ont une probabilité p de se produire par unité de temps et d'espace,
- la chaleur reçue par la couche limite provient d'un nombre fini d'évènements discrets.

Si on considère un domaine spatial ΔA et un intervalle de temps Δt finis, alors le nombre N de tourbillons se produisant dans ce domaine spatio-temporel suit une loi de Poisson :

$$\mathcal{P}(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \text{avec} \quad \lambda = E[N] = p\Delta A\Delta t$$

La chaleur délivrée à la couche limite par un tourbillon individuel est définie par

$$Q_t = \rho C_p \theta_t A_t h,$$

où $A_t h$ est le volume moyen occupé par un tourbillon avec A_t son aire horizontale et h la hauteur de la couche limite. θ_t est la température potentielle représentative du tourbillon et C_p est la capacité thermique de l'air à pression constante. La chaleur totale délivrée à la couche limite sur un domaine ΔA pendant l'intervalle Δt est obtenue en sommant sur tous les tourbillons qui s'y produisent :

$$Q = N Q_t$$

L'évolution moyenne de la température durant l'intervalle Δt , sur le domaine ΔA est donc :

$$\frac{d\theta}{dt} = \frac{N Q_t}{\Delta A \Delta t h \rho C_p}.$$

On peut également définir la moyenne d'ensemble de $d\theta/dt$, c'est-à-dire l'évolution moyenne de la température sur $\Delta A \times \Delta t$, moyennée sur un nombre infini de réalisations. Pour reprendre les notations de CH21, la moyenne d'ensemble est ici désignée par le symbole " $\langle \cdot \rangle$ ". Cette moyenne d'ensemble est donnée par le nombre moyen λ d'occurrence des tourbillons sur $\Delta A \times \Delta t$:

$$\frac{d\langle \theta \rangle}{dt} = \frac{\lambda Q_t}{\Delta A \Delta t h \rho C_p}.$$

Ainsi, il est possible d'écrire

$$\frac{d\theta}{dt} = \frac{N}{\lambda} \frac{d\langle \theta \rangle}{dt} = \frac{d\langle \theta \rangle}{dt} + \left(\frac{N}{\lambda} - 1 \right) \frac{d\langle \theta \rangle}{dt}.$$

Dans le modèle de couche limite choisi par CH21, l'évolution moyenne de θ est modélisée par $d\langle \theta \rangle/dt = \theta^*/\tau^*$, où θ^* est une échelle de fluctuation des températures et τ^* est défini comme

une échelle de temps de renouvellement des tourbillons. L'incrément de température durant un pas de temps δt du modèle peut donc s'écrire

$$\Delta\theta = \Delta\langle\theta\rangle + \eta\delta t\frac{\theta^*}{\tau^*},$$

où $\eta = N/\lambda - 1$ est une variable aléatoire dépendant du processus de Poisson N .

Une deuxième formulation un peu plus générale proposée par CH21 consiste à considérer les incréments de température comme résultant de la contribution d'une part des tourbillons apparus au pas de temps précédent et de la contribution des tourbillons nouvellement créés. Ceci leur permet de remplacer η par un processus autorégressif f (qui dépend lui aussi du processus de Poisson N).

Le schéma stochastique de Kober et Craig (2016)

Le schéma de perturbation stochastique de CH21, fondé sur la variabilité d'occurrence des tourbillons turbulents, présente de fortes similarités avec le schéma proposé par Kober et Craig (2016). Kober et Craig (2016), noté dans la suite KC16, partent eux aussi du constat que lorsque la résolution d'un modèle devient petite devant la taille caractéristique des phénomènes paramétrisés, les schémas déterministes classiques ne représentent pas assez leur variabilité. Bien que ne s'appuyant pas sur une modélisation aussi détaillée que celle de CH21, le schéma stochastique qu'ils proposent est très semblable. Pour une variable pronostique Φ du modèle, le schéma se résume par

$$\left(\frac{\partial\Phi}{\partial t}\right)^{\text{stoch}} = \frac{\partial\Phi}{\partial t} + \alpha \frac{l_{\text{eddy}}}{5\Delta x} \frac{1}{\tau_{\text{eddy}}} \eta \overline{\Phi'^2}^{1/2}, \quad (2.10)$$

où l_{eddy} et τ_{eddy} représentent une taille et un temps de renouvellement caractéristiques des tourbillons, $5\Delta x$ est la résolution effective du modèle, $\overline{\Phi'^2}$ est la variance sous-maille de la variable Φ et η est un champ aléatoire gaussien, dont en tout point l'espérance est nulle et l'écart-type vaut 1.

CH21 présentent une comparaison de la variance des perturbations de température potentielle de leur schéma stochastique avec celle du schéma de KC16. Pour une perturbation $\Delta\theta_{\text{pert}}$ ajoutée lors d'un pas de temps δt du modèle on a :

$$V(\Delta\theta_{\text{pert}}) = \alpha^2 \frac{l_{\text{eddy}}^2}{(5\Delta x)^2} \frac{\delta t^2}{\tau_{\text{eddy}}^2} \overline{\theta'^2} \quad (\text{KC16})$$

$$V(\Delta\theta_{\text{pert}}) = \frac{A_t}{\Delta A} \frac{\delta t^2}{\tau^{*2}} \theta^{*2} \quad (\text{CH21})$$

On retrouve bien, dans les deux formules, des termes semblables :

- une échelle des fluctuations sous-maille : θ^* dans CH21, $\overline{\theta'^2}^{1/2}$ dans KC16 ;
- une taille relative des tourbillons sous-maille par rapport à un domaine de moyennage plus grand : $A_t/\Delta A$ et $l_{\text{eddy}}^2/(5\Delta x)^2$;
- un temps caractéristique des tourbillons : τ_{eddy} et τ^* .

CH21 notent néanmoins que la taille caractéristique des tourbillons dans leur schéma, A_t , est donnée par des longueurs de décorrélation, tandis que l_{eddy} est donnée dans KC16 par une longueur de mélange. Quant au temps caractéristique, τ_{eddy} est fixé dans KC16 à 10 minutes tandis que τ^* est déterminé à partir d'une échelle de vitesse w^* et de la hauteur de la couche limite, et peut donc être variable.

Ces schémas sont construits pour représenter une source d'incertitude qui vient du fait que dans une maille de modèle, on ne moyenne pas sur un nombre infini de processus sous-maille, et donc qu'il existe une variabilité intrinsèque, due à l'échantillonnage, non prise en compte dans les schémas classiques de paramétrisation, et que l'on s'efforce de représenter par un processus stochastique. Dans ce travail de thèse, on souhaite tester une telle approche de représentation de l'erreur modèle, et la méthode de perturbation de KC16 est donc introduite dans le modèle AROME au chapitre 4 dans un cadre simplifié afin d'évaluer son impact sur des simulations de couche limite.

2.2 Paramétrisation des phénomènes convectifs

2.2.1 Principes généraux des schémas de convection

Parmi les mouvements ayant une échelle inférieure à la maille des modèles de PNT, les mouvements convectifs, c'est-à-dire les mouvements essentiellement contrôlés par les forces de flottabilité, peuvent être à l'origine d'un transport vertical de grande ampleur, et présenter une organisation qui ne correspond pas aux hypothèses de Kolmogorov et à celles des schémas en K-gradient. La modélisation décrite dans la section 2.1.3 n'est donc pas adaptée, et des paramétrisations spécifiques ont été construites pour représenter ce type de mouvements : ce sont les schémas de convection.

Le rôle d'un schéma de paramétrisation de la convection est de modéliser les interactions entre des conditions environnementales de grande échelle et un ensemble d'ascendances et de nuages convectifs potentiellement présents dans une maille du modèle. Un schéma de convection est en général actif dans un modèle météorologique dès lors que des conditions de grande échelle sont réunies pour le développement d'une activité convective. Le rôle du schéma de convection est alors de modifier l'état de chaque maille de modèle concernée vers un état ayant subi l'action de la convection. Pour cela, l'approche la plus répandue consiste à utiliser un *schéma de flux de masse*.

Les premiers travaux théoriques concernant ces schémas sont parus dans les années 1970 (Ooyama 1971 ; Yanai et al. 1973 ; Arakawa et Schubert 1974). Comme mentionné dans la section 2.1.2, l'idée principale est de considérer qu'une maille convective peut-être séparée en deux types de sous-domaines : d'un côté les ascendances convectives et de l'autre leur environnement (figure 2.1). Les ascendances sont les parties de la maille où l'air s'élève sous l'action des forces de flottabilité. Elles sont décrites notamment par une équation d'évolution verticale du flux de masse. L'environnement correspond au reste de la maille, non instable du point de vue de la flottabilité et qui comprend également les mouvements d'air descendants, appelés *subsidences*. Chaque sous-domaine possède ses propres caractéristiques de vitesse verticale, température, humidité, mais est en interaction avec l'autre à travers les phénomènes d'*entraînement* et de *déentraînement* qui correspondent à des échanges d'air entre les deux.

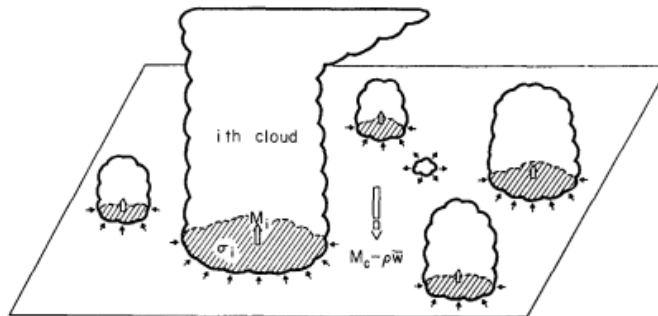


FIGURE 2.1 – Représentation de nuages convectifs dans une maille du modèle (illustration de l'article d'Arakawa et Schubert 1974). La fraction de l'espace occupée par les nuages est supposée très petite.

L'étude fondatrice d'Arakawa et Schubert (1974) présente un schéma de flux de masse complet, avec les équations d'évolution des ascendances et de leur environnement, accompagnées d'hypothèses de fermeture. La principale est l'hypothèse de *quasi-équilibre*, qui permet de déterminer le flux de masse à un niveau de référence pour fermer le système d'équations. Dans leur étude, Arakawa et Schubert (1974) définissent en effet une fonction de travail du nuage, proportionnelle à l'intégrale de la force de flottabilité dans les nuages, qui dépend du flux de masse. Cette fonction évolue dans le temps sous l'action des nuages interagissant les uns avec les autres et sous l'action des forçages environnementaux. L'hypothèse du quasi-équilibre consiste à dire que le temps d'ajustement du système lié à la convection est beaucoup plus rapide que l'échelle de temps des forçages. Ainsi le système connaît une succession de quasi-équilibres, durant lesquels la fonction de travail peut-être considérée comme stationnaire, ce qui permet alors en pratique de déterminer le flux de masse au niveau de référence.

Outre cette séparation des échelles de temps entre les phénomènes convectifs et les forçages de grande échelle qui les provoquent, les schémas de convection font aussi généralement l'hypothèse d'une séparation des échelles d'espace. Ainsi, Arakawa et Schubert (1974) commencent l'introduction de leur schéma par l'hypothèse qu'il existe un domaine de l'espace contenant un grand nombre de phénomènes convectifs mais sur lequel les conditions de forçage de grande échelle sont relativement homogènes. Autrement dit, la taille des ascendances convectives est supposée petite par rapport à l'échelle spatiale à laquelle varient les forçages, et le domaine d'intérêt (qui dans le cadre d'un modèle météorologique est une maille de sa grille de discrétisation) doit contenir un ensemble de nuages suffisamment grand pour être traité d'un point de vue statistique. La fraction de la maille occupée par les ascendances est par ailleurs supposée très petite, ce qui permet d'utiliser l'approximation que les caractéristiques de l'environnement (température humidité, vitesse verticale, etc) sont égales aux valeurs moyennes de la maille.

Même si elles ne se traduisent pas par les mêmes équations et les mêmes critères que ceux d'Arakawa et Schubert (1974), ces hypothèses de séparation d'échelles de temps et d'espace entre les phénomènes convectifs et les conditions qui les gouvernent sont présentes dans la plupart des schémas de flux de masse (e.g. Tiedtke 1989 ; Emanuel 1991 ; Bechtold et al. 2001).

2.2.2 Fluctuations autour de l'équilibre

Avec les hypothèses précédentes, les schémas de convection déterminent les effets de la convection sous-maille uniquement à partir des conditions de grande échelle (l'état explicitement résolu du système). À une condition de grande échelle donnée on associe toujours la même réponse de la convection sous-maille. Celle-ci peut être vue comme une réponse moyenne, qui n'est exacte que si l'on réalise une moyenne sur un nombre infini de phénomènes convectifs sous-maille. Or, en pratique, la maille d'un modèle est un espace fini dans lequel se trouve un nombre fini de nuages ou d'ascendances convectives. Ceci est le constat de départ de Craig et Cohen (2006) qui proposent une théorie de fluctuation autour de l'équilibre convectif.

Le schéma stochastique de Craig et Cohen (2006)

Selon Craig et Cohen (2006), même si les hypothèses de séparation d'échelle et de quasi-équilibre d'Arakawa et Schubert (1974) sont valides, le fait que les nuages soient présents en nombre limité dans une maille de modèle implique une fluctuation statistique autour de cet équilibre, qui s'apparente à un bruit d'échantillonnage. Dans leur article, Craig et Cohen (2006), noté par la suite CC06, proposent quelques bases théoriques pour modéliser ces fluctuations. Ils partent de la situation suivante :

- les hypothèses de séparation de temps et d'espace et l'hypothèse de quasi-équilibre sont vérifiées et un flux de masse total $\langle M \rangle$ peut être défini à partir de conditions de grande échelle avec un schéma de convection classique,
- les nuages peuvent être considérés comme étant indépendants les uns des autres,
- chaque nuage possède un flux de masse m qui lui est propre, et le flux de masse moyen d'un nuage individuel, $\langle m \rangle$, est indépendant des conditions de grande échelle.

Ici, comme dans la section 2.1.4, " $\langle \cdot \rangle$ " désigne une moyenne d'ensemble.

Avec les hypothèses précédentes, CC06 établissent le fait que le nombre de nuages qui contribuent à un flux de masse total M suit une loi de Poisson, et montrent que le flux de masse individuel de ces nuages suit une distribution exponentielle :

$$p(m) = \frac{1}{\langle m \rangle} e^{-m/\langle m \rangle}. \quad (2.11)$$

Ces résultats leur permettent ensuite d'établir une distribution du flux de masse total M dans une maille :

$$p(M) = \left(\frac{\langle N \rangle}{\langle m \rangle} \right)^{1/2} e^{-\langle N \rangle - \frac{M}{\langle m \rangle}} M^{-1/2} I_1 \left(2 \left(\frac{\langle N \rangle}{\langle m \rangle} M \right)^{1/2} \right), \quad (2.12)$$

avec I_1 la fonction de Bessel modifiée du premier ordre¹.

La variance de M est donnée par $2\langle M \rangle^2 / \langle N \rangle$ où $\langle N \rangle$ est le nombre moyen de nuages dans la maille. Ainsi, comme on peut le concevoir intuitivement, cette expression de la variance montre que plus l'échantillon de nuage est grand, plus la variabilité autour de la valeur moyenne $\langle M \rangle$ est réduite. À l'inverse, lorsque l'échantillon est petit, il y a davantage de chance que le flux

1. $I_1(x) = \sum_{n=0}^{\infty} \frac{1}{n!(n+1)!} \left(\frac{x}{2} \right)^{2n+1}$

de masse M sur une maille s'écarte de $\langle M \rangle$, ce qui suggère l'importance de prendre en compte cette variabilité dans les modèles où la résolution horizontale est élevée.

Les travaux théoriques de CC06 ont été utilisés par Plant et Craig (2008) pour construire un schéma de paramétrisation stochastique de la convection profonde, testé dans une version unicolonne du modèle Unified Model (UM) du Met Office. L'implémentation pratique du schéma est la suivante. Le flux de masse moyen $\langle M \rangle$ à un niveau de référence (la base des nuages) est déterminé à partir de conditions de grande échelle, qui correspondent aux variables du modèle moyennées sur un domaine assez grand pour que les hypothèses d'Arakawa et Schubert (1974) semblent justifiées. Une condition de fermeture sur la CAPE permet de calculer la valeur de $\langle M \rangle$. Le flux de masse moyen à la base d'un nuage individuel est considéré constant : $\langle m \rangle = 2.10^7$ kg.s⁻¹. Une probabilité d'occurrence durant un pas de temps δt d'un nuage de rayon r (à sa base) est donnée par Plant et Craig (2008), en faisant une hypothèse sur la relation entre le flux de masse m et le rayon d'un nuage, ainsi qu'en attribuant un temps de vie T constant ($T = 45$ min) à chaque nuage. Des nuages sont générés à chaque pas de temps du modèle et le schéma de convection de Kain-Fritsch (Kain et Fritsch 1990) est utilisé pour déterminer leurs différentes caractéristiques (profils thermodynamiques, etc). À chaque pas de temps du modèle, les nuages présents dans la maille regroupent des nuages nouvellement créés et des nuages apparus aux pas de temps précédents mais dont le temps de vie n'est pas écoulé. Les tendances du schéma de convection stochastique sont alors données par la somme des contributions de chacun de ces nuages. Les résultats obtenus par Plant et Craig (2008) avec leur paramétrisation stochastique montrent une bonne adéquation avec la théorie de CC06.

2.2.3 Remise en question des hypothèses de séparation d'échelles dans les modèles kilométriques

Les travaux de Cohen et Craig (2006), Craig et Cohen (2006) et Plant et Craig (2008) mettent en évidence l'importance des fluctuations à l'échelle de la maille d'un modèle autour d'une réponse convective théorique. Ils montrent bien que les fluctuations sont d'autant plus grandes que la maille est petite, et donc l'importance de représenter cette variabilité si la résolution d'un modèle augmente. En particulier, pour les modèles à aire limitée actuels dont la résolution s'approche du kilomètre et qui ne peuvent donc plus contenir qu'un échantillon restreint de nuages, on peut concevoir l'utilité d'un schéma de convection stochastique.

Le schéma de Plant et Craig (2008) n'est pas applicable tel quel à un modèle kilométrique, car la théorie de CC06 considère essentiellement des modèles globaux et des résolutions assez basses (par exemple entre 1000 et 64 km² dans l'étude de Plant et Craig 2008), et utilise des hypothèses qui peuvent être remises en question à l'échelle kilométrique. Néanmoins, d'autres travaux récents reprennent les principaux arguments de cette théorie et l'étendent aux modèles à haute résolution : il s'agit du schéma de convection peu profonde stochastique de Sakradzija et al. (2015)

À partir d'une résolution de l'ordre de quelques kilomètres, on considère généralement que la convection profonde est explicitement résolue par un modèle (on parle de *convection-permitting models*, e.g. Baldauf et al. 2011; Seity et al. 2011). Il existe néanmoins toujours des mouvements convectifs de moins grande ampleur qui sont d'échelle inférieure à la maille et qui doivent être paramétrés par des schémas de *convection peu profonde*. Mais à mesure que la résolution augmente, ces mouvements peuvent eux aussi finir par être résolus par le modèle. L'intervalle

entre les résolutions pour lesquelles une paramétrisation complète est pertinente et celles pour lesquelles le phénomène en question est entièrement résolu est appelé "terra incognita" (Wyngaard 2004) ou encore "zone grise" (Gerard et al. 2009), et pour les phénomènes convectifs, il concerne les résolutions de l'ordre du kilomètre jusqu'à la centaine de mètres environ (Petch 2006 ; Honnert et al. 2011 ; Sakradzija et al. 2016). Dans la zone grise, la représentation des phénomènes sous-maille est délicate car certaines hypothèses des schémas de paramétrisation classiques peuvent être remises en question (Honnert et al. 2016).

Le schéma stochastique de Sakradzija et al. (2015)

Le schéma de convection stochastique de Sakradzija et al. (2015) propose de répondre à ces différentes limites en étendant la théorie de CC06 des fluctuations autour de l'équilibre aux échelles kilométriques et ainsi à la zone grise de la convection peu profonde. Leurs constats sont les suivants. Premièrement, une maille d'échelle kilométrique contient un trop petit nombre de nuages pour que l'on puisse les considérer comme un ensemble statistiquement représentatif. Ils ne peuvent pas être représentés de façon déterministe, à partir des champs résolus, par des moyennes d'ensemble (cet argument est aussi celui de CC06). Deuxièmement, les modèles de fine résolution spatiale ont également une résolution temporelle fine. On ne peut plus considérer que le temps caractéristique de la réponse convective sous-maille est petit devant le pas de temps du modèle. Ceci implique un effet "mémoire" de la convection entre des pas de temps successifs, ce qui n'est pas prévu par les schémas de convection classiques. Enfin, du fait du manque de séparation des échelles d'espace et de temps entre les phénomènes convectifs et les champs résolus du modèle, l'hypothèse de quasi-équilibre, qui sert de condition de fermeture à la plupart des schémas de convection, n'est plus réaliste à l'échelle kilométrique. Cet aspect est déjà en partie discuté par Plant et Craig (2008) qui précisent que la condition de fermeture en flux de masse $\langle M \rangle$ devrait être calculée sur un domaine suffisamment grand pour que l'hypothèse de quasi-équilibre soit valide, c'est-à-dire sur un domaine potentiellement plus grand que la maille d'un modèle.

Le schéma de Sakradzija et al. (2015) a les caractéristiques suivantes, qui répondent aux limites que l'on vient de lister :

- il s'agit d'un schéma stochastique, où l'activité convective sur une maille va fluctuer en fonction d'un nombre aléatoire de nuages, aux caractéristiques également aléatoires ;
- le schéma possède une mémoire à travers un temps de vie des nuages et de leur caractéristiques qui évoluent temporellement ;
- la condition de fermeture est calculée sur un large domaine, où l'hypothèse de quasi-équilibre est supposée être vérifiée.

Comme CC06 et Plant et Craig (2008), Sakradzija et al. (2015) s'appuient sur deux distributions principales pour construire leur schéma de convection stochastique : la distribution du nombre de nuages et la distribution du flux de masse d'un nuage individuel. Concernant le nombre de nuages, Sakradzija et al. (2015) suivent la théorie de CC06 et considèrent que le nombre de nuages N apparaissant dans un domaine spatio-temporel fini est supposé suivre une loi de Poisson :

$$\mathcal{P}(N = n) = p(n) = \frac{(G\Delta t\Delta A)^n e^{-G\Delta t\Delta A}}{n!}, \quad (2.13)$$

où G est le taux de production de nuages par unité de temps et d'espace et $\Delta t \times \Delta A$ le domaine

spatio-temporel d'intérêt. Cet aspect du schéma est crucial car il lui permet de s'adapter à la résolution du modèle : si le domaine est grand, davantage de nuages sont générés, et inversement, ce qui se répercute sur la variabilité du flux de masse total sur le domaine.

Concernant la distribution du flux de masse d'un nuage, Sakradzija et al. (2015) s'appuient sur l'analyse de LES d'un cas de cumulus peu profonds à l'aide d'un algorithme d'identification et de suivi des nuages. Cette analyse met en évidence l'existence de deux grands types de nuages : les nuages "actifs", qui durant leur cycle de vie atteignent le niveau de convection libre, et les nuages "passifs" qui ne l'atteignent pas. Sakradzija et al. (2015) montrent que la distribution dérivée des LES du flux de masse m d'un nuage individuel, moyenné sur son cycle de vie, peut être modélisée par une loi de Weibull bimodale :

$$p(m) = \sum_{i=1,2} f_i \frac{k}{\lambda_i} \left(\frac{m}{\lambda_i} \right)^{k-1} e^{-\left(\frac{m}{\lambda_i}\right)^k}, \quad (2.14)$$

où les indices 1 et 2 correspondent aux deux groupes identifiés de nuages : f_1 indique par exemple le taux de nuages passifs et $f_2 = 1 - f_1$ le taux de nuages actifs. Cette distribution est plus générale que la distribution exponentielle dérivée par CC06 (équation 2.11), mais elle s'y ramène si $k = 1$. Sakradzija et al. (2015) argumentent que ceci vient du fait que les nuages ont tous des temps de vie différents.

Dans les équations 2.13 et 2.14, les paramètres G et λ sont les inconnues du système (les autres ont une valeur fixe déterminée par les LES). La condition de fermeture permettant de les calculer dépend des moyennes d'ensemble $\langle M \rangle$ et $\langle m \rangle$. Sakradzija et al. (2015) montrent que :

$$\langle M \rangle = G\alpha\lambda^{\beta+1}\Gamma\left(1 + \frac{1}{k} + \frac{\beta}{k}\right),$$

où α et β sont les paramètres d'une loi de puissance reliant le flux de masse m au temps de vie du nuage τ , déterminée à partir des LES : $\tau = \alpha m^\beta$. Par ailleurs, le flux de masse moyen $\langle m \rangle$ d'un nuage est donné par l'espérance de la distribution de Weibull, soit

$$\langle m \rangle = \lambda\Gamma\left(1 + \frac{1}{k}\right).$$

$\langle M \rangle$ et $\langle m \rangle$ sont les conditions de fermeture du système. $\langle M \rangle$ peut être déterminé à partir des hypothèses d'un schéma de convection classique et $\langle m \rangle$ peut être fixé à une valeur constante (Sakradzija et al. 2015) ou bien modélisé par une relation empirique (Sakradzija et Hohenegger 2017).

Le schéma de Sakradzija et al. (2015) est implémenté et testé dans le modèle ICOSahedral Nonhydrostatic model (ICON ; Zängl et al. 2015) par Sakradzija et al. (2016) et Sakradzija et Klocke (2018). On résume ici les différentes étapes du schéma tel que décrit par Sakradzija et Klocke (2018).

1. Détermination des conditions de grande échelle Les variables pronostiques du modèle sont moyennées sur un domaine suffisamment grand pour que l'hypothèse de quasi-équilibre soit considérée comme valide (figure 2.2). En pratique, ce domaine est constitué d'une maille et de ses 12 mailles adjacentes sur la grille triangulaire du modèle ICON, ce qui, pour une résolution de 1.2 km, correspond à un domaine d'environ 20 km². Le schéma de convection de

Tiedke-Bechtold (Bechtold et al. 2014) est employé pour déterminer un flux de masse total, qui est considéré être le flux de masse moyen $\langle M \rangle$. Le flux de masse moyen pour un nuage individuel $\langle m \rangle$ est calculé d'après la relation empirique de Sakradzija et Hohenegger (2017), également à l'aide de variables pronostiques moyennées sur le grand domaine.

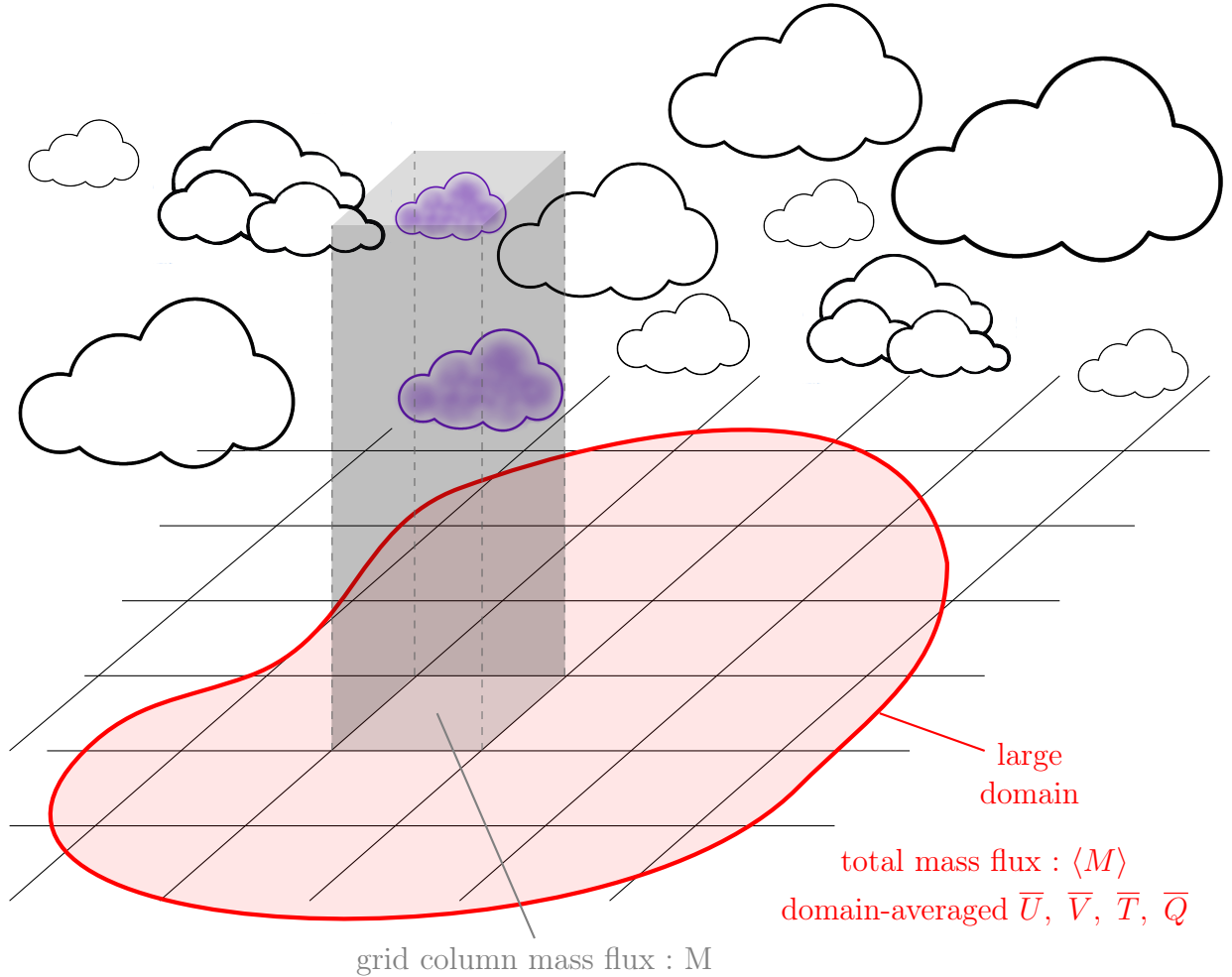


FIGURE 2.2 – Illustration du moyennage sur un grand domaine autour de la maille.

2. Génération d'un échantillon aléatoire de nuages dans une maille Le nombre de nuage apparaissant dans une maille de surface A_{cell} pendant un pas de temps δt du modèle suit la distribution de Poisson 2.13 avec $\Delta t = \delta t$ et $\Delta A = A_{\text{cell}}$. Une méthode d'inversion permet à chaque pas de temps de tirer selon cette distribution un nombre aléatoire de nuages nouvellement créés. Une méthode d'inversion permet également d'attribuer à chaque nuage un flux de masse moyen sur son temps de vie m , suivant la distribution 2.14. Le flux de masse instantané m' est déterminé grâce à une équation de cycle de vie idéalisée (Sakradzija et al. 2015) :

$$m'(t) = m \frac{2}{3} \left| 4 \frac{t}{\tau} \left(\frac{t}{\tau} - 1 \right) \right|, \quad t \in [0, \tau]$$

avec τ la durée de vie du nuage, donnée par la relation empirique $\tau = \alpha m^\beta$.

4. Calcul du flux de masse total dans une maille M À chaque pas de temps, le nombre de nuage total N dans une maille est donc la somme de nouveaux nuages et de nuages encore en vie, chacun ayant son propre flux de masse instantané m' . On calcule donc le flux de masse M en sommant sur tous ces nuages :

$$M = \sum_{i=1}^N m'_i$$

À l'issue de ce dernier calcul, le flux de masse total M peut être utilisé comme condition de fermeture du schéma de Tiedtke-Bechtold. Dans le schéma de convection stochastique de Sakradzija et Klocke (2018), le schéma de Tiedtke-Bechtold est donc utilisé de la même façon que dans une configuration déterministe, la seule différence étant que le flux de masse total M , au lieu d'être donné par la condition de fermeture déterministe, est un flux de masse aléatoire. Cette formulation est intéressante, car elle permet d'adapter le schéma de convection stochastique à n'importe quel schéma de convection déterministe ayant une condition de fermeture en flux de masse.

Bien qu'attrayant car reposant sur de nombreux arguments physiques, le schéma Sakradzija et al. (2015) a le désavantage d'avoir un coût numérique important car il est nécessaire de garder en mémoire un ensemble de nuages et leurs différentes caractéristiques dans chaque maille du modèle. Ceci a motivé les travaux de Machulskaya et Seifert (2019) qui traduisent le schéma de Sakradzija et al. (2015) en terme de deux équations différentielles stochastiques et montrent leur capacité à reproduire de façon satisfaisante le schéma de Sakradzija et al. (2015).

Une partie de ces travaux de thèse a consisté à s'inspirer du schéma de convection stochastique de Sakradzija et Klocke (2018) pour perturber le schéma de convection peu profonde du modèle AROME. Ce dernier est décrit dans la section suivante, et la méthode choisie pour le perturber est présentée dans le chapitre 4.

2.2.4 La paramétrisation de la convection peu profonde dans AROME

Du fait de sa résolution spatiale importante, le modèle AROME fait partie des *convection-permitting models* : la convection profonde est considérée comme explicitement résolue, mais un schéma de convection peu profonde est toujours utilisé pour paramétrer une partie des mouvements convectifs qui restent sous-maille. Dans AROME, il s'agit du schéma de Pergaud et al. (2009). Ce schéma permet de calculer la partie non locale du transport turbulent défini dans l'approche EDMF par l'équation 2.5.

Le schéma de Pergaud et al. (2009) est un schéma de flux de masse où toutes les ascendances convectives, sèches ou humides, sont représentées par une unique ascendance – ou *plume convective* – occupant une fraction a_u de la maille et possédant une vitesse w_u . Le flux de masse convectif M_u sur une maille est défini par :

$$M_u = \rho a_u w_u.$$

L'indice u est utilisé pour indiquer les variables relatives aux ascendances (u pour *updrafts* en anglais).

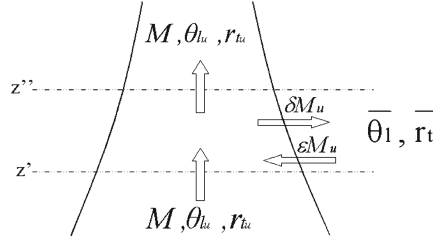


FIGURE 2.3 – Schéma des échanges entre une plume convective et son environnement tiré de Pergaud et al. (2009).

Lorsque l'air s'élève au sein d'une plume convective, des échanges se produisent avec l'air environnant comme indiqué sur la figure 2.3 : des parcelles d'air de l'environnement sont entraînées dans le courant convectif et s'y mélangent (entraînement) ; et à l'inverse de l'air du courant convectif est abandonné dans l'environnement (déentraînement). En définissant un taux d'entraînement ϵ et un taux de déentraînement δ , un bilan de masse donne l'évolution du flux de masse sur la verticale :

$$\frac{1}{M_u} \frac{\partial M_u}{\partial z} = \epsilon - \delta. \quad (2.15)$$

Pour une variable conservative Φ_u dans la plume convective on a :

$$\frac{\partial \Phi_u}{\partial z} = -\epsilon(\Phi_u - \bar{\Phi}), \quad (2.16)$$

avec $\bar{\Phi}$ la valeur moyenne sur la maille. Dans le schéma de Pergaud et al. (2009), les variables conservatives utilisées sont la température potentielle liquide θ_l et le rapport de mélange de l'eau total r_t . Des équations sont également données pour les composantes du vent, avec des termes supplémentaires car ces variables ne sont pas conservatives :

$$w_u \frac{\partial w_u}{\partial z} = \frac{g}{\theta_v} (\theta_{v,u} - \bar{\theta}_v) - \epsilon w_u^2, \quad (2.17)$$

$$\frac{\partial u_u}{\partial z} = -\epsilon(u_u - \bar{u}) + C_u \frac{\partial \bar{u}}{\partial z}, \quad (2.18)$$

$$\frac{\partial v_u}{\partial z} = -\epsilon(v_u - \bar{v}) + C_v \frac{\partial \bar{v}}{\partial z}, \quad (2.19)$$

où θ_v désigne la température potentielle virtuelle, et C_u, C_v sont des constantes.

Les taux d'entraînement et déentraînement, apparaissant dans toutes les équations, sont modélisés différemment selon que l'ascendance est sèche ou nuageuse. Pour la partie sèche, le mélange avec l'air extérieur est supposé proportionnel à un équilibre entre la vitesse verticale w_u et la flottabilité $B_u = \frac{g}{\theta_v} (\theta_{v,u} - \bar{\theta}_v)$. On a

$$\begin{aligned} \epsilon_{dry} &= \max \left[0, C_\epsilon \frac{B_u}{w_u^2} \right], \\ \delta_{dry} &= \max \left[\frac{1}{L_{up} - z}, C_\delta \frac{B_u}{w_u^2} \right]. \end{aligned}$$

Concernant la partie nuageuse de l'ascendance, qui apparaît si le niveau de condensation par ascension est atteint, la formulation de Kain et Fritsch (1990) est utilisée pour déterminer ϵ_{moist}

et δ_{moist} . Cette formulation conceptualise une zone de transition, où l'air de l'ascendance et l'air environnemental se mélangent. Les parcelles d'air de cette zone de transition sont ensuite séparées entre celles dont la flottabilité est positive, et qui sont entraînées dans l'ascendance, et celles dont la flottabilité est négative et qui sont détraînées.

Les ascendances convectives modélisées par le schéma de Pergaud et al. (2009) pouvant être humides, une contribution à l'eau liquide nuageuse sous-maille et à la fraction nuageuse est prévue. La fraction nuageuse f diagnostiquée par le schéma est donnée par :

$$f = C a_u, \quad (2.20)$$

où $C = 2.5$. La contribution à l'eau liquide nuageuse de la maille est déterminée à partir de r_{c_u} , elle-même diagnostiquée à partir de θ_{l_u} et r_{t_u} , selon :

$$\overline{r_{c,\text{conv}}} = r_{c_u} f \quad (2.21)$$

Pour obtenir M_u et les différentes variables Φ_u à tous les niveaux de la plume convective, des conditions de fermeture sont nécessaires. Il s'agit de connaître le flux de masse et les variables caractéristiques des ascendances à un niveau de référence afin de pouvoir ensuite intégrer les équations 2.15-2.19. Dans Pergaud et al. (2009), les conditions de fermeture sont données au niveau du sol par :

$$\begin{aligned} M_u(z_{\text{surf}}) &= C_{M_0} \rho \left(\frac{g}{\theta_{v\text{ref}}} \overline{w' \theta'_v(z_{\text{surf}})} L \right)^{1/3}, \\ \theta_{l_u}(z_{\text{surf}}) &= \overline{\theta}_l(z_{\text{surf}}) + \alpha \frac{\overline{w' \theta}_l(z_{\text{surf}})}{\overline{e}^{1/2}(z_{\text{surf}})}, \\ r_{t_u}(z_{\text{surf}}) &= \overline{r}_t(z_{\text{surf}}) + \alpha \frac{\overline{w' r}_t(z_{\text{surf}})}{\overline{e}^{1/2}(z_{\text{surf}})}, \\ w_u^2(z_{\text{surf}}) &= \frac{2}{3} \overline{e}(z_{\text{surf}}). \end{aligned}$$

avec C_{M_0} et α des constantes fixes, L la longueur de Bougeault-Lacarrère (Bougeault et Lacarrère 1989), et e l'énergie cinétique turbulente (TKE).

2.3 Paramétrisation des nuages

La paramétrisation des nuages regroupe la représentation de tous les phénomènes physico-chimiques liés à la formation et à l'évolution des hydrométéores, c'est-à-dire des particules d'eau en phase liquide ou glacée dans l'atmosphère. On peut globalement distinguer des phénomènes très rapides, presque instantanés à notre échelle comme la condensation de vapeur d'eau en petites gouttelettes liquides, leur évaporation, ou la sublimation de particules glacées, et des phénomènes plus lents comme la croissance de cristaux de glace, la formation de gouttes de pluie, etc. Dans les deux cas il s'agit de phénomènes de microphysique, mais leur traitement dans AROME est séparé : les premiers sont représentés par un schéma d'ajustement à la saturation, aussi appelé schéma de condensation sous-maille, et les seconds forment ce que l'on appelle le schéma microphysique.

2.3.1 Principes de l'ajustement à la saturation

Ce schéma est utilisé dans AROME à chaque pas de temps, avant toutes les autres paramétrisations physiques. Le but est de corriger une éventuelle sursaturation dans chaque maille du modèle en simulant la condensation, l'évaporation et la sublimation qui peuvent se produire selon les conditions, ce qui permet d'adapter les contenus en vapeur d'eau, en hydrométéores non précipitants ainsi que la température. Ces différentes variables "ajustées" sont utilisées ensuite dans tous les autres schémas de paramétrisation.

Pendant longtemps, les schémas d'ajustement à la saturation ont utilisé un diagnostic "all or nothing" (e.g. Steiner 1973; Sommeria 1976) : selon ses conditions moyennes, une maille était considérée soit totalement saturée, donc avec de l'eau condensée (un nuage) présente sur toute la maille, soit pas du tout (voir la figure 2.4). Or, selon la situation météorologique et avec la résolution des modèles de PNT, les nuages sont souvent des phénomènes sous-maille. Ainsi, bien que les conditions moyennes sur la maille soient non saturantes, les fluctuations sous-maille des différentes variables peuvent produire localement des conditions saturantes et l'apparition d'un nuage. De même, des conditions moyennes saturantes ne garantissent pas que le nuage couvre toute la maille. Pour prendre cela en compte, des schémas "statistiques" ont été développés (Sommeria et Deardorff 1977). Le principe général de ces schémas est de faire des hypothèses sur la distribution sous-maille de certaines variables quasi-conservatives et d'en déduire la fraction nuageuse et le contenu en hydrométéores associé dans la maille.

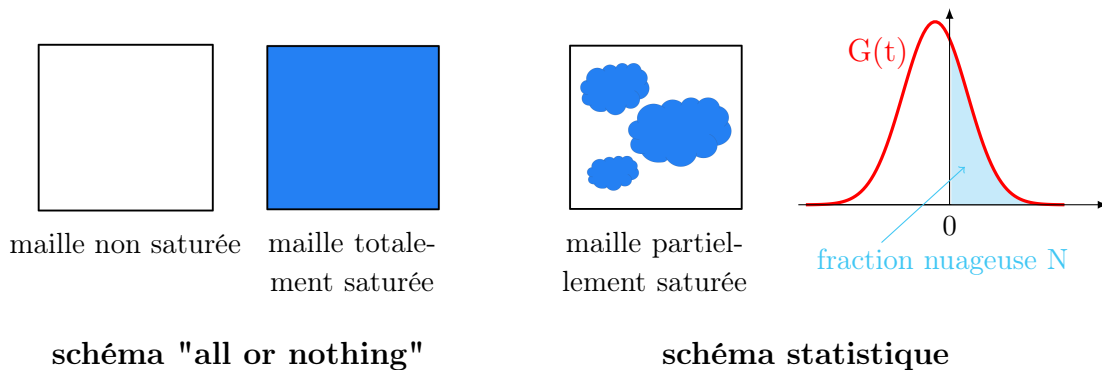


FIGURE 2.4 – Illustration des schémas de condensation "all or nothing" ou statistique.

Le changement de phase d'une partie de la vapeur d'eau peut être déterminé à partir du contenu en vapeur saturante r_s pour l'eau liquide ou la glace. Le rapport de mélange des hydrométéores non précipitants (c'est-à-dire des gouttelettes nuageuses et de la glace nuageuse) provenant de la condensation est donné localement par

$$r_{\text{cond}} = \max(r_{np} - r_s, 0),$$

où r_{np} est le rapport de mélange total de l'eau non précipitante initialement présente, c'est-à-dire la somme des rapports de mélange de la vapeur d'eau, des gouttelettes nuageuses et de la glace nuageuse ($r_{np} = r_v + r_c + r_i$). r_{np} est une variable quasi-conservative pour laquelle on peut faire une hypothèse sur la distribution sous-maille. Le rapport r_s peut lui être exprimé à l'aide de la température potentielle liquide θ_l , qui elle aussi est une variable quasi-conservative. En supposant une distribution jointe bi-gaussienne pour r_{np} et θ_l , Sommeria et Deardorff (1977) dérivent des expressions pour la fraction nuageuse et l'eau nuageuse. D'autres études introduisent la variable s , représentant l'écart à la saturation (Mellor 1977; Bougeault 1981) et qui

dépend de r_{np} et r_s . L'écart-type de s , noté σ_s , peut être défini à partir des mêmes variables et on définit également la variable normalisée $t = s/\sigma_s$. Plusieurs études utilisent des hypothèses de distribution sous-maille directement sur cette variable t (e.g. Bougeault 1981, 1982 ; Bechtold et al. 1995). Si on note $G(t)$ cette distribution, comme la condensation se produit lorsque $r_{np} > r_s \Leftrightarrow s > 0$, on peut calculer la fraction nuageuse et le contenu en eau moyen sur une maille de la façon suivante (Bechtold et al. 1995) :

$$N = \int_0^{+\infty} G(t)dt, \quad (2.22)$$

$$\frac{\bar{r}_{\text{cond}}}{\sigma_s} = \int_0^{+\infty} tG(t)dt, \quad (2.23)$$

où le symbole "-" représente la moyenne sur une maille de modèle. La moyenne de la distribution $G(t)$ est notée $Q_1 = \bar{s}/\sigma_s$.

Selon les études, les hypothèses sur la forme de la distribution $G(t)$ peuvent varier. Par exemple, Mellor (1977) utilise une distribution gaussienne, Bougeault (1981) propose des distributions allant de la gaussienne à l'exponentielle, Bougeault (1982) fait l'hypothèse d'une distribution gamma. Dans le cas de l'étude de Bechtold et al. (1995), des distributions différentes sont choisies en fonction du type de nuage. Dans des travaux ultérieurs, Cuijpers et Bechtold (1995) et Chaboureau et Bechtold (2002) dérivent des relations analytiques entre la fraction nuageuse, l'eau nuageuse et l'écart à la saturation moyen Q_1 à partir de simulations LES de plusieurs types de nuages, s'affranchissant ainsi des hypothèses sur les distributions. Ces relations sont représentées sur la figure 2.5 et sont celles utilisées dans AROME.

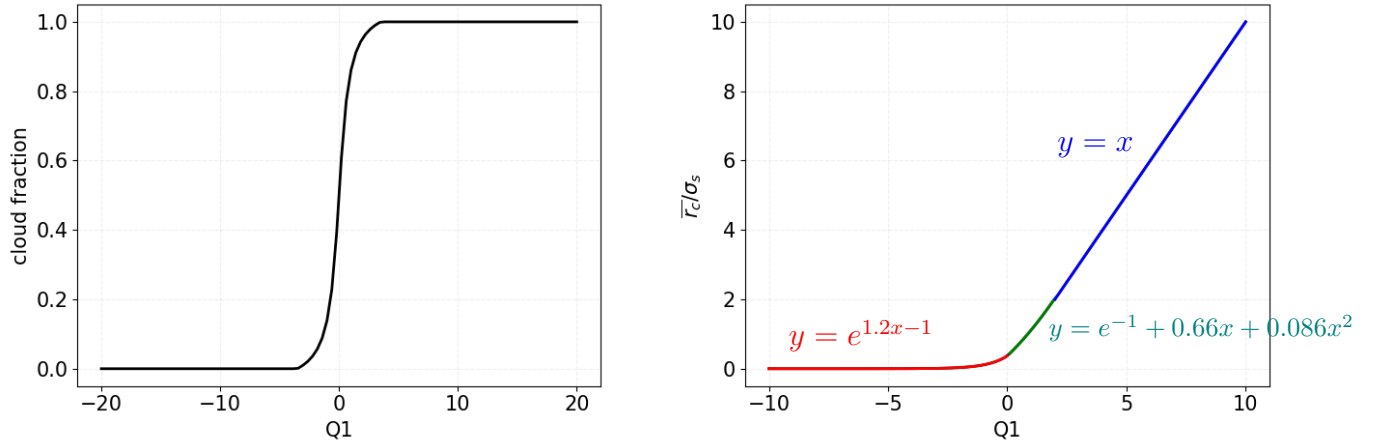


FIGURE 2.5 – Formules de Chaboureau et Bechtold (2002) de diagnostic de la fraction nuageuse et du contenu en eau moyen sur la maille d'après l'écart à la saturation moyen Q_1 .

La valeur de Q_1 est calculée dans AROME lors de l'étape d'ajustement à la saturation. σ_s en revanche est calculé par le schéma de turbulence et transmis d'un pas de temps sur l'autre au schéma d'ajustement. En pratique, le schéma de turbulence produit en général des valeurs non nulles pour ce terme seulement dans la couche limite instable, et ne représente donc pas bien la variabilité dans les nuages troposphériques hauts ou dans les brouillards se développant à partir d'une couche limite nocturne stable par exemple. Un terme dépendant du contenu en vapeur

saturante a donc été ajouté dans AROME pour l'augmenter artificiellement. Il est contrôlé par le coefficient `VSIGQSAT` qui fait partie des 8 paramètres de réglage d'AROME les plus influents d'après l'étude de sensibilité de Wimmer et al. (2022).

Enfin, l'ajustement à la saturation tient compte de la contribution du schéma de convection peu profonde mentionnée dans la section 2.2.4 (voir les équations 2.20 et 2.21). Cette contribution est directement ajoutée à r_{cond} et à la fraction nuageuse diagnostiquée par le schéma statistique.

2.3.2 Le schéma microphysique ICE3

Description générale du schéma

Dans AROME, le schéma microphysique ICE3 (Pinty et Jabouille 1998) sert à représenter les processus physiques "lents" liés aux changements de phase de l'eau et affectant les différentes classes d'hydrométéores du modèle. Une représentation des différents processus en jeu dans ce schéma est donnée par la figure 2.6.

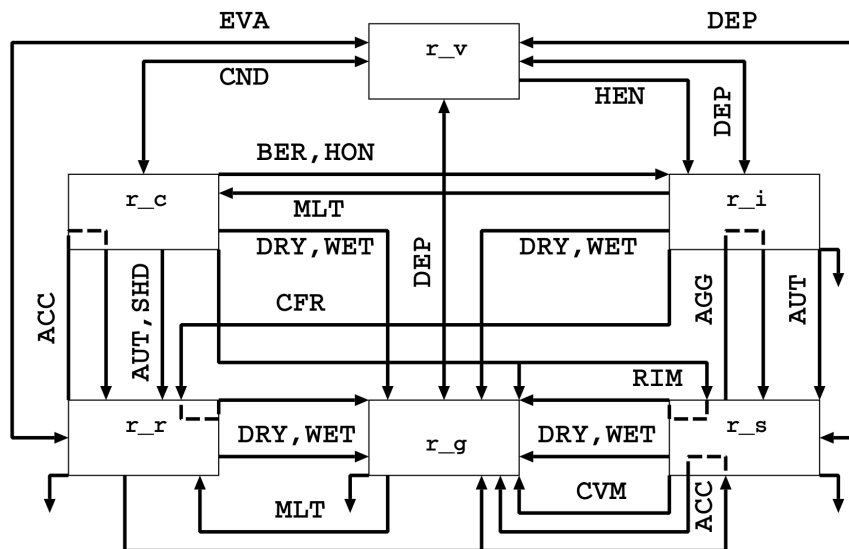


FIGURE 2.6 – Processus microphysiques modélisés par le schéma microphysique ICE3 (illustration de Pinty et Jabouille 1998). Dans AROME, le processus de formation et d'évaporation des gouttelettes liquides nuageuses (flèches "CND") est traité par l'ajustement à la saturation.

Les cinq classes d'hydrométéores d'AROME sont représentées sur ce schéma. Deux catégories non précipitantes, l'eau liquide nuageuse (indice "c") et la glace nuageuse ("i"), et trois catégories d'hydrométéores précipitants : la pluie ("r"), la neige ("s") et la neige roulée ou graupel ("g"). À cela s'ajoute la vapeur d'eau ("v").

Parmi les processus physiques modélisés par ICE3, on trouve :

- des processus de nucléation qui font apparaître de la glace nuageuse (le processus de condensation de gouttelettes liquides est lui entièrement géré par l'ajustement à la saturation décrit précédemment) ;

- des processus d'autoconversion grâce auxquels l'eau et la glace nuageuses forment des hydrométéores précipitants ;
- des processus d'accrétion, d'auto-collection, d'agrégation, de déposition, etc, qui participent à la croissance des hydrométéores précipitants ;
- des processus de désintégration, fonte, etc, qui participent à la diminution de la quantité d'hydrométéores ;
- des processus liés à la gravité terrestre : la sédimentation, qui affecte tous les hydrométéores.

Les processus impliquant l'eau liquide suivent un schéma de Kessler (Kessler 1969). Les processus impliquant les phases glacées sont présentés par exemple dans Pinty et Jabouille (1998). La documentation du modèle de recherche Méso-NH (Lac et al. 2018), dont les schémas de paramétrisation physique sont communs à ceux d'AROME en fournit une description détaillée (<http://mesonh.aero.obs-mip.fr/mesonh54/BooksAndGuides>).

Le schéma ICE3 est qualifié de schéma microphysique à "1 moment" car les différentes catégories d'hydrométéores ne sont représentées de façon pronostique qu'à travers leur rapport de mélange. Ce dernier ne donne pas d'information sur le nombre et la taille des hydrométéores : un même rapport de mélange d'eau liquide nuageuse peut par exemple être obtenu avec de nombreuses petites gouttes ou avec quelques grosses gouttes. Or, la plupart des processus microphysiques sont sensibles à la taille des hydrométéores. Pour avoir cette information, il faut donc faire des hypothèses en considérant que les hydrométéores se répartissent selon une certaine distribution de taille dont on fixe la formule. Dans ICE3, on utilise comme formule générale celle d'une loi gamma généralisée. Si on note D le diamètre d'un hydrométéore et $n(D)$ la distribution en terme de diamètre de cet hydrométéore on a :

$$n(D) = N_{0,j} \frac{\alpha_j}{\Gamma(\nu_j)} \lambda_j^{\alpha_j \nu_j} D^{\alpha_j \nu_j - 1} e^{-(\lambda_j D)^{\alpha_j}}, \quad (2.24)$$

où j désigne ici une catégorie d'hydrométéore, et Γ est la fonction gamma. Cette distribution fait intervenir quatre paramètres : $N_{0,j}$, qui correspond à la concentration totale (c'est-à-dire au nombre total d'hydrométéores par unité de volume), et ν_j , α_j , et λ_j , qui sont des paramètres de forme. Dans ICE3, $N_{0,j}$, ν_j et α_j ont des valeurs fixées, avec seulement une distinction faite en fonction du type de surface (océanique ou continentale). Le paramètre λ_j est lui diagnostiqué de façon à ce que la distribution s'adapte au rapport de mélange de l'hydrométéore. En supposant la relation de puissance $m(D) = aD^b$ entre la masse d'un hydrométéore et son diamètre, on a :

$$\rho_d r_j = \int_0^\infty a D^b n(D) dD \quad \Rightarrow \quad \lambda_j = \left(a \frac{\Gamma(\nu_j + \frac{b}{\alpha_j})}{\Gamma(\nu_j)} \frac{N_{0,j}}{\rho_d r_j} \right)^{1/b}, \quad (2.25)$$

où r_j est le rapport de mélange de la catégorie d'hydrométéores j , et ρ_d la masse volumique de l'air sec.

Pour les catégories d'hydrométéores précipitants, la distribution est simplifiée en une distribution de Marshall-Palmer modifiée (Marshall et Palmer 1948 ; Caniaux et al. 1994) en réglant $\alpha_j = 1$, $\nu_j = 1$ et $N_{0,j} = C_j \lambda^{x_j}$. Pour l'eau et la glace nuageuse, la forme générale de l'équation 2.24 est conservée. Pour l'eau nuageuse, à laquelle on s'intéresse plus particulièrement dans le chapitre 5, les réglages sont les suivants :

- sur les continents : $N_{0,c} = 300$ gouttes/cm³, $\alpha_c = 1$, $\nu_c = 3$

- sur les océans : $N_{0,c} = 100$ gouttes/cm³, $\alpha_c = 3$, $\nu_c = 1$

Il existe des schémas plus sophistiqués que le schéma à 1 moment ICE3, pour lesquels les hypothèses sur la distribution de taille des hydrométéores sont moins restrictives. Les schémas à "2 moments", comme par exemple LIMA (Vié et al. 2016), utilisent comme variables pronostiques supplémentaires les concentrations totales $N_{0,j}$. Il existe des schémas d'ordre encore supérieur, qui prédisent l'évolution de différents moments de la distribution de taille (e.g. Loftus et al. 2014; Morrison et Milbrandt 2015). Le réalisme peut être poussé encore plus loin avec les schémas dits "bin", où les hydrométéores sont séparés en plusieurs classes de tailles qui évoluent (e.g. Hall 1980; Ovtchinnikov et Kogan 2000), ou avec l'approche récente de modèle de nuage lagrangien (Shima et al. 2009; Grabowski et al. 2019; Schwenkel et Maronga 2020) dont le principe est de simuler des "superparticules" qui représentent un ensemble d'hydrométéores aux caractéristiques similaires. C'est deux derniers types de schémas sont néanmoins très coûteux et ne sont en général pas utilisés opérationnellement dans les modèles de PNT.

Détails sur la sédimentation des gouttelettes nuageuses

On présente ici plus en détail le processus de sédimentation de l'eau liquide nuageuse, car il s'agit d'un processus clé dans certaines situations, comme le brouillard, qui ont été étudiées durant la thèse (voir la présentation des cas d'étude dans le chapitre 3 et le chapitre 5).

La sédimentation est l'un des processus microphysiques pouvant affecter le contenu en gouttelettes nuageuses dans une maille de modèle. Elle correspond à un flux vertical de gouttelettes F , qui tombent sous l'effet de la gravité. L'évolution du contenu en eau nuageuse à une altitude donnée dépend du gradient vertical de ce flux, selon :

$$\frac{\partial r_c}{\partial t} = \frac{1}{\rho_d} \frac{\partial F}{\partial z}. \quad (2.26)$$

Le flux dépend de la vitesse terminale de chute des gouttelettes, V_T , et du contenu en eau nuageuse r_c , selon $F = \rho_d r_c V_T$. La vitesse de chute d'une gouttelette individuelle dépend de son diamètre. Comme les gouttelettes ont une grande variété de tailles différentes, V_T est calculée en intégrant sur tout le spectre de taille, de la façon suivante :

$$V_T = \frac{1}{\rho_d r_c} \int_0^\infty \frac{\pi}{6} D^3 \rho_l V(D) n(D) dD \quad (2.27)$$

où $V(D)$ est la vitesse terminale d'une gouttelette de diamètre D et ρ_l la masse volumique de l'eau liquide.

L'intégration de l'équation 2.26 n'est pas triviale, du fait la non-linéarité en r_c . Dans le modèle AROME, on évite sa résolution numérique en utilisant l'approche statistique de Geleyn et al. (2008) et Bouteloup et al. (2011). Celle-ci consiste à calculer, à chaque niveau vertical, des probabilités de passage des gouttelettes d'eau au niveau suivant. Le principe de cette méthode est illustré sur la figure 2.7.

Deux probabilités de passage à travers la couche k sont calculées : la probabilité que les gouttes initialement présentes dans la couche passent dans celle du dessous (P_1) et probabilité que des gouttes provenant des couches supérieures traversent la couche k et passent dans celle du dessous (P_2). P_1 et P_2 dépendent de la vitesse de chute des gouttelettes. Dans Bouteloup et al. (2011), la vitesse terminale V_T est utilisée pour toutes les gouttes. Dans Geleyn et al.

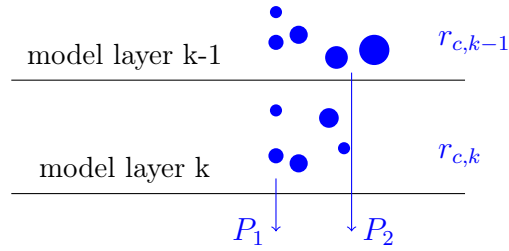


FIGURE 2.7 – Principe du schéma de sédimentation de Bouteloup et al. (2011). Les ronds bleus représentent des gouttelettes nuageuses de différentes tailles, et les deux traits horizontaux délimitent deux couches verticales successives du modèle.

(2008), l’hypothèse d’une distribution exponentielle de la vitesse verticale des gouttes est faite, et les probabilités P_1 et P_2 sont calculées en intégrant sur cette distribution.

2.3.3 Processus microphysiques et incertitude

Les approches de perturbation de la turbulence et de la convection peu profonde présentées dans les sections 2.1.4 et 2.2.2-2.2.3 ne peuvent pas être transposées telles quelles à un schéma microphysique. En effet, les schémas de turbulence et de convection cherchent à modéliser des structures sous-maille liées à des mouvements non résolus par le modèle, et une partie de l’incertitude vient de la présence en nombre limité de ces structures sous-maille. En revanche, un schéma microphysique traite de différents processus diabatiques comme les changements de phase de l’eau, ou de phénomènes tels que la chute des hydrométéores, qui ont des échelles de temps et d’espace beaucoup plus courtes. Ajouter des fluctuations qui représentent, à l’échelle d’une maille, l’effet d’un nombre restreint et aléatoire de structures sous-mailles indépendantes n’a donc pas de sens pour les processus microphysiques en tant que tels. En revanche, étant donné qu’il s’agit de processus locaux, ils peuvent être affectés par la turbulence qui produit des variations de fine échelle des champs de température, d’humidité, etc (Grabowski et Wang 2013). La prise en compte de la turbulence dans les processus microphysiques est d’ailleurs à l’origine de la théorie de la *condensation stochastique*, qui propose de modéliser l’impact des fluctuations sous-maille de supersaturation sur la distribution de taille des hydrométéores moyenne sur la maille (voir par exemple Jeffery et al. 2007). Les schémas de condensation stochastique reposent sur une équation stochastique d’évolution de la distribution de taille des hydrométéores dont par exemple Khvorostyanov et Curry (1999a,b) dérivent des solutions analytiques simplifiées. Leurs solutions donnent la distribution moyenne sur une maille de modèle de la taille des gouttelettes nuageuses. Déterminer les fluctuations possibles autour de cette moyenne à partir de leurs résultats ne semble néanmoins pas trivial.

Une des questions que l’on peut se poser est la façon d’évaluer les schémas microphysiques. Comment estimer leurs erreurs (et donc éventuellement la façon de les perturber) ? En mettant de côté les observations directes, des schémas de paramétrisation comme la turbulence ou la convection présentés précédemment peuvent s’appuyer sur des simulations explicites des phénomènes qu’ils paramétrisent par des modèles à haute résolution (LES, DNS : Direct Numerical Simulations; e.g. Brown 1996; Moin et Mahesh 1998; Couvreur et al. 2010; Garcia et Mellado 2014; Lu et al. 2016; Angevine et al. 2018). Ce type de simulations peut être utilisé comme référence pour calibrer les schémas déterministes, mais aussi, dans le cadre de la prévision d’ensemble, pour régler les perturbations à ajouter au modèle (e.g. Sakradzija et al. 2015; Christensen 2020). Dans le cas des phénomènes microphysiques, il n’existe pas de simulations

aux échelles où se déroulent les diverses transformations physiques et qui permettraient de les représenter explicitement. Néanmoins, comme mentionné dans la section précédente, les schémas microphysiques ont des degrés de complexité divers (schémas à 2, 3, 4 moments, schémas "bin", schémas lagrangiens), et des schémas complexes peuvent servir à évaluer certains aspects de schémas plus simples. Par exemple, l'étude de Igel et van den Heever (2017a,b) utilise des simulations d'un cas de cumulus non précipitant réalisées avec un schéma microphysique "bin", pour montrer la variabilité du paramètre de forme ν utilisé dans les schémas microphysique "bulk". L'étude de Schwenkel et Maronga (2020) quant à elle utilise un modèle de nuage lagrangien pour montrer les différences entre les résultats de ce schéma et ceux d'un schéma "bulk" (voir par exemple leur figure 4 sur la différence de distribution de taille des gouttelettes nuageuses).

À notre connaissance, les méthodes de perturbation de la microphysique utilisées aujourd'hui dans le cadre de la prévision d'ensemble sont soit des méthodes de perturbation des tendances dérivées de SPPT (Christensen et al. 2017; Wastl et al. 2019a), soit des méthodes de perturbation de paramètres (Thompson et al. 2021; Frogner et al. 2022; Tsiringakis et al. 2022). Les méthodes de perturbation de paramètre semblent particulièrement intéressantes pour assouplir certaines hypothèses simplificatrices fortes des schémas microphysiques. Par exemple Wimmer et al. (2022) proposent de perturber les seuils d'autoconversion de la pluie et de la neige dans ICE3, qui sont des seuils de contenu en eau ou glace nuageuse à partir desquels on considère que de la pluie ou de la neige se forment. Thompson et al. (2021) appliquent des perturbations de type SPP à certaines relations ou certains paramètres fixes du schéma microphysique de Thompson et Eidhammer (2014), dont le caractère incertain est suggéré par les observations (paramètre de forme de la loi gamma généralisée servant à modéliser la distribution de taille des gouttelettes nuageuses, vitesse verticale servant à diagnostiquer la fraction d'aérosols activés, paramètre de la distribution de taille de la catégorie mixte d'hydrométéores graupel/grêle). Des tests de perturbation des paramètres α et ν du spectre de taille des gouttelettes nuageuses ont également été réalisés dans l'ensemble HarmonEPS (Tsiringakis et al. 2022), et des tests de sensibilité du modèle HARMONIE-AROME à ces paramètres dans des situations de brouillard sont en cours (Sebastián Contreras, communication personnelle). De façon générale, il semble exister de bonnes indications que la valeur des paramètres réglant la distribution de taille des hydrométéores utilisée par les schémas microphysiques à 1 moment est incertaine. Par exemple, la compilation effectuée par Miles et al. (2000) de plusieurs études observationnelles de nuages stratiformes montre que la concentration totale en nombre de gouttelettes peut varier de 5 à 693 gouttes/cm³ et qu'en modélisant les distributions de taille observées par une distribution de type gamma modifiée, on obtient des valeurs du paramètre de forme ν allant de 0.7 à 44.6. Igel et van den Heever (2017a) présentent également une revue de quelques études observationnelles sur des cumulus peu profonds et rapportent des valeurs de ν globalement comprises entre 2 et 14.

Dans ces travaux de thèse, nous avons choisi de suivre les travaux de Thompson et al. (2021) et Tsiringakis et al. (2022) et perturber les paramètres correspondant aux hypothèses sur la distribution de taille des gouttelettes nuageuses faites par le schéma ICE3, à savoir la concentration en nombre de gouttelettes (N_0) et les paramètres de forme de la distribution (ν et α), étant données les indications que ces paramètres sont variables dans le temps et l'espace et étant donnée leur implication dans de nombreux processus modélisés par le schéma microphysique. Davantage de détails concernant les motivations et l'implémentation de ces perturbations sont donnés dans le chapitre 5.

2.4 Paramétrisation des effets radiatifs

2.4.1 Quelques éléments sur le schéma radiatif d'AROME

Les schémas radiatifs des modèles de PNT servent à modéliser les échanges de chaleur par rayonnement électromagnétique. Dans AROME, le schéma radiatif produit une tendance de température selon :

$$\left(\frac{\partial T}{\partial t}\right)_{\text{rad}} = \frac{g}{c_p} \frac{\partial \mathcal{F}}{\partial p},$$

où g est l'accélération de la pesanteur, c_p est la chaleur spécifique à pression constante de l'air humide, \mathcal{F} désigne le flux radiatif net et p la pression. Le flux radiatif est la quantité d'énergie sous forme de rayonnement qui traverse une unité de surface par unité de temps.

La variable de base pour décrire les phénomènes radiatifs est la luminance L , définie comme l'énergie traversant une unité de surface par unité de temps et par unité d'angle solide (qui définit une direction) pour tout le spectre électromagnétique. Lorsque l'on se restreint à un intervalle élémentaire sur le spectre (un intervalle de fréquence $\delta\nu$ par exemple), on parle de luminance spectrale ou monochromatique L_ν . Il est possible d'établir une *équation de transfert radiatif* pour la luminance spectrale (voir par exemple Morcrette 2000), en prenant en compte trois phénomènes pouvant affecter le rayonnement :

- l'absorption (une partie du rayonnement est absorbée par la matière rencontrée et transformée en agitation moléculaire) ;
- la diffusion (le rayonnement est dévié de sa direction originale) ;
- l'émission d'un rayonnement par la matière, due à sa température.

L'équation obtenue est néanmoins très complexe et les modèles de PNT ont recours à un certain nombre de simplifications pour la résoudre.

On sépare généralement le traitement de l'équation de transfert radiatif en deux domaines du spectre électromagnétique : les longueurs d'ondes "courtes" correspondant au spectre solaire et les longueurs d'ondes "longues" correspondant à l'émission thermique de rayonnement infrarouge par l'atmosphère et la surface terrestre. Dans AROME on a ainsi :

- le schéma de Fouquart et Bonnel (1980) pour les longueurs d'onde "courtes" (SW pour shortwave en anglais), qui comprend le domaine visible, le proche ultraviolet et le proche infrarouge (0.185 à 4 μm) ;
- le schéma Rapid Radiative Transfer Model (RRTM ; Mlawer et al. 1997) pour les longueurs d'onde "longues" (LW pour longwave en anglais), qui comprend le domaine infrarouge d'environ 3 à 1000 μm .

Le schéma SW

La source des flux SW est le rayonnement solaire incident E^0 arrivant au sommet de l'atmosphère. Dans l'atmosphère, une partie de ce rayonnement peut être absorbée (absorption moléculaire, absorption par les hydrométéores) et diffusée (diffusion de Rayleigh par les molécules d'air, diffusion de Mie par les aérosols et les gouttelettes nuageuses). L'équation de transfert radiatif est intégrée sur la verticale, et pour chaque niveau on définit deux facteurs : le facteur de réflexion R au sommet du niveau, qui représente la proportion du rayonnement

incident qui est renvoyée vers le haut, et le facteur de transmission T à la base du niveau, qui représente la proportion du rayonnement incident qui est transmise au niveau inférieur. Pour un niveau j on a :

$$F^\uparrow(j) = F^\downarrow(j)R(j-1)$$

$$F^\downarrow(j) = E^0 \prod_{k=j}^N T(k)$$

L'interaction du rayonnement avec les hydrométéores étant très différente de l'interaction avec les gaz et les aérosols, les facteurs de réflexion et de transmission sont calculés séparément pour les parties nuageuses et les parties en ciel clair de la maille. Dans les deux cas, trois paramètres interviennent dans les calculs : l'épaisseur optique δ , qui correspond à l'intégrale verticale d'un coefficient d'extinction, l'albedo de diffusion simple ω et un facteur d'asymétrie g . Pour les parties nuageuses ces paramètres sont qualifiés de *propriétés optiques* du nuage.

Le schéma LW

Le rayonnement LW provient de l'émission dans le domaine des infrarouges par l'atmosphère et la surface terrestre, du fait de leur température. La luminance spectrale est donnée pour un corps noir par la fonction de Planck :

$$L_{\Omega,\nu}(\nu, T) = \frac{2h\nu^3}{c^2 (e^{h\nu/kT} - 1)},$$

avec ν la fréquence, h la constante de Planck, k la constante de Boltzmann et c la vitesse de la lumière. Les flux radiatifs sont calculés grâce à l'équation de transfert radiatif, calculée pour chaque niveau vertical j , qui fait intervenir $L_{\Omega,\nu}$ au niveau j et une contribution des niveaux inférieurs ou supérieurs via l'intégration verticale d'une transmission t_ν . Là encore, les flux en ciel clair sont calculés séparément des flux au travers des couches nuageuses. Pour les parties nuageuses, le schéma fait intervenir des propriétés optiques, notamment l'émissivité, qui permet de tenir compte du fait que le nuage n'est pas un corps noir (l'émissivité est le rapport entre la luminance du nuage et la luminance d'un corps noir qui aurait la même température), mais également l'épaisseur optique.

Pour les deux schémas, l'équation du transfert radiatif est définie pour un intervalle élémentaire de longueur d'onde (ou, de façon équivalente, de fréquence). Pour calculer les flux SW et LW, elle doit donc être intégrée sur le domaine spectral concerné. Pour cela, les domaines sont séparés en plusieurs *bandes spectrales* : 6 pour le domaine SW et 16 pour le domaine LW.

Les paramétrisations des flux SW et LW dans les couches nuageuses font intervenir différentes propriétés optiques des nuages (épaisseur optique, albedo de diffusion simple, facteur d'asymétrie, émissivité). Ces propriétés sont paramétrées pour l'eau liquide et la glace nuageuse d'après Morcrette et Fouquart (1986) et Ebert et Curry (1992). Elles font intervenir des informations comme le contenu en eau liquide ou glace, et le *rayon effectif* des gouttelettes ou des particules de glace, défini comme le rapport entre les moments d'ordre 3 et d'ordre 2 de leur distribution de taille.

Enfin, il faut noter que les schémas radiatifs sont en général coûteux en calcul et ne sont donc pas appelés à chaque pas de temps des modèles de PNT. Par exemple dans AROME, le calcul des flux radiatifs complet n'est effectué que toutes les 15 minutes (néanmoins les flux SW sont mis à jour à chaque pas de temps du modèle pour tenir compte de l'angle du soleil par rapport au zénith).

2.4.2 L'incertitude liée à la paramétrisation des effets radiatifs

L'énergie thermique de l'atmosphère et de la surface terrestre provient en grande majorité du rayonnement solaire. Celui-ci constitue un moteur de la circulation et des phénomènes atmosphériques, et la paramétrisation des effets radiatifs, qui permet de déterminer la répartition de cette énergie à travers toute l'atmosphère, a donc une grande importance pour la modélisation.

La sensibilité à la paramétrisation des flux radiatifs a surtout été mise en évidence pour les prévisions climatiques, où des différences de quelques W/m^2 dans le bilan énergétique du système Terre peuvent indiquer des changements importants en terme de climat (e.g. Ramanathan et al. 1983; Li et al. 2015). Néanmoins, leur effet sur la dynamique de certains phénomènes à des échelles de temps bien plus courtes (e.g. le refroidissement radiatif au sommet des stratocumulus; Wood 2012, l'impact des flux de surface sur le déclenchement de la convection; Brown et al. 2002), ainsi que le besoin d'estimer précisément les flux radiatifs au sol pour le secteur de l'énergie solaire en expansion (e.g. Kosmopoulos et al. 2015), montrent leur importance également pour les prévisions des modèles de PNT.

En examinant la formulation d'un schéma radiatif, il est possible d'en identifier les principales sources d'incertitude (Schäfer et al. 2021). Par exemple, l'utilisation de moyennes climatologiques pour l'albedo et l'émissivité de la surface terrestre, ou les propriétés optiques des aérosols est une simplification qui peut être questionnée (Nikitidou et al. 2014; Rieger et al. 2017; Sablon 2020). L'intégration spectrale et verticale de l'équation de transfert radiatif est également un point clé des schémas radiatifs. L'intégration spectrale est réalisée en effectuant les calculs sur des bandes spectrales, ce qui rend difficile le traitement de paramètres comme l'absorption moléculaire, très variable selon la longueur d'onde. L'intégration verticale, quant à elle, est reliée au traitement de la répartition des nuages sous-maille, qui serait l'une des sources d'incertitude majeures des schémas radiatifs (Slingo 1990; Barker et al. 1999; Shonk et Hogan 2010; Rontu et Lindfors 2018). D'un niveau vertical à l'autre, les zones nuageuses peuvent en effet se recouvrir plus ou moins. Comme cette information n'est pas disponible dans un modèle, les schémas radiatifs utilisent diverses hypothèses pour déterminer la géométrie verticale des nuages. Par exemple, dans AROME, on suppose un recouvrement maximal pour deux niveaux adjacents nuageux et un recouvrement aléatoire pour deux niveaux nuageux séparés par un ou plusieurs niveaux sans nuages. Une autre incertitude potentielle concernant les nuages sous-maille est leur inhomogénéité horizontale : les caractéristiques d'un nuage comme son contenu en eau, la taille de ses hydrométéores, etc, varient localement, or de nombreux schémas font l'hypothèse d'une partie nuageuse sous-maille homogène. Des représentations plus complexes des nuages sous-maille ont été proposées dans la littérature : par exemple, le schéma "Triple-clouds" (Shonk et Hogan 2008) prévoit de séparer les nuages en deux zones, une zone optiquement épaisse et une zone optiquement fine, tandis que le schéma "McICA" (Pincus et al. 2003) utilise une méthode d'échantillonnage aléatoire d'états nuageux possibles associés à diverses bandes spectrales pour réaliser l'intégration de l'équation de transfert radiatif en introduisant de la variabilité dans les caractéristiques des nuages sous-maille.

La représentation de l'incertitude des schémas radiatifs pour la prévision d'ensemble est à ce jour traitée essentiellement par des méthodes de perturbation des tendances (SPPT) ou de paramètres (Ollinaho et al. 2017; Lock et al. 2019; Frogner et al. 2022). L'utilisation de plusieurs schémas radiatifs dans une approche "multiphysique" est également suggérée par Rontu et Lindfors (2018). De façon générale, certains aspects des schémas radiatifs sont jugés peu incertains, comme l'estimation des flux radiatifs en ciel clair, ce qui a motivé l'atténuation des perturbations dans la stratosphère de la méthode SPPT (Palmer et al. 2009), et plus récemment la suppression des perturbations sur les tendances radiatives en ciel clair (Lock et al. 2019). L'incertitude liée aux nuages, en revanche, est prise en compte : par exemple, Ollinaho et al. (2017) proposent de perturber deux paramètres réglant la variabilité verticale et horizontale des nuages sous-maille, ainsi qu'un paramètre affectant les propriétés optiques des nuages. L'incertitude concernant le calcul des propriétés optiques des nuages a également récemment été étudiée par Jahangir et al. (2021), qui concluent que, pour le rayonnement SW, et dans le cas de nuages liquides, l'incertitude vient principalement de la paramétrisation du rayon effectif des gouttelettes nuageuses.

Étant données les incertitudes sur la répartition en trois dimensions des nuages dans une maille, les méthodes sélectionnant aléatoirement des caractéristiques nuageuses dans des "sous-colonnes" de la maille, comme McICA, paraissent intéressantes. Néanmoins, McICA n'a pas été développée dans une optique de prévision d'ensemble et le bruit généré par l'échantillonnage aléatoire a été démontré comme ayant un impact statistiquement négligeable sur des simulations de modèles globaux (Pincus et al. 2003; Barker et al. 2008).

Dans le chapitre 5, nous présentons des tests de perturbations de paramètre appliquées au schéma radiatif. Nous avons choisi de nous appuyer sur les travaux de Jahangir et al. (2021), qui proposent de mettre en cohérence les hypothèses faites dans le schéma microphysique et le schéma radiatif concernant la distribution de taille des gouttelettes nuageuses, à travers le calcul du rayon effectif. Davantage de détails concernant la méthode de perturbation sont donnés dans la section 5.2 du chapitre.

Chapitre 3

Description du modèle et des simulations numériques

Ce chapitre présente le cadre de travail utilisé durant la thèse. L'intégralité des résultats ont été produits avec le modèle à aire limitée AROME, qui est décrit brièvement dans la section 3.1. Une large part des tests de perturbation ont été effectués avec une version simplifiée unicolonne d'AROME, sur des cas d'étude idéalisés qui sont présentés dans les sections suivantes. Enfin, on présente très succinctement le modèle de recherche Mésos-NH ayant servi à produire des simulations LES sur lesquelles une partie du travail de thèse s'est appuyé.

3.1 Description générale du modèle AROME

AROME est le modèle actuellement utilisé opérationnellement à Météo France pour réaliser les prévisions jusqu'à 2 jours d'échéance. Il s'agit d'un modèle à aire limitée non-hydrostatique, construit sur les bases de deux modèles préexistants : la résolution des équations du mouvement (que l'on appelle la partie "dynamique" du modèle) est tirée du modèle ALADIN dans sa version non-hydrostatique (Bubnová et al. 1995), tandis que les schémas de paramétrisation proviennent du modèle de recherche Mésos-NH (Lafore et al. 1998 ; Lac et al. 2018).

Le modèle possède 12 variables pronostiques tridimensionnelles : la température T , les composantes horizontales du vent u et v , l'humidité spécifique q_v , les contenus spécifiques de différentes catégories d'hydrométéores (l'eau nuageuse q_c , la glace nuageuse q_i , la pluie q_r , la neige q_s et le graupel q_g), l'énergie cinétique turbulente (TKE) ainsi que deux variables \hat{q} et d reliées à la vitesse verticale et à la pression et dont les expressions découlent du choix de la coordonnée verticale dans AROME qui est une coordonnée de pression hybride. Les variables pronostiques sont complétées par une dernière variable bidimensionnelle, Π_s , correspondant à la pression hydrostatique à la surface. Pour la résolution des équations du mouvement, AROME utilise un schéma semi-implicite semi-Lagrangien et les calculs sont réalisés dans le domaine spectral, suite à une transformation de Fourier horizontale des principales variables pronostiques. Les calculs des schémas de paramétrisation sont eux en revanche effectués en point de grille.

La résolution spatiale du modèle est actuellement de 1.3 km. 90 niveaux verticaux sont utilisés, le premier se situant à environ 5 m d'altitude et le dernier à environ 34 km. La résolution temporelle est de 50 s. On peut distinguer plusieurs configurations d'AROME selon son domaine géographique : par exemple, AROME-France est utilisé pour les prévisions opérationnelles sur la France métropolitaine sur un domaine de 1996×1872 km couvrant la France, une

partie de l'Angleterre, de l'Espagne et d'autres pays frontaliers. Mais il existe aussi des versions AROME-OM (AROME-Outre mer) pour différents territoires ultra-marins, AROME-WMED centré sur la Méditerranée, etc. Dans ces travaux de thèse, seule la version AROME-France est utilisée.

AROME-France est actuellement lancé 8 fois par jour, à partir des analyses de 00, 03, 06, 09, 12, 15, 18 et 21 UTC. L'échéance maximale des simulations est de 51 h. L'analyse utilisée pour initialiser le modèle est issue d'un schéma d'assimilation de données 3DVar, et les conditions latérales sont fournies toutes les heures par les prévisions du modèle ARPEGE. Enfin, les conditions à la surface sont issues du couplage avec le modèle SURFEX.

3.2 MUSC : une version 1D du modèle

Une grande partie des travaux de thèse présentés dans ce manuscrit ont été réalisés dans un cadre simplifié, de façon à pouvoir tester et étudier l'impact de différentes méthodes de perturbation du modèle AROME avec un coût numérique relativement faible. Pour cela, une version uni-colonne d'AROME, appelée AROME-MUSC (MUSC : Modèle Unifié Simple Colonne ; Malardel 2008a,b) a été utilisée.

Avec la configuration MUSC, le domaine d'intégration du modèle est réduit à 4 colonnes adjacentes dans la direction méridienne. La présence des 4 colonnes permet aux calculs spectraux de la partie dynamique d'être effectués, mais les tendances produites sont nulles. L'état d'une colonne de modèle ne va donc évoluer que sous l'effet des paramétrisations physiques, de la prescription de conditions de surface ainsi que de termes de forçage qui peuvent être ajoutés aux tendances des variables pronostiques. MUSC est donc une configuration "pseudo-1D" mais si les forçages sont identiques pour les 4 colonnes, alors celles-ci restent identiques les unes aux autres tout au long de l'intégration.

L'évolution d'une variable pronostique X dans une colonne du modèle est déterminée par :

$$\frac{\partial X}{\partial t} = F + P,$$

où P correspond à la tendance calculée par les schémas de paramétrisation et F correspond aux forçages. Ces derniers peuvent être donnés directement sous la forme de tendances, le plus souvent des tendances d'advection (termes bleus dans les équations 3.1-3.4) mais parfois également des tendances radiatives (terme rouge de l'équation 3.3). Des forçages peuvent également être calculés à partir du vent géostrophique (termes en orange des équations 3.1 et 3.2) ou encore à partir d'une vitesse verticale (termes en vert dans les équations 3.1-3.4).

$$\left(\frac{\partial \bar{u}}{\partial t}\right)_{forcing} = \left(\frac{\partial \bar{u}}{\partial t}\right)_{adv} - \bar{w} \frac{\partial \rho \bar{u}}{\rho \partial z} + f(\bar{v} - v_g) \quad (3.1)$$

$$\left(\frac{\partial \bar{v}}{\partial t}\right)_{forcing} = \left(\frac{\partial \bar{v}}{\partial t}\right)_{adv} - \bar{w} \frac{\partial \rho \bar{v}}{\rho \partial z} - f(\bar{u} - u_g) \quad (3.2)$$

$$\left(\frac{\partial \bar{\theta}}{\partial t}\right)_{forcing} = \left(\frac{\partial \bar{\theta}}{\partial t}\right)_{adv} - \bar{w} \frac{\partial \rho \bar{\theta}}{\rho \partial z} + \left(\frac{\partial \bar{\theta}}{\partial t}\right)_{rad} \quad (3.3)$$

$$\left(\frac{\partial \bar{q}}{\partial t}\right)_{forcing} = \left(\frac{\partial \bar{q}}{\partial t}\right)_{adv} - \bar{w} \frac{\partial \rho \bar{q}}{\rho \partial z} \quad (3.4)$$

Souvent, les forçages ne sont donnés que pour un nombre limité d'échéances. Dans ce cas une interpolation linéaire est effectuée pour pouvoir calculer le terme de forçage à chaque pas de temps du modèle.

3.3 Simulations 1D de cas d'étude idéalisés

Plusieurs cas idéalisés de couche limite ont été simulés avec AROME-MUSC. Ces cas d'étude sont bien documentés dans la littérature et ont été utilisés dans le cadre d'intercomparaisons entre des modèles météorologiques (Brown et al. 2002; Siebesma et al. 2003; Chlond et al. 2004; Boutle et al. 2022). Le tableau 3.1 présente les principaux cas étudiés, qui sont décrits plus en détail dans les sections qui suivent.

	ARMCu	BOMEX	FIRE	LANFEX
Niveaux verticaux	90	90	105	90
Durée	15 heures	23 heures	∞	19 heures
Date	1997-06-21	1969-06-24	1987-07-14	2014-11-24
Conditions initiales (profils verticaux)				
Température	x	x	x	x
Humidité	x	x	x	x
Eau nuageuse			x	
Vent zonal	x	x	x	x
Vent méridien	x	x	x	x
Forçages				
Advection de température	x	x	x	
Advection d'humidité	x	x	x	
Vent géostrophique	x	x	x	
Subsidence verticale		x	x	
Forçage radiatif	x	x		
Flux de chaleur sensible	x	x		
Flux de chaleur latente	x	x		
Température de surface			x	x
Articles de référence				
	Brown et al. 2002	Siebesma et al. 2003	Duynkerke et al. 2004	Boutle et al. 2022

TABLE 3.1 – Réglages utilisés pour les simulations uni-colonnes des différents cas d'étude.

Pour chaque cas, les spécifications des conditions initiales et des forçages sont tirées d'articles de référence qui ont été utilisés pour construire les fichiers nécessaires à l'utilisation d'une configuration MUSC. Les fichiers des cas ARMCu, BOMEX et FIRE ont été récupérés à partir de l'environnement EMS (Environment for Musc Simulations) développé au Centre National de Recherches Météorologiques (CNRM, CNRS/Météo France - UMR3589) par Romain Roehrig (disponible sous [GitHub](#)¹), et les fichiers du cas LANFEX ont été obtenus auprès de Ian Boutle (Met Office).

Le choix des différents cas d'étude a été guidé par le type de situation météorologique représentée et par la disponibilité des données et des réglages adéquats d'AROME-MUSC pour effectuer les simulations. Les cas de cumulus ARMCu et BOMEX ont été choisis car ils sont principalement pilotés par les processus de turbulence et de convection peu profonde auxquels on s'intéresse dans le chapitre 4. Le cas de stratocumulus FIRE a permis d'étudier une situation plus complexe, avec davantage de processus physiques en jeu (effets radiatifs et microphysiques). Enfin, le cas LANFEX est un cas de brouillard, soit une situation météorologique à fort enjeu, encore difficilement simulée par les modèles météorologiques et très sensible aux phénomènes sous-maille et à leur paramétrisation (Ryerson et Hacker 2014; Boutle et al. 2022).

3.3.1 ARMCu

Le cas ARMCU est une idéalisation d'observations effectuées le 21 juin 1997 sur une zone d'environ 140 000 km² aux États-Unis, à cheval sur le Kansas et l'Oklahoma, dans le cadre de la campagne Atmospheric Radiation Measurement (ARM). Il s'agit d'un cas de convection peu profonde non précipitante où des cumulus se développent au sommet d'une couche limite convective continentale (Brown et al. 2002). Le cas démarre à 11:30 UTC et dure 15 heures, jusqu'au 22 juin 02:30 UTC, soit en heure locale (LT pour Local Time) : le 21 juin de 05:30 à 20:30.

La figure 3.1 montre l'évolution de plusieurs variables durant la simulation réalisée avec AROME-MUSC. On observe au tout début de la simulation une couche limite stable avec une inversion de température à la surface, puis celle-ci disparaît et la couche se réchauffe et s'étend verticalement jusqu'à 2500 m environ. De l'eau nuageuse apparaît après environ 4 heures de simulation, soit à 15:30 UTC (09:30 LT). Le nuage se développe petit à petit jusqu'à atteindre environ 2500 m d'altitude (et une épaisseur de plus de 1000 m). A partir de 21:30 UTC le nuage décroît et disparaît à 00:00 UTC. Ces résultats sont comparables à ceux des LES présentées par Brown et al. (2002).

Dans le cas ARMCu, le forçage à la surface est donné par des flux de chaleur latente et sensible (figure 3.1c), il n'y a donc pas besoin de couplage avec un modèle de surface. Les forçages restants sont des forçages d'advection et radiatifs (le schéma radiatif d'AROME n'est donc pas activé pour simuler ce cas). D'après Brown et al. (2002), les flux de chaleur à la surface représentent pour ce cas ARMCu le forçage le plus important et ce sont essentiellement eux qui contrôlent le cycle diurne.

La figure 3.2 montre les tendances issues de chaque schéma de paramétrisation d'AROME, moyennées entre +6 et +10 heures de simulation. On constate que les tendances sont maximales près de la surface, en lien avec les importants flux de chaleur à ce niveau. On constate également

1. <https://github.com/romainroehrig/EMS>

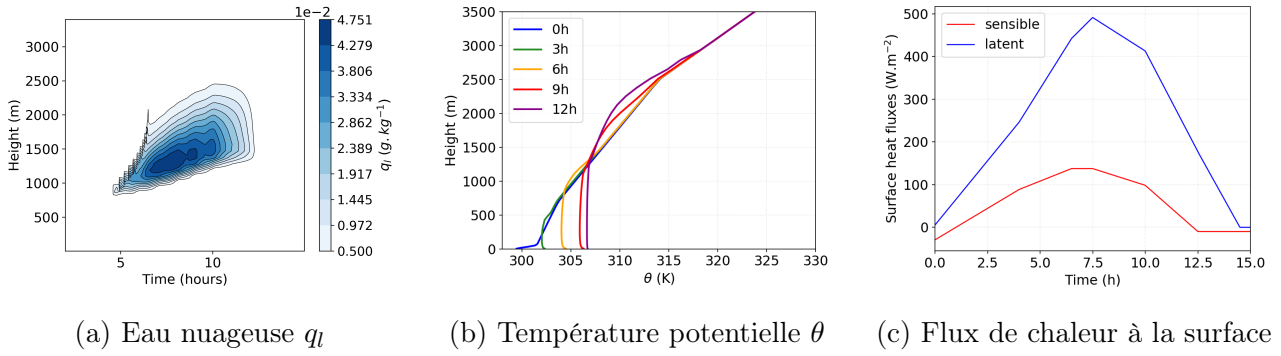


FIGURE 3.1 – Évolution temporelle des profils verticaux du contenu en eau nuageuse, de la température potentielle, ainsi que des flux de chaleur à la surface pour le cas ARMCu.

que le cas ARMCu est essentiellement gouverné par les paramétrisations de la turbulence et de la convection peu profonde. La tendance radiative (pour la température) est nulle car le schéma est désactivé, et les tendances de condensation sous-maille et de microphysique lente sont négligeables par rapport à la tendance totale. La turbulence distribue la chaleur issue des flux de surface, elle contribue donc à réchauffer la couche limite. La convection peu profonde assure un transport vertical de chaleur et d’humidité, elle contribue donc à refroidir la couche limite près de la surface et à la réchauffer plus haut.

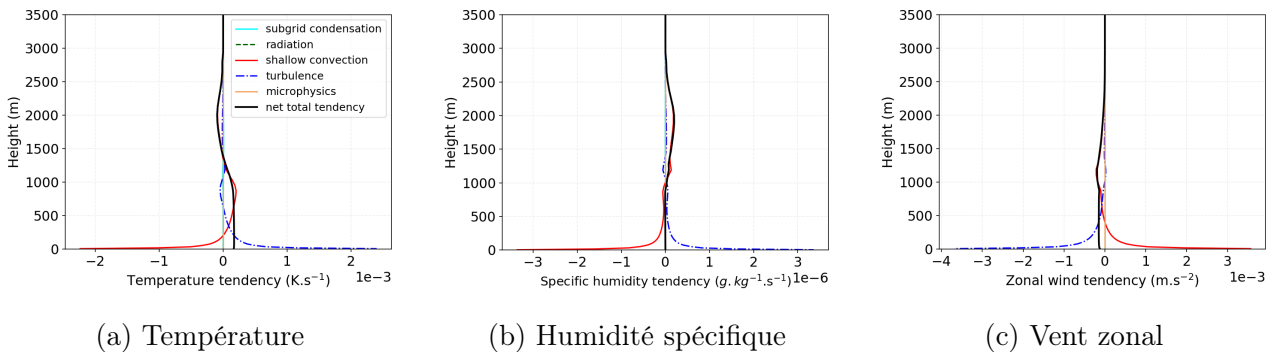


FIGURE 3.2 – Profils verticaux des tendances issues des schémas de paramétrisation d’AROME, moyennés sur l’intervalle +6-10 h pour le cas ARMCu.

3.3.2 BOMEX

BOMEX est un cas de cumulus non précipitants présents au sommet d’une couche limite océanique dans la région des Alizés au large de la Barbade. Ce cas est décrit dans Siebesma et Cuijpers (1995) et Siebesma et al. (2003) et est construit à partir d’observations provenant de la phase 3 de la campagne Barbados Oceanographic and Meteorological Experiment (BOMEX; Holland et Rasmusson 1973). Cette phase a pour particularité de présenter des conditions de grande échelle relativement constantes. Le cas idéalisé est initialisé à partir de radiosondages moyennés sur les journées du 22 et 23 juin 1969, et utilise des termes de forçage constants, correspondant à l’estimation des forçages moyens sur la période du 22 au 26 juin (Siebesma et Cuijpers 1995).

Le cumulus simulé par AROME-MUSC s’étend de 500 à 2000 m environ (figure 3.3a). On

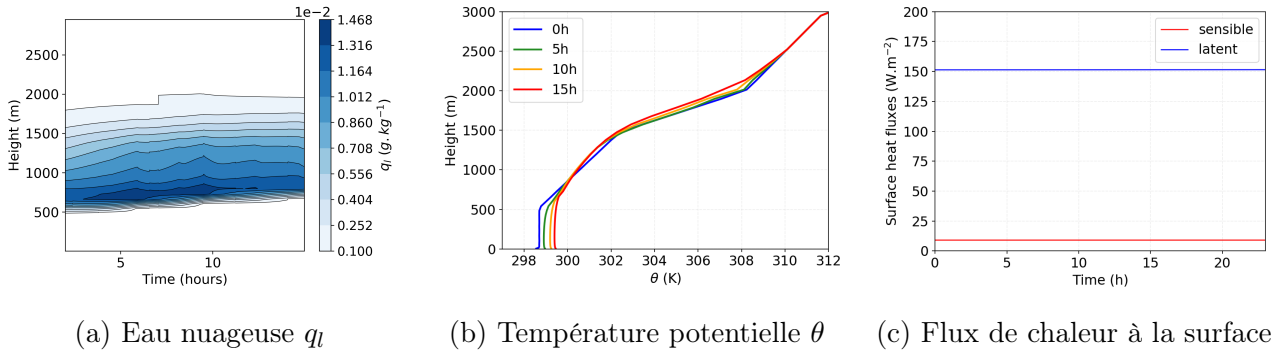


FIGURE 3.3 – Évolution temporelle des profils verticaux du contenu en eau nuageuse, de la température potentielle, ainsi que des flux de chaleur à la surface pour le cas BOMEX.

constate que contrairement au cas ARMCu, ce cas ne présente pas de cycle diurne. Après une rapide période d’ajustement (environ 1 heure, non visible sur la figure 3.3a) durant laquelle le nuage apparaît, le système reste dans un état quasi-stationnaire. Ceci vient du fait que les forçages latéraux et à la surface sont constants (figure 3.3c).

Les tendances des différents schémas de paramétrisation d’AROME sont très semblables à celles du cas ARMCu (figure 3.4). Elles sont néanmoins moins fortes (environ un ordre de grandeur de différence pour la température), ce qui peut s’expliquer par les flux de surface moins élevés pour BOMEX qui est un cas océanique.

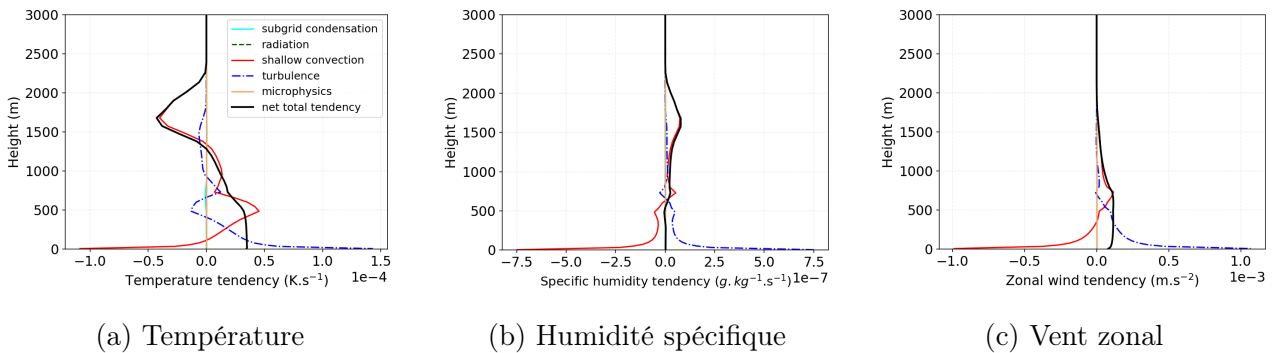


FIGURE 3.4 – Profils verticaux des tendances issues des schémas de paramétrisation d’AROME, moyennés sur l’intervalle +2-23 h pour le cas BOMEX.

3.3.3 FIRE

Le cas suivant correspond à un autre type de situation : une couche limite océanique subtropicale surmontée de stratocumulus. Il a été construit dans le cadre du projet EUROCS (European Project on Cloud Systems in Climate Models), à partir d’observations effectuées au large de la Californie en juillet 1987 durant la campagne First ISCCP (International Satellite Cloud Climatology Project) Regional Experiment (FIRE) (Albrecht et al. 1988). Le cas idéalisé présenté par Duynkerke et al. (2004) s’appuie sur des conditions initiales et des forçages idéalisés, mais proches des conditions observées le 14 et le 15 juillet 1987.

Pour le cas FIRE, les forçages consistent en des forçages d’advection, de vent géostrophique, une subsidence de grande échelle et une température de surface imposée. Ainsi, contrairement aux deux cas précédents, il n’y a pas de flux de surface prescrits pour FIRE, et le modèle SURFEX doit donc être couplé à AROME-MUSC afin de les calculer. Par ailleurs, les forçages atmosphériques ne contiennent pas de forçage radiatif donc le schéma radiatif d’AROME est activé pour simuler ce cas.

Le contenu en eau nuageuse simulé par AROME-MUSC est présenté sur la figure 3.5. Les forçages atmosphériques du cas FIRE étant constants, le cycle diurne que l’on observe est uniquement dû à l’action du schéma radiatif. Ceci permet d’ailleurs de produire une simulation pour un temps arbitrairement long : ainsi la simulation de la figure 3.5a dure 109 heures et présente plusieurs successions de jour/nuit. On constate sur cette simulation que les cycles des premières 48 heures sont différents des cycles suivants qui eux sont pratiquement identiques. Pour toutes les expériences faites avec ce cas, il a donc été décidé de ne pas utiliser ces premières 48 heures. Dans toute la suite, les simulations du cas FIRE sont des simulations de 85 heures que l’on n’étudie qu’à partir de l’échéance +48 h. Pour des raisons de lisibilité, on omettra dans la présentation des résultats les premières 48 heures et les simulations seront considérées commencer à +0 h (comme sur l’axe des abscisses de la figure 3.5b) et aller jusqu’à l’échéance +37 h. Dans le cadre des tests de représentation de l’erreur modèle, si des perturbations sont ajoutées, elles ne le sont qu’à partir de l’échéance +48 h.

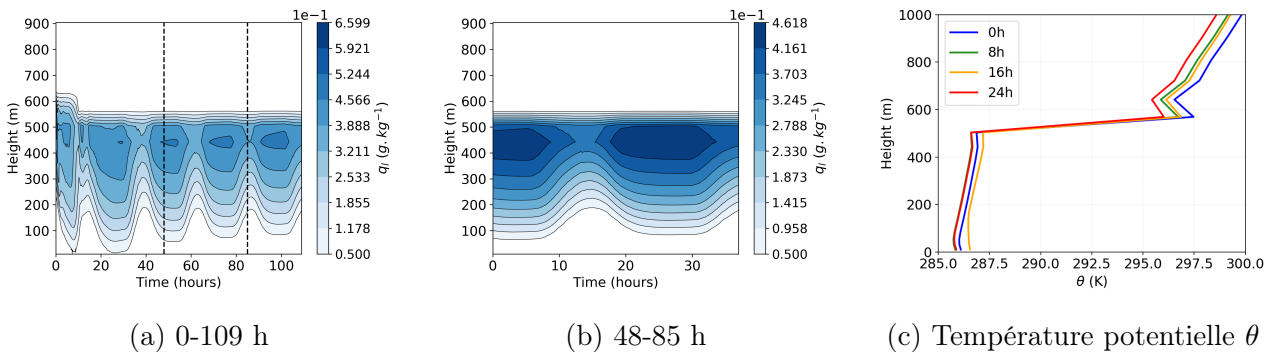


FIGURE 3.5 – Évolution temporelle des profils verticaux d’eau nuageuse et de température potentielle pour le cas FIRE. Les barres verticales sur la figure a indiquent l’intervalle de temps couvert par la figure b.

Les figures 3.5a,b montrent que le sommet des nuages simulé par AROME-MUSC se trouve aux alentours de 550 m, ce qui est inférieur aux observations et aux simulations LES présentées par Chlond et al. (2004) et Duynkerke et al. (2004). Duynkerke et al. (2004) soulignent néanmoins déjà ce défaut dans plusieurs simulations de modèles uni-colonne. La base des nuages est également trop basse dans la simulation AROME-MUSC comparée aux observations.

Les phénomènes physiques en jeu dans le cas FIRE sont différents de ceux à l’origine de la formation des cumulus ARMCu et BOMEX, ce qu’illustrent les profils des tendances de la figure 3.6. On voit que la tendance de température est principalement affectée par le schéma radiatif et le schéma de condensation sous-maille. La formation et le maintien des stratocumulus dépendent en effet du refroidissement radiatif ayant lieu au sommet des nuages qui entraîne un mélange efficace de toute la couche limite (Wood 2012).

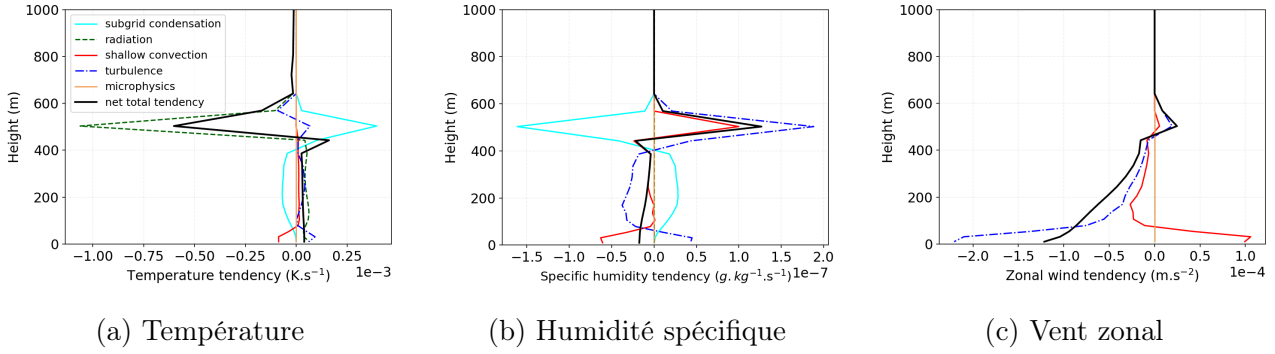


FIGURE 3.6 – Profils verticaux des tendances issues des schémas de paramétrisation d'AROME, moyennés sur l'intervalle +48-85 h pour le cas FIRE.

Les tendances de la figure 3.6 sont moyennées sur 37 heures mais il existe un cycle diurne dans leur évolution. Durant les phases de nuit, le refroidissement radiatif est plus intense, ce qui favorise une couche de nuage plus épaisse. En journée, le réchauffement des nuages par le rayonnement solaire affaiblit les effets du refroidissement radiatif et donc le mélange. Avec des mouvements turbulents plus faibles, les apports en humidité au niveau de la couche nuageuse sont réduits et celle-ci s'affine (Wood 2012).

3.3.4 LANFEX

LANFEX (Local And Non-Local Fog Experiment) est une campagne de mesures et d'expérimentations menée au Royaume-Uni entre 2014 et 2016 (Price et al. 2018).

Le cas d'étude "LANFEX" utilisé pour ces travaux de thèse correspond à la première période d'observation intensive (IOP1 pour Intensive Observation Period 1) de la campagne, qui a eu lieu à Cardington entre le 24 et le 25 novembre 2014. Durant cette période, une couche de brouillard radiatif peu épais s'est développée un peu avant 18:00 UTC le 24 novembre, et a persisté durant la nuit jusqu'au lendemain aux environs de 08:00 UTC (Smith et al. 2018 ; Price 2019). Durant la majeure partie de la nuit, la couche de brouillard est stable et a une épaisseur d'environ 100 mètres. Vers 06:00 UTC, elle devient instable et la turbulence augmente dans la couche de brouillard qui s'épaissit. Le lever du soleil intervient à 07:30 UTC et le brouillard évolue en stratus après 08:00 UTC.

Les profils de température et d'eau nuageuse obtenus avec AROME-MUSC sont présentés sur la figure 3.7. De même que dans les études de Boutle et al. (2018), Smith et al. (2018) et Poku et al. (2019), la simulation AROME-MUSC démarre à 17:00 UTC à partir d'un profil de radiosondage. Elle dure 19 heures et utilise des observations de température à la surface comme forçage. Aucun forçage atmosphérique n'est spécifié en cohérence avec l'étude d'intercomparaison de Boutle et al. (2022), sachant qu'il s'agit d'un cas durant lequel l'advection était faible et les conditions de grande échelle relativement stationnaires, de sorte que leur impact est jugé limité (Boutle et al. 2018 ; Smith et al. 2018). Par ailleurs, dans cette simulation, la concentration totale en gouttelettes nuageuses, réglée par défaut à 300 gouttes/cm³ sur les continents dans AROME, a été abaissée à 50 gouttes/cm³ pour suivre les spécifications de Boutle et al. (2022).

La figure 3.7a montre le développement d'une couche de brouillard, qualitativement simi-

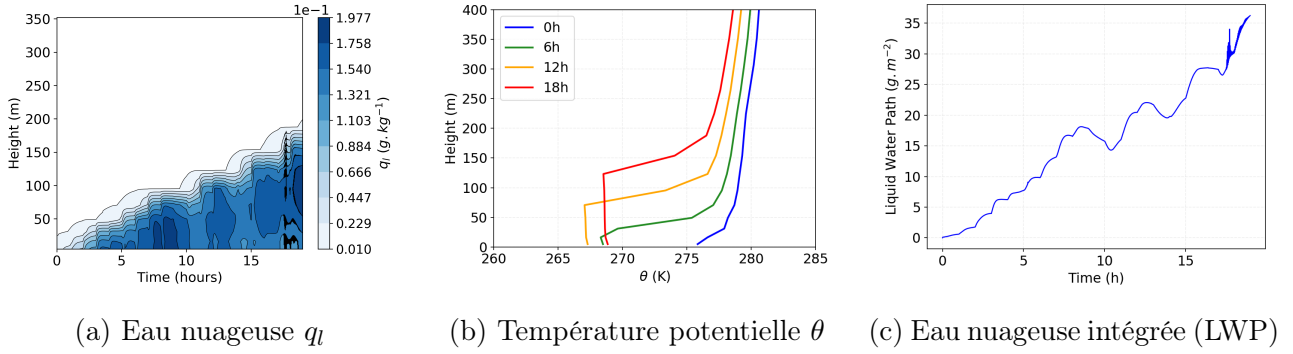


FIGURE 3.7 – Évolution temporelle des profils verticaux d’eau liquide nuageuse et de température potentielle, et évolution du contenu en eau liquide intégré (LWP) pour le cas LANFEX. Les simulations démarrent à 17:00 UTC. Les échéances +6h, +12h et +18h correspondent à 23:00, 05:00 et 11:00 UTC respectivement.

laire à celle obtenue par Smith et al. (2018), bien qu’avec un contenu en eau nuageuse plus faible. L’évolution du contenu en eau intégré verticalement (LWP pour Liquid Water Path) est également cohérente avec les résultats présentés par Boutle et al. (2022) : la simulation AROME-MUSC se trouve dans l’intervalle des valeurs prévues par différents modèles unicolonne et surestime le LWP par rapport aux observations (voir la figure 1d de Boutle et al. 2022). Par ailleurs, on constate sur les profils de température potentielle que la couche de brouillard devient instable après environ 6 heures de simulation, soit à 23:00 UTC, ce qui est trop tôt par rapport aux observations. Cette tendance à former une couche de brouillard instable et mélangée trop tôt a déjà été observée avec d’autres modèles unicolonne (Boutle et al. 2018 ; Boutle et al. 2022).

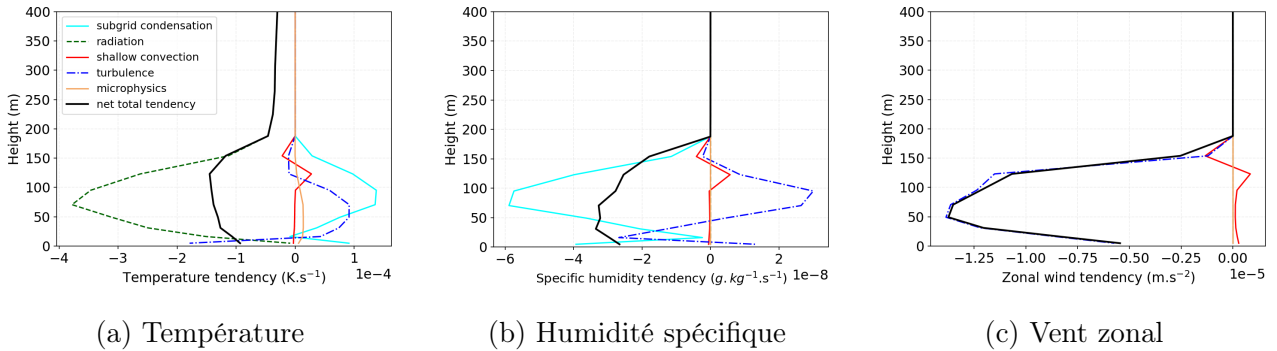


FIGURE 3.8 – Profils verticaux des tendances issues des schémas de paramétrisation d’AROME, moyennés sur l’intervalle +0-19 h pour le cas LANFEX.

La figure 3.8 illustre les différents processus en jeu dans la formation du brouillard radiatif. Les principaux schémas de paramétrisation contribuant à l’évolution de la température sont le schéma radiatif, le schéma de condensation sous-maille et la turbulence. Le schéma radiatif modélise le refroidissement radiatif ayant lieu à la surface puis au sommet de la couche de brouillard au fur et à mesure qu’elle se développe. La baisse de température entraîne une saturation de la masse d’air et l’apparition de gouttelettes nuageuses est modélisée par le schéma de condensation sous-maille. Enfin, le schéma de turbulence tend à homogénéiser la couche de brouillard. On constate également une tendance moyenne non nulle du schéma de

convection peu profonde. Ceci correspond en réalité à une activation épisodique de ce schéma lorsque la couche de brouillard devient instable, qui cause les variations rapides du contenu en eau que l'on peut observer sur les figures 3.7a,c.

3.4 Le modèle Méso-NH et les simulations LES

Méso-NH est un modèle atmosphérique initialement développé et maintenu conjointement par le Laboratoire d'Aérodynamique (CNRS/UPS/IRD - UMR 5560) et le CNRM, et qui est dédié à la recherche météorologique (Lafore et al. 1998 ; Lac et al. 2018). Il s'agit d'un modèle à aire limitée non-hydrostatique, prévu pour pouvoir être utilisé à des résolutions pouvant aller de quelques dizaines de kilomètres à la dizaine de centimètres. Le modèle est discrétisé selon une grille Arakawa C et les équations sont résolues en point de grille (il n'y a pas de représentation spectrale comme dans AROME). Méso-NH possède de nombreuses options de paramétrisation possibles, parmi lesquelles on trouve les schémas de paramétrisation utilisés opérationnellement dans le modèle AROME.

Dans ces travaux de thèse, des LES produites par le modèle Méso-NH ont été utilisées afin d'évaluer et calibrer certaines méthodes de perturbation appliquées au modèle AROME. Elles concernent trois cas idéalisés décrits précédemment (ARMCu, BOMEX et FIRE) et ont déjà été utilisées dans le cadre de différentes publications (Couvreur et al. 2010 ; Honnert et al. 2011 ; Perraud et al. 2011 ; Jam et al. 2013 ; Brient et al. 2019). Les données ont toutes été fournies par Fleur Couvreur (CNRM), et les caractéristiques des différentes simulations sont résumées dans le tableau 3.2.

	ARMCu	BOMEX	FIRE
Domaine	$6.4 \times 6.4 \times 4 \text{ km}^3$	$12.8 \times 12.8 \times 4 \text{ km}^3$	$25.6 \times 25.6 \times 1.2 \text{ km}^3$
Résolution	$\Delta x = \Delta y = \Delta z = 25 \text{ m}$	$\Delta x = \Delta y = \Delta z = 25 \text{ m}$	$\Delta x = \Delta y = 50 \text{ m}, \Delta z = 10 \text{ m}$
Durée	15 h (05:30 – 20:30 LT)	14 h	24 h (00:00 – 24:00 LT)
Disponibilité des fichiers	5 min	horaire	horaire

TABLE 3.2 – Caractéristiques des différentes LES utilisées.

Le contenu en eau nuageuse moyenné sur tout le domaine pour chaque LES est présenté sur la figure 3.9. On constate que les résultats sont plutôt cohérents avec ceux d'AROME-MUSC : l'extension verticale des nuages est similaire pour les cas ARMCu et BOMEX, ainsi que les cycles diurnes d'ARMCu et FIRE. Le contenu en eau moyen est en revanche un peu plus faible pour BOMEX dans la LES qu'avec AROME-MUSC, et le stratocumulus de la LES du cas FIRE est plus élevé que celui d'AROME-MUSC (ce qui est davantage en accord avec les observations et d'autres LES de ce cas d'étude ; Chlond et al. 2004 ; Duijkerke et al. 2004).

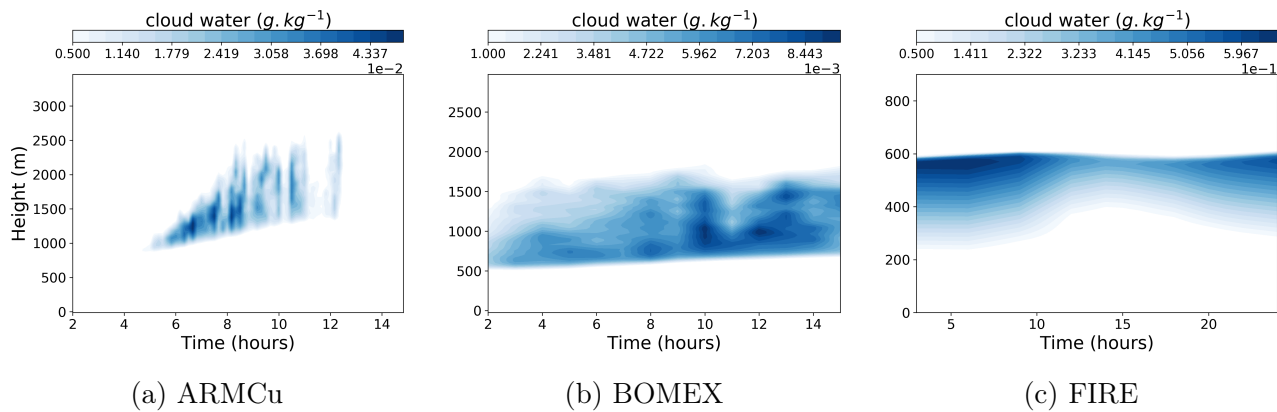


FIGURE 3.9 – Graphes hovmoller de l'eau liquide nuageuse moyennée sur tout le domaine des LES des cas ARMCu, BOMEX et FIRE.

3.5 Méthodologie pour la production d'ensembles de simulations 1D

Pour répondre à la problématique de recherche, cette thèse s'appuie en grande partie sur l'étude d'ensembles de simulations 1D des cas d'étude précédemment décrits, produits grâce à différentes méthodes de perturbation du modèle AROME. Il faut noter qu'aucune perturbation des conditions initiales, des forçages atmosphériques ou des conditions de surface n'a été introduite. La différence entre les simulations provient seulement des perturbations appliquées au modèle. Pour tous les résultats présentés dans les chapitres suivants 4 et 5, les ensembles sont en général constitués de 50 à 100 simulations d'un même cas.

Chapitre 4

Perturbation des processus physiques d'AROME à l'aide de simulations LES

Ce chapitre présente les résultats de l'implémentation, dans AROME-MUSC, de méthodes POP (voir section 1.4.4) pour perturber les schémas de paramétrisation de la turbulence et de la convection peu profonde. Des travaux exploratoires sont également rapportés concernant la perturbation de l'ajustement à la saturation. La particularité des méthodes présentées dans ce chapitre est qu'elles s'appuient fortement sur les données de simulations LES, soit pour le réglage des perturbations, soit pour l'évaluation des résultats obtenus.

4.1 Perturbation de la turbulence et de la convection peu profonde

Dans cette première section, on rapporte l'implémentation de deux méthodes POP, décrites dans les chapitres 1 et 2, pour perturber la paramétrisation de la turbulence et de la convection peu profonde d'AROME. Il s'agit du schéma de perturbation de la turbulence de Kober et Craig (2016) et du schéma de convection stochastique de Sakradzija et al. (2015). Ces travaux ont été publiés dans le *Quarterly Journal of the Royal Meteorological Society*.

4.1.1 Introduction et résumé de l'article

Comme on l'a vu dans le chapitre 2, les méthodes de Kober et Craig (2016) et Sakradzija et al. (2015) visent à représenter un "bruit d'échantillonnage" dû à la présence en nombre limité des structures sous-maille turbulentes et convectives. Les perturbations qu'elles proposent ont des caractéristiques (occurrence, amplitude, corrélations) qui proviennent des choix de modélisation des phénomènes sous-maille (événements aléatoires et indépendants, temps de vie, etc), contrairement aux méthodes comme SPPT, SPP, où elles sont arbitraires. L'objectif de l'article est d'étudier l'impact de ces deux méthodes POP sur le modèle AROME-MUSC et de le comparer à celui de la méthode opérationnelle SPPT. Étant donné que SPPT est une méthode généraliste, alors qu'au contraire les méthodes POP ciblent des sources d'incertitude précises, une référence est par ailleurs construite à l'aide de simulations LES pour tenter d'évaluer si les méthodes POP atteignent ou non leurs objectifs.

Définition d'une référence

L'utilisation de LES comme référence repose sur les constats suivants :

- les LES décrites dans le chapitre 3 ont une résolution suffisante pour résoudre complètement ou en partie les tourbillons turbulents et les ascendances convectives ;
- leurs domaines sont tous plus grands que la maille d'AROME ;
- les forçages sont homogènes sur tout le domaine et sont les mêmes que ceux prescrits à AROME-MUSC.

Ainsi, si l'on découpe les domaines LES en petits sous-domaines de la taille de la maille d'AROME, on obtient des "mailles" ayant toutes subi le même forçage mais qui sont néanmoins différentes du fait de la variabilité de fine échelle résolue par les LES. Chaque "maille" représente une version possible de la réponse turbulente et convective associée au forçage, et nous considérons que c'est ce que cherchent à représenter les deux méthodes POP. La variabilité entre les valeurs de T , q , u et v moyennées sur les sous-domaines est donc utilisée comme référence de la variabilité que cherchent à représenter les deux méthodes POP à travers leurs perturbations stochastiques. Ce découpage et ce moyennage sur des petits sous-domaines est une procédure appelée "coarse-graining", que l'on illustre sur la figure 4.1.

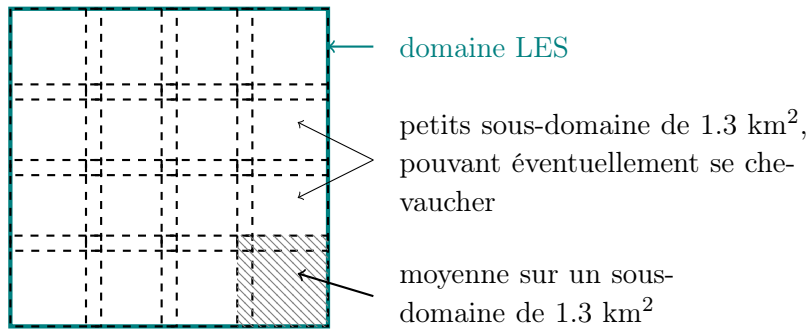


FIGURE 4.1 – Procédure de coarse-graining appliquée aux champs LES.

Implémentation de la méthode de Kober et Craig (2016) (KC16)

L'implémentation de cette méthode dans AROME-MUSC est assez directe, car toutes les informations nécessaires sont présentes dans le modèle. Le réglages des paramètres l_{eddy} et τ_{eddy} (voir l'équation 2.10) suivent ceux de Hirt et al. (2019). La variance sous-maille, qui règle l'amplitude des perturbations, est calculée par le schéma de turbulence de Cuxart et al. (2000) dans sa version 1D. En turbulence 1D, les gradients horizontaux sont négligés, et, dans le cadre de simulations uni-colonne, les phénomènes de convergence ou de divergence horizontale du vent ne sont pas permis, de sorte que la composante verticale du vent résolue est nulle (ce qui n'empêche pas l'existence de gradients et de mouvements verticaux sous-maille). Par conséquent, la variance des composantes horizontales du vent se réduit à :

$$\overline{u'^2} = \overline{v'^2} = \frac{2}{3}\bar{e}$$

Pour la température et l'humidité, les équations 2.8 et 2.9 sont utilisées, avec des gradients verticaux uniquement. Par ailleurs, le bruit η de l'équation 2.10 est choisi comme étant un processus autorégressif d'ordre 1, de variance 1 et avec une corrélation temporelle de 10 min, en accord avec Hirt et al. (2019).

Implémentation de la convection stochastique

Comme mentionné dans le chapitre 2, le schéma proposé par Sakradzija et al. (2015) est relativement complexe et coûteux numériquement. Pour la réalisation de premiers tests dans AROME, nous avons donc choisi de simplifier grandement la méthode. Nous sommes partis du constat que le schéma de convection stochastique de Sakradzija et al. (2015) diffère d'un schéma déterministe seulement par la condition de fermeture en flux de masse, qui est rendue stochastique grâce à la population aléatoire de nuages. Une version simplifiée de ce schéma consisterait donc à perturber la condition de fermeture directement, sans passer par la génération de nuages. Pour cela, les distributions du flux de masse total obtenues par Sakradzija et al. (2015) (voir leur figure 9) auraient pu être utilisées, mais elles correspondent à un niveau de référence situé à la base des nuages, et non à la surface comme la condition de fermeture du schéma de Pergaud et al. (2009) utilisé par AROME. Nous avons donc choisi d'estimer ces distributions à l'aide des données LES, afin de régler les perturbations à appliquer à la condition de fermeture M_{surf} de Pergaud et al. (2009). Pour cela, la même procédure de coarse-graining que décrite précédemment a été employée pour évaluer le flux de masse moyen sur des petits sous-domaines de la taille de la maille d'AROME. Le calcul du flux de masse en lui-même a été effectué grâce à la méthode d'identification d'updrafts de Couvreux et al. (2010), et l'algorithme développé par Najda Villefranque et Florent Briant (CNRM; <https://gitlab.com/tropics/objects>). La méthode consiste à utiliser dans les simulations LES un traceur passif émis à la surface avec un temps de vie limité. Ce traceur est transporté par les ascendances convectives et le calcul de l'anomalie de concentration en traceur par rapport à la moyenne du domaine permet d'identifier ces dernières, comme cela est illustré sur la figure 4.2.

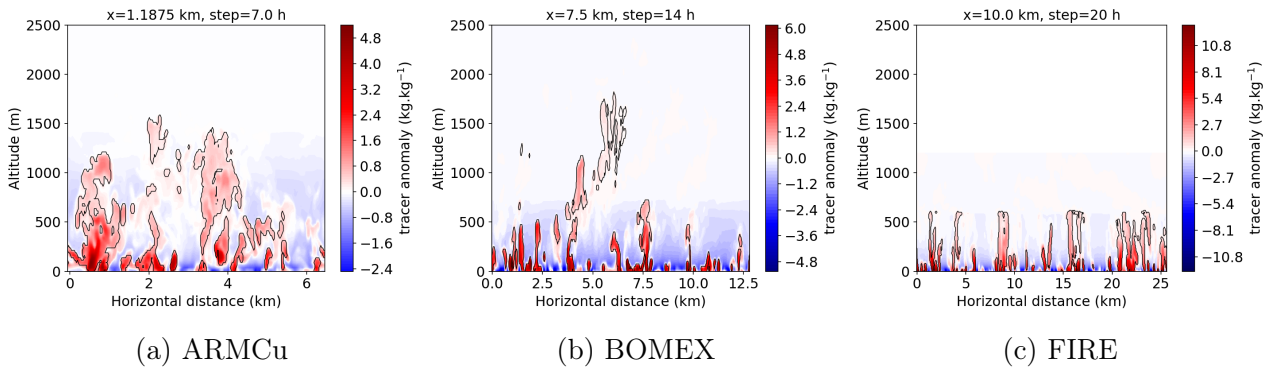


FIGURE 4.2 – Coupes verticales de l'anomalie de concentration en traceur avec les contours des objets identifiés grâce à l'algorithme de N. Villefranque et F. Briant pour les cas d'étude ARMCu, BOMEX et FIRE.

L'identification des ascendances permet de calculer leur vitesse verticale moyenne et la fraction de la maille qu'elles occupent, ce qui permet d'obtenir le flux de masse moyen, et ce pour chaque sous-domaine et chaque niveau vertical. L'histogramme des valeurs de flux de masse proche de la surface permet alors d'estimer la distribution des perturbations à appliquer à la condition de fermeture M_{surf} (figure 2 et équation 3 de l'article).

Synthèse des résultats

Les résultats présentés dans l'article proviennent de différents ensembles de simulations AROME-MUSC qui ont été produits pour les trois cas d'étude ARMCu, BOMEX et FIRE

avec différentes méthodes de perturbation :

- SPPT sur les tendances de turbulence et de convection peu profonde
- KC16
- convection stochastique
- KC16 + convection stochastique

Étant donné que les méthodes POP n'agissent que sur deux schémas de paramétrisation (turbulence et convection peu profonde), le choix a été fait de restreindre les perturbations SPPT à la somme des tendances de ces deux schémas pour rendre la comparaison plus équitable.

Les principaux résultats de l'étude montrent que les deux méthodes POP atteignent partiellement leurs objectifs en terme de dispersion produite dans les ensembles de simulations AROME-MUSC. La comparaison avec la référence construite à partir des LES montre trop peu de dispersion de température et d'humidité, notamment à l'intérieur des nuages. En revanche, la méthode KC16 produit des dispersions de vent fortes, comparables à la référence des LES, et comparables voir supérieures à la méthode SPPT. La méthode de convection stochastique quant à elle produit des dispersions intéressantes en terme d'eau nuageuse. La méthode SPPT reste globalement la plus efficace pour produire de la variabilité entre les simulations AROME-MUSC, mais comme cela est souligné dans l'article, cette méthode est construite pour représenter de multiples sources d'incertitude contrairement aux méthodes POP, ce résultat n'est donc pas surprenant.

4.1.2 Article

Les pages suivantes correspondent à la version de l'article tel qu'accepté par l'éditeur. La version mise en forme et définitive peut-être consultée via le lien suivant :

<https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/qj.4242>

Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction.

Axelle Fleury^{1*}, François Bouttier¹, Fleur Couvreur¹

¹ CNRM, University of Toulouse, Météo-France, CNRS, Toulouse, France

* Correspondence to: Axelle Fleury, CNRM, Météo-France, 42 Av. Gaspard Coriolis, 31057 Toulouse, France. E-mail: axelle.fleury@meteo.fr

Key-words: coarse-graining, ensemble weather prediction, large-eddy simulations, model uncertainty, shallow convection, stochastic parametrization, turbulence.

Acknowledgments: This study was financed by the French Government through Météo-France, CNRS and the University of Toulouse. It benefited from the LEFE/MANU IPSIPE research grant of CNRS/INSU.

Weather forecasting nowadays often requires some estimation of uncertainties associated with the output of meteorological models, in order to better inform decision making, especially in the context of intense weather events. Ensemble prediction systems provide such information through sets of possible scenarios which are designed to represent various uncertainty sources, including model uncertainties. A wide variety of methods have been proposed to estimate model uncertainties, among which perturbation methods targeting uncertain processes are a promising research field. In this study, we focus on the representation of small-scale variability by process-oriented perturbation schemes applied to two key physical processes, namely turbulence and shallow convection. The perturbations are applied to a single-column version of the convection-permitting AROME model, in three idealized boundary-layer cases. Large-eddy simulations (LESs) of the same cases serve as a reference for the subgrid variability that has to be represented, and the results are also compared to those given by the Stochastically Perturbed Parametrization Tendencies (SPPT) method, which is a method commonly used by weather forecast centers to represent model uncertainty. The spread produced by our process-oriented perturbations of turbulence and shallow convection does not represent all the small-scale variability implied by the LESs for temperature and humidity. However, it is of a similar order of magnitude for the wind, thanks to perturbations generated by the stochastic turbulence scheme. The dispersion is structurally different from what is obtained with SPPT. It is non-negligible in the lower levels, where SPPT perturbations are usually suppressed because of numerical instabilities, indicating a possible complementarity between the schemes.

1. Introduction

Probabilistic forecasts constitute one of the products commonly used to assess the forthcoming weather, and are becoming increasingly valuable in many areas, including the prediction of extreme meteorological events. They result from the concept of ensemble prediction, which has been developed over several decades as a way to estimate the degree of confidence one can have in a forecast. Predictions of weather forecasting systems are

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/qj.4242](https://doi.org/10.1002/qj.4242)

This article is protected by copyright. All rights reserved.

indeed necessarily imperfect because of various errors arising, for example, from incomplete knowledge of the initial state of the atmosphere, or from approximations in the modeling of atmospheric processes. They can have a greater or lesser impact depending on the situation, time and location. Instead of a single forecast, an ensemble prediction system provides several forecasts, called members, thanks to random perturbations and therefore provides a continuous, prognostic measure of uncertainty. The random perturbations are designed to represent known error sources, among which those contained in initial and boundary model conditions have been shown to play a major role (Lorenz 1965). In addition, we know that the errors of the model design itself, called “model errors”, must be addressed as well in order to account for the entire forecast uncertainty (Buizza et al. 2005, Wilks 2005, Houtekamer 2009). They are currently the target of active research.

One important source of model uncertainty lies in the parametrization of subgrid processes (Palmer et al. 2009), which are all the phenomena whose time and space scales are smaller than those used by the model but need to be parametrized because their effect on the resolved fields is not negligible. Such phenomena can be complex, imperfectly known and thus challenging to represent. Different strategies have been proposed to deal with parametrization uncertainty, like the ‘multiphysics’ approach, which is based on the idea of combining different sets of existing parametrizations (Stensrud et al. 2000, Charon et al. 2010, Berner et al. 2011, Duda et al. 2014, Jankov et al. 2017), or the broad category of stochastic perturbations, where stochastic terms are introduced in the model, with various purposes and behaviors (Buizza et al. 1999, Lin and Neelin 2002, Shutts 2005, Dorrestijn et al. 2013, Baker et al. 2014, among others). The common feature of these methods is that they can produce, for the same resolved state, a set of parametrized subgrid effects, thus allowing the uncertainty of the parametrizations to be represented. Nevertheless, each method targets the sources of uncertainty with different degrees of precision, introducing noise at various levels of the parametrization schemes.

A commonly used stochastic perturbation method is the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme introduced by Buizza et al. (1999). In this method, a ‘bulk’ uncertainty coming from the parametrizations is represented by multiplying the total parametrization tendency by a random field, with temporal and spatial correlations. It has proved efficient in producing ensemble spread in several ensemble systems (Palmer et al. 2009, Bouttier et al. 2012, Batté and Doblas-Reyes 2015, Berner et al. 2015), and has the advantage of remaining relatively simple, and keeping the balance between the tendencies of the different parametrization schemes. However, its formulation implies hypotheses that may not always be justified, such as the proportionality of the error to the tendency, the correlation of the errors between variables, and the compensation effects induced by the use of the sum of all the parametrization tendencies.

More recently, new stochastic perturbation schemes have been developed which seek a more physically justified approach. For example, efforts have been undertaken to perturb free input parameters, instead of the parametrization output tendencies (Baker et al. 2014, Christensen et al. 2015, Ollinaho et al. 2017, Lang et al. 2021). These perturbation schemes, called Random Parameters (RP) or Stochastically Perturbed Parametrizations (SPP)

schemes, address the problem of parameter uncertainty inside the parametrization schemes, and have the advantage to act on specific physical processes.

We can identify another category of ‘physically-based’ perturbation schemes which aims specifically at representing ‘random errors’ coming from the spatial and temporal heterogeneity of the physical processes taking place in the atmosphere, and their representation in a model with a given grid size. They address the assumption of a deterministic link between resolved state and some subgrid state characteristics, which is made in many parametrization schemes. Indeed, because of the apparent randomness of some fine-scale phenomena, such as turbulence, and the multiple interactions between physical processes, two grid boxes with the same resolved state may not have the exact same subgrid state and feedback from subgrid processes. These fluctuations around an average contribution from subgrid processes may be negligible in the case of low resolutions, where the sample size of the subgrid phenomena present in a model grid box is large. With finer resolutions however, the sample size can become very small (e.g. the number of cumulus clouds in a 1 km² grid box) and the deterministic link between the resolved and subgrid state is more clearly violated. These subgrid state fluctuations may have an impact on resolved scales and therefore are likely to be a source of uncertainty (Lin and Neelin 2002). Stochastic schemes have therefore been proposed to represent it.

Some schemes target the variability linked to convective updrafts and downdrafts: they can consist of adding a noise to a variable diagnosed by a deterministic scheme (Lin and Neelin 2000) or of rendering key aspects of the parametrization stochastic: a stochastic trigger for deep convection is studied by Rochetin et al. (2014), Bengtsson et al. (2021) propose to use a cellular automata to generate subgrid plumes and derive stochastic formulations of the entrainment, triggering and mass-flux closure, and Craig and Cohen (2006) and Sakradzija et al. (2015) propose a stochastic mass-flux closure condition for deep and shallow convection respectively, further developed or tested in subsequent studies (Plant and Craig 2008, Groenemeijer et al. 2012, Sakradzija et al. 2016, Sakradzija and Klocke 2018, Machulskaya and Seifert 2019). These last schemes explicitly consider the limited number of clouds inside a model grid box and their heterogeneous characteristics.

Other schemes target the heterogeneity of small-scale eddies represented by turbulence schemes. Kober and Craig (2016) and Hirt et al. (2019) propose to perturb the output tendencies of the turbulence scheme in the COSMO model by adding a noise controlled by subgrid variances diagnosed by the turbulence scheme. Clark et al. (2021) study the representation of eddies as random events following a Poisson distribution inside a bulk model of the convective boundary layer.

For simplicity, these kinds of schemes will be referred to as ‘process-oriented’ schemes in this paper. They have the advantage of relying on physical arguments and of targeting a source of uncertainty – small-scale variability of physical processes – which is likely to become increasingly important to consider as the resolution of numerical weather prediction (NWP) models increases.

The evaluation of perturbation methods used to produce ensembles can be difficult, depending on the kind of uncertainty they are designed to represent. They can be validated by computing statistical scores on large data sets of real case simulations, but this does not allow one to look in detail at the effects of the perturbations on the model physics, and can

Accepted Article

hide contrasting results between different meteorological situations and compensation effects. In this study, we choose to work in an idealized framework using single-column model (SCM) simulations of several idealized cases, compared to large-eddy simulations (LESs). Taking LESs as a reference to investigate boundary-layer situations and evaluate or improve parametrization schemes has already been done in a number of studies (Ayotte et al. 1996, Duynkerke et al. 2004, Cuxart et al. 2006, Nie and Kuang 2012, Lu et al. 2016, Angevine et al. 2018, Couvreux et al. 2021, to cite only a few). The high resolution of LESs allows one to resolve a substantial part of the phenomena parametrized in NWP models, and LESs have the advantage to provide continuous 3D variable fields from which various diagnostics, including process-oriented ones, can be derived and compared to SCM outputs. LESs have also been validated on several occasions against observations and have been shown to realistically represent organized structures and small-scale variability (Neggers et al. 2003, Couvreux et al. 2005, Heus et al. 2009). In our case, we are particularly interested in the information provided on the small-scale variability of variables. By applying, on a LES domain large enough to be several times the size of the SCM grid box, the same homogeneous initial conditions and large-scale forcing as the ones prescribed in the SCM, the variability of the LES fields coarse-grained to the SCM resolution can be taken as the reference for the variability we wish to represent with process-oriented perturbations. Our first question therefore is: are process-oriented perturbation schemes able to represent the small-scale variability of an LES domain?

To be useful in ensemble forecasts, perturbation schemes should be able to produce sufficient ensemble spread. The process-oriented perturbation schemes target one source of uncertainty among several, therefore it may be interesting to compare their results in terms of dispersion with methods that use a more general approach. This is our second question: how does the dispersion produced by process-oriented perturbation schemes compare with the dispersion produced by the SPPT method?

Two process-oriented perturbation schemes are tested in this study in a single-column version of the convection-permitting AROME model. They are based on schemes proposed in the literature and target the representation of turbulent motions, which are known to lack variability in their current parametrization in AROME at resolutions belonging to the grey-zone of turbulence (Honnert et al. 2016). Our perturbation schemes are closely related to Kober and Craig (2016), hereafter referred to as KC16, for the perturbation of turbulence, and we build on the work of Sakradzija et al. (2015, 2016) and Sakradzija and Klocke (2018) to design a very simplified version of their stochastic shallow-convection scheme, based on the analysis of LES fields.

The paper is organized as follow. The process-oriented perturbation schemes applied to turbulence and shallow convection are described in section two. The analysis of LESs used to design the perturbation of shallow convection is presented in section three. Section four documents the dispersion of different ensembles constructed with the new perturbation schemes, evaluated against the dispersion inside the LES domain, and also compared with ensemble spread produced by SPPT. Section five is dedicated to discussion and conclusion.

2. Description of the perturbation schemes and simulation setup

2.1. The mesoscale AROME model

The perturbation schemes tested in this study are applied to the convection-permitting limited area model AROME (Seity et al. 2011) used operationally at Météo-France for weather prediction at kilometric scales. The experiments are performed in a one-dimensional (1D) framework, using a single-column version of the model (AROME-SCM), using 90 atmospheric vertical levels. Prognostic variables are used for temperature, wind, water vapor, five condensed water species and the turbulent kinetic energy (TKE). In the 1D framework, the tendencies produced by the dynamical part of the model are replaced by specified large-scale and surface forcing. The rest of the evolution is due to the parametrization of subgrid processes. In AROME, the parametrization schemes include radiation, surface, microphysics, turbulence and shallow convection. These last two schemes follow the EDMF (Eddy Diffusivity/Mass Flux) approach (Soares et al. 2004) to parametrize turbulent motions: the turbulence scheme is a classical eddy-diffusivity scheme (Cuxart et al. 2000), used to represent unorganized, local turbulence, while the shallow-convection scheme is a mass-flux scheme (Pergaud et al. 2009) used for the representation of non local turbulence in the form of a single convective plume. In this work, we propose to perturb each of these two schemes.

2.2. Implementation of stochastic turbulence in AROME

The KC16 scheme is designed as a situation- and scale-adaptive perturbation scheme that consists of an additive noise applied to the output tendencies of a turbulence scheme (Kober and Craig 2016). Its adaptability comes from the formulation of the noise amplitude which depends on subgrid variances diagnosed by the turbulence scheme and on a measure of the relative size of the eddies compared to the horizontal resolution of the model. The formulation of the scheme is summarized in Equation 1.

$$\left(\frac{\partial\phi}{\partial t}\right)^{\text{stoch}} = \frac{\partial\phi}{\partial t} + \alpha \frac{1}{\Delta t \Delta x_{\text{eff}}} l_{\text{eddy}} \left(\overline{\phi'^2}\right)^{\frac{1}{2}} \eta. \quad (1)$$

ϕ is a model prognostic variable, $\partial\phi/\partial t$ is the tendency produced by the deterministic turbulence scheme and other symbols are defined below. The stochastic version of the tendency, $(\partial\phi/\partial t)^{\text{stoch}}$, is obtained by adding a random perturbation composed of a Gaussian random number field η , multiplied by the square root of the subgrid variance of the variable $(\overline{\phi'^2})^{1/2}$, a scaling factor depending on time and length scales, and a scaling parameter α . This formulation has similarities with the stochastic perturbation scheme of Clark et al. (2021), where, starting from a simplified boundary layer model and the assumption that turbulent eddies can be considered as random events following a Poisson distribution, they mathematically derive a perturbation of the temperature tendency representing the fluctuations around the mean due to the random presence of eddies.

The implementation of the KC16 scheme in AROME is rather straightforward. The tendencies produced by the turbulence scheme of Cuxart et al. (2000) are perturbed following Equation 1. The key components of the method, the subgrid variances, are available in the scheme of Cuxart et al. (2000), which provides diagnostic equations for them. An example is given for potential temperature in Equation 2.

$$\overline{\theta'^2} = \frac{2}{3} \frac{1}{C_s} \frac{L^2}{C_\theta} \left(\frac{\partial \overline{\theta}}{\partial z} \right)^2 \varphi, \quad (2)$$

where C_s and C_θ are constants, φ is a stability function and L is the Bougeault-Lacarrère mixing length (Bougeault and Lacarrère 1989). In AROME, the scheme is employed in its one-dimensional version, where only the vertical gradients are taken into account, as shown in Equation 2.

The other components of the perturbation are taken as in Hirt et al. (2019):

- Δt is the characteristic lifetime of an eddy, set to 10 minutes,
- l_{eddy} is the characteristic horizontal size of an eddy, set to 1000 m,
- Δx_{eff} is the effective resolution of the model, here set to $5\Delta x = 5 \times 1.3$ km,

except for the tuning parameter α which here is set to 1.

Since we work in a 1D framework, the random field η is specified as a random scalar, taken from an autoregressive process of order 1 with a time constant set to 10 minutes. This temporal correlation of the noise has been chosen by Kober and Craig (2016) and Hirt et al. (2019) considering the characteristic lifetime of an eddy. The perturbations are added at each time step to the tendency of temperature T , specific humidity q_v and the two components of horizontal wind u and v . The same η is used for the four variables, which means that at a given time step, the perturbations added to T , q_v , u and v tendencies always have the same sign. This may not be ideal and one could think of using different random processes for each variable in order to decorrelate the perturbations. However, in this work we choose to remain close to the scheme original setup and therefore use a unique η . No parametrization tendency is produced for the vertical wind in AROME, hence it is not perturbed as in the original scheme of Kober and Craig (2016). We acknowledge that the interpretation of horizontal wind perturbations in a single-column framework without any interaction with the model dynamics may be limited, but as the results show that KC16 scheme has the ability to produce significant wind perturbations, they are kept in our simulations. A more careful design may however be necessary in future 3D experiments to prevent undesirable behavior of the perturbations such as a fast decay due to the appearance of acoustic modes reported by Hirt et al. (2019).

2.3. Design of a simple stochastic convection scheme

Stochastic convection schemes such as proposed by Plant and Craig (2008) or Sakradzija and Klocke (2018) rely on the generation of a random population of clouds inside each model grid box, constrained by large-scale conditions. In the case of coarse resolutions

studied by Cohen and Craig (2006), it allows one to study statistical fluctuations around a bulk equilibrium state due to the finite size of cloud samples inside model grid boxes. In mesoscale models with kilometric-scale horizontal resolution, the cloud population is small so the scheme is further used to relax the assumption that a statistically representative sample of clouds, well represented by its averaged characteristics, is present in each model grid box (Sakradzija et al. 2015). In addition to generating a cloud population, each cloud is assigned a mass flux randomly chosen according to a statistical distribution, and a lifetime. This enables one to compute the total mass flux on a grid box at a given time step as the sum of the mass fluxes of all the clouds existing in this grid box at this time, which can then be used as a closure condition for the rest of the convection scheme.

The generation of clouds in each model grid box and their tracking over several time steps can be computationally expensive. Machulskaya and Seifert (2019) recently demonstrated that they could reproduce the essential behavior of the stochastic convection schemes through a set of stochastic differential equations (SDEs), and thus reduce the computational cost. In both cases, the key feature of the scheme is the stochastic mass-flux closure given to an otherwise deterministic convection scheme. Here, we choose to try a very simplified approach, where the mass-flux closure is directly perturbed according to a statistical distribution.

The mass-flux scheme of Pergaud et al. (2009), hereafter referred to as PM09, is used in AROME to represent dry and moist convective thermals in the boundary layer. A single ascending plume represents the bulk effects of all the updrafts within a grid box. It is initialized at the surface thanks to closure conditions on the mass flux, vertical velocity and thermodynamic characteristics. In order to try our simplified version of a stochastic convection scheme, we thus need to find an appropriate distribution of the mass flux at the surface. We cannot directly use the empirical distributions found by Sakradzija et al. (2015) as a result of their cloud generation scheme, for they are located at the cloud base where their convection scheme has its closure condition. Therefore, in this study we choose to rely on the analysis of LESs to determine the likely distributions of the surface mass fluxes. The LES analysis and the final formulation of the stochastic convection scheme are presented in section three.

2.4. The SPPT scheme

SPPT is the current perturbation scheme used to represent model errors for ensemble prediction at Météo-France. Its implementation in the AROME-EPS ensemble system is described in Bouttier et al. (2012). In this study, we use a configuration of the SPPT scheme slightly modified from the operational one. In the following, every mention of SPPT will refer to our modified configuration. It has the same temporal correlation of the noise, set to 6 hours, and the same standard deviation, set to 0.3, as the current operational configuration. However, the tapering function used to damp the perturbations inside the boundary layer in the operational configuration is deactivated, as we are mainly interested in perturbations of boundary-layer processes in this study. In addition, the perturbations are not added to the total net physical tendencies but to the sum of the tendencies produced by the turbulence and shallow-convection schemes only. This ensures a fair comparison between the ensemble spread produced by SPPT and the ensemble spread

produced by the two process-oriented schemes previously described, which are only perturbing these two parametrizations.

2.5. Cases

Three well-documented cases are examined, representing different boundary-layer situations:

- the ARMCu case is an idealization of observations from the Atmospheric Radiation Measurement (ARM) campaign carried out on June 21, 1997 in the Great Plains region of the United States (Brown et al. 2002). It features the development of shallow non-precipitating cumulus clouds at the top of a continental convective boundary layer ;
- the BOMEX case features a marine boundary layer topped by cumulus clouds under steady conditions in the trade-wind region (east of Barbados). This case is described in Siebesma et al. (2003) and is based on observations from the Barbados Oceanographic and Meteorological Experiment (Holland and Rasmusson 1973) ;
- the FIRE case represents the diurnal cycle of a stratocumulus-topped marine boundary layer. It is taken from the European Projection Cloud Systems in Climate Models (EUROCS) project and based on observations made off the coast of California in July 1987 (Albrecht et al. 1988, Duynkerke et al., 2004).

For each case, both SCM simulations and LESs are used, whose characteristics are summarized in Table 1. The LESs are made with the research model Méso-NH (Lac et al. 2018) and use the same initial profiles and large-scale forcing as the SCM simulations. Large-scale forcing include advective and, for some cases, radiation forcing, given directly to the model as tendency terms. A large-scale subsidence can also be prescribed (BOMEX, FIRE) through a large-scale vertical velocity, which is used by the model together with the vertical gradients of the variables to compute the corresponding tendency term. For each case, the geostrophic wind is also specified.

For the FIRE case, the large-scale forcing and the sea surface temperature are time-independent and the diurnal cycle is entirely controlled by the cyclic solar radiations provided by the radiation scheme. Hence, this case can be run for an arbitrary long time with AROME-SCM. After 48 hours of simulation, the cloud evolution simulated by AROME becomes perfectly regular between successive 24 hour-intervals. We thus choose to systematically drop the first 48 hours in our simulations of the FIRE case. All the perturbations applied to the model start at 48h. For the sake of readability, the first 48 hours are never plotted in the figures and the time axis starts at 0h.

The evolution of the domain-averaged cloud liquid water in the LES fields, as well as the cloud liquid water inside the column simulated by AROME-SCM are presented in Figure 1.

We can note that AROME-SCM over-estimates the cloud water for BOMEX with regard to the LES, and that the FIRE stratocumulus cloud top and base heights are too low. Otherwise, AROME-SCM and averaged LES fields are in relatively good agreement. These cases have already been used in a number of studies comparing SCM and LESs (Sušelj et al.

2013, De Rooy and Siebesma 2008, Huang et al. 2013, Rio and Hourdin 2008, among others), including the work of Pergaud et al. (2009) on the mass-flux scheme currently used in AROME for the parametrization of shallow convection.

For each case, four ensembles of 100 members are produced using AROME-SCM, with the different perturbation methods previously described: KC16 (perturbation of turbulence) only, stochastic convection only, both KC16 and stochastic convection, and SPPT. No initial or boundary perturbations are applied in these ensembles.

The small-scale variability inside the LES fields is evaluated with a coarse-graining procedure: small subdomains of 1.3km^2 are defined, and the variables of interest are averaged over each subdomain. The dispersion between subdomains provides an indication on the subgrid variability we seek to represent.

3. Calibration of the stochastic convection method

As indicated in section 2.3, we wish to test a simple stochastic convection scheme where the surface mass-flux closure of the PM09 scheme is perturbed according to a statistical distribution. To determine the appropriate distribution to use for the three boundary-layer cases described in section 2.5, we choose to rely on the corresponding LESs. Our strategy is the following: for each LES field we calculate a coarse-grained mass-flux field by dividing the LES domain into small subdomains of the same size as the AROME grid box (1.3 km^2), and estimating the total mass flux of each subdomain at each vertical level. The estimation of the mass flux is achieved in the following way:

1. updrafts objects are identified in the LES fields, according to the method of Couvreux et al. (2010) based on the emission of a passive tracer at the surface, with limited lifespan. Points are identified as part of an updraft when they exceed a certain threshold of positive tracer concentration anomaly relative to the field average, together with a positive vertical velocity ;
2. the fractional area of each updraft is computed, as well as its average vertical velocity (at each vertical level), and both quantities are multiplied in order to get an approximation of the updraft mass flux (density variations are neglected) ;
3. on each subdomain the total mass flux is obtained by summing over all the updrafts it contains.

These mass-flux values can then be summarized into histograms that inform us on the shape of the distribution, which is done here in Figure 2, for three vertical levels relative to the cloud base height.

The distribution of the mass flux is found to be highly dependent on the height. It has an exponential shape at cloud base, which is consistent with the distribution obtained by Sakradzija et al. (2015) for a resolution of 1.6 km. Near the surface, Gaussian distribution functions fit reasonably well the results.

Accepted Article

According to these results, we want our stochastic convection scheme to produce surface mass-flux values following a Gaussian distribution. In order to avoid introducing bias into the model because of systematic differences between AROME-SCM and the LESs, the surface mass-flux values are not sampled directly from the LES distribution, but a Gaussian noise η is applied to the AROME surface mass-flux initialization, following Equation 3. The standard deviation of the noise is set to $\frac{\sigma_{LES}}{\mu_{LES}}$, where σ_{LES} is the standard deviation and μ_{LES} the experimental mean of the LES coarse-grained mass-flux distributions shown in Figure 2. Their values are given for each case in the appendix. The noise distribution has a zero mean so that the stochastic surface mass flux M_{surf}^{stoch} distribution should be centered around the deterministic surface mass flux M_{surf} diagnosed by the shallow-convection scheme.

$$M_{surf}^{stoch} = M_{surf} * (1 + \eta), \text{ with } \eta \sim N\left(0, \frac{\sigma_{LES}}{\mu_{LES}}\right) \quad (3)$$

Here, η is again an autoregressive process of order 1. By analogy with the KC16 scheme, we choose the time constant to be approximately the lifetime of a convective updraft. In our experiments, it is set to 40 minutes, close to the 45 minutes chosen by Plant and Craig (2008). However, this lifetime may be overestimated in the case of shallow convection, and other values around 20 minutes could be considered as well (Zhao and Austin 2005, Sakradzija et al. 2015).

4. Evaluation of the perturbation schemes

In this section, the SCM ensembles produced with the process-oriented perturbation schemes (KC16 for the perturbation of turbulence and stochastic convection) are compared with the coarse-grained LES fields to examine whether or not their dispersion is realistic. They are also compared with ensembles produced with SPPT, and we study how the perturbation methods differ from each other. We emphasize that no initial or boundary perturbations are used in our study, therefore the spread between ensemble members is solely due to the perturbations applied to the parametrization schemes. The control simulation that is referred to in the following is, for each case, one AROME-SCM simulation without any perturbations (the deterministic simulation).

4.1. KC16 turbulence perturbation

First, we examine the impact of the KC16 scheme. We start by looking at the amplitude and location of the perturbations added to the tendencies (second term on the right hand side of Equation 1), and then we look at the ensemble spread induced by these perturbations.

4.1.1. A limited impact of KC16 on temperature and humidity

The KC16 perturbation amplitude is controlled by the square-root of subgrid variances, which are given for temperature in Figure 3, for the three cases.

In the ARMCu case, most of the variance is located in the first levels above the ground, which is linked to strong heat fluxes from the continental surface. The variance close to the

Accepted Article

surface is lower for BOMEX and FIRE which are oceanic cases. The profiles are otherwise similar between ARMCu and BOMEX, with significant variances in the cloud layer, except that the ARMCu cloud extends higher. The FIRE profile is very different, with high variances around the stratocumulus top due to the radiative cooling process occurring there.

To see what the perturbations controlled by these variances look like, we present in Figure 4 instantaneous vertical profiles of their standard deviation. In the same graphs are shown the standard deviations of SPPT perturbations, and the temperature absolute tendency of the control simulation. We recall that the tendency here is the sum of the tendencies produced by the turbulence and the shallow-convection schemes, instead of the total net tendency.

As shown by the red and black profiles, the standard deviation of SPPT perturbations follows the tendency profile, which is expected since the perturbations are multiplicative. The standard deviation is about 30% of the tendency, consistent with the settings. KC16 perturbations on the contrary are not linked to the tendency values and their profiles are logically close to those of the subgrid standard deviation shown in Figure 3. The perturbation values are in general lower than the SPPT perturbations, and rather small compared to the tendency.

The temperature spread in the 100-member ensembles obtained with these perturbations is illustrated in Figure 5. It is quantified by the tenth to ninetieth percentile ensemble range, which will be used throughout this study as a measure of the ensemble spread. In Figures 5b and 5c we can see that the vertical structure of the SPPT ensemble spread globally follows that of the perturbations shown in Figures 4a and 4b. For the KC16 ensemble, where the perturbation peaks are more localized, the perturbations seem to have spread vertically. The local perturbation peak at the surface visible in Figures 4a and 4b does not result in high temperature spread at this level ; the spread is rather homogeneous throughout all the subcloud layer suggesting a propagation of the perturbation by the turbulent transport in this layer. Furthermore, we observe temperature spread up to 3000 m in ARMCu after 9 hours (Figure 5b) and at 2000 m for BOMEX after 14 hours (Figure 5c), whereas no perturbations have been introduced at this level at this time (Figures 4a and 4b). This also shows that the perturbations can propagate upward and even be amplified in the cloud as the peak around 2000 m for BOMEX suggests. The results of FIRE (Figure 5d) also indicate that the perturbations propagate but no amplification can be seen.

Compared to the variability of the coarse-grained LES fields, the SPPT ensemble spread is generally too large and KC16 ensemble spread too low. For the cumulus cases, SPPT ensemble spread is especially large in the subcloud layer, whereas the LES variability there is smaller than above, probably because of turbulent mixing which tends to homogenize the layer. The KC16 ensemble spread in this layer is closer to the LESs. In the cloud layer however, the KC16 method does not seem to be able to produce as much spread as observed in the LESs, contrary to SPPT.

Accepted Article

For FIRE, both SPPT and KC16 produce very similar temperature dispersion during both night (Figure 5d) and day (not shown), of the same order of magnitude as the LES dispersion, except at the top of the clouds. The LES dispersion has a strong peak there not at all captured by any of our ensembles. This is probably because radiation and condensation processes play a critical role at this level, and these schemes should be perturbed as well in order to get significant dispersion.

Similar observations can be made for the specific humidity (Figure 6). The main difference is that specific humidity tendencies are low in the subcloud layer, therefore the SPPT ensemble spread is reduced at this level contrary to the previous results on temperature. Otherwise the SPPT ensemble is over-dispersive in the cumulus clouds compared to the LES variability, whereas KC16 ensemble tends to be under-dispersive. The humidity dispersion profile for FIRE is homogeneous throughout the first 500 m, similarly to temperature.

4.1.2. Significant wind spread produced by KC16

Contrary to the thermodynamic variables, we observe important wind disturbance with KC16. The stochastic term added to the tendency with KC16 method is large (not shown), because of large wind subgrid standard deviation values. This leads to the high zonal wind spread in the KC16 ensembles seen in Figure 7 (similar results are obtained for the meridional wind), which can sometimes exceed that of the SPPT ensembles. We can also note that the KC16 perturbations, which are added mainly in the subcloud layer (not shown), slightly propagate vertically but do not amplify in the cloud layer, which suggests that the amplification observed for temperature and humidity resulted from interaction with the microphysics.

To summarize, the KC16 method seems to produce limited temperature and humidity spread, especially in the case of the continental cumulus ARMCu. On the other hand, it can produce significant wind spread, of the same order of magnitude as the LES variability in the low levels. The SPPT method on the contrary produces high temperature and humidity dispersion in the cumulus cases, but it is often over-estimated compared to the LES variability. The dispersion obtained for the stratocumulus is more similar between KC16 and SPPT, except for the wind, for which KC16 method leads to greater spread.

The low spread observed in the cloud layer when using KC16 method is a motivation for investigating other perturbation methods targeting different physical processes more likely to bring perturbations in clouds, such as shallow convection.

4.2. Impact of the stochastic convection scheme

Ensembles produced with the stochastic convection scheme are now examined. We first investigate the physical consistency of our scheme by examining the parametrized updraft characteristics compared to the LES updrafts, and then look at the spread of SCM ensembles.

4.2.1. Impact of the scheme on the updraft characteristics

The mass-flux profile of the SCM ensemble made with the stochastic convection scheme is presented in Figure 8a for the ARMCu case.

The range of mass-flux values in the ensemble is large below 1000 m, but decreases sharply toward the top of the profile suggesting that all the members of our ensemble have updrafts that reach similar altitude, but with various strengths inside the first 1000 m. This is linked to the equation governing the mass flux in the shallow convection scheme:

$$\frac{1}{M_{\text{up}}} \frac{\partial M_{\text{up}}}{\partial z} = \epsilon - \delta, \quad (4)$$

where M_{up} is the mass flux of the updraft, ϵ the entrainment rate and δ the detrainment rate. The exponential form of the solution of this equation and the evolution of the sign of $(\epsilon - \delta)$ along the vertical (Figure 8b) imply that any perturbation at the surface will be amplified in the middle part of the updraft and will shrink again toward the updraft top. Thus, to have additional spread around the updraft top, modifying the surface mass flux is not enough, the entrainment and detrainment rate should vary as well. We see very little spread in the entrainment and detrainment values of our SCM ensemble (Figure 8b), indicating that the perturbations of the mass-flux closure do not have a strong feedback on the entrainment and detrainment rates. This suggests the need for further research on the perturbation of these two variables, which are known to be difficult to accurately model (Rio et al 2019).

Figure 8a also shows that the mass-flux values of the SCM ensemble mean, as well as that of the control simulation, are strongly biased with respect to the LES. This over-estimation of the mass flux was observed by Honnert et al. (2016) when using PM09 shallow-convection scheme at resolutions belonging to the grey-zone of convective boundary-layer thermals. The use of the stochastic mass-flux closure does not change the average behavior of PM09 scheme, but it enables one to explore more values belonging to the LES range. The distribution of the values, however, is not entirely satisfactory. Having been perturbed by a Gaussian perturbation at the surface, the mass-flux values keep a Gaussian-like distribution at higher levels, whereas in the LES they evolve toward an exponential distribution (see Figure 2). Similar observations can be made in BOMEX and FIRE (not shown).

4.2.2. Impact of the scheme on the main variables tendencies

The ensemble spread generated by the stochastic convection scheme is of the same order of magnitude (ARMCu) or lower (BOMEX, FIRE) than the spread generated by the KC16 scheme for humidity (Figure 9) and temperature (not shown). The vertical structure however can be rather different. In the FIRE case for example, there is a spread peak for specific humidity at the top of the stratocumulus, not observed in the KC16 ensemble spread profile.

As for the wind, the spread range with the stochastic convection scheme does not exceed half that of the KC16 ensemble, and is negligible in the FIRE case. Its vertical structure is also quite different (Figures 9d-f): stochastic convection has an impact on the wind mainly

near the ground and at cloud base, whereas KC16 perturbations act on all the mixed layer, with a maximum impact in the middle of the subcloud layer.

4.2.3. Impact of the scheme on cloud water

Although the impact of the stochastic convection scheme on the resolved temperature, humidity and wind is not greater than that of KC16 method, there is a strong effect on the cloud liquid water. This is due to the parametrization of subgrid condensation and subgrid cloud in AROME which includes a direct contribution from the mass-flux scheme, of the form:

$$r_c = r_{c_{up}} C_{cf} a_{up}, \quad (5)$$

where $r_{c_{up}}$ is the updraft liquid water mixing ratio, a_{up} the updraft area and C_{cf} a constant set to 2.5. The updraft area is directly linked to the updraft mass flux (along with the updraft vertical velocity), so that a_{up} in our ensemble can vary by more than 100%. This has a significant impact on the cloud water as can be seen in Figure 10 representing the vertical profiles of cloud liquid water spread in the three cases.

The majority of the spread is located in the lower part of the cumulus clouds as for KC16, but it is greater and extends higher in the cloud layer. It has a more realistic shape, for ARMCu and BOMEX, than the SPPT ensemble which has a spread peak in the upper part of the cloud not observed in the LESs. For the stratocumulus case, the stochastic convection ensemble has more spread in the upper part of the cloud, and is again more realistic when compared to the LES than the other ensembles. However, even if the spread of cloud water is improved with the stochastic convection scheme, it does not cover all the variability observed in the LES field, which is linked to the high spatial heterogeneity of the clouds.

4.3. Combination of KC16 and stochastic convection schemes

In the previous sections, KC16 and the stochastic convection schemes have been tested separately on each boundary layer case in order to understand their impact on the model variables and their ability to produce ensemble spread. However, the aim is to use them jointly in the same ensemble since they are designed to represent distinct small-scale variability sources. Here, we look at the spread of ensembles made with both KC16 and stochastic convection schemes activated. They will be denoted as POP (for Process-Oriented Perturbations) ensembles in the following.

For temperature and humidity spread, the interaction between the two schemes is not trivial (Figure 9). Overall, it is difficult to say whether the differences in the spread profiles are statistically significant, given that we are working with ensembles of finite size. Re-running the four ensembles for the ARMCu case shows that we can not conclude on any increase or decrease in the temperature and specific humidity spread when combining KC16 and stochastic convection schemes. For the FIRE case however, the dispersion from KC16 and the dispersion from stochastic convection seem to add up to each other. We also observe that, generally, the POP ensemble wind spread is lower than the KC16 ensemble wind spread (see Figure 9 for the zonal wind). Even if the vertical structure of the wind

spread in the KC16 ensembles and stochastic convection ensembles is very different, combining both schemes does not seem to bring additional spread and there even seems to be some form of compensation.

We finally summarize the results of POP and SPPT ensembles and of the coarse-grained LES in Figure 11. The time-height diagrams represent the tenth to ninetieth percentile range of the cloud water, potential temperature, humidity and wind for the ARMCu case.

In the LES domain, we observe two local maximums of temperature and humidity spread on the vertical, whose altitudes evolve during the simulation. The first peak appears two hours after the beginning of the simulation, and rises to about 2500 m after 15 hours. It can be associated with the top of the convective boundary layer, reached by many updrafts that bring to this altitude air moister and cooler than the environment. The second peak appears later, after about 6 hours, and is located around 1000 m height which corresponds to the cloud base height. It may therefore be associated with some variability in the condensation level height. The SCM ensembles reproduce reasonably well the position of the highest peak, although with either too much (SPPT ensemble) or too little (POP ensemble) spread. The second peak of temperature and humidity dispersion is less visible in the SCM ensembles. They rather produce dispersion within the subcloud layer for the temperature, because of the positive tendencies in the case of the SPPT ensemble, and because of the propagation of surface perturbations in the case of the POP ensemble. This is not necessarily a desirable feature, as we observe that in the LES, on the contrary, the subcloud layer is more homogeneous, probably because of active turbulent mixing.

Contrary to temperature and humidity, the wind spread range is similar between the LES and the SCM ensembles, but the vertical structure shows some important differences. The wind spread in the LES is large in the entire subcloud layer, especially near the surface. The SPPT ensemble does not reproduce this surface dispersion, contrary to the POP ensemble. However, in the POP ensemble the large wind spread values extend higher up in the subcloud layer because of high KC16 perturbations following the TKE profile, which may be less realistic.

The same kind of analysis for BOMEX (not shown here) leads to very similar results. The results for FIRE are however more difficult to interpret, because of the low spread generated by both SPPT and POP ensembles, except for the wind spread of the POP ensemble which has the same order of magnitude as the LES spread.

5. Discussion and conclusions

Two process-oriented perturbation schemes have been introduced in the AROME model, and used to produce single-column simulation ensembles of three idealized boundary-layer cases. The objective was to introduce perturbations representing the subgrid variability of turbulent and convective processes which is missing in deterministic parametrization schemes. This is achieved by different means in the two perturbation schemes. For the turbulence, the method of Kober and Craig (2016), which has been adapted here for AROME, relies on the use of second-order moments, diagnosed by the deterministic

scheme, to perturb its grid to subgrid relation. For shallow convection, the work of Sakradzija et al. (2015) enables one to relax some assumptions of classical convection schemes, such as the existence of a statistical sample of clouds inside a model grid box, by building a stochastic mass-flux closure condition. We have followed the same strategy, in a very simplified way, by using LESs to determine statistical distributions of the mass flux close to the surface in our experiments.

To evaluate the ability of these schemes to represent small-scale variability, we have relied on SCM/LES comparison for three well-known boundary-layer cases. Identical homogeneous forcing are applied to LES and SCM and the heterogeneity of the LES fields after the application of a coarse-graining procedure is taken as a reference for the subgrid state variability that the perturbation schemes should represent. Additional SCM ensembles have been performed with the 'bulk' perturbation scheme SPPT, which is an operational method to represent model uncertainty, modified in this study to target only turbulence and shallow convection.

The perturbation scheme targeting turbulence (KC16) proved to be effective in disturbing the horizontal wind variables in the three cases, which is a promising result. It is yet limited by the fact that our experiments only included single-column simulations where the horizontal advection is prescribed, hence no interaction with the dynamics could be studied, although we know that it can have important impacts (Lin and Neelin 2002). Future experiments with 3D simulations of real cases are planned, where the wind perturbation impact on the model will be further examined. The scheme may also have to be adapted in the light of Hirt et al. (2019) findings on the importance of having balanced 3D wind perturbations to prevent the appearance of acoustic modes leading to a fast decay of the perturbations.

The second process-oriented perturbation scheme, stochastic convection, had a much lower impact on the wind, but was able to generate substantial cloud water spread. Both stochastic convection and KC16 schemes, however, did not give satisfactory results regarding temperature and humidity spread. It was generally under-estimated, except for the temperature spread in the subcloud layer which could be too large. For KC16, the under-estimation is linked to the small values of temperature and humidity subgrid standard deviations, leading to perturbations that are small compared to the tendencies. As subgrid standard deviations are the central part of the KC16 scheme for controlling the perturbation amplitude, improving the scheme so that it produces more spread is not trivial. One could think of increasing the scale parameter α of Equation 1, however Kober and Craig (2016) emphasize that this parameter has no physical meaning, and should not be too different from 1. One possibility for future tests would be to decorrelate temperature, humidity (and wind) perturbations by using different random processes for each variable instead of a single one. On the other hand, Clark et al. (2021) argue that KC16 perturbations are, at least for temperature, too large at the top and bottom of the boundary layer. In our study, we found indeed too large temperature spread in the subcloud layer in some of our ensembles. Clark et al. (2021) suggest that a height-dependent length scale could be used in the scheme. We did some tests where we replaced the constant eddy length scale of 1000 m by the Bougeault-Lacarrère mixing length. The perturbations were indeed reduced at the surface, but it decreased the impact of the scheme on all the model

variables.

Regarding the stochastic convection scheme, in this study only the mass-flux closure condition was perturbed. However, we saw that the impact on the updraft characteristics was limited, therefore, perturbing other elements of the parametrization scheme such as the entrainment and detrainment rates should be considered. Another limitation of the stochastic convection scheme in this study is the determination of an appropriate distribution for the surface mass-flux perturbations. The results of the LES analysis showed that for the three cases it can be chosen as a Gaussian distribution function, and that its standard distribution was not too different between the cases. Some tests showed that replacing the σ_{LES}/μ_{LES} value of 0.33 by 0.4 for the BOMEX case resulted in ensembles with very similar dispersion. However, more boundary-layer situations should be examined to generalize these results and to check the behavior of the scheme when simulating real cases.

In this study, the SPPT method has been applied on the net tendencies of turbulence and shallow-convection parametrizations, instead of the total net tendencies. Therefore, both SPPT and POP methods perturb the same processes, but not in the same way. SPPT is globally a more efficient method to produce ensemble spread, which can be linked to the fact that it is designed to represent various errors coming from the parametrization schemes, whereas the process-oriented perturbation schemes studied here target only one possible source of error coming from the random occurrence of small-scale processes. The differences between POP and SPPT ensembles spread can also be explained by the temporal correlation of the noise which is set to 10 minutes in the KC16 scheme and 40 minutes in the stochastic convection scheme, whereas it is 6 hours in SPPT. SPPT perturbations thus accumulate on a longer time period. In several stochastic perturbation schemes, the temporal and spatial correlation scales of the uncertainty to be represented can not be easily determined. The noise correlation scales are thus chosen according to practical reasons and can be enlarged to ensure a significant impact on the resolved flow (Buizza et al. 1999, Ollinaho et al. 2017). In the process-oriented schemes presented here however, the temporal correlation of the perturbations is set following physical considerations: 10 minutes in KC16 scheme is the characteristic lifetime of an eddy, and 40 minutes for stochastic convection is chosen to represent the lifetime of a convective updraft.

In our experiments, all the perturbations were applied on turbulence and shallow-convection processes only. Although they are the main physical processes governing the evolution of the boundary layer in ARMCu and BOMEX cases, other processes are parametrized whose subgrid variability should also be accounted for. It appears to be critical in the FIRE case, where subgrid condensation and radiation are major contributors to the physical tendencies, and could also have an impact in the cloud layer of cumulus cases. Currently, the representation of uncertainty in these processes is done through random perturbations of some of their uncertain parameters (with RP, SPP schemes), and to the authors' knowledge, no process-oriented schemes similar to those presented in this study have been tested. Further development on the representation of subgrid variability for these processes would therefore be interesting, as well as the possible combination of RP and process-oriented perturbations. More generally, combining different perturbation

Accepted Article

methods to produce ensembles seems to be an interesting – yet challenging – approach. It is probably difficult to combine perturbations that are not redundant, but it offers the possibility to take advantage of the benefits of each type of scheme. For example, the process-oriented schemes tested in this study produced non-negligible spread in the boundary layer, where SPPT perturbations are operationally switched off to avoid instabilities. However, by design they produce perturbations only once the processes in question have started, and seem less effective in modifying the timing of their triggering than SPPT. Christensen (2020) on the other hand argues that a multiplicative noise may not be suited to represent the uncertainty of all the parametrizations, and that a diversity of approaches could be beneficial. The iSPPT scheme (Christensen et al. 2017), where each parametrization scheme is independently perturbed instead of the total tendencies, could be a convenient framework to test such combinations.

Appendix

The mean and standard deviation of the surface mass-flux distribution presented in section 3 are given in Table A1. These values have been used to draw the red curves on Figure 2.

References

- Albrecht, B. A., Randall, D. A., & Nicholls, S. (1988). Observations of marine stratocumulus clouds during FIRE. *Bulletin of the American Meteorological Society*, 69(6), 618-626.
- Angevine, W. M., Olson, J., Kenyon, J., Gustafson, W. I., Endo, S., Suselj, K., & Turner, D. D. (2018). Shallow cumulus in WRF parameterizations evaluated against LASSO large-eddy simulations. *Monthly Weather Review*, 146(12), 4303-4322.
- Ayotte, K. W., Sullivan, P. P., Andren, A., Doney, S. C., Holtslag, A. A., Large, W. G., ... & Wyngaard, J. C. (1996). An evaluation of neutral and convective planetary boundary-layer parameterizations relative to large eddy simulations. *Boundary-Layer Meteorology*, 79(1), 131-175.
- Baker, L. H., Rudd, A. C., Migliorini, S., & Bannister, R. N. (2014). Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, 21(1), 19-39.
- Batté, L., & Doblas-Reyes, F. J. (2015). Stochastic atmospheric perturbations in the EC-Earth3 global coupled model: Impact of SPPT on seasonal forecast quality. *Climate Dynamics*, 45(11), 3419-3439.
- Bengtsson, L., Dias, J., Tulich, S., Gehne, M., & Bao, J. W. (2021). A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System. *Journal of Advances in Modeling Earth Systems*, 13(1), e2020MS002260.

Berner, J., Ha, S. Y., Hacker, J. P., Fournier, A., & Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review*, 139(6), 1972-1995.

Berner, J., Fossell, K. R., Ha, S. Y., Hacker, J. P., & Snyder, C. (2015). Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review*, 143(4), 1295-1320.

Bougeault, P., & Lacarrere, P. (1989). Parameterization of orography-induced turbulence in a mesobeta-scale model. *Monthly weather review*, 117(8), 1872-1890.

Bouttier, F., Vié, B., Nuissier, O., & Raynaud, L. (2012). Impact of stochastic physics in a convection-permitting ensemble. *Monthly Weather Review*, 140(11), 3706-3721.

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J. C., Khairoutdinov, M., ... & Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1075-1093.

Buizza, R., Miller, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560), 2887-2908.

Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y., & Wei, M. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133(5), 1076-1097.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gagnon, N., Mitchell, H. L., & Michelin, L. (2010). Toward random sampling of model error in the Canadian ensemble prediction system. *Monthly Weather Review*, 138(5), 1877-1901.

Cheng, Y., Canuto, V. M., & Howard, A. M. (2002). An improved model for the turbulent PBL. *Journal of the Atmospheric sciences*, 59(9), 1550-1565.

Christensen, H. M., Moroz, I. M., & Palmer, T. N. (2015). Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *Journal of the Atmospheric Sciences*, 72(6), 2525-2544.

Christensen, H. M., Lock, S. J., Moroz, I. M., & Palmer, T. N. (2017). Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society*, 143(706), 2168-2181.

Christensen, H. M. (2020). Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society*, 146(727), 938-962.

Clark, P. A., Halliwell, C. E., & Flack, D. L. A. (2021). A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part I: Formulation and Evaluation in a Convection-Permitting Model. *Journal of the Atmospheric Sciences*, 78(3), 727-746.

Cohen, B. G., & Craig, G. C. (2006). Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *Journal of the Atmospheric Sciences*, 63(8), 2005-2015.

Couvreur, F., Guichard, F., Redelsperger, J. L., Kiemle, C., Masson, V., Lafore, J. P., & Flamant, C. (2005). Water-vapour variability within a convective boundary-layer assessed by large-eddy simulations and IHOP_2002 observations. *Quarterly Journal of the Royal Meteorological Society*, 131(611), 2665-2693.

Couvreur, F., Hourdin, F., & Rio, C. (2010). Resolved versus parametrized boundary-layer plumes. Part I: A parametrization-oriented conditional sampling in large-eddy simulations. *Boundary-layer meteorology*, 134(3), 441-458.

Couvreur, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., ... & Xu, W. (2021). Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, 13(3), e2020MS002217.

Craig, G. C., & Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences*, 63(8), 1996-2004.

Cuxart, J., Bougeault, P., & Redelsperger, J. L. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, 126(562), 1-30.

Cuxart, J., Holtslag, A. A., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., ... & Xu, K. M. (2006). Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Boundary-Layer Meteorology*, 118(2), 273-303.

De Rooy, W. C., & Siebesma, A. P. (2008). A simple parameterization for detrainment in shallow cumulus. *Monthly weather review*, 136(2), 560-576.

Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., & Jonker, H. J. (2013). Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theoretical and Computational Fluid Dynamics*, 27(1), 133-148.

Duda, J. D., Wang, X., Kong, F., & Xue, M. (2014). Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Monthly Weather Review*, 142(6), 2198-2219.

Duynkerke, P. G., de Roode, S. R., van Zanten, M. C., Calvo, J., Cuxart, J., Cheinet, S., ... & Sednev, I. (2004). Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3269-3296.

Groenemeijer, P., & Craig, G. C. (2012). Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Atmospheric Chemistry and Physics*, 12(10), 4555-4565.

Heus, T., J. Pols, C. F. J. Jonker, H. J., A. Van den Akker, H. E., & H. Lenschow, D. (2009). Observational validation of the compensating mass flux through the shell around cumulus clouds. *Quarterly Journal of the Royal Meteorological Society*, 135(638), 101-112.

Hirt, M., Rasp, S., Blahak, U., & Craig, G. C. (2019). Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Monthly Weather Review*, 147(11), 3917-3934.

Holland, J. Z., & Rasmusson, E. M. (1973). Measurements of the atmospheric mass, energy, and momentum budgets over a 500-kilometer square of tropical ocean. *Monthly Weather Review*, 101(1), 44-55.

Honnert, R., Masson, V., & Couvreux, F. (2011). A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometeric scale. *Journal of the Atmospheric Sciences*, 68(12), 3112-3131.

Honnert, R., Couvreux, F., Masson, V., & Lancz, D. (2016). Sampling the structure of convective turbulence and implications for grey-zone parametrizations. *Boundary-layer meteorology*, 160(1), 133-156.

Houtekamer, P. L., Mitchell, H. L., & Deng, X. (2009). Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, 137(7), 2126-2143.

Huang, H. Y., Hall, A., & Teixeira, J. (2013). Evaluation of the WRF PBL parameterizations for marine boundary layer clouds: Cumulus and stratocumulus. *Monthly weather review*, 141(7), 2265-2271.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., ... & Brown, J. M. (2017). A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Monthly Weather Review*, 145(4), 1161-1179.

Kober, K., & Craig, G. C. (2016). Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of Atmospheric Sciences*, 73(7), 2893-2911.

Lac, C., Chaboureaud, J. P., Masson, V., Pinty, J. P., Tulet, P., Escobar, J., ... & Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, 11(5), 1929-1969.

Lang, S. T., Lock, S. J., Leutbecher, M., Bechtold, P., & Forbes, R. M. (2021). Revision of the Stochastically Perturbed Parametrisations model uncertainty scheme in the Integrated Forecasting System. *Quarterly Journal of the Royal Meteorological Society*, 147(735), 1364-1381.

Lin, J. W. B., & Neelin, J. D. (2000). Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophysical research letters*, 27(22), 3691-3694.

Lin, J. W. B., & Neelin, J. D. (2002). Considerations for stochastic convective parameterization. *Journal of the Atmospheric Sciences*, 59(5), 959-975.

Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17(3), 321-333.

Lu, C., Liu, Y., Zhang, G. J., Wu, X., Endo, S., Cao, L., ... & Guo, X. (2016). Improving parameterization of entrainment rate for shallow convection with aircraft measurements and large-eddy simulation. *Journal of Atmospheric Sciences*, 73(2), 761-773.

Machulskaya, E., & Seifert, A. (2019). Stochastic differential equations for the variability of atmospheric convection fluctuating around the equilibrium. *Journal of Advances in Modeling Earth Systems*, 11(8), 2708-2727.

Neggers, R. A. J., Duynkerke, P. G., & Rodts, S. M. A. (2003). Shallow cumulus convection: A validation of large-eddy simulation against aircraft and Landsat observations. *Quarterly Journal of the Royal Meteorological Society*, 129(593), 2671-2696.

Nie, J., & Kuang, Z. (2012). Responses of shallow cumulus convection to large-scale temperature and moisture perturbations: A comparison of large-eddy simulations and a convective parameterization based on stochastically entraining parcels. *Journal of the atmospheric sciences*, 69(6), 1936-1956.

Ollinaho, P., Lock, S. J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., ... & Sandu, I. (2017). Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 408-422.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., ... & Weisheimer, A. (2009). Stochastic parametrization and model uncertainty.

Pergaud, J., Masson, V., Malardel, S., & Couvreux, F. (2009). A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-layer meteorology*, 132(1), 83-106.

Plant, R. S., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, 65(1), 87-105.

Redelsperger, J. L., & Sommeria, G. (1981). Méthode de représentation de la turbulence d'échelle inférieure à la maille pour un modèle tri-dimensionnel de convection nuageuse. *Boundary-Layer Meteorology*, 21(4), 509-530.

Rio, C., & Hourdin, F. (2008). A thermal plume model for the convective boundary layer: Representation of cumulus clouds. *Journal of the atmospheric sciences*, 65(2), 407-425.

Rio, C., Del Genio, A. D., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Current Climate Change Reports*, 5(2), 95-111.

Rochetin, N., Couvreux, F., Grandpeix, J. Y., & Rio, C. (2014). Deep convection triggering by boundary layer thermals. Part I: LES analysis and stochastic triggering formulation. *Journal of the Atmospheric Sciences*, 71(2), 496-514.

Sakradzija, M., Seifert, A., & Heus, T. (2015). Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, 22(1), 65-85.

Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, 8(2), 786-812.

Sakradzija, M., & Klocke, D. (2018). Physically constrained stochastic shallow convection in realistic kilometer-scale simulations. *Journal of Advances in Modeling Earth Systems*, 10(11), 2755-2776.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., ... & Masson, V. (2011). The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3), 976-991.

Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131(612), 3079-3102.

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., ... & Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, 60(10), 1201-1219.

Soares, P. M. M., Miranda, P. M. A., Siebesma, A. P., & Teixeira, J. (2004). An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3365-3383.

Stensrud, D. J., Bao, J. W., & Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, 128(7), 2077-2107.

Sušelj, K., Teixeira, J., & Chung, D. (2013). A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *Journal of the Atmospheric Sciences*, 70(7), 1929-1953.

Wilks, D. S. (2005). Effects of stochastic parametrizations in the Lorenz'96 system. *Quarterly Journal of the Royal Meteorological Society*, 131(606), 389-407.

Zhao, M., & Austin, P. H. (2005). Life cycle of numerically simulated shallow cumulus clouds. Part I: Transport. *Journal of the atmospheric sciences*, 62(5), 1269-1290.

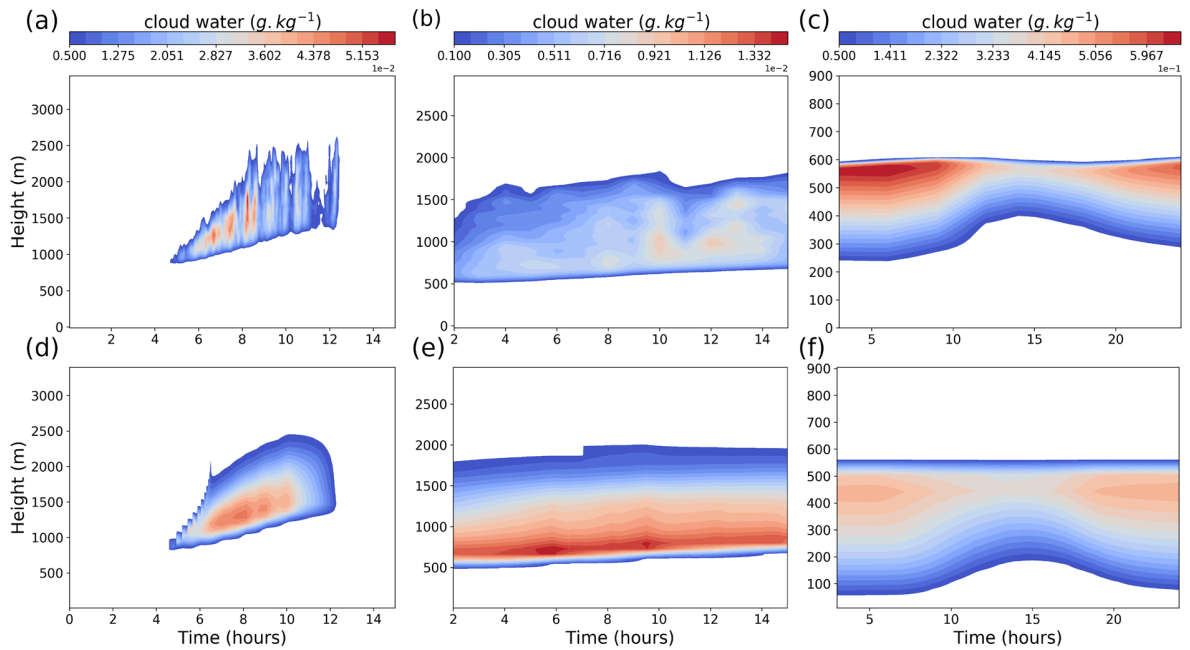


FIGURE 1 Time-height diagrams of the specific mass of cloud liquid water, averaged on the LES domain (a, b, c) and in the column of AROME-SCM (d, e, f) for ARMCu (a, d), BOMEX (b, e) and FIRE (c, f). Note that the y-axis does not extend to the same altitude for the three cases and each case has a specific color bar.

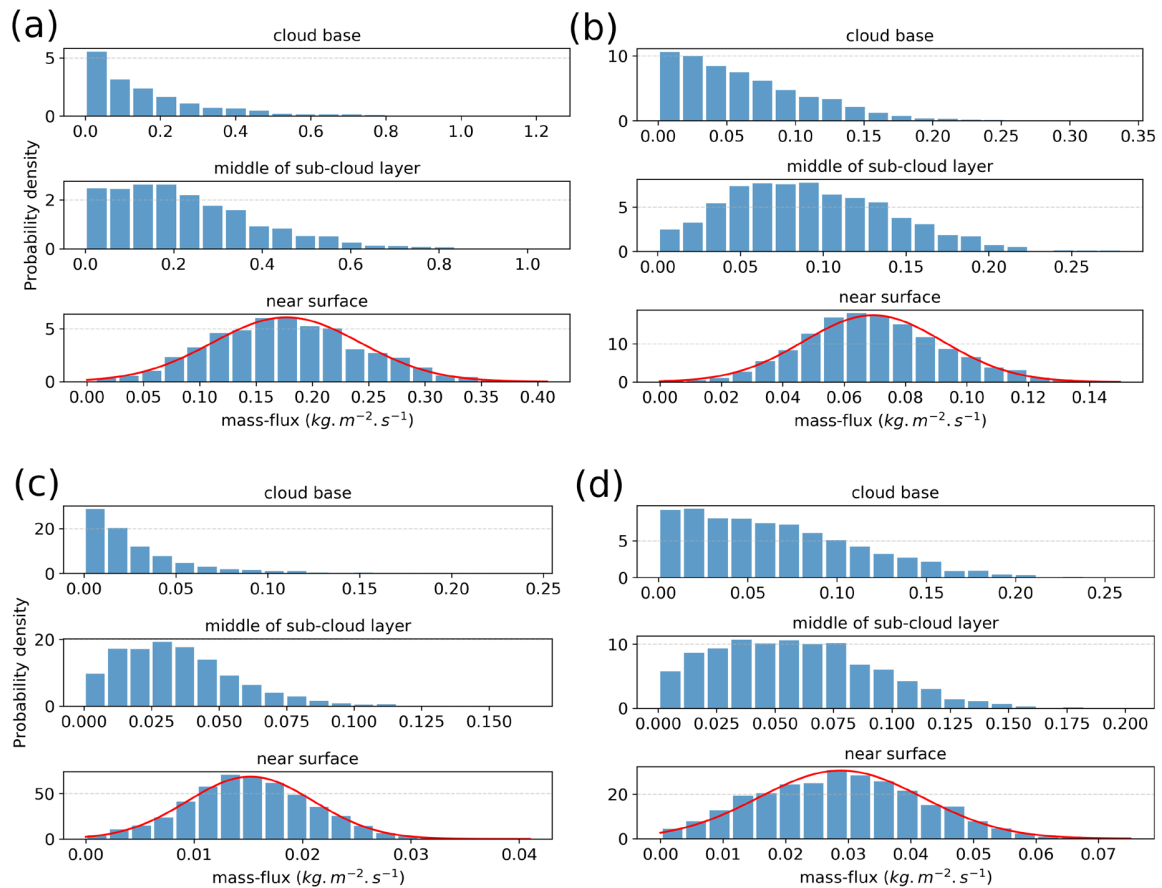


FIGURE 2 Histograms of mass-flux values on the LES domains at three different levels: cloud base, middle of the subcloud layer ($0.5 \times$ cloud base height) and near surface ($0.1 \times$ cloud base height), for (a) ARMCu between +5 and +10h, (b) BOMEX between +2 and +15h, (c) FIRE between +12 and +16h (day) and (d) FIRE between +20 and +24h (night). The results of the FIRE case have been split in two periods corresponding to day and night because the behavior of the updrafts changes according to a diurnal cycle. The mass-flux values have been computed assuming a constant density $\rho=1$. The red curve in the bottom histograms corresponds to a Gaussian fit to the data.

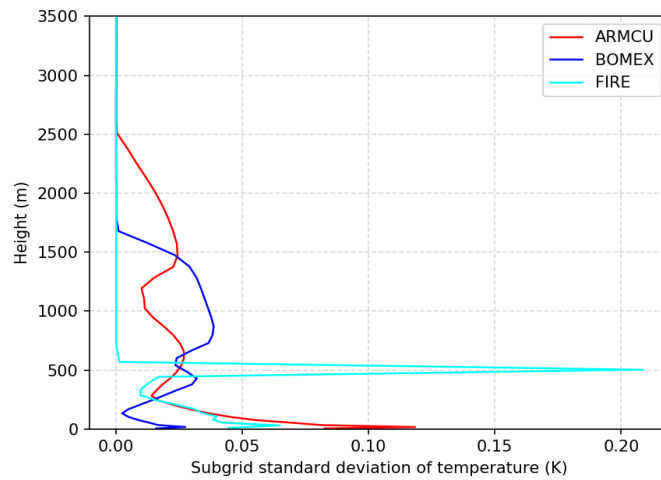


FIGURE 3 Time-averaged vertical profiles of temperature subgrid standard deviation in the deterministic AROME-SCM simulation for ARMCu, BOMEX and FIRE.

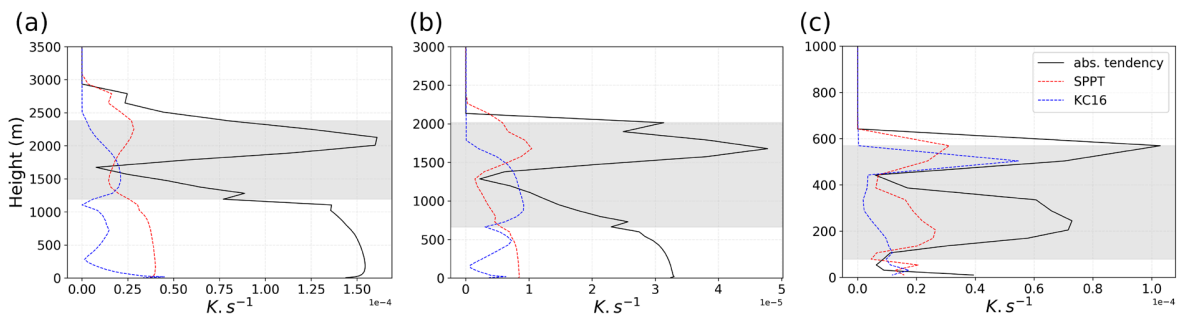


FIGURE 4 Vertical profiles of the absolute net temperature tendency from turbulence and shallow convection in the control simulation (black line), the standard deviation of KC16 perturbations (blue-dashed line) in a 100-member ensemble, and the standard deviation of SPPT perturbations (red-dashed line) in a 100-member ensemble. SPPT perturbations are defined as the tendency difference before/after the application of the SPPT scheme. The grey shaded area represents the vertical extension of the cloud in the control simulation. Profiles are shown for (a) ARMCu at +9h, (b) BOMEX at +14h and (c) FIRE at +24h.

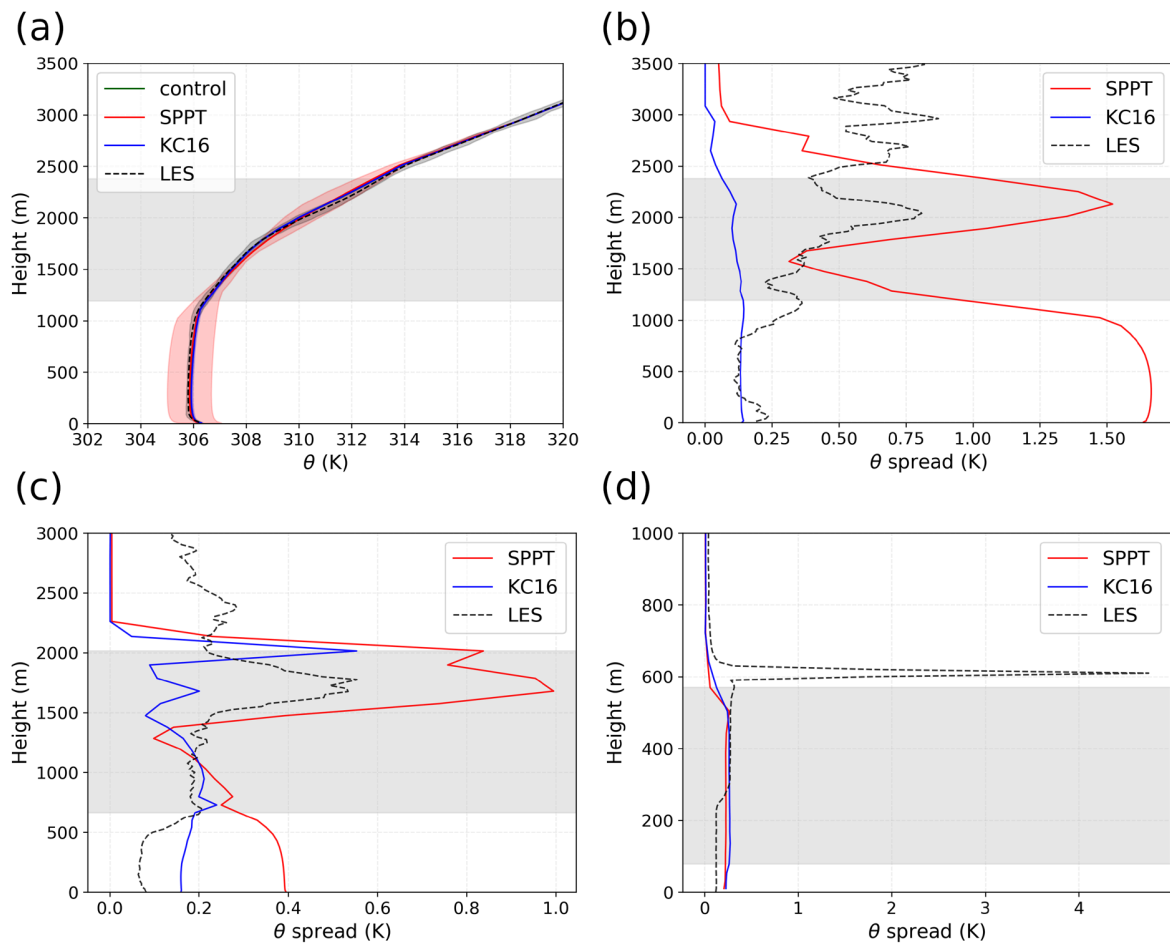


FIGURE 5 (a) Potential temperature profiles of the control simulation (green line), the SPPT ensemble mean (red line), the KC16 ensemble mean (blue line) and the coarse-grained LES mean (black dashed line) for ARMCu at +9h. The red shaded area is the tenth to ninetyth percentile range of the SPPT ensemble. This range is much smaller for the KC16 ensemble and for the coarse-grained LES so the corresponding shaded areas are not visible on the graph. (b,c,d) Vertical profiles of the potential temperature spread in SCM ensembles and coarse-grained LES for (b) ARMCu at +9h, (c) BOMEX at +14h and (d) FIRE at +24h. The spread is calculated as the tenth to ninetyth percentile range. The horizontal grey shaded area represents the vertical extension of the cloud in the control simulation.

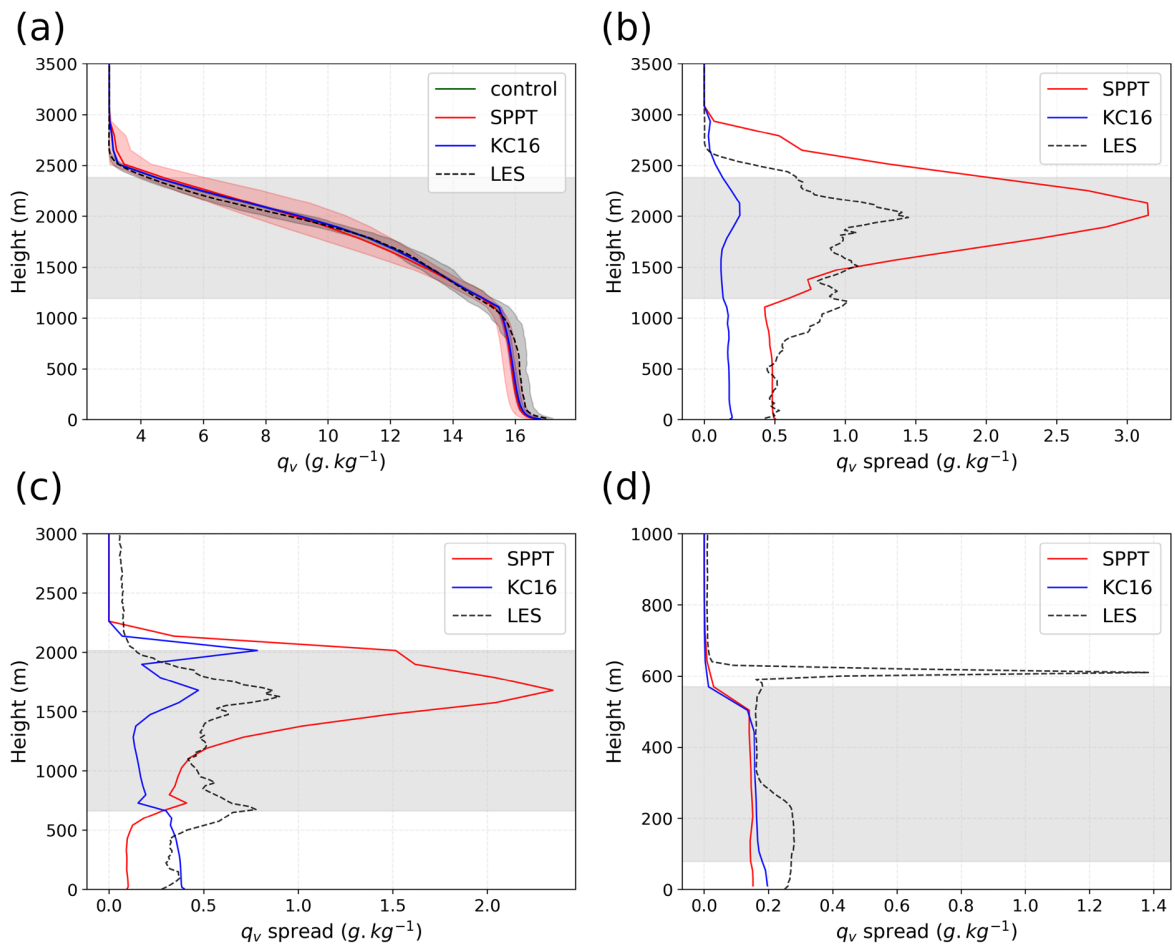


FIGURE 6 Same as Figure 5 but for specific humidity.

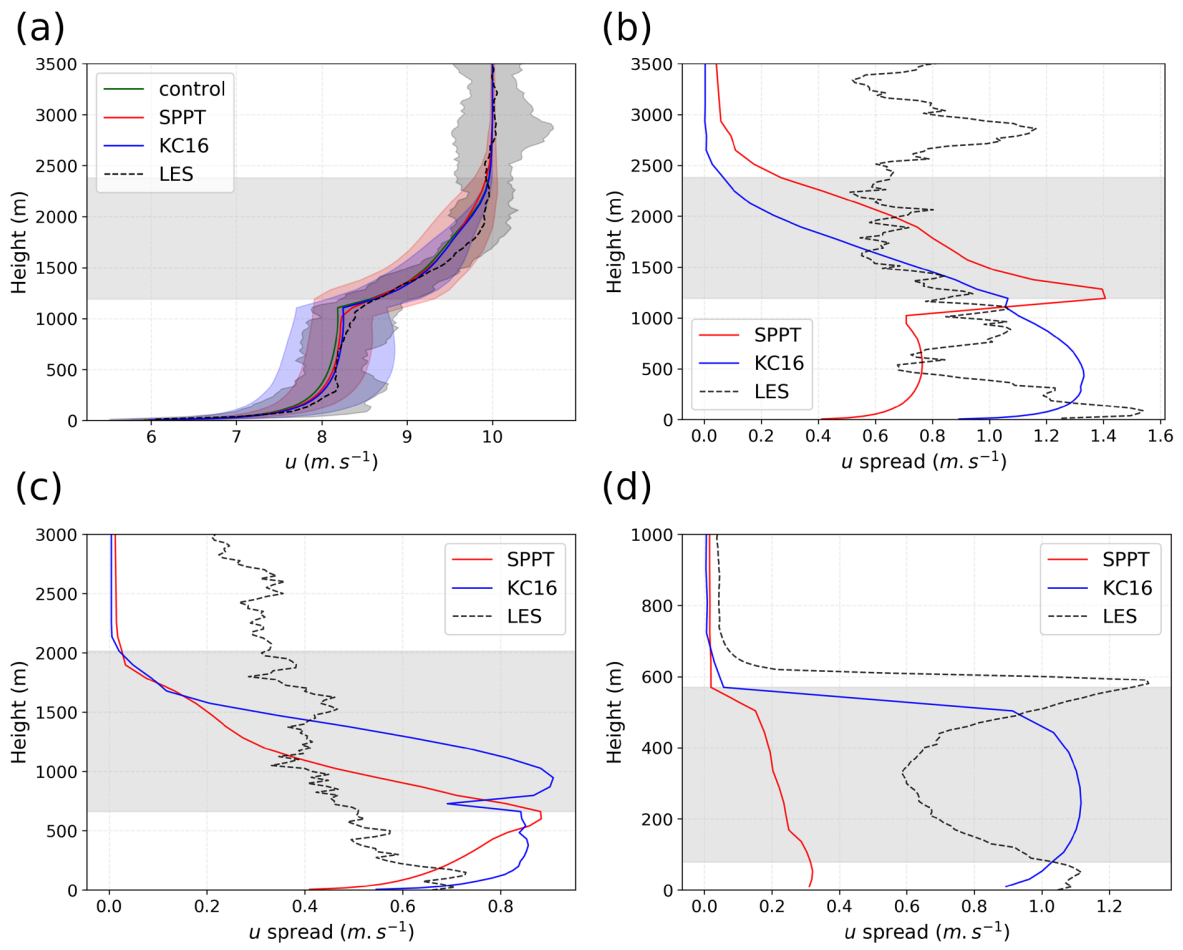


FIGURE 7 Same as Figure 5 but for zonal wind.

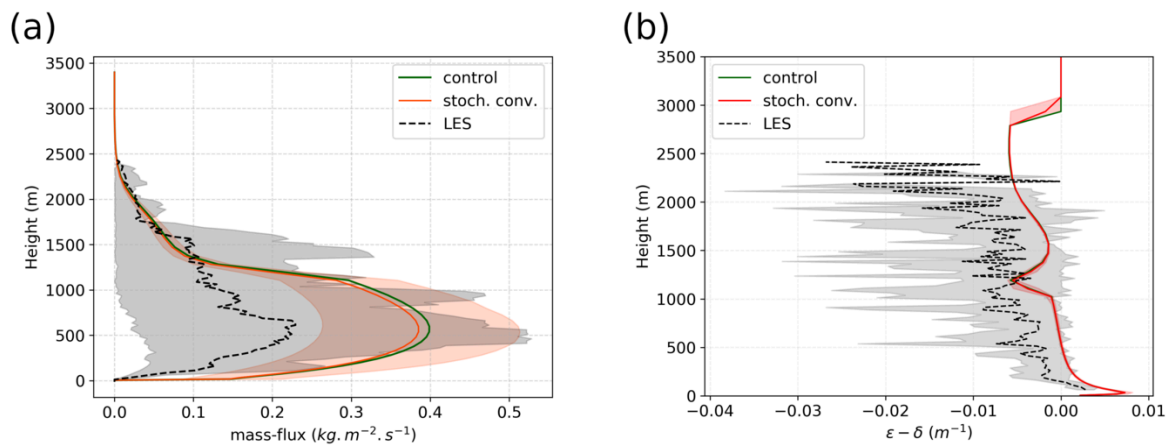


FIGURE 8 Vertical profiles of (a) mass flux and (b) entrainment (ϵ) minus detrainment (δ) rate for a 100-member AROME-SCM ensemble produced with stochastic convection (in red) and coarse-grained LES (in grey), for the ARMCu case at +9h. Lines are ensemble averages and shaded areas represent tenth to ninetieth percentile ranges. The green lines are the profiles of the control simulation.

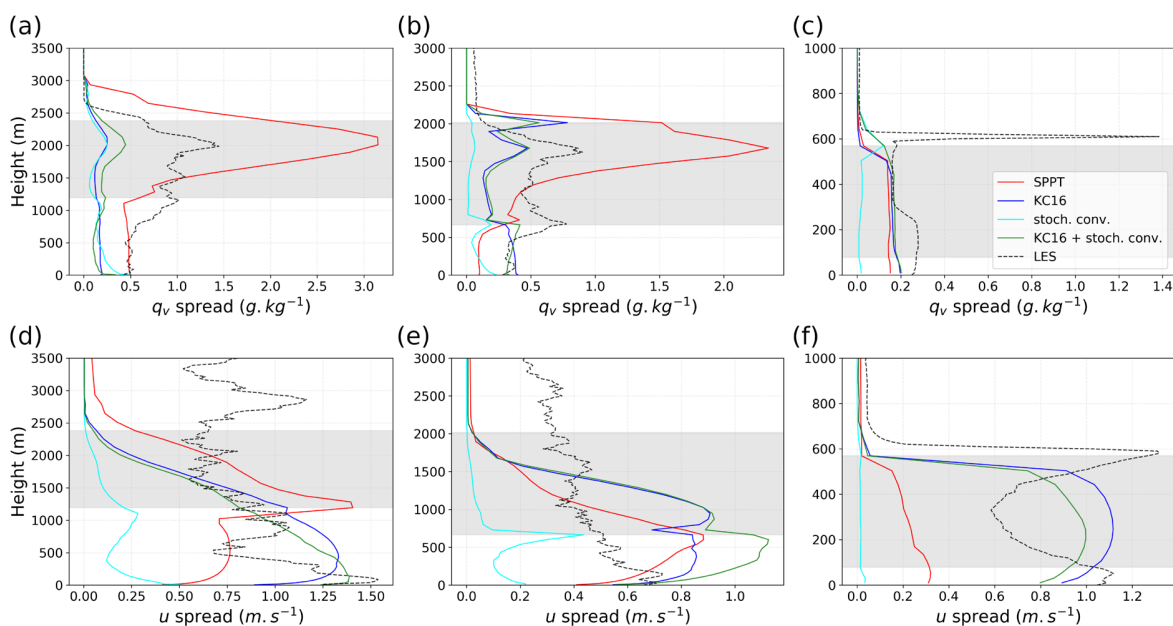


FIGURE 9 Tenth to ninetieth percentile range profiles of (a,b,c) specific humidity and (d,e,f) zonal wind for the AROME-SCM ensembles and the coarse-grained LES, for (a,d) ARMCu at +9h, (b,e) BOMEX at +14h, (c,f) and FIRE at +24h.

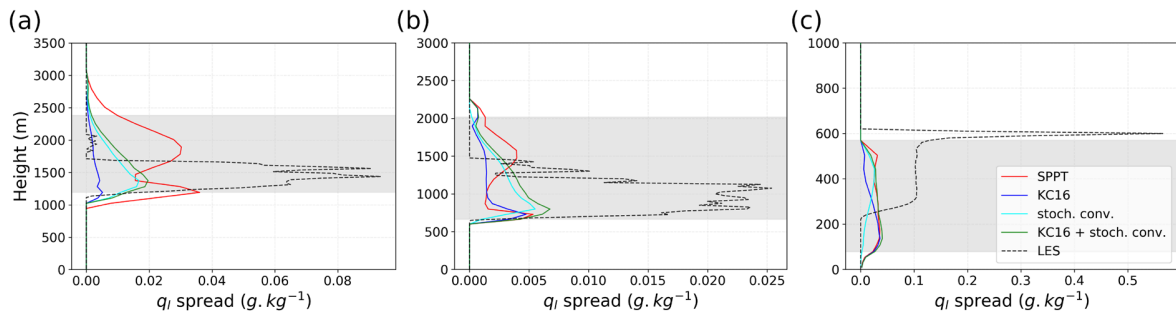


FIGURE 10 Same as Figure 9 but for cloud liquid water.

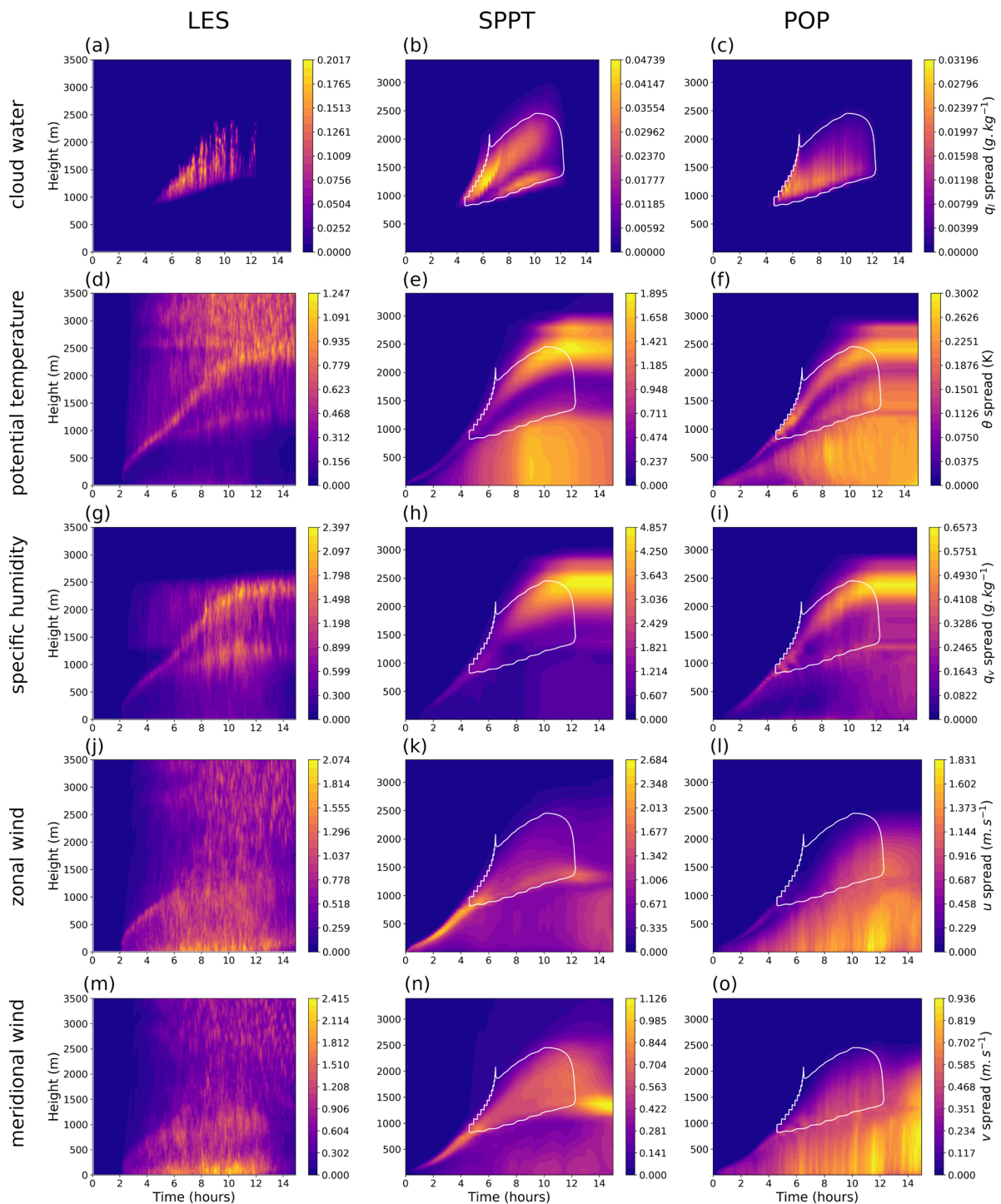


FIGURE 11 Time-height diagrams of the tenth to ninetieth percentile range of cloud water (a, b, c), potential temperature (d, e, f), specific humidity (g, h, i), zonal wind (j, k, l) and meridional wind (m, n, o) in the LES field (a, d, g, j, m), the SPPT ensemble (b, e, h, k, n) and the

POP ensemble (c, f, i, l, o), for the ARMCu case. The white contour corresponds to the cloud limits in the control simulation.

TABLE 1 Description of LES and SCM simulation setup for each case.

	ARMCU	BOMEX	FIRE
LES domain	6.4 x 6.4 x 4 km ³	12.8 x 12.8 x 4 km ³	25.6 x 25.6 x 1.2 km ³
LES resolution	$\Delta x = \Delta y = \Delta z = 25$ m	$\Delta x = \Delta y = \Delta z = 25$ m	$\Delta x = \Delta y = 50$ m, $\Delta z = 10$ m
Time range	15 hours: 05:30 LT – 20:30 LT	14 hours	24 hours: 00:00 LT – 24:00 LT
Time availability of LES fields	every five minutes	hourly	hourly
Large-scale forcing	evolving temperature and humidity advection, radiative forcing, steady geostrophic wind	steady temperature radiative forcing, humidity advection, subsidence and geostrophic wind	steady temperature advection, humidity advection, subsidence and geostrophic wind
Surface forcing	sensible and latent heat fluxes	sensible and latent heat fluxes	sea surface temperature
Parametrization schemes activated in AROME-SCM	shallow convection, turbulence, condensation and cloud microphysics	shallow convection, turbulence, condensation and cloud microphysics	radiation, shallow convection, turbulence, condensation and cloud microphysics

TABLE A1 Empirical mean and standard deviation values of the mass-flux distribution near the surface, for the three cases.

	ARMCU	BOMEX	FIRE	
			day	night
μ_{LES} (kg.m ⁻² .s ⁻¹)	0.18	0.07	0.02	0.03
σ_{LES} (kg.m ⁻² .s ⁻¹)	0.07	0.02	0.01	0.01
σ_{LES}/μ_{LES}	0.37	0.33	0.4	0.45

4.2 Recherches sur la perturbation de l'ajustement à la saturation

4.2.1 Motivations

Comme mentionné dans le chapitre 2, les problématiques de représentation de l'incertitude dans les schémas de paramétrisation des phénomènes microphysiques sont différentes de celles des schémas de turbulence ou de convection. Les LES précédemment décrites utilisent le même schéma d'ajustement à la saturation et le même schéma microphysique qu'AROME, donc ne simulent pas, a priori, les processus microphysiques de façon plus réaliste. Néanmoins, nous avons vu dans le chapitre 2 que l'ajustement à la saturation repose sur des hypothèses de distribution sous-maille de variables comme le contenu en eau et la température, ou l'écart à la saturation. Ces distributions sous-maille sont affectées par la variabilité de petite échelle, provoquée notamment par la turbulence et les mouvements convectifs que les LES peuvent résoudre en partie. Les LES peuvent donc fournir des informations sur ces distributions, ce qui semble intéressant étant donnée l'incertitude de modélisation suggérée par les nombreuses distributions théoriques différentes utilisées dans la littérature : gaussienne (Sommeria et Deardorff 1977 ; Bougeault 1981), exponentielle (Bougeault 1981), gamma (Bougeault 1982), beta (Tompkins 2002), triangulaire (Smith 1990), bi-gaussienne (Lewellen et Yoh 1993 ; Golaz et al. 2002 ; Neggers 2009 ; Jam et al. 2013), etc. Dans Perraud et al. (2011), plusieurs de ces distributions sont évaluées à partir de LES de quatre cas de cumulus et de stratocumulus idéalisés, et l'étude montre que l'utilisation de certaines distributions peut mener à de larges sous-estimations de la fraction nuageuse et du contenu en eau moyen sur un domaine.

Cette section présente donc quelques résultats sur la distribution du contenu en eau non précipitante r_{np} et de l'écart à la saturation s dans les domaines LES. On examine également à l'aide des LES les erreurs potentiellement commises par la paramétrisation de Chaboureau et Bechtold 2002 (dans la suite notée CB02) du contenu en eau nuageuse, avec pour objectif de savoir si des perturbations de cette paramétrisation peuvent être envisagées.

4.2.2 Distributions dans les LES

Les figures suivantes présentent les histogrammes des valeurs de r_{np} ou s dans tout le domaine des LES des cas ARMCu et FIRE, pour différentes échéances et niveaux verticaux. On étudie la possibilité de simuler les distributions observées avec deux approches : celle de Tompkins (2002) utilisant une distribution beta, et celle de Jam et al. (2013) basée sur une distribution bi-gaussienne.

Tompkins (2002) propose de modéliser la distribution de l'eau totale r_{np} par une distribution beta définie par :

$$f(x) = \frac{1}{B(p, q)} \frac{(x - a)^{p-1} (b - x)^{q-1}}{(b - a)^{p+q-1}}, \quad \text{avec } a \leq x \leq b, \quad p > 0, \quad q > 0,$$

où $B(p, q)$ est la fonction beta définie par

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)},$$

avec Γ la fonction gamma. Les quatre coefficients a, b, p, q sont choisis de la façon suivante :

- a et b sont respectivement les valeurs minimales et maximales de r_{np} sur le domaine LES à l'instant et l'altitude considérés,
- p est la moyenne de r_{np} sur le domaine et q est sa variance.

Néanmoins, ces réglages ne permettent pas d'obtenir des résultats satisfaisants pour nos cas d'étude, et la façon d'estimer p et q est donc modifiée de façon à ce que la moyenne et l'écart-type de la distribution beta correspondent à ceux observés dans les LES. La moyenne et l'écart-type d'une distribution beta définie sur l'intervalle $[a, b]$ sont rappelées par Tompkins (2002) :

$$\mu_{\text{beta}} = (b - a) \frac{p}{p + q} + a$$

$$\sigma_{\text{beta}} = \frac{b - a}{p + q} + \sqrt{\frac{pq}{p + q + 1}}$$

Connaissant a , b , et en prenant μ_{beta} et σ_{beta} comme la moyenne et l'écart-type de r_{np} observés dans les LES, il est possible de calculer p et q :

$$p = \frac{\mu_{\text{beta}} - a}{b - a} \left(\frac{(b - \mu_{\text{beta}})(\mu_{\text{beta}} - a)}{\sigma_{\text{beta}}^2} - 1 \right)$$

$$q = \frac{b - \mu_{\text{beta}}}{b - a} \left(\frac{(b - \mu_{\text{beta}})(\mu_{\text{beta}} - a)}{\sigma_{\text{beta}}^2} - 1 \right)$$

Les figures 4.3a et 4.4a montrent les résultats obtenus avec ce réglage.

La méthode de Jam et al. (2013) cherche à représenter la bimodalité des distributions de l'eau totale ou de l'écart à la saturation observée dans de nombreuses études. Pour cela, leur méthode se base sur la séparation entre ascendances et environnement faite par les schémas de convection en flux de masse. Pour chacune de ces deux parties, une distribution gaussienne peut être définie et une distribution bi-gaussienne est obtenue avec la combinaison des deux. Pour l'écart à la saturation, on a ainsi :

$$f(s) = (1 - \alpha)g(s, \overline{s_{\text{env}}}, \sigma_{\text{env}}) + \alpha g(s, \overline{s_{\text{up}}}, \sigma_{\text{up}}),$$

avec $\overline{s_{\text{env}}}$ (respectivement $\overline{s_{\text{up}}}$) désignant l'écart à la saturation moyen de l'environnement (respectivement des ascendances convectives), σ_{env} (respectivement σ_{up}) l'écart-type de l'écart à la saturation de l'environnement (respectivement des ascendances convectives) et g la densité de probabilité d'une loi normale :

$$g(s, \bar{s}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(s - \bar{s})^2}{2\sigma^2}\right).$$

Les figures 4.3b et 4.4b montrent les résultats obtenus avec cette méthode.

Avec les figures 4.3 et 4.4 on constate qu'au-dessus du nuage, les distributions de r_{np} et s sont bien modélisées par des gaussiennes, que sont capables de reproduire la méthode de Tompkins (2002) et la méthode de Jam et al. (2013). Dans le nuage ou dans la couche sous-nuageuse en revanche, les distributions ne sont plus symétriques, et parfois bimodales. La méthode de Tompkins (2002) permet bien de reproduire les distributions ayant des queues de distribution importantes à droite ou à gauche, mais ne peut reproduire la bimodalité. La méthode de Jam et al. (2013) permet dans une certaine mesure de reproduire les deux pics observés par exemple dans le cas ARMCu à +8 et +11 h à 1000 m ou dans le cas FIRE à +6 h à 600 m. Néanmoins

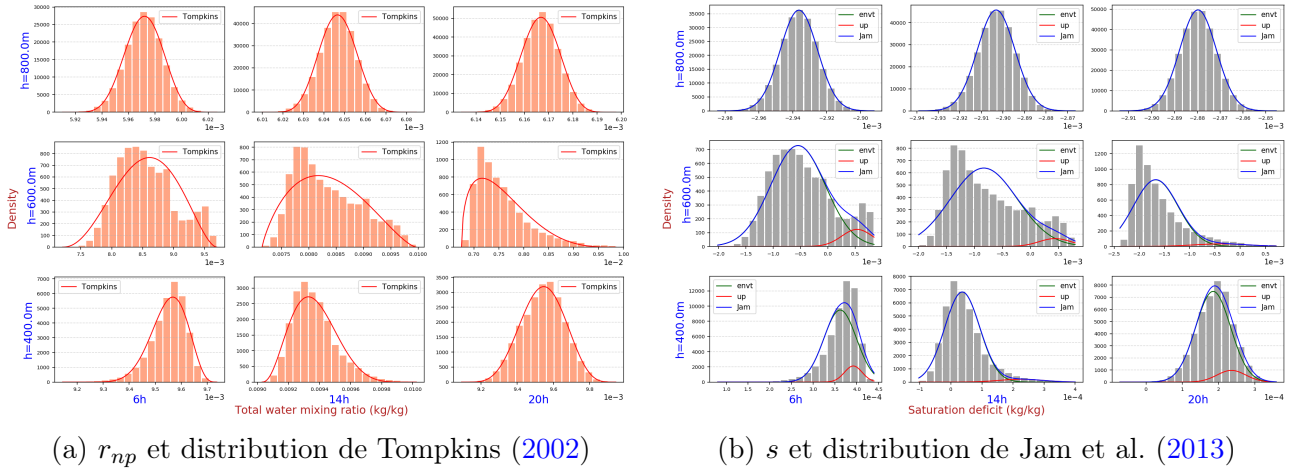


FIGURE 4.3 – Histogrammes des valeurs de r_{np} et s observées dans le domaine LES et distributions théoriques pour le cas FIRE, pour trois échéances (+6, +14, +20 h) et trois niveaux verticaux (400, 600, 800 m).

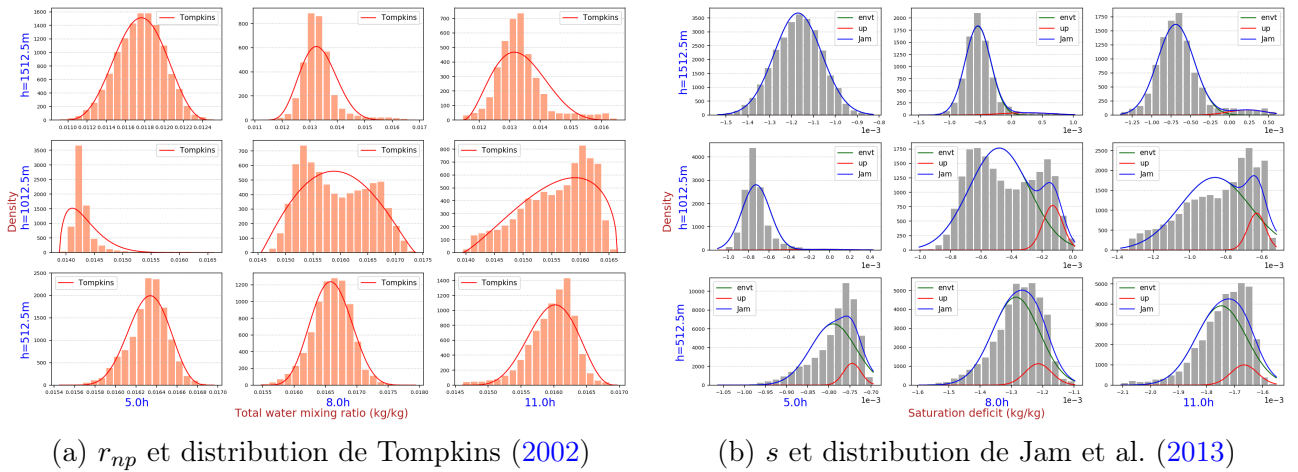


FIGURE 4.4 – Histogrammes des valeurs de r_{np} et s observées dans le domaine LES et distributions théoriques pour le cas ARMCu, pour trois échéances (+5, +8, +11 h) et trois niveaux verticaux (environ 500, 1000 et 1500 m).

elle semble moins adaptée dans le cas de distributions asymétriques (par exemple dans le cas FIRE à +20 h à 600 m). Globalement, on peut constater que la forme des distributions est variable selon l'altitude et l'échéance, et qu'il semble difficile de trouver une formulation unique pour toutes les modéliser. Les hypothèses faites par les schémas de condensation statistiques cités précédemment concernant les distributions de variables comme r_{np} ou s ne sont donc probablement pas tout le temps réalistes et pourraient mener à des erreurs dans l'estimation de l'eau et la glace nuageuse sous-maille.

Le schéma de condensation de CB02 utilisé dans AROME a la particularité d'utiliser des relations empiriques entre Q_1 et r_{cond} (qui correspond au contenu total en hydrométéores non précipitants liquides et glacés) qui permettent de s'affranchir de l'hypothèse sur la forme de la distribution. Il est néanmoins possible d'examiner les effets de ce choix de relations empiriques à partir des données LES. Pour cela, on applique la procédure de coarse-graining décrite dans la section 4.1.1 pour calculer l'écart à la saturation moyen Q_1 , son écart-type σ_s et le contenu

en eau nuageuse moyen (il n'y a pas de glace nuageuse dans les cas ARMCu, BOMEX et FIRE donc $\bar{r}_{\text{cond}}/\sigma_s = \bar{r}_c/\sigma_s$) sur des petits sous-domaines de la taille de la maille d'AROME. Ceci permet de tracer les points rouges sur les graphes de la figure 4.5.

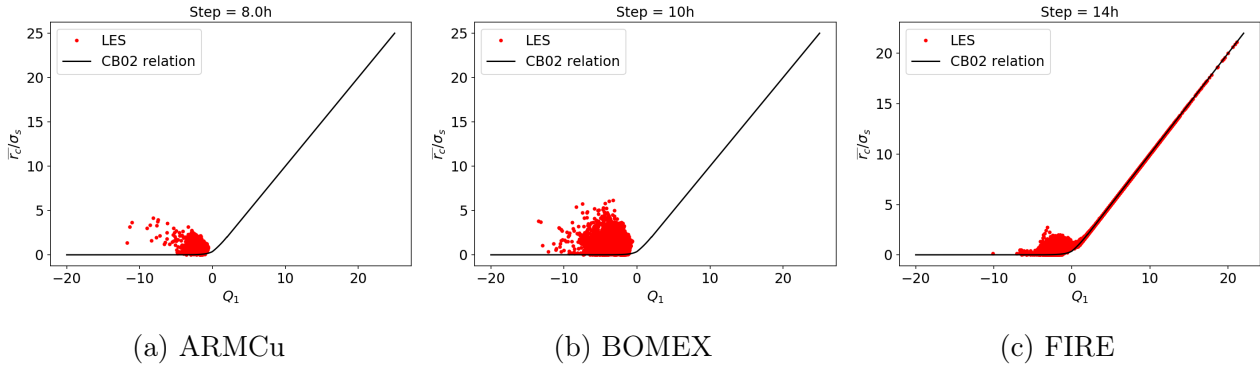


FIGURE 4.5 – Correspondance entre Q_1 et \bar{r}_c/σ_s calculés dans les LES des cas ARMCu à +8 h, BOMEX à +10 h et FIRE à +14 h, avec la procédure de coarse-graining. Chaque point rouge correspond à un sous-domaine et à un niveau vertical.

On constate que lorsque l'écart à la saturation moyen Q_1 est très élevé, le contenu en eau nuageuse \bar{r}_c/σ_s est bien égal à Q_1 (figure 4.5c, points le long de la droite $y = x$). En revanche, lorsque les valeurs de Q_1 sont comprises entre 0 et -10 environ, \bar{r}_c/σ_s s'éloigne davantage de la relation empirique de CB02. Ceci pourrait indiquer que la relation de CB02 n'est pas adaptée pour représenter toute la variabilité possible du contenu en eau nuageuse à l'échelle d'une maille. Néanmoins, il faut remarquer que dans AROME, une part de la formation d'eau nuageuse est assurée par la contribution du schéma de convection peu profonde, qui est ajoutée à l'eau nuageuse diagnostiquée par la relation de CB02 (voir la section 2.3.1). Ainsi, une partie des points rouges est potentiellement modélisée par AROME grâce à cette contribution, qui est probablement importante dans les cas de cumulus ARMCu et BOMEX. L'information précise n'ayant pas été extraite, il est néanmoins difficile d'estimer si cette contribution serait capable de modéliser toute la variabilité observée dans les LES. Dans la section suivante, on effectue donc des premiers tests simples de perturbation de la relation CB02, de façon à étudier la variabilité d'eau nuageuse que l'on peut produire indépendamment de perturbations du schéma de convection.

4.2.3 Perturbations de la relation de CB02

On présente ici un premier test de perturbation de la relation CB02, appliqué au cas FIRE. Une perturbation gaussienne est ajoutée au contenu \bar{r}_c/σ_s lorsque Q_1 est compris dans l'intervalle $[-5, 2]$, en accord avec les données LES de la figure 4.5c, et l'écart-type des perturbations est réglé à 0.3, également d'après les LES. Utiliser une perturbation gaussienne n'est en revanche pas en accord avec les LES qui indiquent plutôt des distributions fortement asymétriques, mais ce choix a été fait dans un premier temps par simplicité. Dans le cas où les perturbations aboutissent à un contenu en eau négatif, celui-ci est considéré nul.

Les résultats sont présentés sur la figure 4.6. Le graphe 4.6a présente les profils d'eau nuageuse à +14 h de 50 simulations AROME-MUSC et on constate que la dispersion de cet ensemble est faible, excepté à la base du stratocumulus et dans la couche sous-nuageuse. Ceci pourrait en

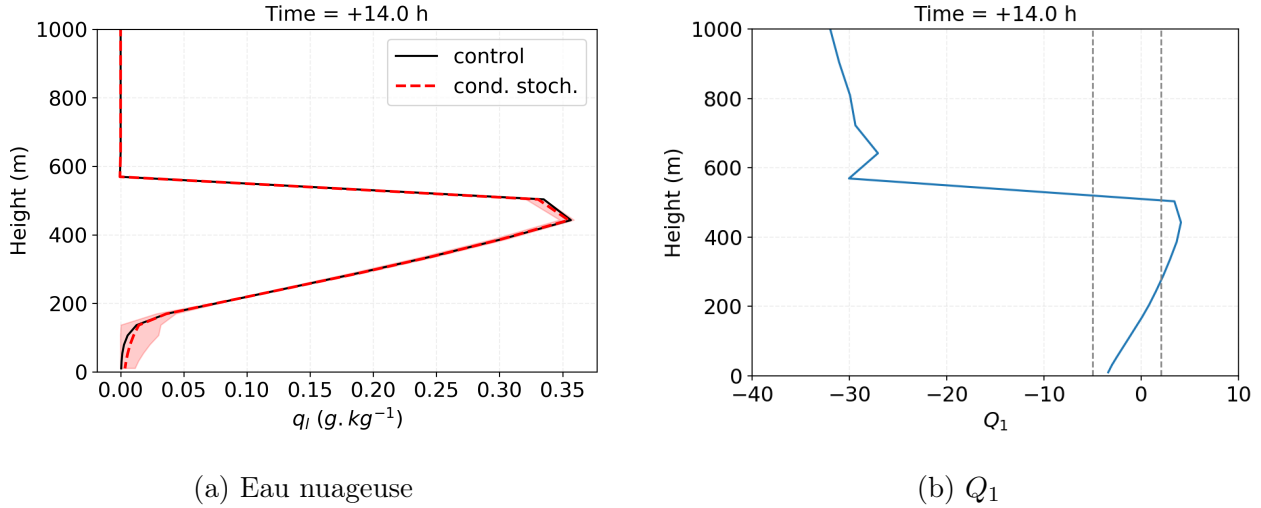


FIGURE 4.6 – Profils verticaux de l’eau nuageuse et de Q_1 à +14 h dans les simulations AROME-MUSC du cas FIRE. La courbe en pointillé rouge de la figure a correspond à la moyenne des 50 simulations AROME-MUSC et la zone en rouge correspond à l’intervalle entre les percentiles 10 et 90 de l’ensemble. La courbe noire de la figure a et la courbe de la figure b proviennent d’une simulation AROME-MUSC non perturbée.

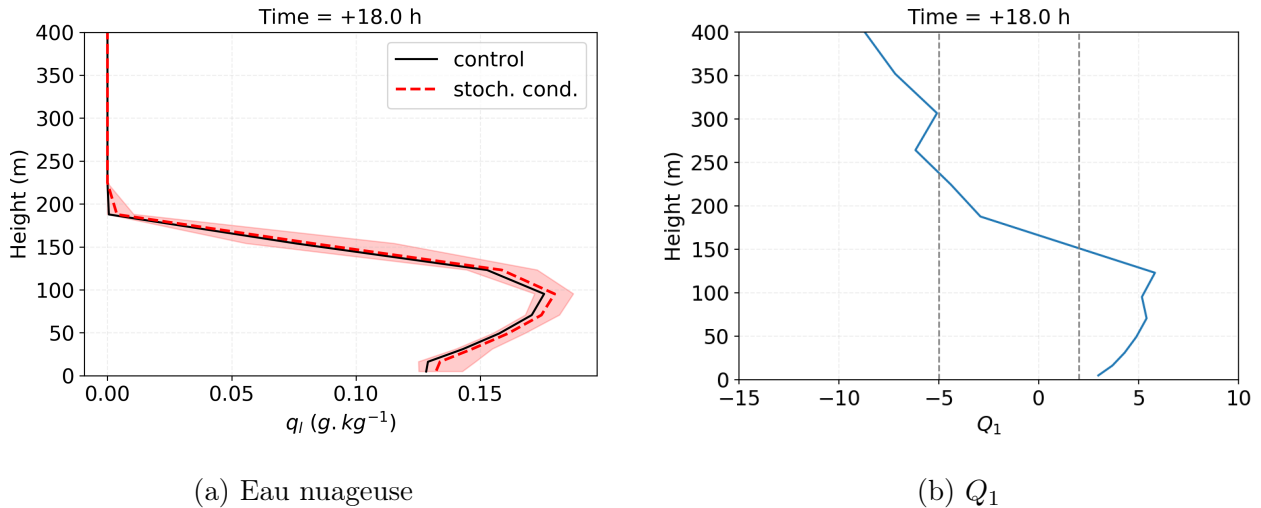


FIGURE 4.7 – Identique à la figure 4.6 pour le cas LANFEX à +18 h.

partie s’expliquer par le profil de Q_1 présenté sur le graphe 4.6b : les valeurs comprises entre -5 et 2 sont seulement présentes dans les premiers 300 m, ainsi qu’au sommet du stratocumulus. La zone perturbée ne concerne donc pas tout le nuage. À titre de comparaison, on présente dans la figure 4.7 l’impact des mêmes perturbations sur le cas de brouillard LANFEX. L’analyse de simulations LES n’ayant pas été réalisée pour ce cas, les perturbations sont réglées de la même façon que pour FIRE. On constate sur le graphe 4.7a que tout le profil d’eau nuageuse est perturbé. Des différences jusqu’à environ 10% concernant le maximum d’eau nuageuse sur le profil vertical sont observées par rapport à la référence. Les valeurs de Q_1 présentées sur le graphe 4.7b montrent qu’à +18 h, la plage de valeur $[-5, 2]$ concerne le haut de la couche de brouillard, mais, du fait du cycle de vie du brouillard qui se développe à partir du sol, le profil d’eau nuageuse serait plus efficacement perturbé que pour FIRE. Le fait que LANFEX soit

davantage déterminé par la condensation sous-maille donnée par la relation de CB02 et moins par la contribution du schéma de convection peu profonde que le cas FIRE est également une explication possible. On constate par ailleurs un biais positif sur les profils d'eau nuageuse : ceci est probablement dû au fait que les perturbations gaussiennes sont tronquées pour éviter de produire des contenus en eau négatifs. Ceci met en évidence le fait qu'une distribution plus appropriée (par exemple une distribution gamma) pourrait être choisie.

La relation empirique de CB02 n'est pas la seule source d'incertitude du schéma d'ajustement à la saturation. Cette relation permet de diagnostiquer un contenu en eau et glace nuageuse *normalisé*, et les valeurs de \bar{r}_c et \bar{r}_i sont ensuite obtenues en le multipliant par l'écart-type σ_s . Or la paramétrisation de cet écart-type est également sujette à des incertitudes (Perez et al. 2014; de Rooy et al. 2022). Dans AROME, σ_s est diagnostiqué par le schéma de turbulence. Dans les situations où la turbulence est peu active, sa valeur est trop faible, ce qui est en partie corrigé par l'ajout d'un terme proportionnel au contenu de vapeur saturante. Le coefficient de proportionnalité est le paramètre **VSIGQSAT**, qui fait partie des 8 paramètres d'AROME identifiés comme étant les plus influents par Wimmer et al. (2022). On présente sur la figure 4.8 les profils d'eau nuageuse de 50 simulations du cas FIRE où la valeur du paramètre **VSIGQSAT** a été tirée dans une distribution uniforme sur l'intervalle $[0, 0.1]$ préconisé par Wimmer et al. (2022).

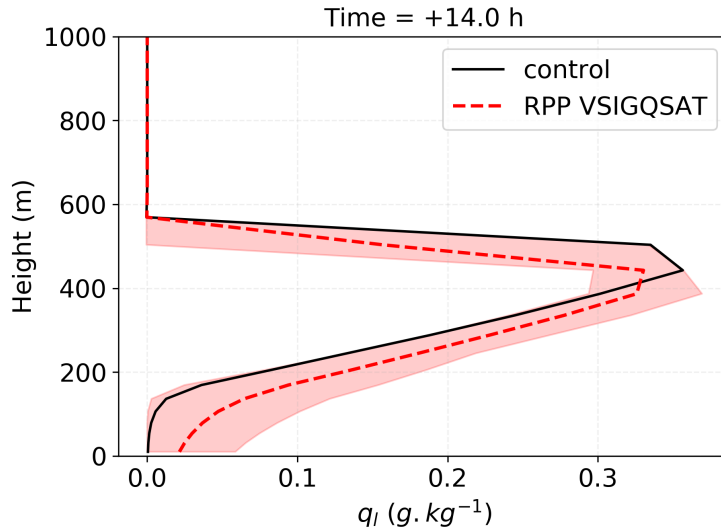


FIGURE 4.8 – Identique à la figure 4.6a, mais pour un ensemble de simulations AROME-MUSC du cas FIRE où le paramètre **VSIGQSAT** est perturbé.

On constate que la dispersion du contenu en eau nuageuse est ici plus importante qu'avec les perturbations de la relation CB02. Comme le terme impliquant **VSIGQSAT** est systématiquement ajouté à σ_s , les perturbations concernent tous les niveaux verticaux et on observe une dispersion sur tout le profil. En revanche, on observe également un biais : la moyenne de l'ensemble de 50 simulations s'écarte de la simulation non perturbée. Ceci peut être dû au choix d'une distribution uniforme sur l'intervalle $[0, 0.1]$, dont la moyenne (0.05) est différente du réglage par défaut dans AROME (0.02), ainsi qu'à l'éventuelle non-linéarité du modèle vis-à-vis de ce paramètre. Ce problème de biais d'un ensemble obtenu avec une méthode de perturbation de paramètre est abordé dans le chapitre 5.

Chapitre 5

Combinaison avec d'autres méthodes de perturbation

Ce chapitre traite de l'application de méthodes de perturbation de paramètres aux schémas microphysique et radiatif d'AROME. Certaines difficultés propres à ces méthodes, comme l'apparition de biais dans les ensembles, sont étudiées et des solutions sont proposées pour y remédier dans le cadre de travail simplifié uni-colonne. Enfin, la combinaison avec les méthodes introduites au chapitre précédent est abordée et les résultats sont comparés à d'autres approches publiées dans la littérature.

5.1 Perturbation du schéma microphysique ICE3 à l'aide d'une méthode de perturbation de paramètres

5.1.1 Motivations

Les méthodes de perturbation de type POP et basées sur des LES présentées dans le chapitre précédent concernent trois schémas de paramétrisation : la turbulence, la convection peu profonde et l'ajustement à la saturation. Comme illustré sur la figure 5.1 pour la température, ces méthodes ont un impact sur les simulations AROME-MUSC mais la dispersion produite est souvent nettement inférieure à celle générée avec SPPT, ce qui motive l'exploration d'autres sources d'incertitude. Ici, on s'intéresse à deux schémas de paramétrisation non abordés au chapitre précédent et dont dépendent fortement les cas FIRE et LANFEX : le schéma microphysique et le schéma radiatif.

Cette première section traite de perturbations appliquées au schéma microphysique ICE3 d'AROME. Comme mentionné dans le chapitre 2, l'incertitude des schémas microphysiques provient en partie de paramètres incertains. Plusieurs études de sensibilité montrent qu'en particulier, les paramètres réglant la distribution de taille des hydrométéores peuvent avoir un impact important sur des simulations de nuages et d'orages (e.g. Grabowski 2003; Gilmore et al. 2004; van Den Heever et Cotton 2004; Posselt et Vukicevic 2010; van Weverberg et al. 2011), et de récents travaux proposent de perturber ce type de paramètres dans le cadre de la prévision d'ensemble (Thompson et al. 2021; Tsiringakis et al. 2022).

Les cas d'étude que nous utilisons étant des cas non précipitants, nous choisissons de nous intéresser aux paramètres réglant la distribution de taille des gouttelettes nuageuses dans ICE3 : $N_{0,c}$ (la concentration en nombre de gouttelettes), ν_c et α_c (deux paramètres de forme de la

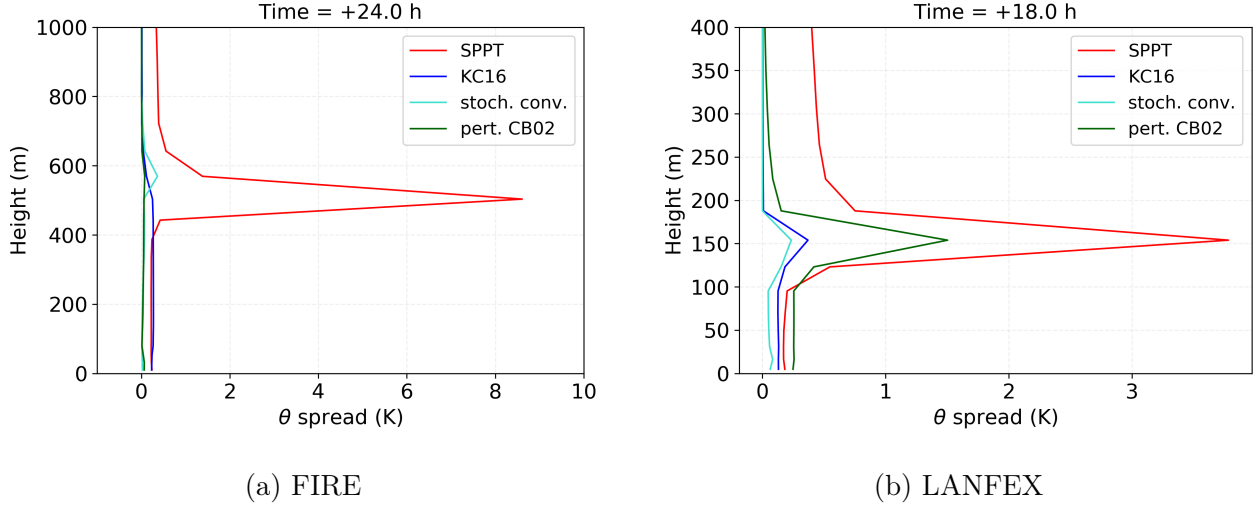


FIGURE 5.1 – Profils verticaux de dispersion de température potentielle dans des ensembles de simulations AROME-MUSC des cas FIRE à +24 h et LANFEX à +18 h. La dispersion est mesurée ici comme l’intervalle entre les percentiles 10 et 90 des ensembles. Les méthodes de perturbation utilisées pour construire les ensembles sont la méthode SPPT (courbe rouge), la méthode de KC16 (courbe bleu foncée), la convection stochastique (courbe cyan) et la perturbation de la relation de CB02 (courbe verte).

distribution, voir la section 2.3.2). Pour des raisons de simplicité, l’étude est restreinte à deux paramètres, $N_{0,c}$ et ν_c , mais le travail présenté dans la suite pourrait être étendu au paramètre α_c . Étant donné que $N_{0,c}$ et ν_c ne font pas partie de la liste des paramètres potentiellement incertains d’AROME étudiés par Wimmer et al. (2022), nous commençons cette section par une étude de sensibilité, réalisée avec le cas de brouillard LANFEX, de façon à évaluer qualitativement et quantitativement leur impact sur le modèle.

Dans la suite de ce chapitre, l’indice "c" est omis pour alléger les notations. Toutes les mentions des paramètres N_0 , ν ou α feront référence à $N_{0,c}$, ν_c , α_c .

5.1.2 Analyse de sensibilité du cas LANFEX aux paramètres N_0 et ν

L’analyse de sensibilité présentée dans cette section consiste simplement à produire plusieurs simulations du cas LANFEX avec différents réglages de N_0 et ν et à étudier la différence de contenu en eau nuageuse entre les simulations.

Sensibilité au nombre de gouttelettes N_0

On étudie ici trois simulations du cas LANFEX avec les valeurs suivantes pour le paramètre N_0 : 10, 50 et 300 gouttes/cm³. La valeur de 50 gouttes/cm³ correspond à la valeur de référence choisie pour le cas LANFEX, comme mentionné au chapitre 3, car elle correspond aux observations faites durant la campagne de mesure (voir par exemple Poku et al. 2019) et permet d’obtenir une simulation comparable à celles de l’étude de Boutle et al. (2022). La valeur plus faible de 10 gouttes/cm³ est une autre valeur utilisée par Boutle et al. (2022) à des fins de comparaison. On choisit enfin la valeur plus élevée de 300 gouttes/cm³, qui est normalement la valeur par défaut pour les surfaces continentales dans AROME.

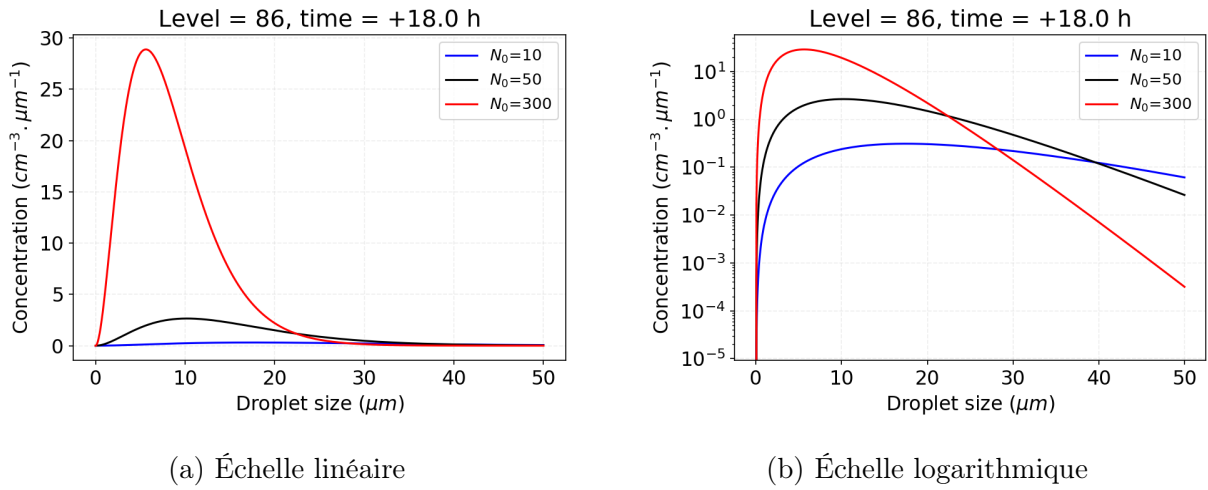


FIGURE 5.2 – Distributions de taille des gouttelettes pour trois valeurs de N_0 .

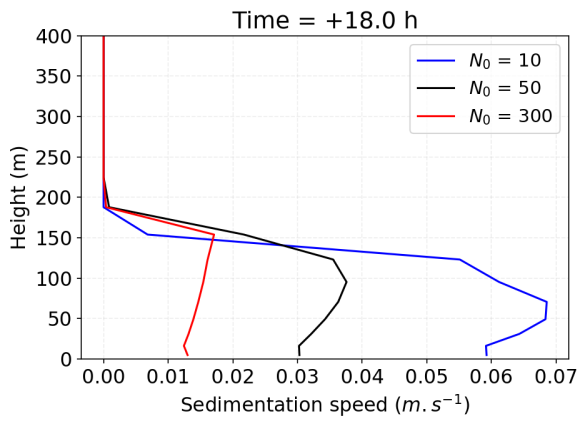


FIGURE 5.3 – Profils verticaux de la vitesse de sédimentation.

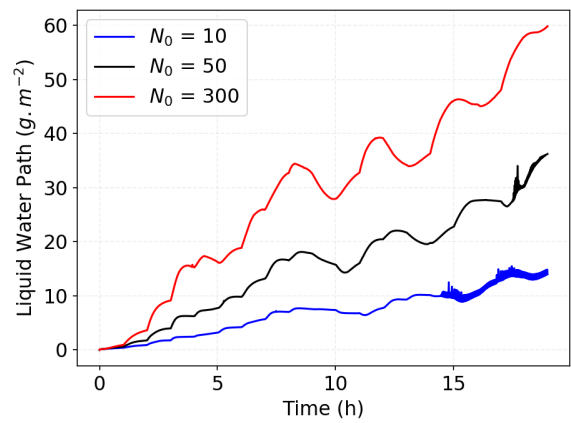
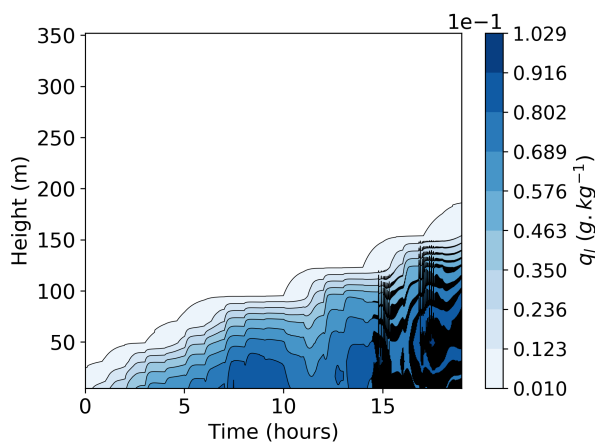
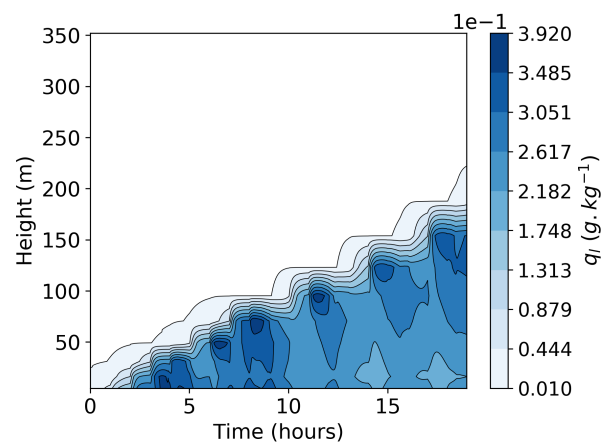


FIGURE 5.4 – Évolution temporelle du contenu en eau intégré verticalement (LWP).



(a) $N_0 = 10$ gouttes/cm³



(b) $N_0 = 300$ gouttes/cm³

FIGURE 5.5 – Diagrammes hovmoller du contenu en eau nuageuse

La figure 5.2 illustre la distribution de taille des gouttelettes que l'on obtient avec les trois valeurs de N_0 . Les courbes ont été tracées d'après l'équation 2.24 (voir le chapitre 2), en calculant λ d'après l'équation 2.25, avec le contenu en eau nuageuse de la simulation LANFEX de référence ($N_0 = 50$ gouttes/cm³) à +18h et au niveau 86 (environ 50 m d'altitude). Cette figure indique que pour un même contenu en eau, un faible nombre de gouttelettes implique logiquement une distribution décalée vers les plus grandes valeurs de tailles de gouttes – la queue de la distribution est plus importante lorsque N_0 diminue, ce que l'on peut voir sur le graphe 5.2b en échelle logarithmique.

La figure 5.3 présente la vitesse de sédimentation, calculée par le schéma de Bouteloup et al. (2011) pour chaque niveau vertical. On constate que lorsque N_0 diminue, la vitesse de sédimentation augmente. Ceci semble logique : lorsque N_0 diminue, il y a davantage de grosses gouttelettes et celles-ci sédimentent plus vite. La vitesse de sédimentation moyenne, calculée en intégrant sur tout le spectre de taille des gouttes, augmente donc. Les différences de vitesse de sédimentation de la figure 5.3 peuvent à leur tour expliquer les résultats obtenus pour le contenu en eau nuageuse intégré sur la verticale (LWP pour Liquid Water Path en anglais) : il est quatre à cinq fois plus faible lorsque $N_0 = 10$ gouttes/cm³ que lorsque $N_0 = 300$ gouttes/cm³ (figure 5.4). En effet, une vitesse de sédimentation plus élevée fait augmenter le flux descendant de gouttelettes, qui est un terme puits dans l'équation d'évolution de l'eau nuageuse. Il y a donc une diminution du contenu en eau nuageuse à tous les niveaux. Les diagrammes hovmoller de la figure 5.5 montrent d'ailleurs que le contenu en eau maximal se trouve proche de la surface dans le cas $N_0 = 10$, tandis qu'il est situé au sommet de la couche de brouillard dans la simulation $N_0 = 300$.

Sensibilité au paramètre de forme ν

On s'intéresse à présent à l'impact du paramètre ν . De même que précédemment, on étudie trois simulations où la valeur de ce paramètre est différente : 1, 3 et 15. $\nu = 3$ est la valeur par défaut d'AROME sur les surfaces continentales. Les deux autres valeurs ont été choisies d'après les intervalles de valeurs possibles pour ce paramètre rapportés par Miles et al. (2000) et Igel et van den Heever (2017a), et d'après les valeurs choisies dans l'étude de Jahangir et al. (2021).

La figure 5.6 est construite de la même façon que la figure 5.2, avec cette fois des valeurs de ν différentes. On constate que modifier la valeur du paramètre ν provoque un changement de la forme de la distribution. Elle tend vers une exponentielle pour les petites valeurs et son maximum est au contraire décalé vers la droite lorsque ν augmente (graphe 5.6a). Néanmoins, bien que le maximum de la distribution corresponde à des tailles de gouttes plus élevées lorsque ν augmente, on peut aussi constater, avec le graphe 5.6b en échelle logarithmique, que la queue de la distribution à droite diminue, c'est-à-dire que le nombre des plus grosses gouttes diminue.

La figure 5.7 montre que lorsque ν augmente, la vitesse de sédimentation est plus faible. Ainsi, bien que le maximum de la distribution corresponde à des gouttes plus grosses lorsque ν augmente, le fait que la queue de distribution soit moins importante aboutit finalement à une vitesse de sédimentation moyenne plus faible. Ceci peut se vérifier en calculant explicitement V_T (voir l'annexe A). L'équation obtenue et la figure A.1 en annexe montrent bien que V_T diminue lorsque la valeur du paramètre ν augmente, et inversement.

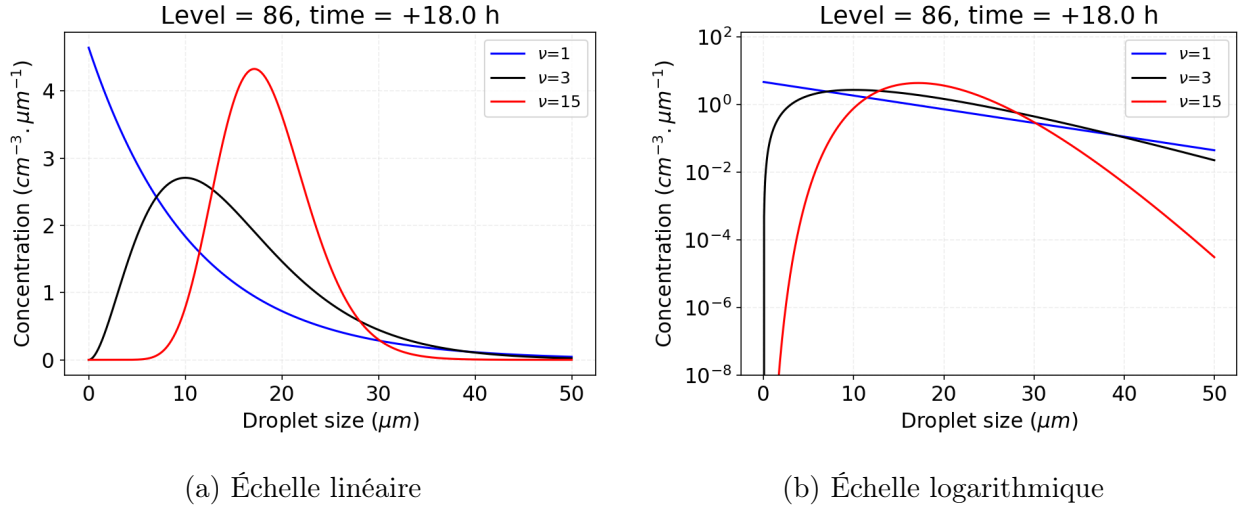
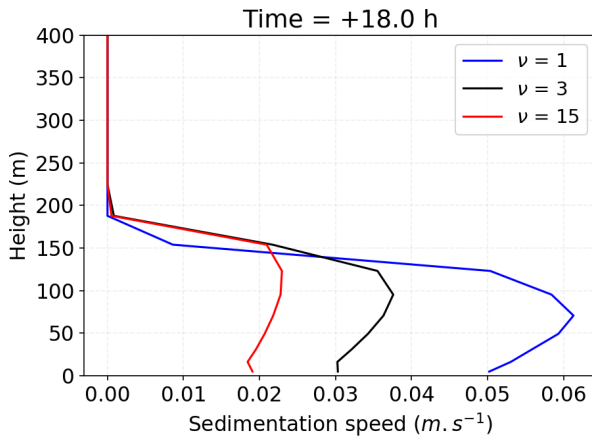

 FIGURE 5.6 – Distributions de taille des gouttelettes pour trois valeurs de ν .


FIGURE 5.7 – Vitesse de sédimentation

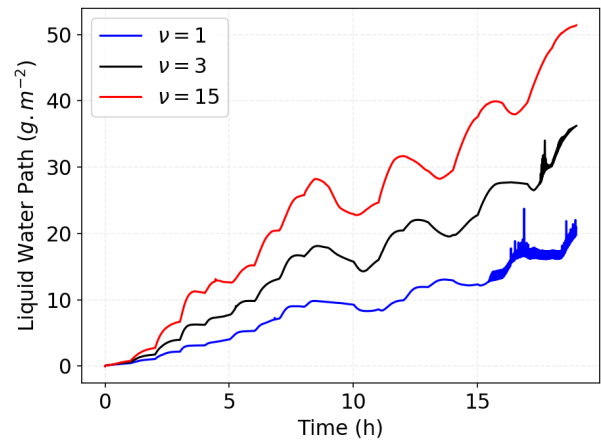


FIGURE 5.8 – Évolution temporelle du LWP.

Les résultats en terme de contenu en eau nuageuse (figure 5.8) sont cohérents avec ceux de la vitesse de sédimentation : lorsque ν augmente, les gouttelettes sédimentent en moyenne moins vite, donc le contenu en eau nuageuse de la simulation est plus élevé. Ces derniers résultats sont en accord avec Boutle et al. 2022 (voir leur figure 12).

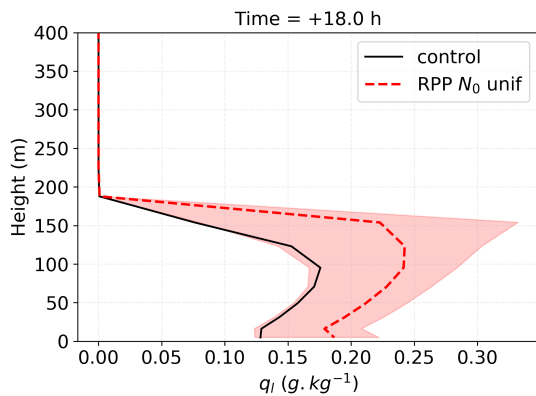
Les différentes simulations présentées ci-dessus montrent bien que les valeurs de N_0 et ν ont un impact significatif sur la simulation de brouillard LANFEX : le LWP à +18 h peut être multiplié par cinq selon les valeurs choisies. Ceci confirme donc l'intérêt potentiel de perturber ces deux paramètres dans AROME.

5.1.3 Premiers tests de perturbation des paramètres N_0 et ν

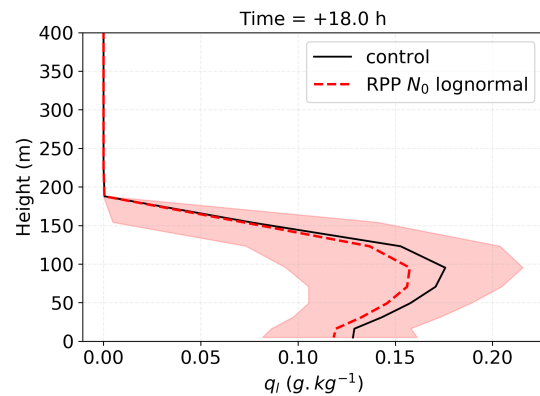
Étant donnée la sensibilité des simulations LANFEX, on souhaite à présent produire des ensembles de simulations à l'aide d'une méthode de perturbation de paramètre appliquée à N_0 et ν . Différentes méthodes existent dans la littérature, selon que la valeur du paramètre change ou non durant la simulation et dans l'espace (voir le chapitre 1). Ici, dans un souci de simplicité, nous choisissons d'appliquer une méthode "RPP" (Random Parameter Perturbation), d'après

la définition de Wimmer et al. (2022) : les valeurs des paramètres perturbés sont choisies aléatoirement pour chaque membre d'un ensemble, mais sont gardées constantes pendant toute la durée des simulations.

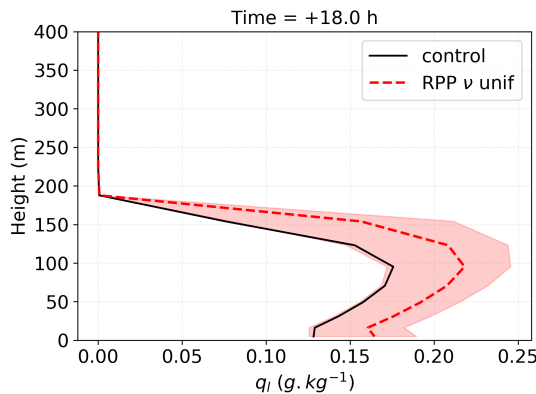
Pour appliquer la méthode de perturbation RPP, il est nécessaire de connaître la distribution dans laquelle piocher les valeurs de N_0 et ν . Dans la littérature, plusieurs types de distributions sont utilisés : distribution uniforme (Murphy et al. 2004; Tsiringakis et al. 2022), gaussienne (Christensen et al. 2015; Jankov et al. 2017, 2019), gamma (Hacker et al. 2011a), ou lognormale (Ollinaho et al. 2017; Lang et al. 2021; Frogner et al. 2022). Dans la figure 5.9, on présente les profils et la dispersion d'eau nuageuse pour 4 ensembles de simulations LANFEX où les paramètres N_0 ou ν ont été tirés dans une distribution uniforme ou lognormale. Dans le cas de la distribution uniforme, l'intervalle de valeurs choisi est $[10, 350]$ pour N_0 (en gouttes/cm³) et $[1.5, 15]$ pour ν . La valeur minimale de 1.5 a été choisie pour ν plutôt que la valeur 1 testée dans la section précédente, car celle-ci donne une distribution exponentielle qui ne semble pas réaliste pour modéliser le spectre de taille des gouttelettes nuageuses. Les distributions lognormales quant à elles sont réglées de façon à être centrées sur la valeur de référence de chaque paramètre, et à couvrir approximativement les intervalles de valeurs mentionnés.



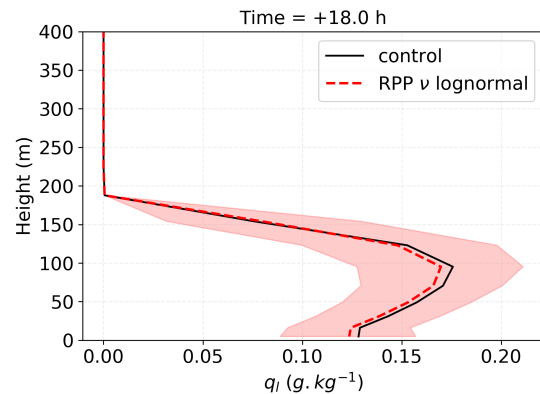
(a) Distribution uniforme de N_0



(b) Distribution lognormale de N_0



(c) Distribution uniforme de ν



(d) Distribution lognormale de ν

FIGURE 5.9 – Profils verticaux du contenu en eau nuageuse à +18 h pour quatre ensembles de simulations LANFEX faisant varier la valeur des paramètres N_0 et ν selon une distribution uniforme ou lognormale. La zone en rouge correspond à l'intervalle entre les percentiles 10 et 90 de l'ensemble. La courbe noire provient d'une simulation AROME-MUSC non perturbée.

On voit immédiatement sur les graphes 5.9a et 5.9c que les distributions uniformes produisent des ensembles dont la moyenne s'éloigne significativement de la simulation de référence. Ce résultat n'est pas surprenant, étant donné que la moyenne de ces distributions ne correspond pas aux valeurs de référence de N_0 et ν (le problème a déjà été évoqué au chapitre précédent avec la perturbation du paramètre VSIGQSAT). Les ensembles obtenus avec les distributions lognormales sont beaucoup mieux répartis autour de la simulation de référence, mais on observe quand même un biais négatif sur tout le profil d'eau nuageuse. Ce biais est observé pour toutes les échéances (et on observe également des biais sur la température et l'humidité). Ainsi, même avec des distributions centrées sur la valeur de référence de N_0 et ν , les ensembles sont biaisés. Ceci peut être expliqué par la non-linéarité du spectre de taille des gouttelettes vis-à-vis de ces deux paramètres (voir les équations 2.24 et 2.25), ce qui contribue probablement à une réponse non linéaire du modèle.

Dans ce travail de thèse, on souhaite tester des méthodes de perturbation permettant d'explorer des scénarios alternatifs à la prévision d'un modèle déterministe, sans chercher à corriger ce modèle déterministe. On souhaite que la moyenne des ensembles produits soit aussi proche que possible de la simulation de référence non perturbée. Le biais observé dans les ensembles RPP de la figure 5.9 a donc motivé la recherche de distributions alternatives pour N_0 et ν , réglées de façon à produire des ensembles non biaisés, pour ne pas changer la climatologie du modèle. Les paragraphes suivants montrent que ce problème peut être envisagé sous la forme d'un *problème inverse*.

5.1.4 Réglage des distributions de N_0 et ν par la résolution d'un problème inverse

Dans un problème inverse, on cherche à estimer un ou plusieurs paramètres d'un modèle à partir de l'observation de leurs effets. On part de la situation suivante : on dispose d'un ensemble de *mesures* y et d'un ensemble de *paramètres* λ qui sont reliés par un *modèle direct* G (*forward model*) :

$$y = G(\lambda)$$

La résolution du problème direct permet, connaissant la valeur des paramètres λ , de trouver la valeur de y . Si l'on dispose au contraire d'une série d'observations de y , et que l'on souhaite trouver la valeur la plus probable des paramètres λ qui aurait permis d'obtenir ces observations, en tenant compte de leur possible incertitude, il s'agit alors de résoudre un problème inverse.

La résolution de problèmes inverses est une technique utilisée dans de nombreux domaines scientifiques, comme en géophysique pour déterminer l'origine d'un séisme ou en imagerie médicale (Aster et al. 2018). En météorologie, elle est également utilisée, notamment dans le domaine de l'assimilation de données (Vukicevic et Posselt 2008 ; Posselt et Vukicevic 2010).

Dans notre cas d'étude, le problème inverse consiste à trouver une distribution appropriée des paramètres N_0 et ν , de façon à obtenir un ensemble de simulations du cas LANFEX non biaisé vis-à-vis du contenu en eau nuageuse (on souhaite que la moyenne de l'ensemble corresponde au contenu en eau nuageuse de la simulation de référence non perturbée).

Approche bayésienne

Le formalisme bayésien est souvent utilisé pour traiter des problèmes inverses. Le théorème de Bayes, dans sa formulation la plus générale, permet de déterminer la probabilité qu'un évènement A se produise, en ayant connaissance de la réalisation d'un autre évènement, B .

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad \text{si } P(B) \neq 0$$

où

- $P(A|B)$ est la probabilité conditionnelle que l'évènement A se produise, sachant que l'évènement B a eu lieu ; c'est la probabilité que l'on recherche ;
- $P(B|A)$ est la probabilité conditionnelle que l'évènement B ait lieu, sachant que l'évènement A s'est réalisé ;
- $P(A)$ et $P(B)$ sont les probabilités marginales des évènements A et B .

Ce théorème peut également s'écrire pour des variables aléatoires continues, grâce à la notion de densité de probabilité conditionnelle. Si l'on considère deux variables aléatoires réelles continues Y et Λ avec $f_{Y|\Lambda}$ la densité de probabilité conditionnelle de Y sachant Λ et $f_{\Lambda|Y}$ la densité de probabilité conditionnelle de Λ sachant Y , alors le théorème de Bayes s'écrit

$$f_{\Lambda|Y}(\lambda|y) = \frac{f_{Y|\Lambda}(y|\lambda)f_{\Lambda}(\lambda)}{f_Y(y)}, \quad (5.1)$$

où f_Y et f_{Λ} sont les densités de probabilité marginales des variables Y et Λ . Dans cette formulation, $f_{\Lambda|Y}$ est la densité de probabilité que l'on cherche à calculer, aussi appelée densité de probabilité *a posteriori* de Λ . f_{Λ} est au contraire la connaissance *a priori* que l'on a de la distribution de Λ . Quant à la densité de probabilité conditionnelle $f_{Y|\Lambda}$, on la désigne souvent sous le nom de fonction de *vraisemblance*. Trouver $f_{\Lambda|Y}$ nécessite de trouver $f_{Y|\Lambda}$ et f_{Λ} . f_Y est en revanche identique quelle que soit la valeur de Λ et peut donc être vue comme une constante de normalisation.

Approche de Mosegaard et Tarantola (2002)

D'autres formalismes existent, comme celui décrit par Mosegaard et Tarantola (2002) et Tarantola (2005) qui utilise le concept d'"état d'information". Cette approche a été utilisée par exemple par van Lier-Walqui et al. (2012) pour décrire l'incertitude d'un schéma de paramétrisation de la microphysique, et plus récemment pour le réglage d'un schéma de paramétrisation de la pluie (van Lier-Walqui et al. 2020).

Dans cette approche, on considère trois sources d'information : les observations d'une ou plusieurs mesures, une connaissance *a priori* sur les paramètres et un modèle reliant les paramètres aux mesures. On définit ainsi trois espaces différents, de dimensions finies : l'espace des paramètres Π , l'espace des mesures observées O , et l'espace des mesures simulées par le modèle S , ainsi que plusieurs densités de probabilité jointes : $f_{\Lambda,Y}$, définie sur $\Pi \times O$ et $g_{\Lambda,Y}$ définie sur $\Pi \times S$. $f_{\Lambda,Y}$ correspond à une information expérimentale, tandis que $g_{\Lambda,Y}$ correspond à une information théorique. On obtient une nouvelle information par la conjonction de ces deux densités de probabilité jointes (Mosegaard et Tarantola 2002), définie sur $\Pi \times D$, avec $D = O \cup S$:

$$p(\lambda, y) = \frac{1}{k} \frac{f_{\Lambda,Y}(\lambda, y)g_{\Lambda,Y}(\lambda, y)}{\mu(\lambda, y)}, \quad (5.2)$$

où μ est la densité de probabilité homogène, définie par Mosegaard et Tarantola (2002) comme une densité de probabilité proportionnelle au volume d'une région d'un espace donnée et qui se réduit à une constante dans le cas d'un espace euclidien et de l'utilisation de coordonnées cartésiennes. La densité de probabilité que l'on cherche dans le cadre d'un problème inverse est la densité de probabilité marginale de la densité jointe $p(\lambda, y)$:

$$p_{\Lambda}(\lambda) = \int_D p(\lambda, y) dy$$

On considère que les mesures observées sont indépendantes des paramètres, de sorte que $f_{\Lambda, Y}(\lambda, y) = f_{\Lambda}(\lambda)f_Y(y)$. Au contraire, les mesures simulées sont dépendantes des paramètres à travers le modèle G , on peut donc décomposer $g_{\Lambda, Y}(\lambda, y) = g_{Y|\Lambda}(y|\lambda)g_{\Lambda}(\lambda)$. Avec $g_{\Lambda}(\lambda) = \mu(\lambda)$ et $\mu(\lambda, y) = \mu(\lambda)\mu(y)$, on obtient la formule permettant de calculer p_{Λ} :

$$p_{\Lambda}(\lambda) = \int_D \frac{1}{\gamma} f_{\Lambda}(\lambda) f_Y(y) g_{Y|\Lambda}(y|\lambda) dy,$$

où γ est une constante.

Résolution du problème inverse avec l'approche de Mosegaard et Tarantola (2002)

L'approche bayésienne ainsi que celle de Mosegaard et Tarantola (2002) ont dans un premier temps été testées et aboutissent à des résultats similaires. Pour ces premiers tests, on choisit :

- la mesure $y = q_l(t, k)$, c'est-à-dire le contenu en eau nuageuse à un instant t et à un niveau vertical k donnés, avec ici $t = +18$ h et $k = 100$ m ;
- le paramètre N_0 comme seul paramètre ($\lambda = N_0$).

Le modèle G est le modèle AROME lui-même : la solution du problème direct correspond donc au contenu en eau nuageuse $q_l(t, k)$ d'une simulation AROME-MUSC du cas LANFEX où la valeur de N_0 a été définie.

On illustre ici seulement les résultats obtenus avec l'approche de Mosegaard et Tarantola (2002). La méthodologie détaillée dans Vukicevic et Posselt (2008) a été suivie pour obtenir la figure 5.10. On la résume brièvement ici :

- génération d'observations : tirages dans une distribution gaussienne centrée sur la valeur $q_l(t, k)$ de la simulation de référence LANFEX non perturbée. On obtient ainsi la distribution f_Y ;
- réalisation de simulations AROME-MUSC où l'on fait varier la valeur de N_0 entre 10 et 350 gouttes/cm³. À chaque valeur de $q_l(t, k)$ obtenue, on associe une incertitude gaussienne centrée et d'écart-type arbitrairement choisi. Ceci permet d'obtenir la distribution jointe $g_{\Lambda, Y}$ de la figure 5.10 ;
- la distribution *a priori* du paramètre (f_{Λ}) est choisie uniforme de façon à donner le moins d'information possible ;
- application de l'équation 5.2 pour trouver la distribution jointe *a posteriori* p ;
- calcul de la distribution marginale p_{Λ} .

La figure 5.10 illustre bien le passage de la distribution théorique $g_{\Lambda, Y}$ à la distribution jointe *a posteriori* p qui intègre l'information de la distribution gaussienne des observations.

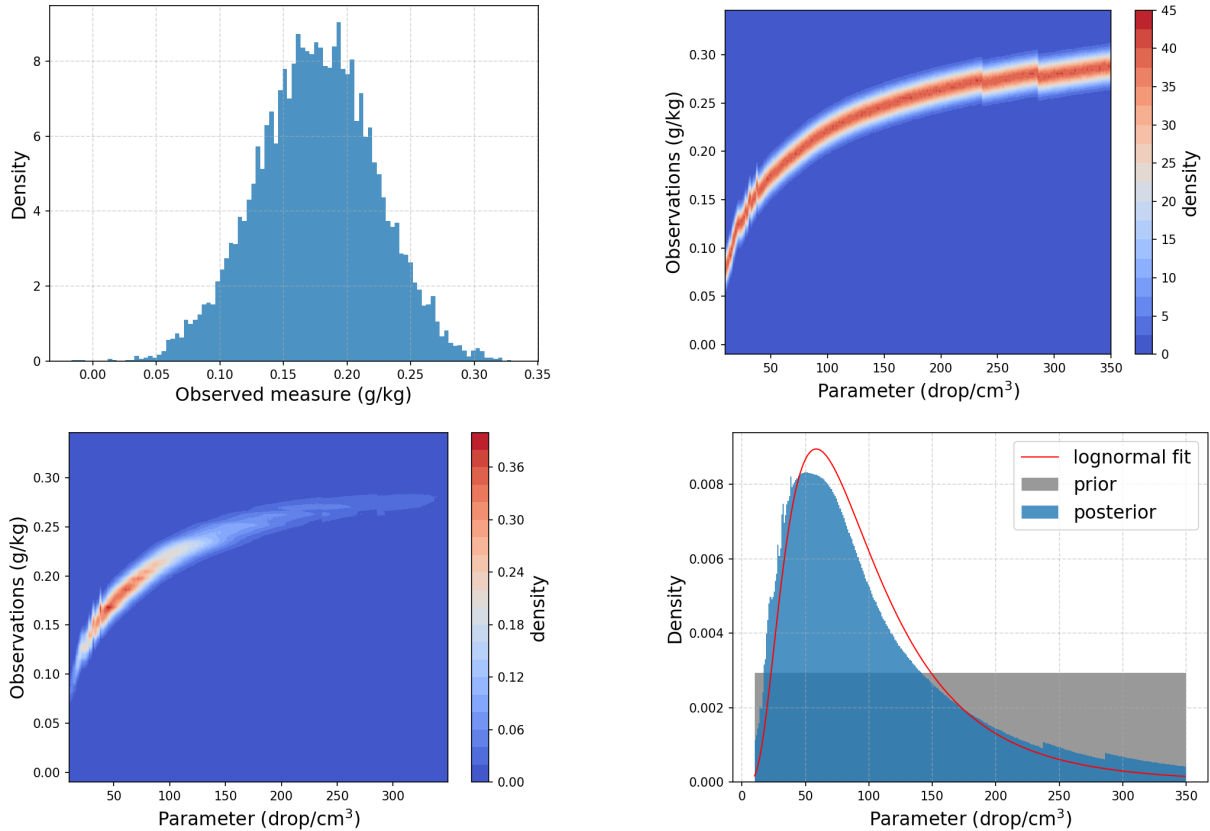


FIGURE 5.10 – De haut en bas et de gauche à droite : distribution des observations (f_Y), distribution jointe théorique ($g_{\Lambda, Y}$), la distribution jointe *a posteriori* (p) et la distribution marginale *a posteriori* (p_{Λ}).

La distribution marginale *a posteriori* du paramètre (p_{Λ}) obtenue est une distribution pouvant être raisonnablement modélisée par une loi lognormale. Le calcul de la moyenne expérimentale de la distribution p_{Λ} indique une valeur d'environ 107 gouttes/cm³, ce qui est très supérieur à la valeur de 50 gouttes/cm³ utilisée dans la simulation de référence. L'écart-type quant à lui est de 75 gouttes/cm³. Si l'on produit un ensemble en tirant les valeurs de N_0 dans une distribution lognormale réglée d'après ces résultats, on obtient un ensemble avec cette fois un biais positif sur le contenu en eau nuageuse. Comme indiqué dans l'article de Vukicevic et Posselt (2008), il est possible d'affiner l'estimation du paramètre en utilisant plusieurs mesures observées. Pour cela, les étapes décrites ci-dessus sont répétées, mais en choisissant comme distribution *a priori* du paramètre (f_{Λ}) la distribution *a posteriori* (p_{Λ}) de l'itération précédente. Les distributions *a posteriori* successivement obtenues sont ainsi de plus en plus étroites, comme illustré sur la figure 5.11, pour laquelle les contenus en eau nuageuse à +18h de 5 niveaux verticaux différents (entre 20 et 100 m) ont été utilisés comme mesures. A la fin des itérations, on obtient cette fois une moyenne expérimentale de la distribution marginale de 56 gouttes/cm³ et un écart-type de 17 gouttes/cm³. La figure 5.12 montre l'évolution de la moyenne expérimentale de la distribution marginale obtenue au fil des itérations. Il semble que cette valeur converge et qu'avec davantage de mesures on obtient une estimation de plus en plus précise. Néanmoins, ces résultats ne répondent pas vraiment au problème de départ : on souhaite trouver une distribution de paramètre N_0 , qui, lorsqu'elle est utilisée pour réaliser des simulations AROME-MUSC, donne un ensemble dont le contenu en eau nuageuse suit une distribution voulue (non biaisée et avec une dispersion choisie). Autrement dit, on peut définir une distribution de la

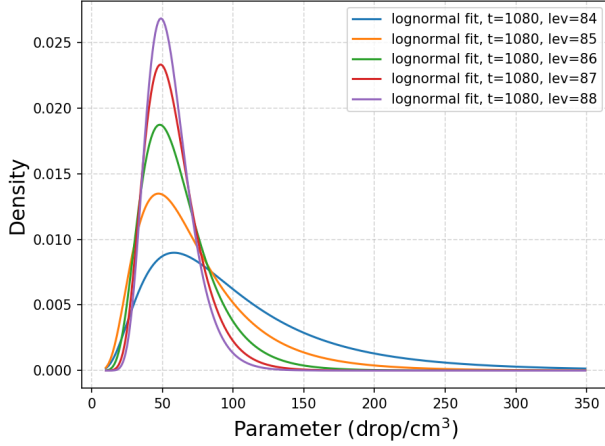


FIGURE 5.11 – Modélisation de la distribution *a posteriori* p_Λ par une loi lognormale au fil des itérations de la méthode de Vukicevic et Posselt (2008)

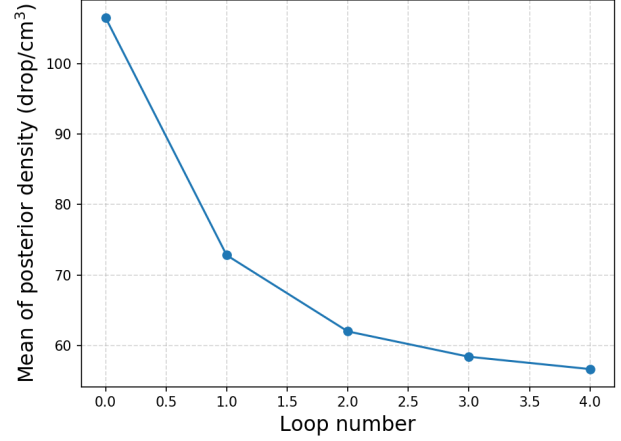


FIGURE 5.12 – Évolution de la moyenne expérimentale de la distribution *a posteriori* p_Λ obtenue au fil des itérations.

mesure f_Y , et l'on souhaite que la distribution *a posteriori* p_Λ , une fois propagée à travers le modèle, reproduise cette distribution f_Y . On souhaite donc trouver les caractéristiques d'une distribution appropriée de N_0 et non pas seulement la valeur la plus probable de N_0 . Ceci nous a mené à envisager une autre méthode, proposée par Butler et al. (2018) et nommée "*Bayesian Push-forward based Inference*".

Approche de Butler et al. (2018) : Bayesian Push-forward based Inference

La méthode proposée par Butler et al. (2018) a pour but de trouver une distribution *a posteriori* des paramètres cohérente avec la distribution des observations de départ et le modèle. C'est-à-dire que la distribution *a posteriori* des paramètres, une fois passée dans le modèle, donne une distribution, caractérisée de "push-forward", qui correspond bien à la distribution des observations utilisée pour réaliser l'inférence.

Butler et al. (2018) proposent une nouvelle équation pour le calcul de la distribution *a posteriori* des paramètres, que l'on donne ci-dessous en adaptant légèrement les notations de leur article :

$$\pi_\Lambda^{\text{post}}(\lambda) = \pi_\Lambda^{\text{prior}}(\lambda) \frac{\pi_Y^{\text{obs}}(G(\lambda))}{\pi_Y^{G(\text{prior})}(G(\lambda))} \quad (5.3)$$

$\pi_\Lambda^{\text{post}}$ correspond à ce qui a précédemment été noté p_Λ , soit la distribution *a posteriori* des paramètres que l'on cherche. $\pi_\Lambda^{\text{prior}}$ correspond à f_Λ , soit la distribution *a priori* de ces paramètres. π_Y^{obs} est la vraisemblance des résultats du modèle G par rapport aux observations. Elle présente des similarités avec la fonction de vraisemblance utilisée dans l'approche bayésienne classique ($f_{Y|\Lambda}$ dans l'équation 5.1) mais n'est pas construite de la même manière : dans l'approche bayésienne classique, on suppose que les observations sont bruitées par rapport au modèle et on fait une hypothèse sur la distribution du bruit de façon à calculer la vraisemblance. Dans l'approche de Butler et al. (2018), on part de la distribution observée de la mesure et on calcule ainsi la probabilité, avec cette distribution, d'une valeur $G(\lambda)$. Enfin, la distribution $\pi_Y^{G(\text{prior})}$ n'est présente dans aucune des approches présentées précédemment. Il s'agit de la distribution dans l'espace des observations que l'on obtient après avoir propagé la distribution *a priori* des

paramètres à travers le modèle G . Elle est ainsi qualifiée de distribution "push-forward". Dans la section suivante, on détaille l'implémentation de cette méthode pour obtenir une distribution des paramètres N_0 et ν cohérente avec une distribution d'eau nuageuse prédéfinie du cas LANFEX.

Résolution du problème inverse avec l'approche de Butler et al. (2018)

De même que précédemment, on choisit comme mesure le contenu en eau nuageuse $q_l(t, k)$ avec $t = +18$ h et $k = 100$ m. De façon à rester dans un cas simple unidimensionnel, on commence par traiter les paramètres N_0 et ν séparément.

La distribution *a priori* de chaque paramètre est choisie comme étant une distribution uniforme, définie sur les intervalles $[10, 350]$ gouttes/cm³ et $[1.5, 15]$ pour N_0 et ν respectivement.

On ne dispose pas d'observations réelles de $q_l(t, k)$, mais on souhaite que la distribution π_Y^{obs} de l'eau nuageuse soit centrée sur la valeur de la simulation de référence et ait une dispersion non négligeable. Ici, on décide de considérer une distribution gaussienne, centrée sur la valeur de référence et avec un écart-type arbitrairement choisi qui correspond à la dispersion obtenue dans les ensembles de simulations LANFEX où N_0 et ν sont tirés dans une loi lognormale non calibrée (voir les figures 5.9b et 5.9d). Pour l'ensemble où N_0 est perturbé, la dispersion de l'eau nuageuse à $t = +18$ h et $k = 100$ m est d'environ 0.046 g.kg⁻¹. C'est l'écart-type choisi pour la distribution π_Y^{obs} tracée sur la figure 5.14. Pour l'ensemble où ν est perturbé, la dispersion de l'eau nuageuse à $t = +18$ h et $k = 100$ m est d'environ 0.028 g.kg⁻¹.

La distribution "push-forward" est obtenue en suivant l'algorithme 1 de Butler et al. (2018). Pour cela, un ensemble de valeurs du paramètre, uniformément réparties dans l'intervalle de définition, est choisi. Des simulations du cas LANFEX avec chacune des valeurs sont réalisées. Les sorties du modèle (le contenu $q_l(t, k)$ de chaque simulation) permet alors d'obtenir, à l'aide d'une méthode d'estimation par noyau, la distribution "push-forward". La figure 5.13 illustre cette étape. Une fois la distribution "push-forward" obtenue, on peut calculer la distribution *a posteriori* en utilisant l'équation 5.3, ce qu'illustre la figure 5.14.

Ces étapes sont réalisées indépendamment pour le paramètre N_0 et pour le paramètre ν . Pour chacun, on obtient une distribution *a posteriori* pouvant être modélisée de façon satisfaisante par une loi lognormale (figure 5.15). La moyenne et l'écart-type expérimentaux de la distribution *a posteriori* sont indiqués sur le tableau 5.1. On constate que la moyenne expérimentale diffère de la valeur par défaut des deux paramètres.

	moyenne	écart-type
N_0 (gouttes/cm ³)	62	49
ν	3.4	2.6

TABLE 5.1 – Moyenne et écart-type expérimentaux des distributions *a posteriori* obtenues avec la méthode de Butler et al. (2018).

Ces résultats ont été utilisés pour produire des ensembles RPP de simulations du cas LANFEX : pour chaque simulation, une valeur du paramètre N_0 et/ou du paramètre ν est tirée

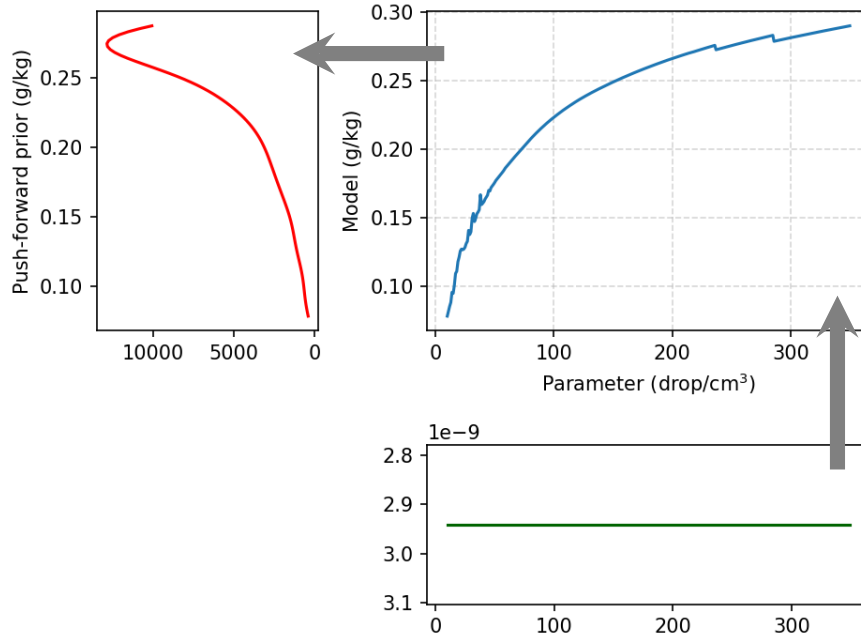


FIGURE 5.13 – Construction de la distribution $\pi_Y^{G(\text{prior})}$

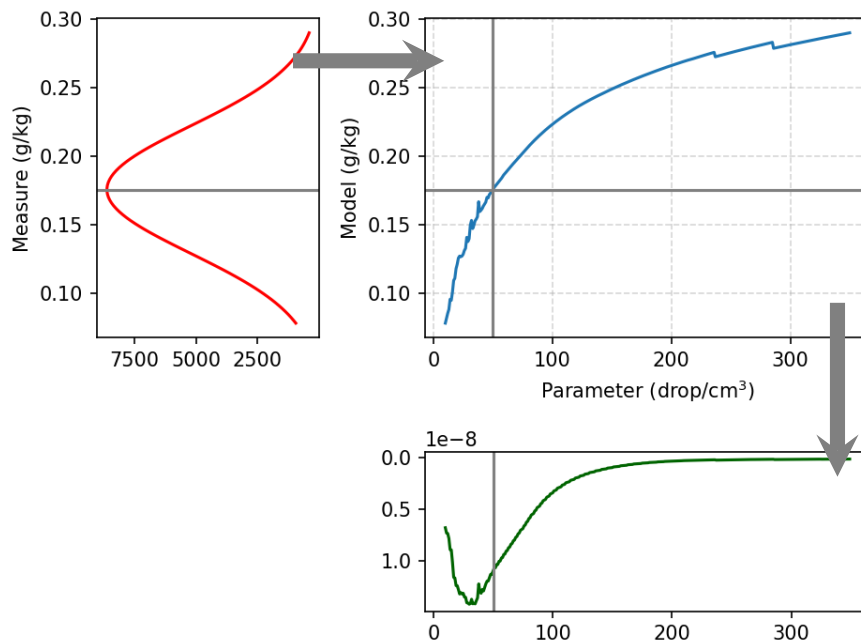
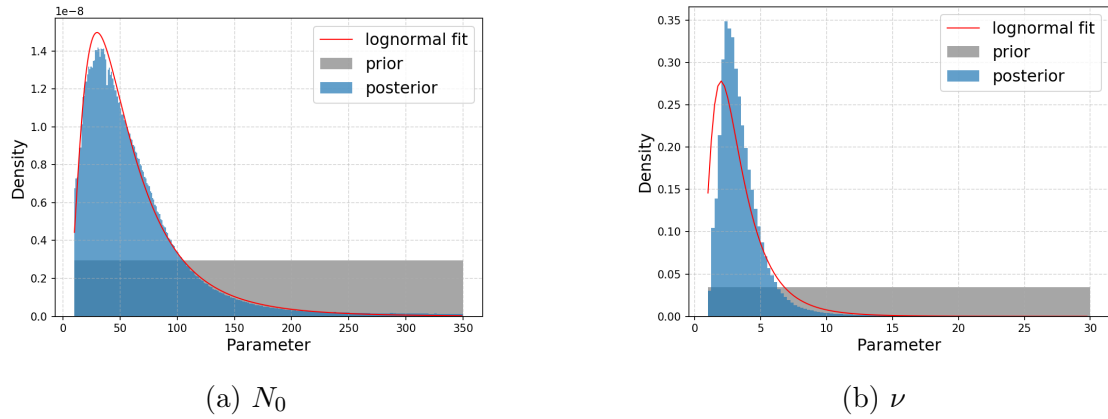
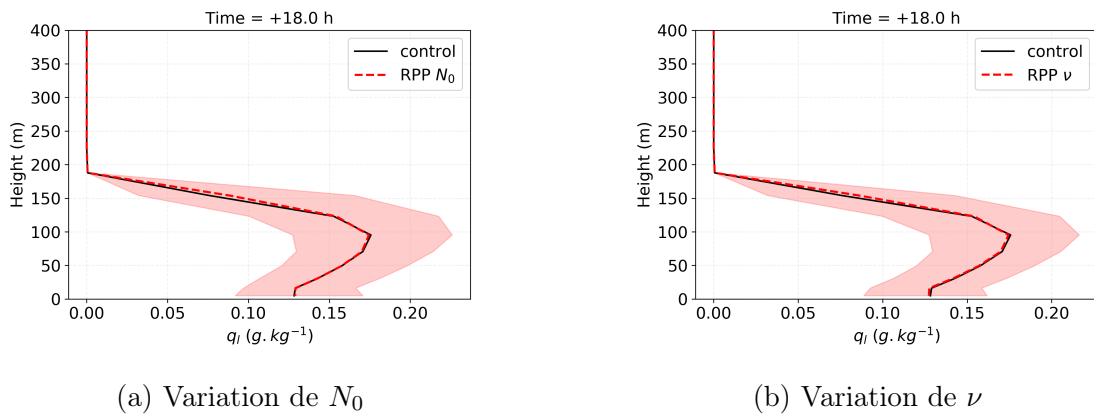


FIGURE 5.14 – Calcul de la distribution $\pi_\Lambda^{\text{post}}$

dans une distribution lognormale, dont la moyenne et l'écart-type sont ceux du tableau 5.1. La figure 5.16 présente le profil d'eau nuageuse à +18 h pour des ensembles de 100 simulations où un seul des deux paramètres est perturbé. On constate que le biais présent sur les profils de la figure 5.9 a disparu pour la plupart des niveaux verticaux. Ainsi, l'utilisation d'une seule mesure à un niveau vertical donné permet, avec la résolution du problème inverse, de supprimer le biais sur la plupart des niveaux verticaux du profil.


 FIGURE 5.15 – Distributions *a posteriori* $\pi_{\Lambda}^{\text{post}}$ obtenues avec la méthode de Butler et al. (2018).

 FIGURE 5.16 – Profils verticaux du contenu en eau nuageuse à +18h pour deux ensembles de simulations faisant varier la valeur des paramètres N_0 ou ν selon une loi lognormale réglée d'après les résultats de la résolution du problème inverse.

Les distributions précédentes ont été obtenues en résolvant le problème inverse indépendamment pour les paramètres N_0 et ν , mais il est également possible de résoudre le problème pour les deux paramètres simultanément pour tenir compte des possibles interactions entre eux. On a alors $\lambda = (N_0, \nu)$ et l'espace des paramètres est de dimension 2. Le nombre de points nécessaires pour échantillonner correctement cet espace est donc plus important que précédemment. La figure 5.17 représentant le modèle direct ($G(\lambda)$) a ainsi été construite avec 1995 simulations du cas LANFEX en faisant varier simultanément N_0 et ν de façon uniforme sur leurs intervalles de définition.

De même que précédemment, ces simulations sont utilisées pour calculer la distribution "push-forward" $\pi_Y^{G(\text{prior})}$ selon l'algorithme 1 de Butler et al. (2018). Ensuite, la distribution *a posteriori* est calculée en sachant que $\pi_{\Lambda}^{\text{prior}}(\lambda) = \pi^{\text{prior}}(N_0)\pi^{\text{prior}}(\nu)$ puisqu'on fait varier les valeurs des deux paramètres indépendamment l'un de l'autre. Elle est représentée sur la figure 5.18. Les marginales de cette distribution sont indiquées sur les côtés du graphe. De même que précédemment, la distribution marginale de N_0 peut être approchée de façon satisfaisante avec une loi lognormale. En revanche, la distribution marginale de ν se rapproche davantage d'une loi exponentielle, mais sa queue de distribution importante la rend difficile à modéliser. On choisit donc cette fois de sélectionner des valeurs de N_0 et ν directement à partir de la distribution expérimentale, grâce à l'algorithme 2 de Butler et al. (2018) qui est un algo-

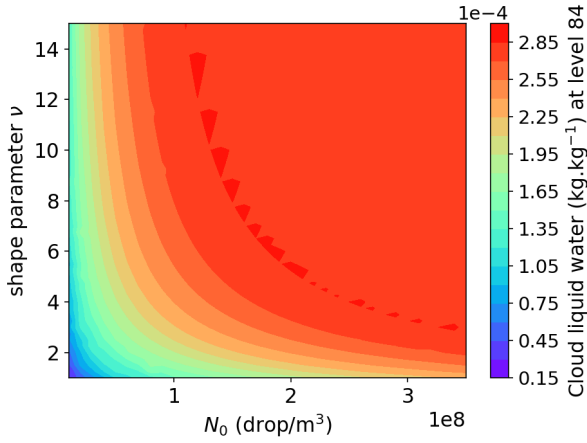


FIGURE 5.17 – Estimation du modèle direct $G(\lambda)$ pour $\lambda = (N_0, \nu)$ à l'aide de 1995 simulations du cas LANFEX.

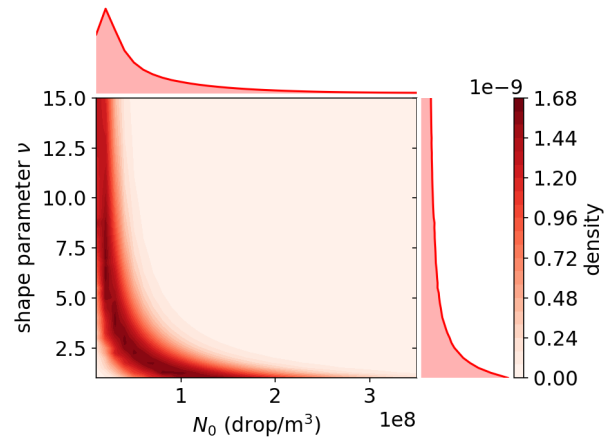


FIGURE 5.18 – Distribution *a posteriori* $\pi_{\Lambda}^{\text{post}}(\lambda)$ pour $\lambda = (N_0, \nu)$.

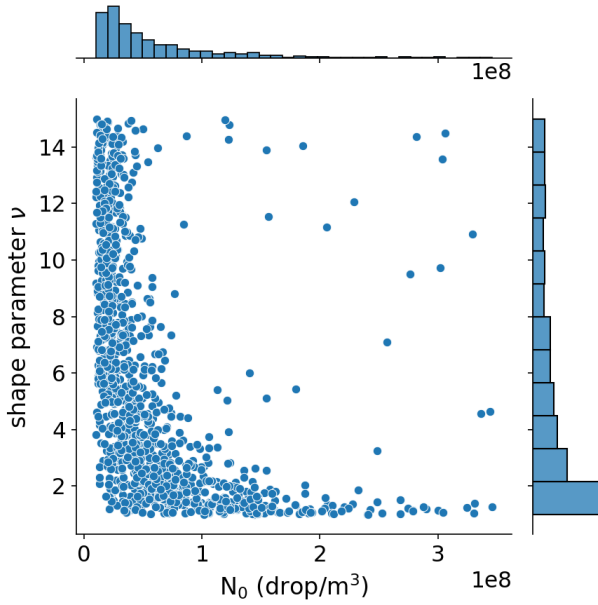


FIGURE 5.19 – Échantillonnage de 10000 valeurs de la distribution *a posteriori* à l'aide d'une méthode de rejet.

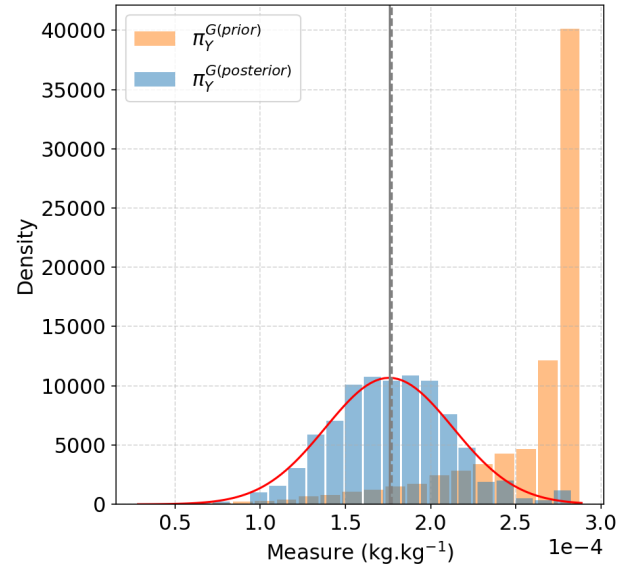


FIGURE 5.20 – Distribution de la mesure obtenue avec la distributions de paramètre *a priori* ("push-forward of the prior") et avec la distribution *a posteriori* ("push-forward of the posterior").

ritme de rejet. Le résultat d'un échantillonnage de 10 000 valeurs est présenté sur la figure 5.19.

Enfin la figure 5.20 permet de vérifier que l'objectif de la méthode de Butler et al. (2018) est atteint, c'est-à-dire que la distribution *a posteriori* permet d'obtenir une distribution de la mesure conforme à π_Y^{obs} . Pour cela, on sélectionne aléatoirement des valeurs de (N_0, ν) dans la distribution *a posteriori* avec l'algorithme de rejet, puis on calcule la mesure $q_l(t, k)$ correspondante grâce à une interpolation bilinéaire de la fonction à deux variables représentant le modèle direct (figure 5.17). L'histogramme des mesures ainsi obtenues est représenté en bleu

sur la figure 5.20 et on constate qu'il s'approche de la courbe rouge représentant π_Y^{obs} . À titre de comparaison, on indique en orange la distribution de la mesure obtenue avec la distribution *a priori* uniforme : cette distribution est fortement asymétrique, du fait de la forme du modèle direct (figure 5.17).

La figure 5.21 présente enfin les profils d'eau nuageuse (en bleu) d'un ensemble RPP où les valeurs de N_0 et ν ont été tirées dans la distribution *a posteriori* de la figure 5.18. L'ensemble produit avec des valeurs de N_0 et ν tirées dans des distributions lognormales centrées sur la valeur de référence de chaque paramètre est représenté en rouge. On constate que la calibration des distributions grâce au problème inverse réduit bien le biais d'eau nuageuse sur presque tout le profil. On voit également que la dispersion de l'ensemble calibré est plus faible que celle de l'ensemble utilisant des distributions lognormales non calibrées. Ceci est dû à la spécification de π_Y^{obs} dans le problème inverse : l'écart-type de cette distribution a été réduit par rapport à l'écart-type de la distribution de l'ensemble biaisé (il a été réglé à environ 0.04 g.kg^{-1} , soit 70% de la valeur de l'écart-type de l'ensemble biaisé). Ce choix a été fait car avec l'écart-type de l'ensemble biaisé, la distribution gaussienne π_Y^{obs} couvre des valeurs de $q_l(t, k)$ non prévues par le modèle direct $G(\lambda)$ tel qu'estimé à partir de 1995 simulations où $N_0, \nu \in [10, 350] \times [1, 15]$, ce qui aboutit à une distribution "push-forward of the posterior" toujours biaisée vis-à-vis de la référence. Pour conserver un écart-type égal à celui de l'ensemble biaisé, il faudrait probablement étendre les intervalles de N_0 et ν utilisés pour construire le modèle direct, au risque d'inclure des valeurs non physiques, ou bien choisir une autre forme de distribution pour π_Y^{obs} .

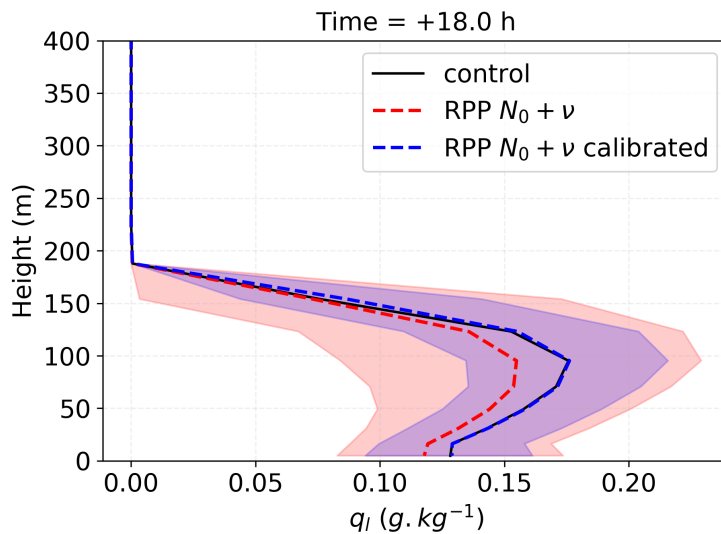


FIGURE 5.21 – Profils verticaux du contenu en eau nuageuse à +18 h dans la simulation de référence LANFEX (courbe noire) et deux ensembles RPP de 50 simulations : en rouge, un ensemble où les valeurs de N_0 et ν sont tirées chacune dans une distribution lognormale centrée sur la valeur de référence du paramètre, et en bleu un ensemble où les valeurs de N_0 et ν proviennent de l'échantillonnage de la distribution *a posteriori* du problème inverse. Les courbes en pointillé correspondent à la moyenne de chaque ensemble, et les zones colorées représentent l'intervalle entre les percentiles 10 et 90.

Conclusions sur le problème inverse

Les résultats présentés dans cette partie montrent que la résolution d'un problème inverse peut permettre de calibrer les distributions de paramètres à perturber, de façon à obtenir des ensembles RPP non biaisés. Ici la méthode de *bayesian push-forward based inference* de Butler et al. (2018) a été utilisée pour calibrer les distributions de deux paramètres microphysiques : N_0 et ν . Cette méthode pourrait en théorie être étendue à davantage de paramètres. Néanmoins, l'augmentation de taille de l'espace des paramètres rend plus difficile son échantillonnage efficace, et représente un coût numérique plus important pour produire toutes les simulations qui donnent le résultat du problème direct $G(\lambda)$. Il n'est en outre pas garanti que l'optimisation effectuée sur ce cas de brouillard produise d'aussi bons résultats sur d'autres situations météorologiques. Des tests sur d'autres cas d'étude et avec un modèle 3D seraient nécessaires pour le prouver.

5.2 Couplage entre perturbation microphysique et perturbation du schéma radiatif

5.2.1 Principe du couplage

Comme vu dans le chapitre 2, la paramétrisation des flux radiatifs dépend notamment des propriétés optiques des nuages. Les propriétés optiques peuvent être calculées à l'échelle d'une gouttelette nuageuse, en fonction de son diamètre, par exemple grâce à la théorie de Lorenz-Mie. Pour estimer ces propriétés à l'échelle d'une couche nuageuse dans une maille de modèle, il faut donc tenir compte de toutes les tailles des gouttes présentes dans le nuage. Dans de nombreux schémas, l'intégration sur la distribution de taille est évitée grâce à des approximations et l'utilisation d'un *rayon effectif* (Stephens 1978), défini par :

$$r_{\text{eff}} = \frac{\int_0^\infty n(r)r^3 dr}{\int_0^\infty n(r)r^2 dr}, \quad (5.4)$$

soit le rapport entre les moments d'ordre 3 et 2 de la distribution de taille des gouttes.

Les schémas microphysique et radiatif sont donc en théorie liés par leur dépendance à la distribution de taille des gouttelettes nuageuses. Néanmoins, dans AROME, les deux schémas utilisent actuellement des hypothèses incohérentes quant à cette distribution. En effet, dans le schéma radiatif, le rayon effectif n'est pas calculé directement à partir de l'équation 5.4 mais grâce à la paramétrisation de Martin et al. (1994), qui ne fait aucune hypothèse sur la forme de la distribution et propose la relation suivante, réglée à partir d'observations aéroportées de stratocumulus :

$$r_{\text{eff}} = \left(\frac{1}{k}\right)^{1/3} \left(\frac{3LWC}{4\pi\rho_l N_0}\right)^{1/3}, \quad (5.5)$$

avec LWC un contenu en eau nuageuse défini comme une masse volumique et ρ_l la masse volumique de l'eau liquide. La relation dépend de deux paramètres : N_0 , qui est la concentration en nombre de gouttelettes vue dans les sections précédentes, et k .

La valeur de N_0 dans le schéma radiatif est différente de celle du schéma microphysique. Elle est réglée d'après Martin et al. (1994) en fonction de concentrations en aérosols (CCN pour Cloud Condensation Nuclei) :

- $N_0 = -1.15 \times 10^{-3} \text{CCN}^2 + 0.963 \text{CCN} + 5.30$ pour les surfaces océaniques,
- $N_0 = -2.10 \times 10^{-4} \text{CCN}^2 + 0.568 \text{CCN} - 27.9$ pour les surfaces continentales.

Dans AROME, pour les surfaces océaniques, $\text{CCN} = 50 \text{ cm}^{-3}$, ce qui donne $N_0 \approx 50$ gouttes/ cm^3 . Pour les surfaces continentales, CCN est réglé à 900 cm^{-3} , ce qui donne $N_0 \approx 313$ gouttes/ cm^3 . L'hypothèse du schéma radiatif concernant N_0 est donc plutôt en accord avec celle du schéma microphysique pour les surfaces continentales, en revanche, sur les océans l'hypothèse du schéma radiatif est deux fois plus faible.

Le paramètre k est lui donné par la formule qui le relie à la dispersion spectrale d :

$$k = \frac{(1 + d^2)^3}{(1 + 3d^2)^2} \quad (5.6)$$

d est définie dans Martin et al. (1994) comme le rapport entre l'écart-type de la distribution de taille des gouttelettes et leur rayon moyen, mais en pratique sa valeur est fixée à :

- $d = 0.33$ pour les surfaces océaniques,
- $d = 0.43$ pour les surfaces continentales.

Avec la relation 5.5, le rayon effectif ne dépend pas explicitement de la forme de la distribution de taille des gouttelettes nuageuses. Ce problème est souligné par Jahangir et al. (2021) qui déterminent une nouvelle expression du rayon effectif en repartant de sa définition initiale 5.4. L'équation proposée est très proche de celle de Martin et al. (1994), mais le paramètre k est remplacé par une expression qui dépend explicitement de la forme de la distribution. Pour une distribution de type gamma généralisée, on a (voir Jahangir et al. (2021), table 1) :

$$k = \frac{(\nu^2 + \nu)}{(\nu + 2)^2} \quad (5.7)$$

Cette relation n'est valable que dans le cas où le paramètre α de la loi gamma généralisée est égal à 1, mais elle peut être généralisée au cas $\alpha > 0$ (voir l'annexe B) :

$$k = \frac{\Gamma(\nu + 2/\alpha)^3}{\Gamma(\nu)\Gamma(\nu + 3/\alpha)^2} \quad (5.8)$$

Si l'on fixe α à 1 ou 3, l'évolution de k en fonction de ν est donnée respectivement par les courbes rouge et bleue sur la figure 5.22. La valeur de k dans la paramétrisation de Martin et al. (1994) (donnée par l'équation 5.6) est indiquée par les traits horizontaux rouge et bleu pour les surfaces continentales et océaniques respectivement. Cela correspond à des valeurs de ν d'environ 7 et 1.22 respectivement, ce qui est proche de l'hypothèse microphysique pour les océans ($\nu = 1$) mais significativement différent dans le cas des continents ($\nu = 3$).

Les relations précédentes montrent qu'il est possible d'unifier les hypothèses du schéma microphysique et du schéma radiatif concernant la distribution de taille des gouttelettes nuageuses. On peut en effet régler N_0 dans le schéma radiatif avec les mêmes valeurs que dans le schéma microphysique, et la relation 5.8 peut être utilisée pour calculer k d'après les valeurs de ν et α du schéma microphysique. Dans la section suivante, on étudie l'impact d'une telle modification sur les cas LANFEX et FIRE, ainsi que la sensibilité de la paramétrisation des flux SW et LW à des modifications de N_0 et ν .

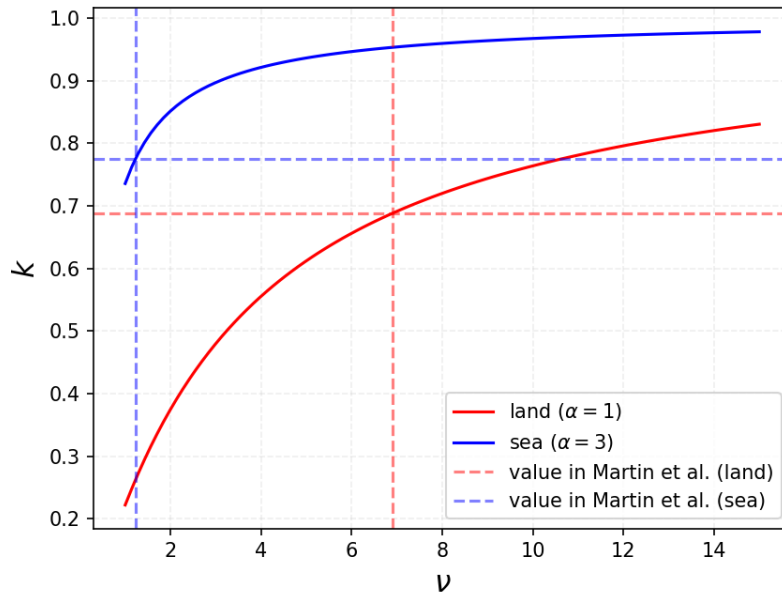


FIGURE 5.22 – Évolution du coefficient k en fonction de ν d'après l'équation 5.8 pour deux valeurs de α .

5.2.2 Étude de sensibilité

Comme on souhaite mettre en cohérence les hypothèses faites dans les schémas microphysique et radiatif quant aux paramètres N_0 , α et ν , une nouvelle configuration d'AROME est choisie comme référence dans cette section. Au lieu de la configuration "oper", où le rayon effectif est calculé avec la formule originale de Martin et al. (1994), on utilise une version modifiée où k est donné par l'équation 5.8, avec les valeurs de α et ν par défaut du schéma microphysique ($\alpha = 1$, $\nu = 3$ sur continent, $\alpha = 3$, $\nu = 1$ sur océan). N_0 est réglé à 50 gouttes/cm³ pour le cas LANFEX et à 100 gouttes/cm³ pour le cas FIRE. Cette nouvelle configuration donne des résultats légèrement différents de la configuration opérationnelle pour le cas FIRE, mais pratiquement identiques pour le cas LANFEX (figure 5.23).

La figure 5.24a, réalisée à partir de plusieurs simulations du cas LANFEX dans lesquelles la valeur de N_0 ou ν a été modifiée dans le schéma radiatif uniquement, permet de vérifier que N_0 et ν ont bien un impact sur le calcul du rayon effectif : passer de 50 gouttes/cm³ à 300 gouttes/cm³ (à ν constant) divise environ par deux le rayon effectif (les gouttes sont plus petites, ce qui est cohérent avec une augmentation de leur nombre total) et choisir $\nu = 15$ au lieu de $\nu = 3$ (à N_0 constant) a également pour effet de diminuer le rayon effectif. On observe l'effet inverse lorsque l'on augmente la valeur de N_0 ou ν , sachant qu'il existe une limite maximale du rayon effectif à 16 μm imposée dans AROME.

Les figures 5.24b et c montrent l'impact sur les flux LW et SW. N_0 et ν semblent avoir un impact uniquement sur les flux SW : on observe une diminution de l'ordre de 20% du flux SW au sol lorsque l'on passe de 50 gouttes/cm³ à 300 gouttes/cm³, mais les modifications du flux LW sont elles négligeables. Ceci semble pouvoir s'expliquer par la sensibilité de l'épaisseur optique, paramètre-clé pour le calcul des flux radiatifs, à un changement du rayon effectif. Dans AROME, la contribution de l'eau liquide nuageuse à l'épaisseur optique d'un nuage dans

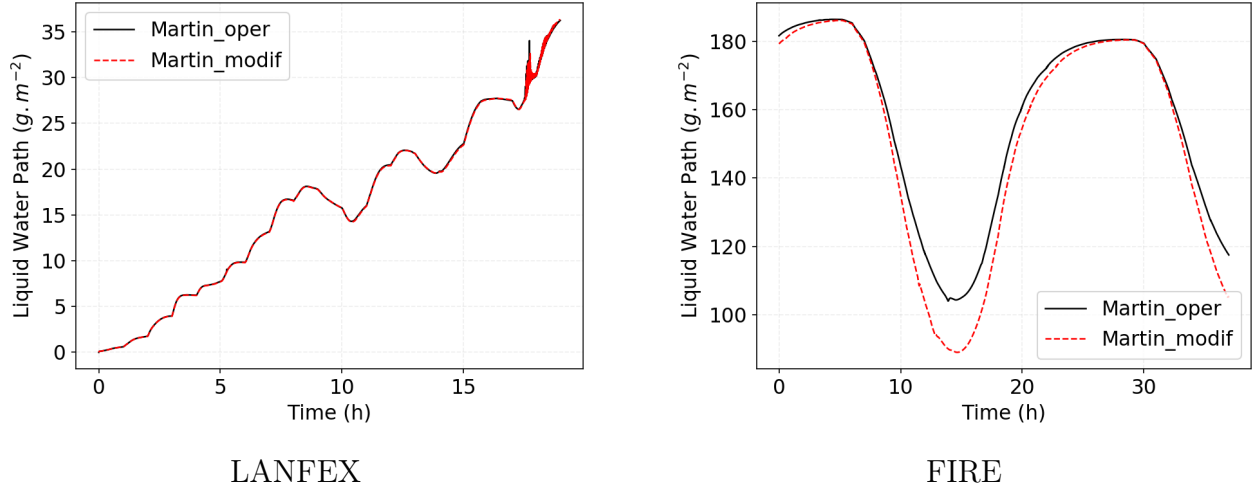


FIGURE 5.23 – Évolution temporelle du LWP dans les cas d'étude LANFEX et FIRE pour deux configurations différentes du schéma radiatif. Les courbes noires correspondent à la configuration originale du schéma de Martin et al. (1994), utilisée opérationnellement dans AROME, tandis que les courbes rouges en pointillé correspondent à la version modifiée de ce schéma, où les hypothèses concernant le spectre de taille des gouttelettes sont cohérentes avec celles du schéma microphysique.

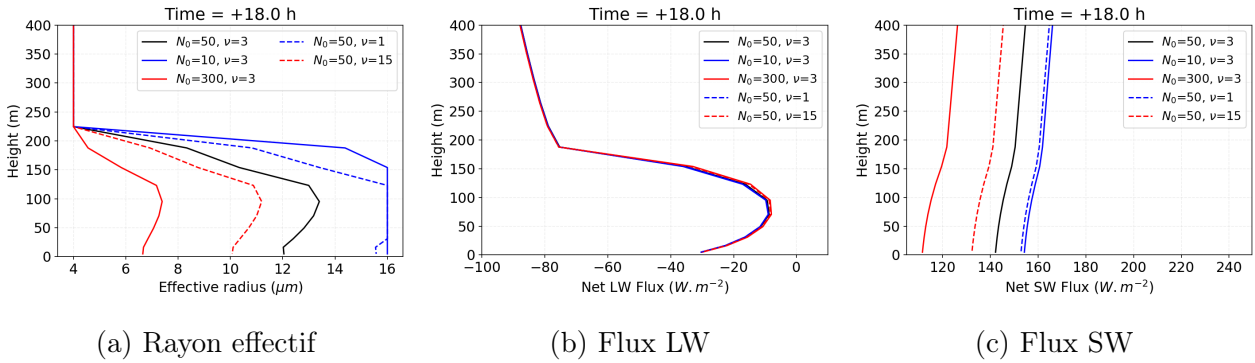


FIGURE 5.24 – Profils verticaux à +18h de plusieurs variables calculées par le schéma radiatif pour des simulations du cas LANFEX où la valeur de N_0 ou ν a été modifiée (dans le schéma radiatif uniquement).

le domaine SW est calculée d'après Fouquart et Bonnel (1980) :

$$\tau_{SW} = \frac{3LWP}{2r_{\text{eff}}},$$

où LWP est le contenu en eau liquide intégré sur l'épaisseur de la couche nuageuse. Dans le domaine LW, le calcul est dérivé de Smith et Shi (1992) :

$$\tau_{LW} = K(1.2 - 0.006r_{\text{eff}})LWP,$$

avec K un coefficient d'absorption. Avec ces formules, passer par exemple d'un rayon effectif de 16 μm à 8 μm multiplie par deux τ_{SW} tandis que τ_{LW} n'augmente que de 4%. La sensibilité de l'épaisseur optique au rayon effectif des gouttelettes nuageuses est donc plus importante dans le SW que dans le LW, ce que confirme la figure 5.25 où τ_{SW} et τ_{LW} sont tracées pour trois simulations LANFEX dans lesquelles la valeur de N_0 est différente. τ_{SW} et τ_{LW} ont des valeurs

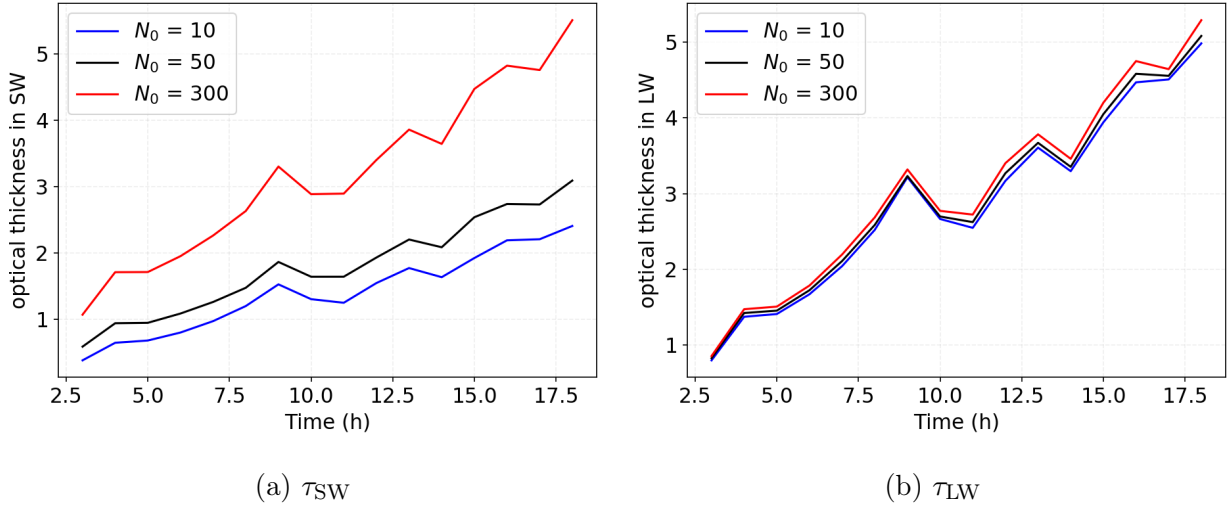


FIGURE 5.25 – Évolution temporelle de l'épaisseur optique du brouillard du cas LANFEX. Il s'agit de la somme des épaisseurs optiques calculées pour chaque niveau vertical du modèle. La courbe noire est la simulation de référence, et les courbes bleue et rouge correspondent à des simulations où la valeur de N_0 a été réglée à 10 et 300 gouttes/cm³ respectivement.

similaires, néanmoins la différence entre les trois simulations est nettement plus importante dans le domaine SW que dans le domaine LW.

Les modifications de N_0 et ν affectant essentiellement les flux SW et LANFEX étant un cas de brouillard radiatif se développant la nuit, les résultats de la figure 5.27, montrant que l'impact sur l'eau nuageuse est négligeable, ne sont pas étonnants (en l'absence de soleil, les flux SW sont nuls).

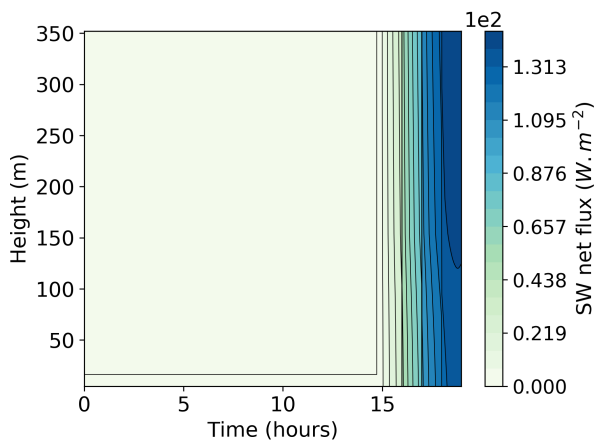


FIGURE 5.26 – Diagramme hovmoller des flux nets SW dans la simulation de référence du cas LANFEX.

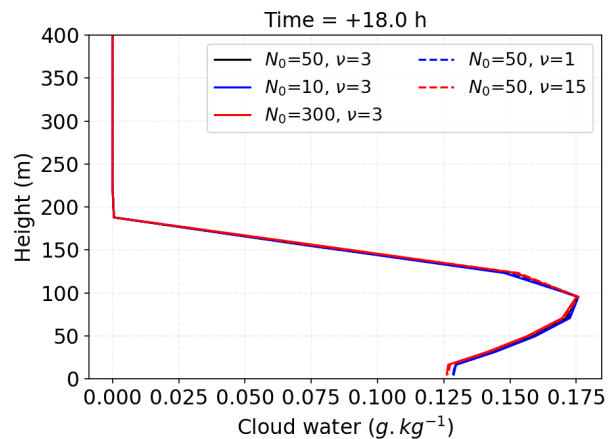


FIGURE 5.27 – Profil vertical à +18h de l'eau liquide nuageuse dans cinq simulations du cas LANFEX.

Dans le cas de stratocumulus FIRE où la simulation peut être lancée pour un temps arbitrairement long, il est possible de simuler plusieurs cycles diurnes. Durant les périodes de jour, où les flux SW sont positifs (figure 5.28), on observe une différence dans le profil de contenu en

eau nuageuse entre la simulation de référence et celles où la valeur de N_0 ou ν a été modifiée (figure 5.29). Pendant les périodes de nuit en revanche, toutes les simulations redeviennent semblables.

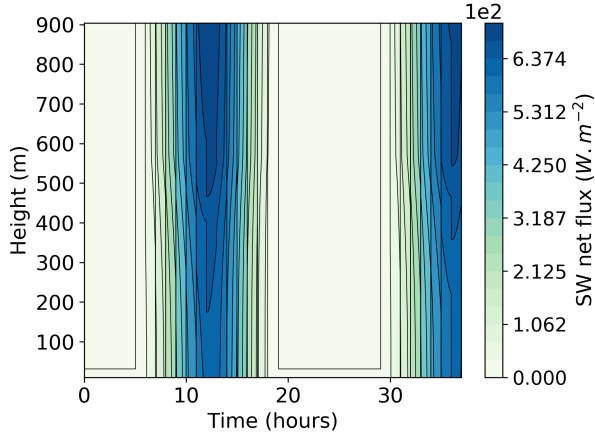


FIGURE 5.28 – Diagramme hovmoller des flux nets SW dans la simulation de référence du cas FIRE.

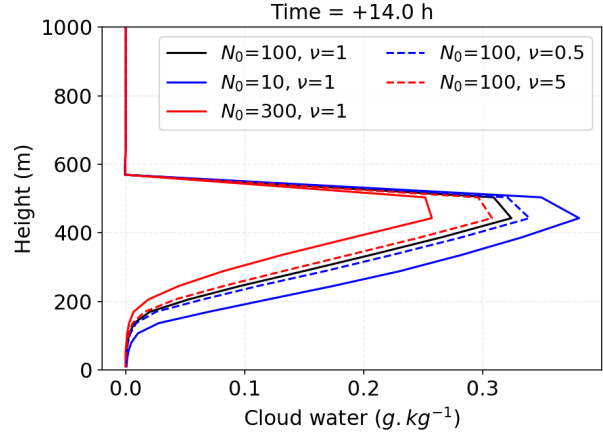
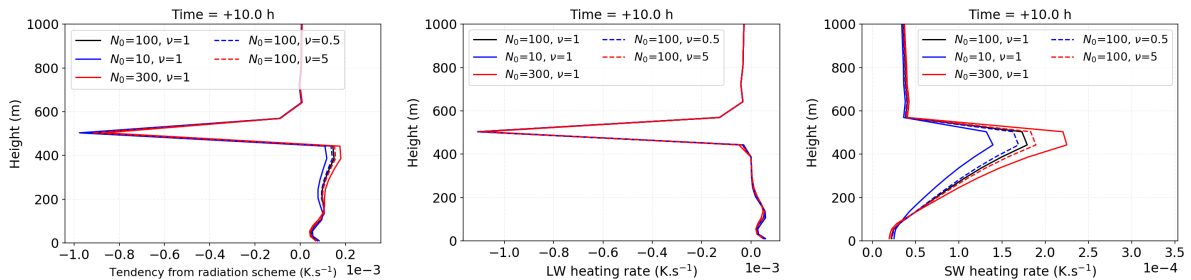


FIGURE 5.29 – Profil vertical à +14h de l'eau liquide nuageuse dans cinq simulations du cas FIRE.

On constate par ailleurs avec la figure 5.29 que lorsque N_0 ou ν augmente, le contenu en eau nuageuse diminue, ce qui est un effet inverse à celui observé avec le schéma microphysique. Lorsque N_0 ou ν augmente (ce qui conceptuellement traduit des gouttelettes nuageuses plus petites), le chauffage radiatif dans le domaine visible augmente au sein du nuage (figure 5.30c) tandis que dans le domaine infrarouge, il reste assez similaire (5.30b). Ceci aboutit à une tendance de température plus élevée dans le nuage (sous 450 m environ, figure 5.30a), et donc à un nuage légèrement plus chaud ce qui pourrait expliquer la diminution de son contenu en eau nuageuse. Les profils de la figure 5.30 sont donnés à +10 h car leurs différences sont alors plus marquées qu'à +14 h.



(a) Total radiation tendency (b) Chauffage ou refroidissement radiatif dû aux flux LW (c) Chauffage ou refroidissement radiatif dû aux flux SW

FIGURE 5.30 – Profils verticaux à +10h de la tendance de température du schéma radiatif, de la contribution des flux SW et des flux LW, pour des simulations du cas FIRE où la valeur de N_0 ou ν a été modifiée (dans le schéma radiatif uniquement).

Cette observation d'un chauffage radiatif plus important lorsque les gouttelettes nuageuses sont plus petites dépend néanmoins peut-être du type de nuage et de son épaisseur optique, et

davantage de cas pourraient être étudiés pour mieux caractériser l'impact de N_0 ou ν dans le schéma radiatif.

5.2.3 Test sur le cas FIRE

Les résultats précédents indiquent que modifier N_0 et ν dans le schéma radiatif a un impact sur le modèle, mais, pour les cas étudiés, seulement à travers une modification des flux SW. La figure 5.31 présente la dispersion d'eau nuageuse obtenue dans des ensembles de 100 simulations du cas FIRE où les paramètres N_0 et ν ont été perturbés, soit dans le schéma microphysique, soit dans le schéma radiatif, soit dans les deux. Pour ces ensembles, les valeurs des paramètres ont été tirées dans des distributions lognormales centrées sur les valeurs $N_0 = 100$ gouttes /cm³ et $\nu = 1$ et d'écart-type respectivement 50 et 2.

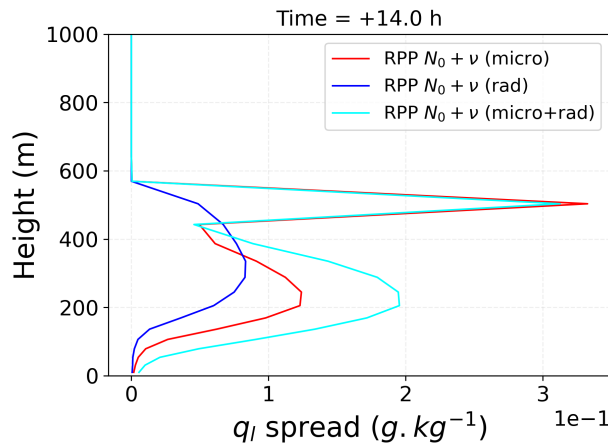


FIGURE 5.31 – Profils verticaux de la dispersion d'eau nuageuse dans différents ensembles de simulations AROME-MUSC du cas FIRE à +14 h. La dispersion est calculée comme l'intervalle entre les percentiles 10 et 90 des ensembles.

Pour ce cas d'étude, la dispersion de l'eau nuageuse (ainsi que des autres variables du modèle) est globalement plus importante lorsque les paramètres N_0 et ν sont perturbés dans le schéma microphysique que lorsqu'ils sont perturbés dans le schéma radiatif. Dans le premier cas, les perturbations peuvent affecter l'extension verticale du nuage (le pic de dispersion vers 500 m sur le graphe 5.31 est dû au fait que dans certaines simulations, le sommet du stratocumulus change de niveau vertical, et la dispersion vers 200 m indique que le nuage s'étend plus ou moins bas). Dans le second cas, les perturbations affectent essentiellement le contenu en eau au sein du stratocumulus. Il faut noter que la figure 5.31 présente des profils à +14 h, soit durant une période de jour. La nuit, la dispersion de l'ensemble où le schéma radiatif seulement est perturbé devient presque nulle. On a vu précédemment que N_0 et ν semblent avoir un impact opposé sur le contenu en eau nuageuse selon qu'ils sont modifiés dans le schéma microphysique ou dans le schéma radiatif. On pourrait donc s'attendre à un effet de compensation, néanmoins, la dispersion obtenue lorsque les deux schémas sont perturbés en même temps semble être équivalente (voire supérieure dans la partie basse du profil) à celle des ensembles où seulement l'un ou l'autre des schémas est perturbé. Davantage de simulations et d'investigations seraient néanmoins nécessaires pour confirmer ce résultat.

5.3 Intercomparaison entre les différentes méthodes : SPPT, iSPPT, RPP, POP

5.3.1 Combinaison des méthodes proposées dans les chapitres 4 et 5

Les sections précédentes de ce chapitre ont introduit une méthode de perturbation de paramètre, appliquée aux schémas microphysique et radiatif, capable de modifier significativement le contenu en eau nuageuse des cas LANFEX et FIRE (figures 5.21, 5.31). On s'intéresse à présent à la combinaison de ces perturbations avec celles présentées dans le chapitre 4. Le tableau 5.2 résume les caractéristiques de chaque méthode de perturbation employée. Pour le cas LANFEX, les valeurs de N_0 et ν sont tirées dans la distribution *a posteriori* du problème inverse. Pour les cas ARMCu, BOMEX et FIRE en revanche, la résolution du problème inverse n'a pas été effectuée et les paramètres sont donc tirés dans des distributions lognormales. L'écart-type des distributions lognormales a été choisi différemment pour les cas océaniques et continentaux dont les valeurs par défaut de N_0 et ν sont différentes, de façon à rester dans des intervalles de valeurs acceptables d'après la littérature.

Méthode de perturbation	Schéma de paramétrisation	Distribution des perturbations	Corrélation temporelle	Ecart-type
KC16	turbulence	gaussienne	10 min	1
Convection stochastique	convection peu profonde	gaussienne	40 min	0.4
Perturbation CB02	ajustement à la saturation	gaussienne	6 h	0.3
RPP N_0	microphysique et radiatif	lognormale ou $\pi_{\Lambda}^{\text{post}}$	∞	25 ou 50 gouttes/cm ³ *
RPP ν	microphysique et radiatif	lognormale ou $\pi_{\Lambda}^{\text{post}}$	∞	0.5 ou 2*

TABLE 5.2 – Caractéristiques des différentes perturbations utilisées. (*) les écart-types de 25 (ARMCu) et 50 (FIRE, BOMEX), et 0.5 (FIRE, BOMEX) et 2 (ARMCu) sont utilisés lorsque N_0 et ν sont tirés dans une distribution lognormale et non dans la distribution *a posteriori* du problème inverse.

La figure 5.32 présente les profils d'eau nuageuse obtenus avec la combinaison des méthodes de perturbation listées dans le tableau 5.2 (ensemble "COMB"), ainsi que ceux obtenus avec la méthode SPPT, pour les quatre cas d'étude. On constate que pour tous les cas, excepté ARMCu, la dispersion de l'eau nuageuse dans les ensembles COMB est comparable voire supérieure à celle des ensembles SPPT sur presque tout le profil vertical. Le biais de la moyenne des ensembles par rapport à la simulation de référence est le plus souvent inférieur d'un ordre de grandeur à la dispersion.

Sur les figures 5.33 et 5.34 on présente directement les profils verticaux de dispersion pour d'autres variables du modèle : la température potentielle et le vent zonal.

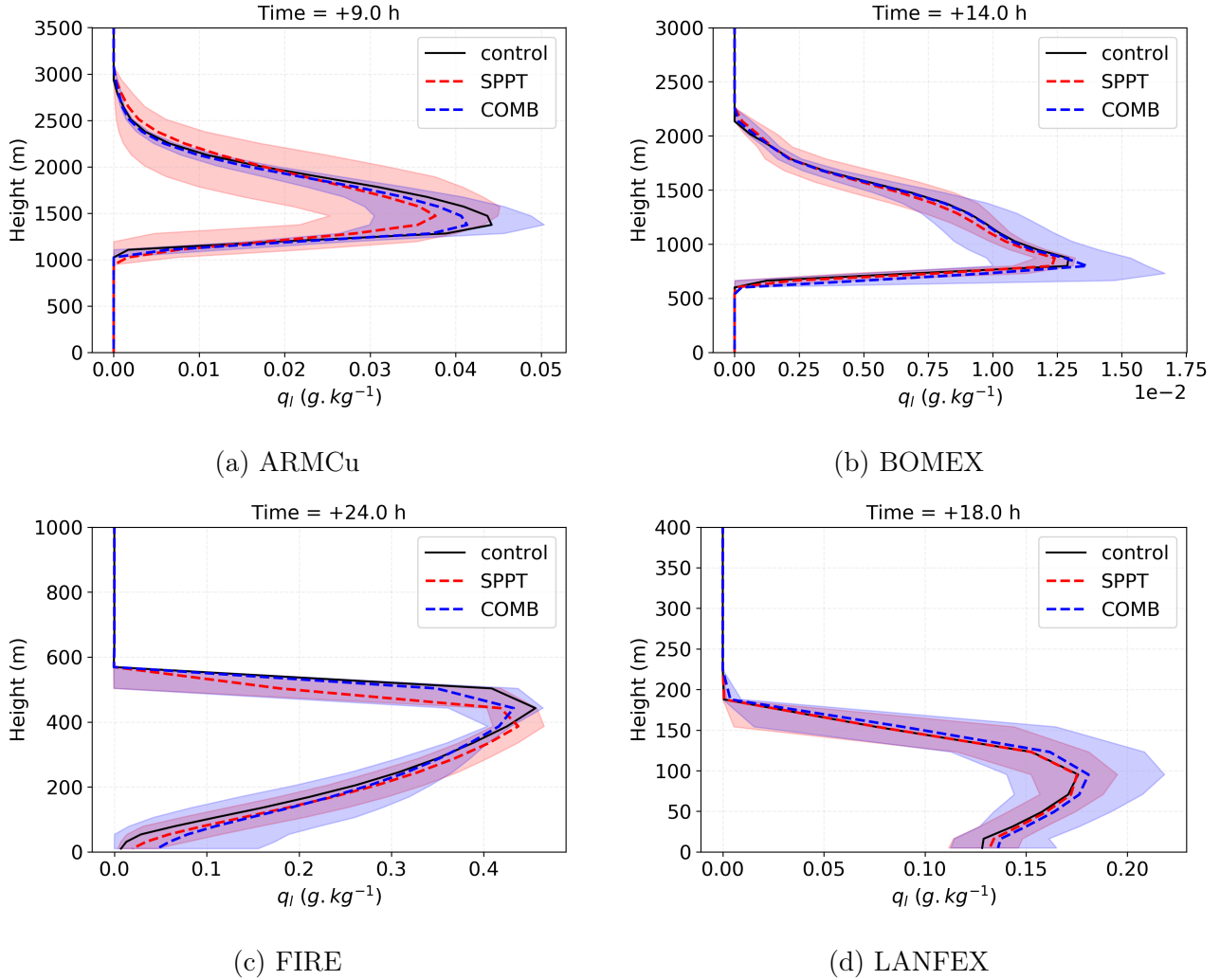


FIGURE 5.32 – Profils verticaux d’eau nuageuse pour les quatre cas d’étude, et deux types d’ensembles : un construit avec la méthode SPPT (en rouge) et un construit avec les perturbations listées dans le tableau 5.2 (COMB, en bleu). Les courbes noires représentent les simulations de référence pour chaque cas. Les courbes en pointillé sont les moyennes des ensembles. Les zones colorées représentent l’intervalle entre les percentiles 10 et 90 des ensembles.

Pour les cas de cumulus ARMCCu et BOMEX, les résultats sont très similaires avec ceux de l’article (Fleury et al. 2022, présenté au chapitre 4). En effet, ces deux cas sont essentiellement gouvernés par les schémas de turbulence et de convection peu profonde, donc l’ajout des perturbations sur les autres schémas n’a pas beaucoup d’impact. Pour ARMCCu, sur les 1000 premiers mètres, la dispersion de température est environ 10 fois supérieure dans l’ensemble SPPT par rapport à l’ensemble COMB. Pour le cas BOMEX, la dispersion de l’ensemble COMB est davantage comparable à SPPT. Pour les cas LANFEX et FIRE, l’ensemble COMB produit une dispersion très similaire à SPPT pour la température, mais également l’humidité (non montré). Ceci représente une amélioration par rapport aux ensembles utilisant seulement les méthodes POP de perturbation de la turbulence et de la convection ou la perturbation de l’ajustement à la saturation (voir la figure 5.1 au début du chapitre). Concernant le vent zonal, l’ensemble COMB produit une dispersion équivalente voire supérieure à SPPT, surtout dans les niveaux proches de la surface, et ce pour tous les cas d’étude. Ceci est essentiellement dû à la méthode KC16 qui introduit des perturbations du vent proportionnelles à la TKE.

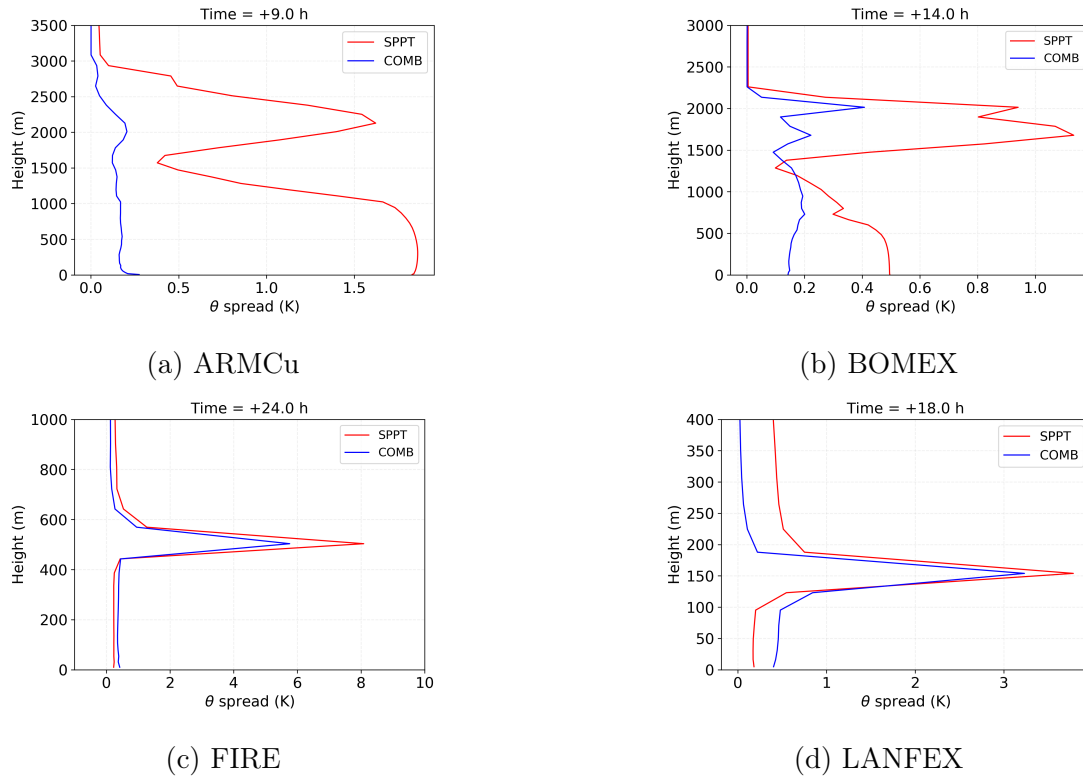


FIGURE 5.33 – Profils verticaux de dispersion de température potentielle pour les quatre cas d'étude, et les deux types d'ensembles. La dispersion est calculée comme l'intervalle entre les percentiles 10 et 90 des ensembles.

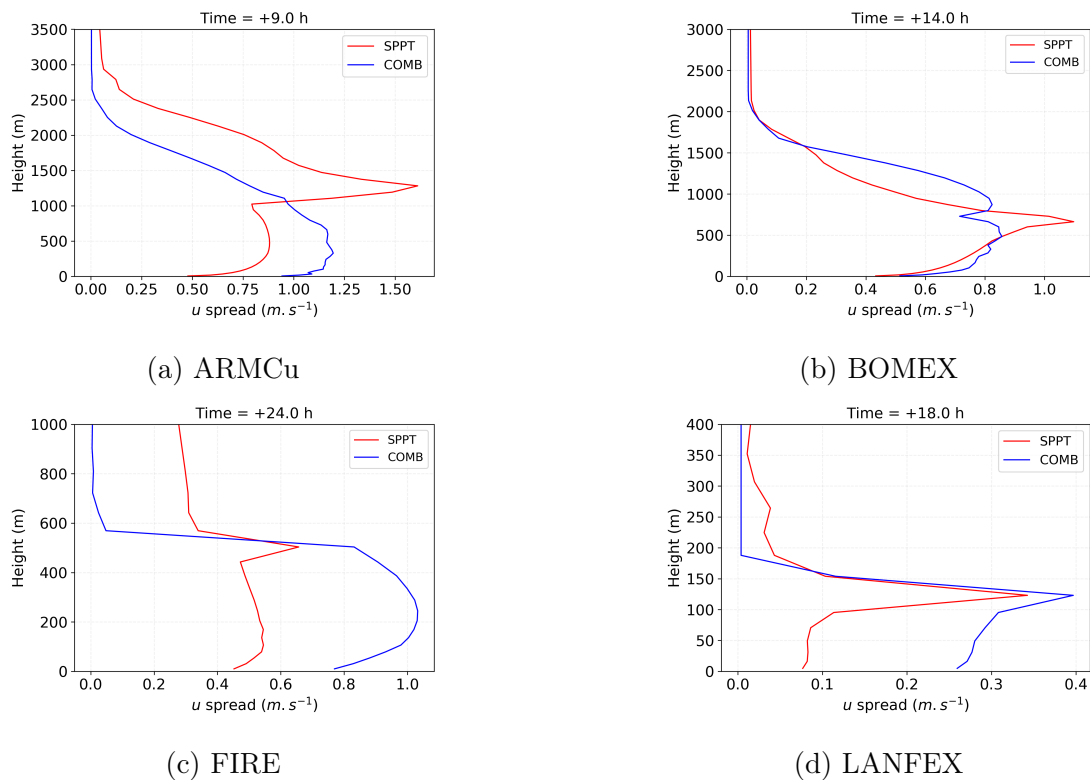


FIGURE 5.34 – Identique à la figure 5.33 mais pour le vent zonal.

Pour les quatre cas d'étude, le biais des ensembles COMB et SPPT concernant la température, l'humidité et le vent n'est pas indiqué sur les figures 5.33 et 5.34 mais est un ordre de grandeur inférieur à la dispersion. Il est similaire entre les ensembles SPPT et COMB, voire légèrement inférieur pour ce dernier.

5.3.2 Comparaison avec iSPPT, RPP

Dans cette dernière section, on prolonge la comparaison en étudiant des ensembles produits avec deux autres méthodes de perturbation. La première est une méthode de perturbation de paramètre "RPP", similaire à celle employée pour perturber les schémas microphysique et radiatifs (sections 5.1.3 et 5.2.3), mais appliquée aux paramètres d'AROME identifiés par l'étude de Wimmer et al. (2022). La deuxième est une méthode dérivée de SPPT, nommée iSPPT, et mentionnée dans le chapitre 1 (voir section 1.4.2 et Christensen et al. 2017).

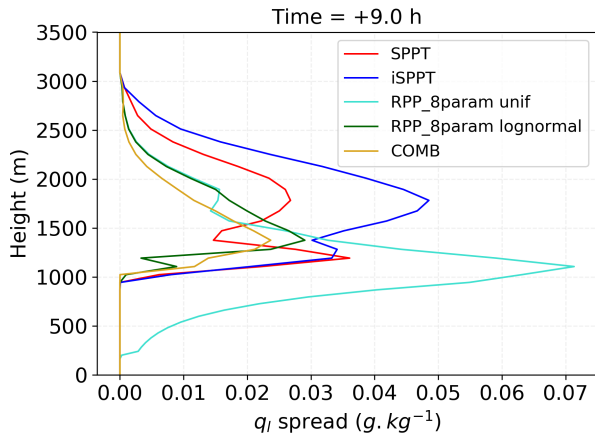
La méthode iSPPT est implémentée de la façon suivante : quatre champs aléatoires gaussiens 2D indépendants, tous d'écart-type 0.3 et de corrélation temporelle 6 h, sont utilisés pour perturber indépendamment les tendances des schémas radiatif, de convection peu profonde, de turbulence et de microphysique d'AROME.

La méthode "RPP", appliquée aux 8 paramètres d'AROME les plus influents identifiés par Wimmer et al. 2022 (voir le tableau 1.1), est déclinée en deux variantes :

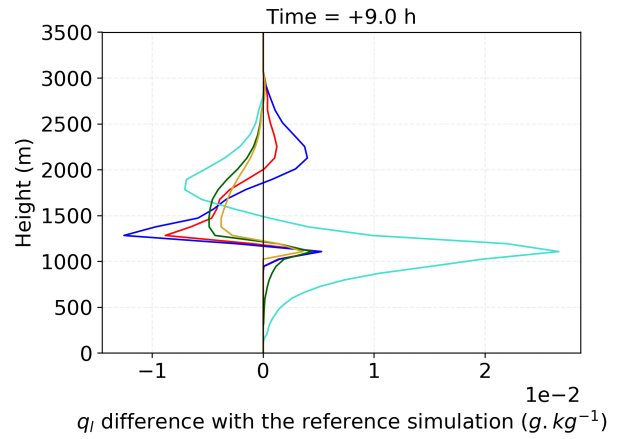
- "RPP_8param unif" : les valeurs des paramètres sont tirées dans des distributions uniformes sur les intervalles donnés dans le tableau 1 de Wimmer et al. (2022) ;
- "RPP_8param lognormal" : les valeurs des paramètres sont tirées dans des distributions lognormales, centrées sur les valeurs par défaut. L'écart-type des distributions est réglé pour chaque paramètre comme la moitié de son intervalle donné dans le tableau 1 de Wimmer et al. (2022).

La figure 5.35 présente, pour le cas ARMCu, les profils verticaux de la dispersion de différents ensembles, ainsi que leur biais par rapport à la simulation de référence non perturbée.

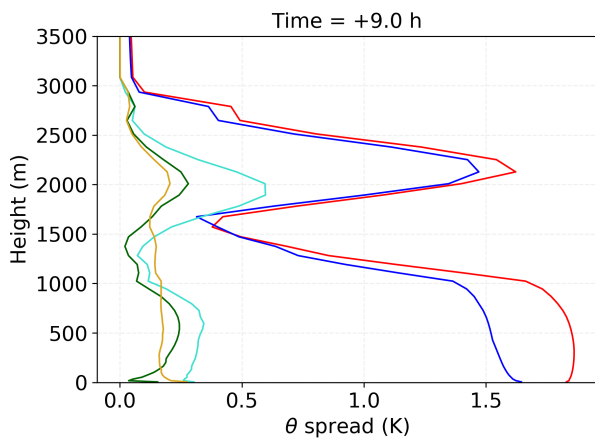
On constate que pour l'eau nuageuse, les ensembles iSPPT et "RPP_8param unif" produisent la dispersion la plus large, mais également le plus fort biais par rapport à la simulation de référence. Celui-ci est en particulier marqué pour l'ensemble RPP qui produit en moyenne un nuage 500 mètres plus bas que dans la référence. L'utilisation de distributions lognormales pour perturber les paramètres permet de réduire fortement le biais (courbe verte représentant l'ensemble "RPP_8param lognormal"). Avec les réglages choisis ici concernant l'écart-type de ces distributions lognormales, la dispersion est réduite par rapport à l'ensemble "RPP_8param unif" ; elles est en revanche comparable à celle de l'ensemble COMB. Concernant la température, les méthodes SPPT et iSPPT produisent une large dispersion du fait des tendances élevées dans ce cas d'étude. Cette dispersion n'est pas reproduite par les autres méthodes, dont les résultats sont relativement similaires. Les mêmes observations peuvent être faites pour l'humidité (non montré). Pour le vent zonal, les méthodes SPPT et iSPPT produisent là encore une dispersion importante mais dans les 1000 premiers mètres, elle est maximale pour l'ensemble COMB, très probablement grâce aux perturbations de la turbulence. Les mêmes observations peuvent être faites pour le vent méridien (non montré).



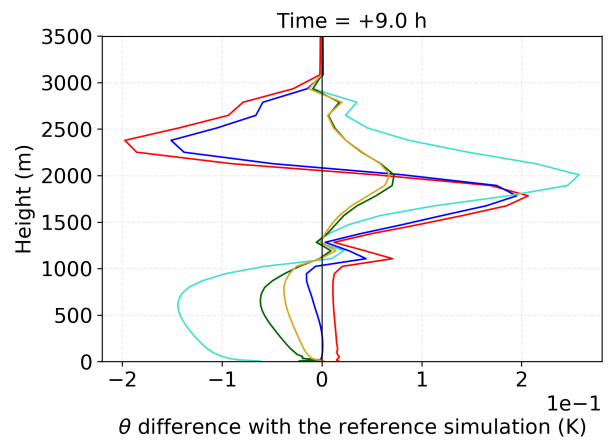
(a) Dispersion d'eau nuageuse



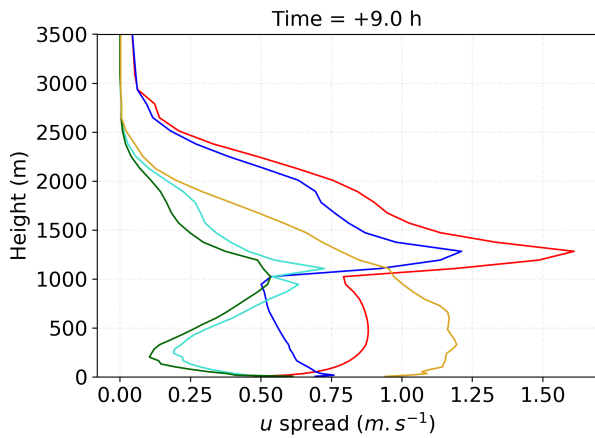
(b) Biais d'eau nuageuse



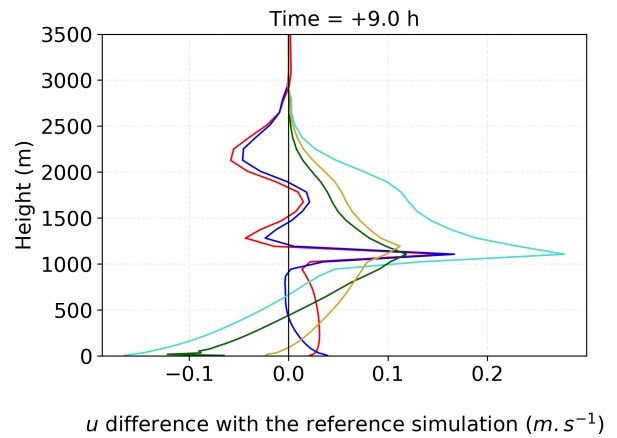
(c) Dispersion de température



(d) Biais de température



(e) Dispersion du vent zonal



(f) Biais de vent zonal

FIGURE 5.35 – Profils verticaux de la dispersion et du biais dans plusieurs ensembles de simulations AROME-MUSC du cas ARMCu à +9 h.

Une comparaison plus complète est présentée dans l'étude publiée dans le journal *Boundary-layer meteorology* par Bouttier et al., et incluse dans l'annexe C. Cette étude compare des ensembles SPPT, iSPPT et des ensembles RPP appliquant des perturbations uniformes à 11

paramètres issus de la liste de paramètres étudiés par Wimmer et al. (2022) (voir leur tableau 1). Les ensembles sont constitués de simulations AROME-MUSC des cas d'étude ARMCu, FIRE et LANFEX. Les principaux résultats de cette étude sont que la méthode iSPPT produit davantage de dispersion que SPPT. La méthode RPP est elle aussi capable de produire une dispersion importante, mais s'accompagne de l'apparition de biais, parfois du même ordre de grandeur que la dispersion. Ceci est probablement dû à l'utilisation d'une distribution uniforme, ce qui souligne l'importance du réglage des distributions dans le cadre des méthodes de perturbation de paramètres.

Chapitre 6

Évaluation des méthodes de perturbation dans des simulations 3D du modèle AROME

Dans ce dernier chapitre, on teste certaines des méthodes de perturbation étudiées précédemment dans une configuration plus proche des conditions opérationnelles. Pour cela, des simulations de cas réels sont réalisées avec la version 3D du modèle déterministe AROME et avec le système de prévision d'ensemble PEARO. Certaines journées de 2020 ou 2021 sont étudiées en détail et des ensembles de simulations produits sur un mois complet permettent de réaliser une première évaluation à l'aide de scores probabilistes.

6.1 Tests sur quelques journées en 2020 et 2021

6.1.1 Sensibilité aux perturbations des paramètres de distribution de taille des gouttelettes nuageuses N_0 et ν

Dans cette première partie, on examine l'effet d'une modification des paramètres N_0 (concentration en nombre de gouttelettes nuageuses) et ν (paramètre de forme de la distribution de taille des gouttelettes), étudiés dans le chapitre 5, sur la simulation d'une situation propice à la formation de brouillard durant l'année 2020. L'objectif est de vérifier si l'impact de ces paramètres, caractérisé dans un cadre simplifié uni-colonne dans les sections 5.1.2 et 5.2.2, est similaire dans une simulation 3D.

Cas d'étude

Le cas d'étude choisi est la nuit du 5 au 6 janvier 2020, qui correspond à la sixième période d'observation intensive (POI) de la campagne de mesure du projet SoFog3D (Burnet et al. 2020). Cette campagne s'est déroulée d'octobre 2019 à mars 2020 sur plusieurs sites situés dans le sud-ouest de la France, avec pour but d'améliorer la compréhension des processus de fine échelle liés au brouillard, de tester une instrumentation innovante et d'évaluer la capacité du modèle AROME à prévoir le brouillard (voir le site du projet [ici](#)¹).

Le choix fait ici de simuler l'une des POIs de la campagne SoFog3D vient essentiellement du fait qu'il s'agit de situations qui ont été bien documentées. La POI-6 couvre une période

1. <https://www.umr-cnrm.fr/spip.php?article1086&lang=en>

allant du 4 au 7 janvier 2020. Des occurrences de brouillard sont observées dans les nuits du 4, du 5 et du 6 janvier sur les différents sites instrumentés. Durant la nuit du 5 janvier, des zones de brouillard sont présentes dans le sud-ouest ainsi que dans d'autres régions, et évoluent en stratus dans la matinée du 6. La figure 6.1 présente des cartes de contenu en eau liquide nuageuse à 10 mètres au-dessus du sol issues d'une simulation AROME initialisée avec l'analyse du 05/01/2020 03:00 UTC. Les zones couvertes par le brouillard le 5 janvier au matin diminuent jusqu'à environ 15:00 UTC, puis s'étendent à nouveau durant toute la soirée et la nuit suivantes.

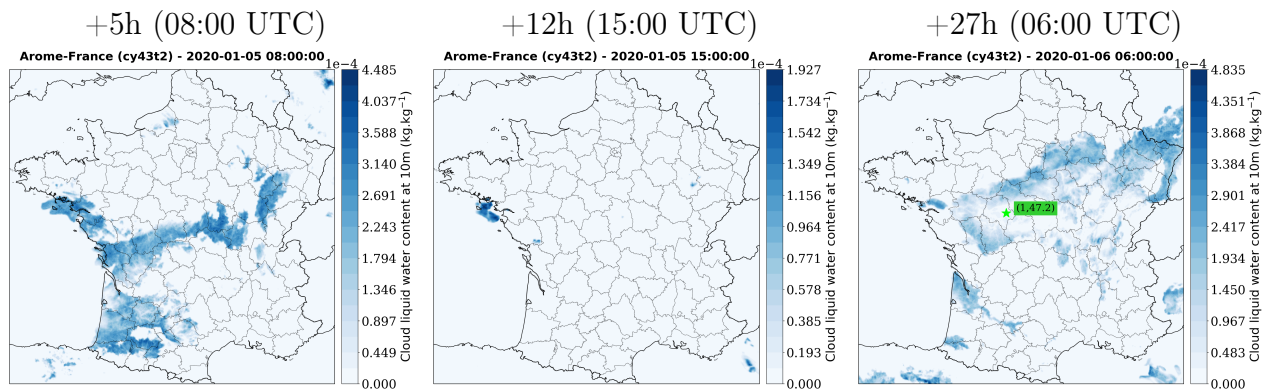


FIGURE 6.1 – Contenu en eau liquide nuageuse à 10 mètres au-dessus du sol à différentes échéances d'une simulation AROME initialisée avec l'analyse du 05/01/2020 03:00 UTC. L'étoile sur la troisième carte correspond au point pour lequel le diagramme hovmoller de la figure 6.3 est tracé.

Dans la nuit du 5 au 6 janvier, on observe deux zones principales où le brouillard est simulé par AROME : une première dans le sud-ouest de la France, notamment dans l'estuaire de la Gironde et la vallée de la Garonne, et une deuxième zone très étirée au centre de la France. Dans le sud-ouest, le brouillard s'étend progressivement dans la nuit, colonisant petit à petit la vallée de la Garonne. La figure 6.2 indique des valeurs de vent très faibles dans cette zone.

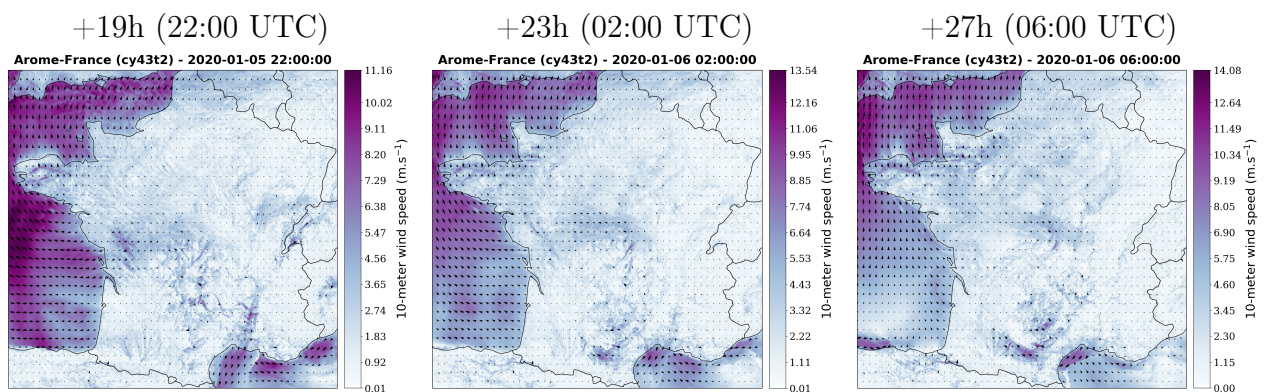


FIGURE 6.2 – Direction et force du vent à 10 m à différentes échéances d'une simulation AROME initialisée avec l'analyse du 05/01/2020 03:00 UTC.

Pour la deuxième zone en revanche, l'évolution du contenu en eau nuageuse ainsi que les cartes de vent suggèrent que des phénomènes d'advection entrent en jeu. Le brouillard apparaît le 5 janvier en fin d'après-midi au nord du Massif Central au niveau des départements de

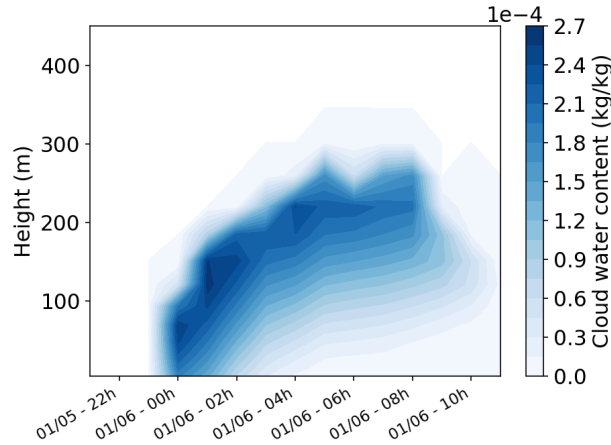


FIGURE 6.3 – Diagramme hovmoller de l'eau liquide nuageuse durant la nuit du 5 au 6 janvier 2020 simulée par AROME au point indiqué sur la figure 6.1.

l'Allier, du Cher et de la Nièvre. Dans les heures suivantes il se propage vers le nord et l'ouest tout en disparaissant des zones où il est apparu en premier. Dans le département de l'Indre par exemple, presque entièrement recouvert par le brouillard à 23:00 UTC, le contenu en eau nuageuse à 10 mètres devient négligeable à partir de 03:00 UTC. La propagation de cette zone de brouillard pourrait être expliquée par la présence de vent au nord du Massif Central, avec des valeurs de l'ordre de 5 m/s et dirigé d'abord vers l'ouest puis progressivement le nord-ouest durant la nuit. L'advection de masses d'air plus humides pourrait contribuer à faire apparaître du brouillard dans ces zones, qui évoluerait néanmoins rapidement en stratus du fait de la turbulence engendrée par ce vent. La figure 6.3 qui présente un diagramme hovmoller en un point du domaine situé en Indre-et-Loire montre bien que le développement du brouillard est rapidement suivi par une élévation en stratus bas.

Expériences

Pour tester l'impact des paramètres N_0 et ν sur la simulation de ce cas d'étude, on réalise plusieurs expériences avec le modèle déterministe AROME-France, dans la version qui a été opérationnelle jusqu'en juin 2022. Toutes les simulations sont réalisées à partir de l'analyse du 05/01/2020 03:00 UTC, et sont résumées dans le tableau 6.1. Dans ce tableau, $N_{0,\text{land}}$ et $N_{0,\text{sea}}$ correspondent aux valeurs de réglage du nombre de gouttelettes N_0 sur continents et océans respectivement. ν_{land} et ν_{sea} de même représentent la valeur de réglage du paramètre de forme ν sur continents et océans. Dans la suite, on utilisera " N_0 " et " ν " pour désigner uniquement $N_{0,\text{land}}$ et ν_{land} . La modification des valeurs sur les surfaces océaniques n'a pas été testée dans l'étude de sensibilité présentée dans cette section.

Les couleurs bleues du tableau 6.1 indiquent que la configuration par défaut du schéma microphysique ou du schéma radiatif a été conservée. L'expérience A est donc l'expérience contrôle, où aucune modification n'a été apportée au modèle. Les cases "default" dans la colonne du schéma radiatif indiquent que pour ces expériences, le lien n'est pas fait entre les hypothèses du schéma microphysique et du schéma radiatif : la paramétrisation de Martin et al. (1994) n'est pas modifiée, donc les valeurs de N_0 sur continent et sur océan utilisées par le schéma radiatif sont respectivement d'environ de 313 et 50 gouttes/cm³ et ν n'apparaît pas directement dans le calcul du rayon effectif. Pour l'expérience D en revanche, les hypothèses des deux schémas

	Microphysics scheme				Radiation scheme			
	$N_{0,land}$	$N_{0,sea}$	ν_{land}	ν_{sea}	$N_{0,land}$	$N_{0,sea}$	ν_{land}	ν_{sea}
Exp A	300	100	3	1	default	default	default	default
Exp B	300	100	1.5	1	default	default	default	default
Exp C	300	100	15	1	default	default	default	default
Exp D	300	100	3	1	300	100	3	1
Exp E	300	100	3	1	300	100	1.5	1
Exp F	300	100	3	1	300	100	15	1
Exp G	100	100	3	1	default	default	default	default
Exp H	50	100	3	1	default	default	default	default
Exp I	300	100	3	1	100	100	3	1
Exp J	300	100	3	1	50	100	3	1

TABLE 6.1 – Description des expériences 3D réalisées pour tester l’impact de ν et N_0 sur la simulation du brouillard par AROME. Les valeurs de N_0 sont données en gouttes/cm³.

concernant le spectre de taille des gouttelettes sont harmonisées et cette expérience sert de référence pour étudier l’impact d’une modification de N_0 ou ν dans le schéma radiatif. Enfin les valeurs en rouge indiquent les modifications d’intérêt, celles dont on veut évaluer l’impact. Elles sont introduites indépendamment dans les schémas microphysique et radiatif pour distinguer leurs effets sur chacun des schémas. Pour le paramètre de forme ν , on choisit de tester les valeurs 1.5 et 15 qui sont proches de celles utilisées dans l’étude de sensibilité 1D (sections 5.1.2 et 5.2.2). La valeur 1.5 est préférée à 1 car, sur continent, $\alpha = 1$ donc régler ν à 1 correspondrait à une distribution exponentielle, ce qui ne semble pas réaliste pour le spectre de taille des gouttelettes nuageuses. Pour le paramètre N_0 , on teste deux valeurs : 100 et 50 gouttes/cm³, en cohérence avec les tests en 1D.

Sensibilité d’AROME au paramètre ν dans le schéma microphysique

On s’intéresse dans un premier temps aux expériences A, B et C, c’est-à-dire à la modification de la valeur de ν dans le schéma microphysique.

La figure 6.4 présente le diagnostic de visibilité d’AROME à l’échéance +25h. La visibilité est définie ici à l’aide du modèle de Koschmieder (1924) :

$$\text{visibility} = -\ln(\epsilon)/\beta,$$

où ϵ est un seuil de contraste fixé à 0.02 et β un coefficient d’extinction. Pour l’extinction due aux hydrométéores non précipitants, β est donné, d’après Kunkel (1984), par une fonction puissance des contenus en eau et glace nuageuses (Dombrowski-Etchevers et al. 2018).

Le diagnostic de visibilité permet d’identifier les zones de brouillard. Les couleurs foncées sur les cartes des expériences B et C (figure 6.4) indiquent des zones où la différence de visibilité avec

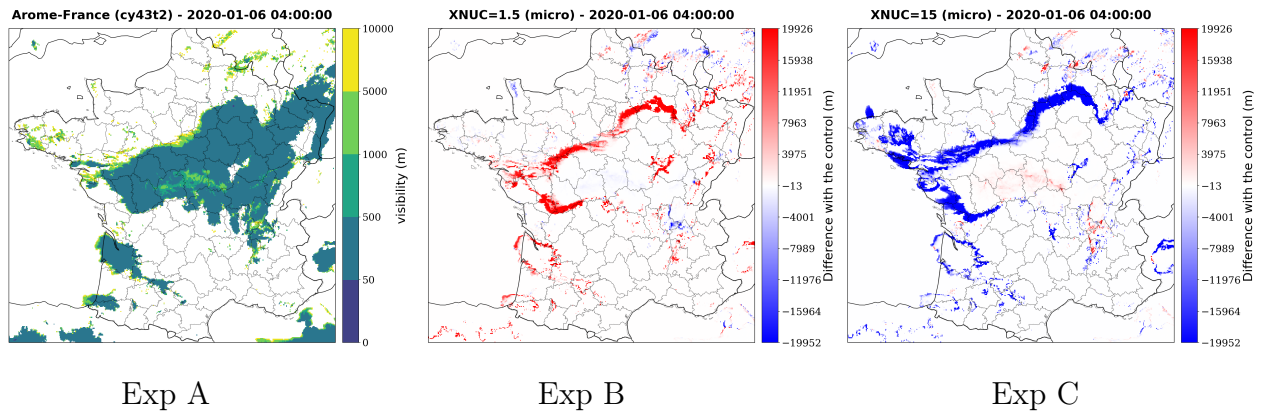


FIGURE 6.4 – Diagnostic de visibilité à +25h dans l’expérience A, et différence par rapport à cette référence pour les expériences B et C.

la référence est de quelques dizaines de kilomètres, soit des zones où le brouillard est apparu ou bien a disparu par rapport à la référence. Dans l’expérience B, les différences positives indiquent que dans de nombreuses zones, le brouillard a disparu par rapport à la référence (augmentation de la visibilité) et on constate l’inverse pour l’expérience C. Ainsi, une plus forte valeur de ν semble aboutir à un brouillard plus étendu, surtout dans la zone du centre de la France. L’impact dans le sud-ouest, sur la zone couverte par la campagne SoFog3D, est qualitativement le même bien que plus limité.

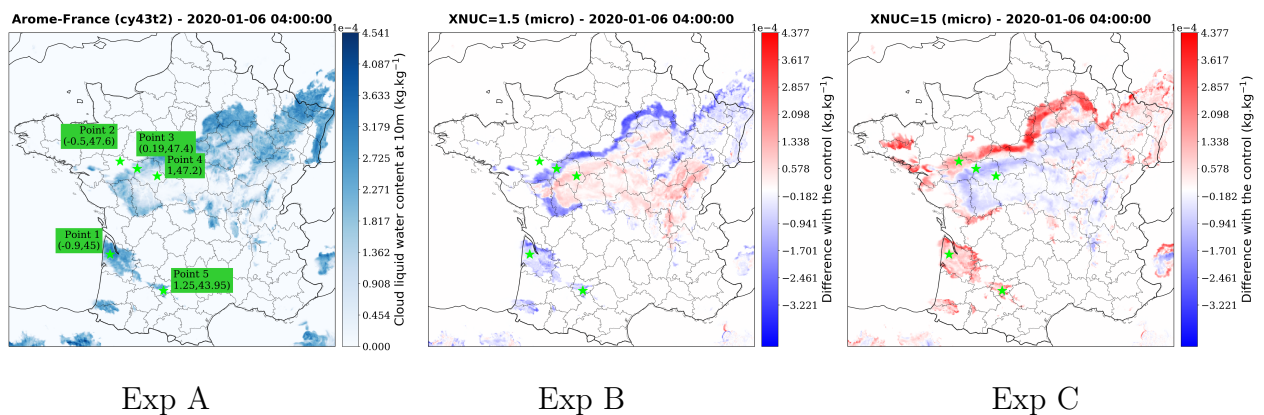


FIGURE 6.5 – Contenu en eau nuageuse à 10 m à +25h dans l’expérience A, et différence par rapport à cette référence pour les expériences B et C.

Avec la figure 6.5 représentant le contenu en eau nuageuse à 10 mètres, on retrouve bien les zones où le brouillard apparaît ou disparaît par rapport à la référence, mais on peut voir également qu’au sein des zones couvertes par le brouillard dans les trois expériences, le contenu en eau nuageuse varie. Dans la région du sud-ouest autour de Bordeaux, il augmente lorsque ν augmente (Exp C) et inversement (Exp B). En revanche, dans le centre de la France, bien que sur le pourtour de la zone de brouillard le contenu en eau nuageuse augmente dans l’expérience C, au sein de la zone de brouillard il est plus faible. Pour compléter ces observations, on s’intéresse aux profils verticaux d’eau nuageuse qui sont présentés sur la figure 6.6 pour les cinq points indiqués sur les cartes de la figure 6.5. On constate une augmentation du contenu en eau nuageuse sur tout le profil lorsque ν augmente pour les points 1 et 5 (autour de Bordeaux et

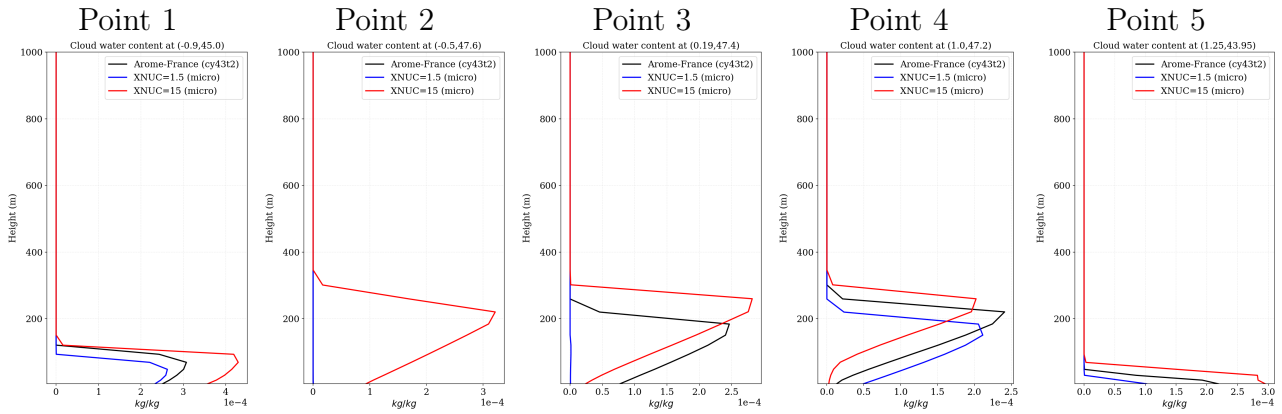


FIGURE 6.6 – Profils verticaux du contenu en eau nuageuse à +25h sur cinq points du domaine. La courbe noire est l’expérience contrôlée (Exp A), la courbe bleue est l’expérience B et la rouge est l’expérience C.

Toulouse respectivement). Le point 2 se trouve dans une zone où seule l’expérience C prévoit du brouillard. Au point 3 on constate que la couche de brouillard est plus élevée en hauteur et au point 4 elle semble évoluer vers une couche de stratus. Ceci explique les différences négatives de contenu en eau nuageuse à 10 mètres visibles sur la figure 6.5 pour Exp C : la couche nuageuse totale n’est pas nécessairement moins dense que la référence mais elle se trouve plus haut. Ceci s’observe également sur la figure 6.7 représentant le contenu en eau nuageuse intégré verticalement (LWP) entre 0 et environ 500 m (afin de filtrer seulement les zones de brouillard ou de stratus bas). La différence avec la simulation de référence est globalement positive pour l’Exp C, donc augmenter la valeur de ν semble avoir pour effet de produire un brouillard plus dense. On peut voir néanmoins des différences légèrement négatives dans certaines zones. Elles correspondent au profil du point 4 (voir figure 6.6) : le nuage est plus élevé, mais son contenu en eau est aussi un peu plus faible que celui de la simulation de référence.

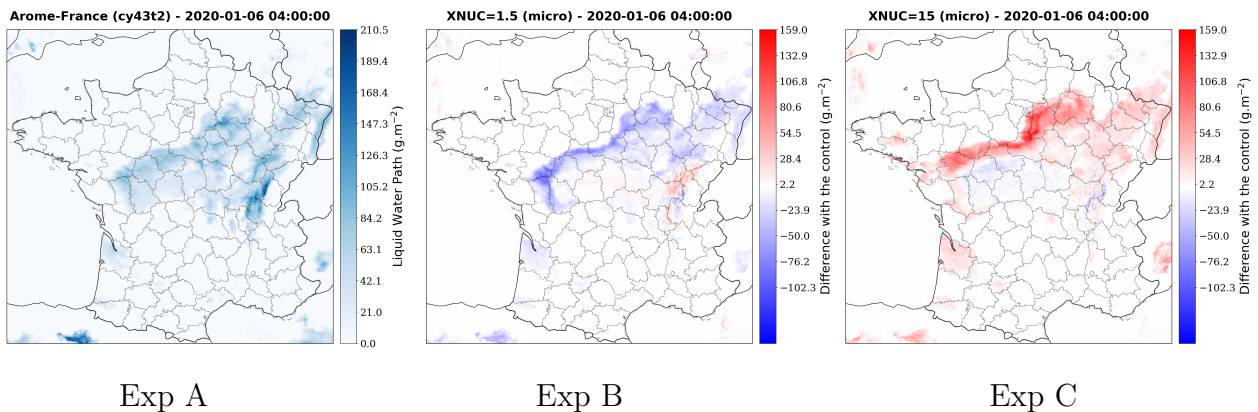


FIGURE 6.7 – LWP à +25h dans l’expérience A, et différence par rapport à cette référence pour les expériences B et C.

Ces différentes observations sont globalement cohérentes avec les résultats obtenus en 1D avec AROME-MUSC (section 5.1.2) : lorsque le paramètre de forme ν augmente, le contenu en eau nuageuse du brouillard augmente, en lien avec la diminution du nombre de grosses gouttelettes et donc de la vitesse de sédimentation. Ceci est particulièrement visible dans la

zone du sud-ouest. Dans la zone du centre de la France en revanche, l'interprétation est un peu plus difficile à cause d'effets indirects dus à l'advection. La plus grande extension horizontale du brouillard pourrait être expliquée par davantage d'advection d'eau liquide. Par ailleurs, l'augmentation de ν , qui correspond à des gouttelettes plus petites et donc une vitesse de sédimentation réduite, pourrait favoriser la transition du brouillard en stratus dans cette zone.

Étant donné que l'on constate que des zones sont couvertes par le brouillard dans certaines simulations et pas dans d'autres, on peut s'intéresser également au cycle de vie du brouillard, pour voir si ces différences sont dues à des instants d'apparition ou de disparition différents selon les simulations. Les figures 6.8 et 6.9 présentent l'échéance à laquelle du brouillard est apparu ou a disparu en chaque point du modèle. Le calcul a été effectué entre les échéances +12 et +36 qui correspondent à 15:00 UTC le 05/01/2020 et 15:00 UTC le 06/01/2020 respectivement.

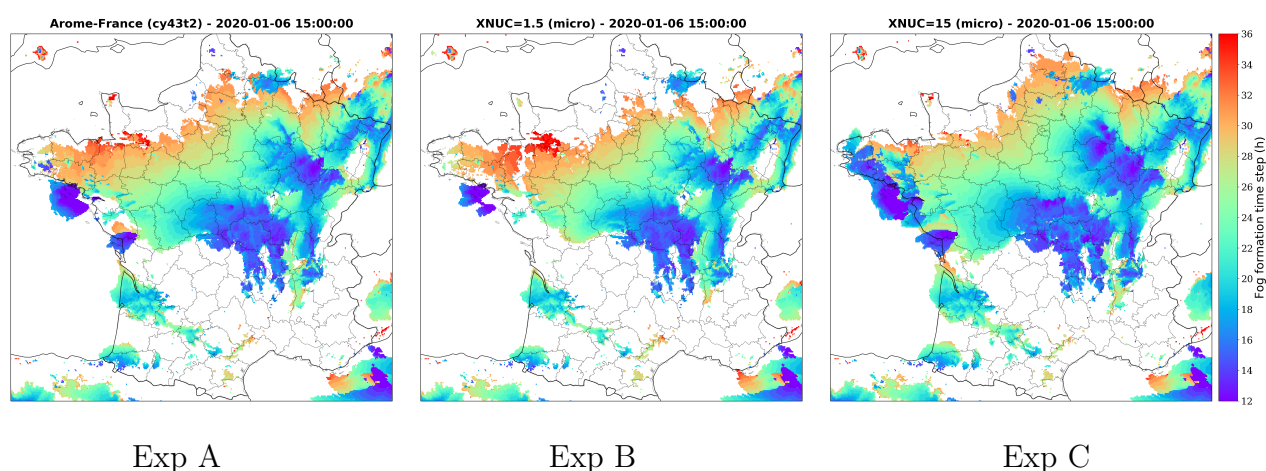


FIGURE 6.8 – Échéance à laquelle le brouillard apparaît dans les expériences A, B, C. Le calcul a été effectué entre les échéances +12 et +36 du modèle.

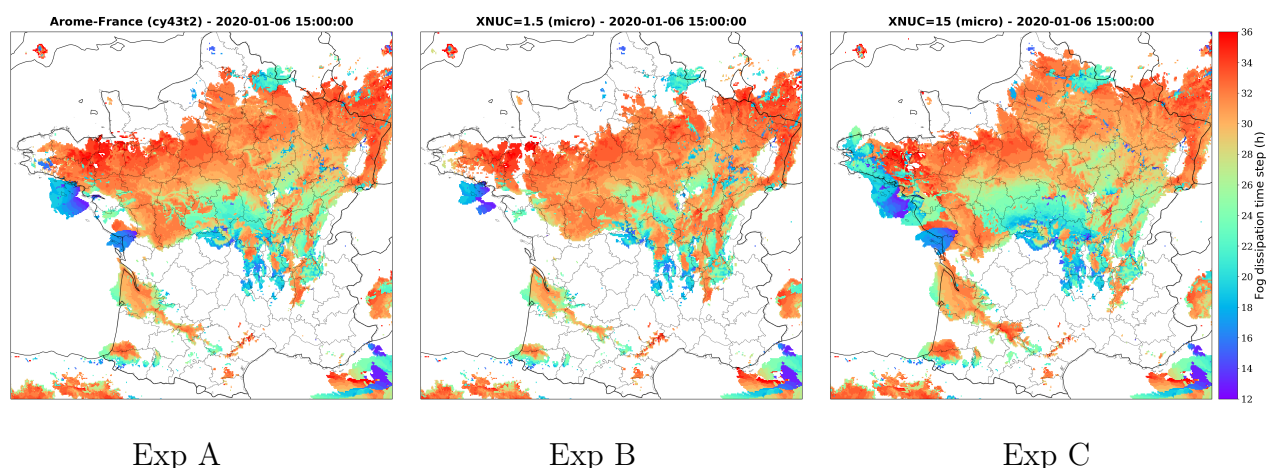


FIGURE 6.9 – Échéance à laquelle le brouillard disparaît dans les expériences A, B, C. Le calcul a été effectué entre les échéances +12 et +36 du modèle.

Dans les zones situées autour de Bordeaux et Toulouse, l'heure d'apparition du brouillard semble globalement identique pour les trois expériences. L'heure de dissipation semble un peu

plus tardive dans l'expérience C. Dans le centre et le nord de la France en revanche, le brouillard progresse plus vite et s'étend davantage dans l'expérience C, mais il se dissipe aussi plus tôt. Ces observations sont confirmées par les séries temporelles d'eau nuageuse à 10 mètres présentées sur les figures 6.10 et 6.11. Aux points 1 et 5, la différence entre les trois expériences

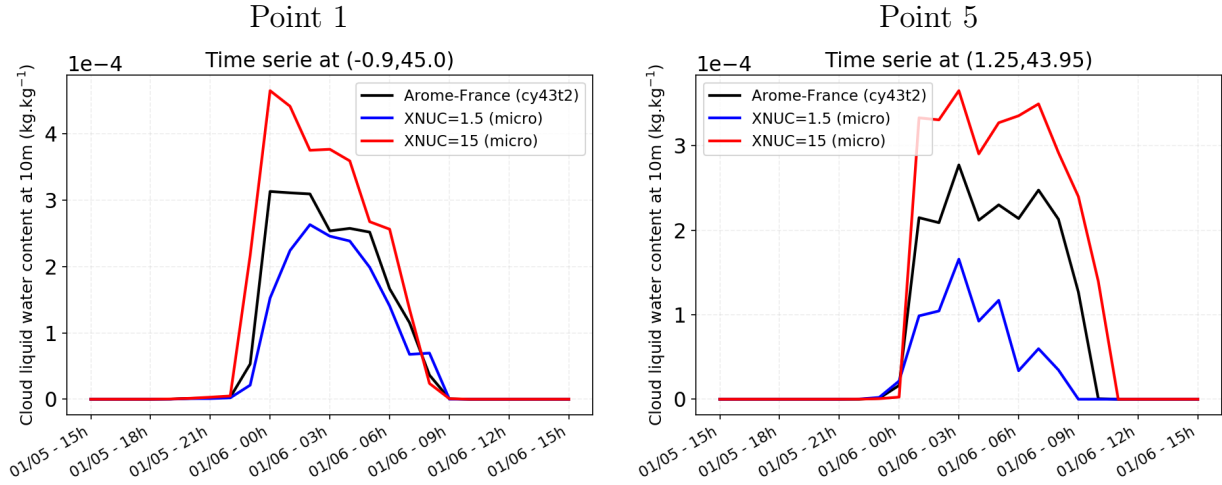


FIGURE 6.10 – Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés près de Bordeaux et Toulouse respectivement. La courbe noire est l'expérience contrôle (Exp A), la courbe bleue est l'expérience B et la rouge est l'expérience C.

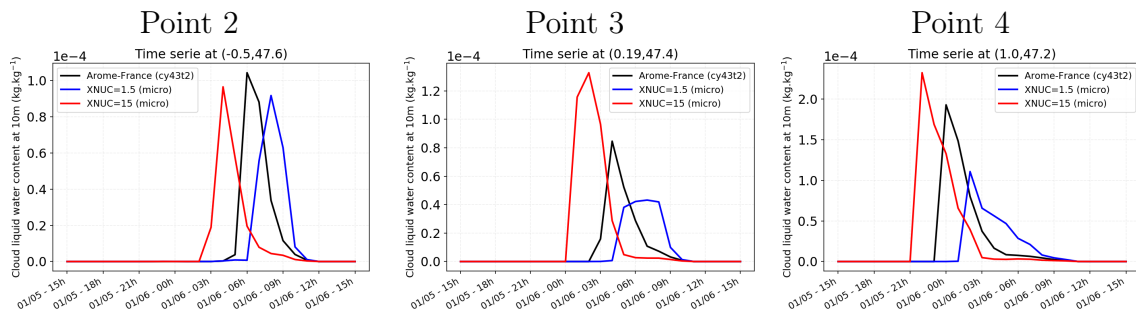


FIGURE 6.11 – Identique à la figure 6.10 pour les points 2, 3 et 4.

est essentiellement la quantité d'eau nuageuse, le cycle de vie du brouillard est peu affecté. Au point 5, on observe néanmoins différentes échéances de dissipation. L'évolution aux points 2, 3 et 4 indique que le scénario est le même pour les trois expériences, mais il est décalé dans le temps.

Ainsi, dans les zones où l'advection est faible (dans le sud-ouest, aux alentours de Toulouse et Bordeaux), le paramètre ν n'a pas d'impact sur l'heure de formation du brouillard, ce qui peut s'expliquer par le fait que le paramètre n'affecte ici que le schéma microphysique, qui n'est actif qu'à partir du moment où de l'eau liquide s'est formée. Le développement du contenu en eau nuageuse est en revanche plus rapide quand ν augmente. L'heure de dissipation semble peu affectée. Des différences de l'ordre d'une heure sont simulées pour le point situé vers Toulouse, mais une meilleure résolution temporelle serait nécessaire pour affiner cette conclusion.

Dans les zones où l'advection prédomine (dans le centre de la France), on observe en revanche un décalage dans la formation et la dissipation du brouillard entre les simulations. L'explication

pourrait être qu’au moment où le brouillard apparaît en premier, il est plus épais et plus dense dans l’expérience C, où ν est plus grand, que dans les autres. La couche de brouillard évolue rapidement en stratus et l’advection horizontale amène ainsi de l’humidité vers le nord-ouest, où les conditions deviennent favorables. Dans le cas de l’expérience C, davantage d’eau peut être advectée ce qui a pour effet de favoriser la formation de brouillard ; celui-ci se forme donc plus tôt et progresse plus loin.

Sensibilité d’AROME au paramètre N_0 dans le schéma microphysique

La valeur du paramètre N_0 est modifiée dans les expériences G (100 gouttes/cm³) et H (50 gouttes/cm³) du tableau 6.1. Ces deux expériences ont donc toutes les deux une valeur de N_0 plus faible que l’expérience contrôle A. Les cartes de LWP présentées sur la figure 6.12 montrent une diminution du LWP lorsque N_0 diminue, en accord avec les résultats des tests 1D (section 5.1.2). Cette diminution est plus importante dans l’expérience H que dans l’expérience G.

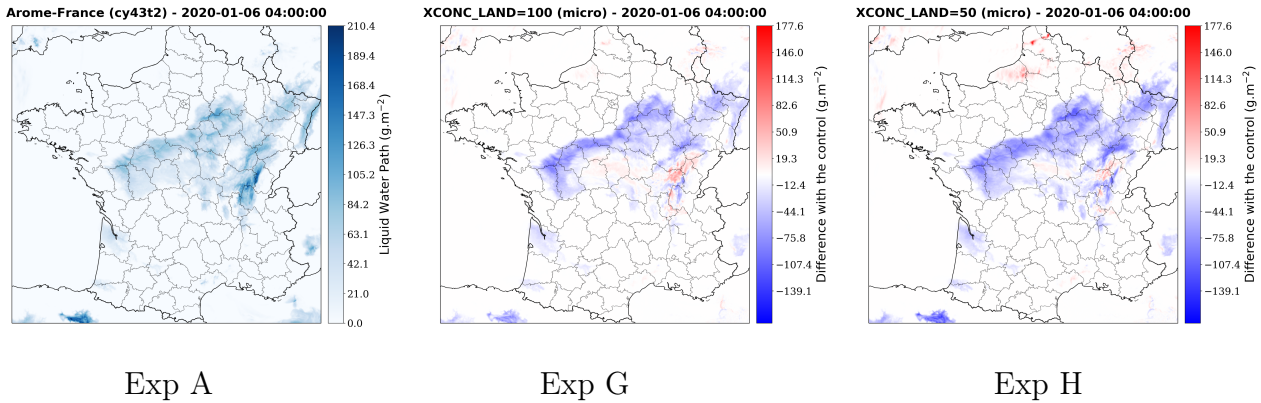


FIGURE 6.12 – LWP à +25h dans l’expérience A, et différence par rapport à cette référence pour les expériences G et H.

De façon générale, les résultats obtenus avec les expériences G et H sont les mêmes que ceux décrits dans la section précédente : de même que pour une diminution de la valeur de ν , une diminution de N_0 aboutit à un brouillard moins dense, moins élevé et réduit la progression de la zone de brouillard dans le centre de la France.

Sensibilité d’AROME au paramètre ν dans le schéma radiatif

On s’intéresse à présent aux simulations obtenues en modifiant la valeur du paramètre ν dans le schéma radiatif uniquement, soit les expériences E et F du tableau 6.1. Dans ces expériences, le calcul du rayon effectif des gouttelettes nuageuses dans le schéma radiatif est réalisé en remplaçant le coefficient k de la formule de Martin et al. (1994) par l’expression dépendant de ν de Jahangir et al. (2021), définie dans le chapitre 5 (équation 5.7). Pour effectuer une comparaison entre les différentes valeurs de ν , on utilise une nouvelle expérience de référence : l’expérience D, où les valeurs $N_0 = 300$ gouttes/cm³ et $\nu = 3$ sont prescrites aux deux schémas (microphysique et radiatif).

Les résultats obtenus sont qualitativement les mêmes que ceux des expériences B et C, mais l’impact sur la quantité d’eau nuageuse est bien moindre. On n’observe pas non plus d’impact significatif sur le cycle de vie du brouillard dans le sud-ouest comme le montre la figure 6.13.

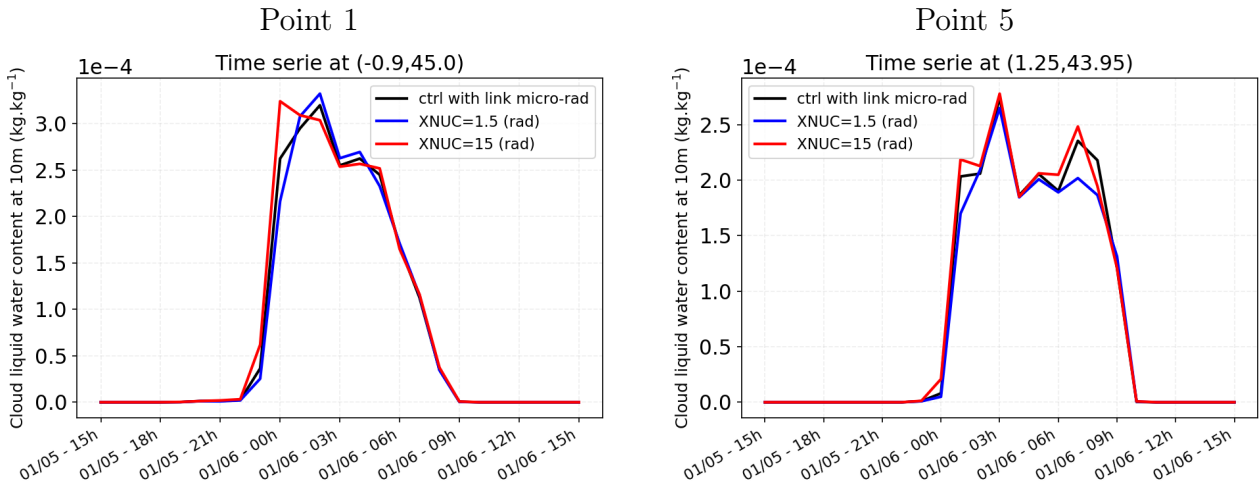


FIGURE 6.13 – Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés près de Bordeaux et Toulouse respectivement. La courbe noire est l’expérience contrôle (Exp D), la courbe bleue est l’expérience E et la rouge est l’expérience F.

Comme suggéré par les expériences 1D, on pourrait expliquer cela par le fait que l’impact de ν se produit essentiellement au travers des flux SW, or ceux-ci sont nuls durant le développement du brouillard qui a lieu la nuit.

Sensibilité d’AROME au paramètre N_0 dans le schéma radiatif

Les expériences I et J permettent enfin d’étudier l’impact d’une modification de N_0 dans le schéma radiatif. Diminuer la valeur de N_0 dans le schéma radiatif semble légèrement diminuer le contenu en eau du brouillard dans le sud-ouest et retarder la propagation du brouillard dans le centre de la France (figure 6.14), mais cet impact est moins prononcé que lorsque la valeur de N_0 est modifiée dans le schéma microphysique.

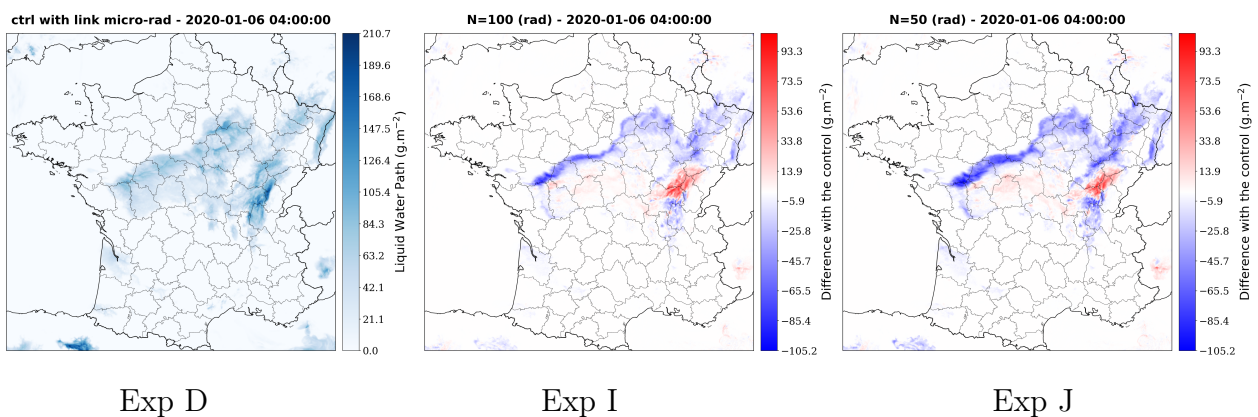


FIGURE 6.14 – LWP à +25h dans l’expérience D, et différence par rapport à cette référence pour les expériences I et J.

Concernant le cycle de vie du brouillard, de même que pour les expériences E et F, on ne constate pas de différences importantes (figure 6.15). L’étude d’une autre situation (par exemple une situation de stratocumulus) serait souhaitable afin d’étudier s’il y a davantage d’impact

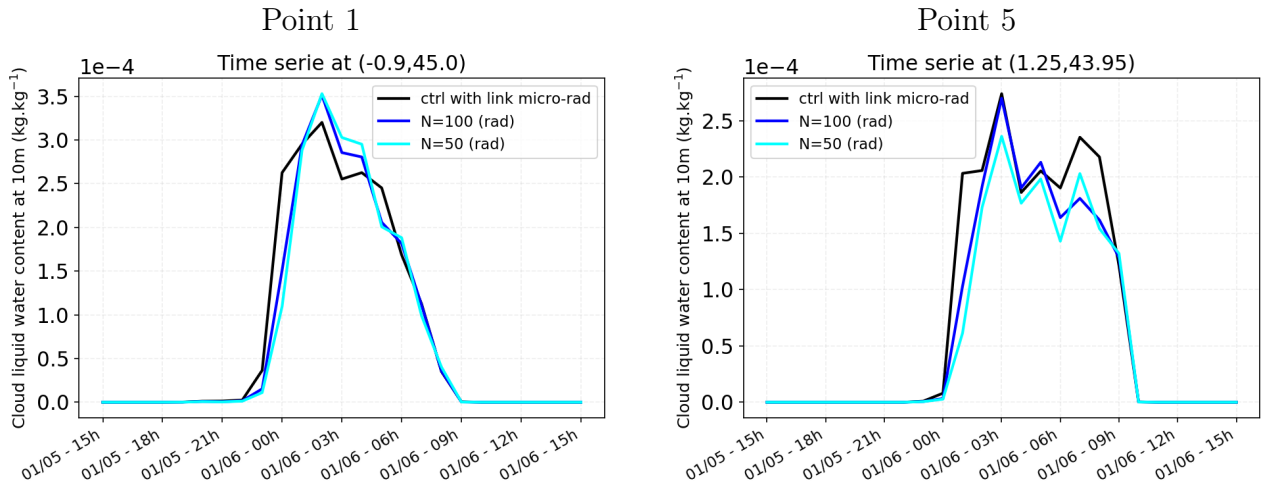


FIGURE 6.15 – Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés aux environs de Bordeaux et Toulouse respectivement. La courbe noire est l’expérience contrôle (Exp D), la courbe bleu foncé est l’expérience I et la courbe bleu clair est l’expérience J.

sur le cycle de vie d’un nuage persistant la journée.

Même si les différences sont moins marquées que pour le schéma microphysique, il semblerait qu’augmenter la valeur de N_0 ou ν dans le schéma radiatif aboutisse à davantage d’eau nuageuse ce qui est un effet inverse à celui observé dans les simulations 1D du cas de stratocumulus FIRE (section 5.2.2). Ceci confirme que l’impact de la modification de N_0 et ν dans le schéma radiatif n’est pas trivial et qu’il pourrait dépendre notamment du type de nuage.

6.1.2 Premiers tests des méthodes de perturbation de la turbulence et de la convection peu profonde

Cette deuxième étude de cas vise à étudier l’impact des méthodes de perturbation de la turbulence et de la convection peu profonde présentées dans le chapitre 4 dans une configuration 3D avec le système de prévision d’ensemble PEARO. La journée du 4 septembre 2021 a été choisie du fait du contexte synoptique suggérant des forçages de grande échelle relativement faibles (figure 6.16). Durant cette journée, des petites cellules convectives se forment en fin de matinée sur une large partie de la France (voir la carte 6.17 à 14:00 UTC) et persistent jusqu’en fin de soirée. Ce type de situation correspond au cas de forçage synoptique faible testé par Kober et Craig (2016) et pour lequel leur méthode a un impact positif sur les précipitations.

Plusieurs expériences PEARO ont été réalisées pour simuler cette journée. Toutes utilisent les mêmes perturbations initiales et de bord, mais différentes méthodes pour la représentation de l’erreur modèle :

- un ensemble de contrôle utilisant la méthode opérationnelle SPPT,
- un ensemble sans représentation de l’erreur modèle,
- un ensemble utilisant seulement la méthode de Kober et Craig (2016) (KC16),
- un ensemble utilisant seulement le schéma de convection stochastique,
- un ensemble utilisant la méthode KC16 et le schéma de convection stochastique.

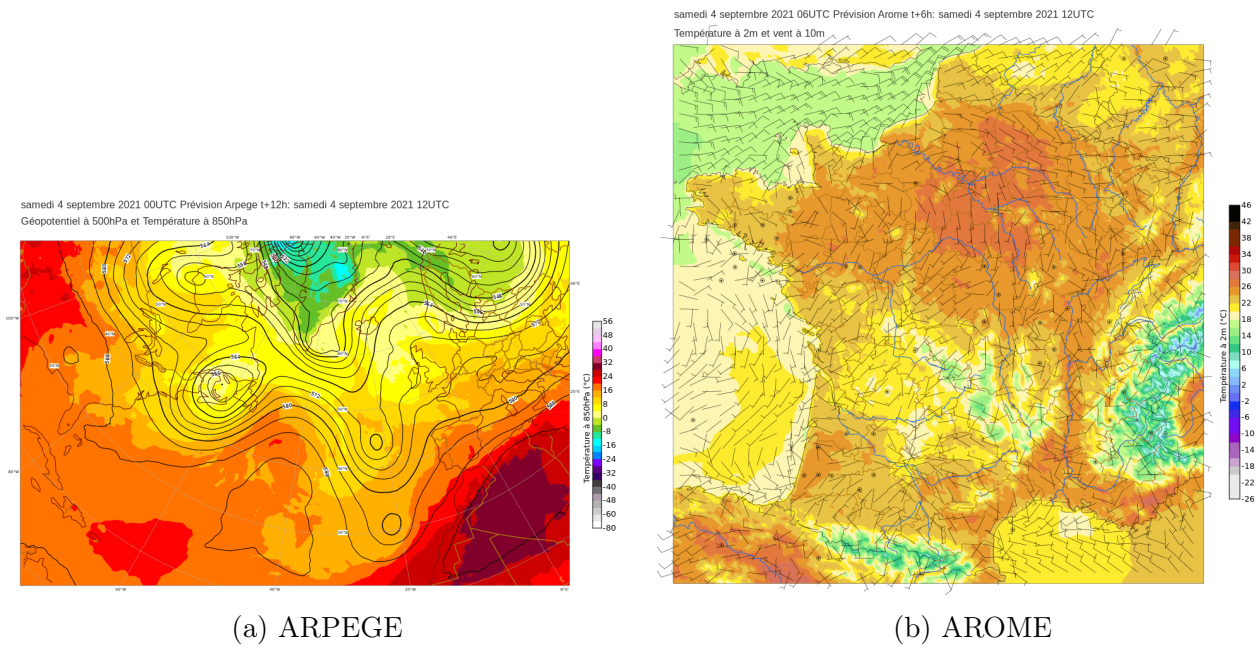


FIGURE 6.16 – Visualisation des données de simulation des modèles ARPEGE et AROME le 04/09/2021 à 12:00 UTC. (a) Géopotential à 500 hPa et température à 850 hPa simulés par le modèle global ARPEGE. (b) Température à 2 m et vent à 10 m simulés par le modèle AROME. Les cartes ont été extraites depuis <http://intra.cnrm.meteo.fr/visumodeles> (site interne).

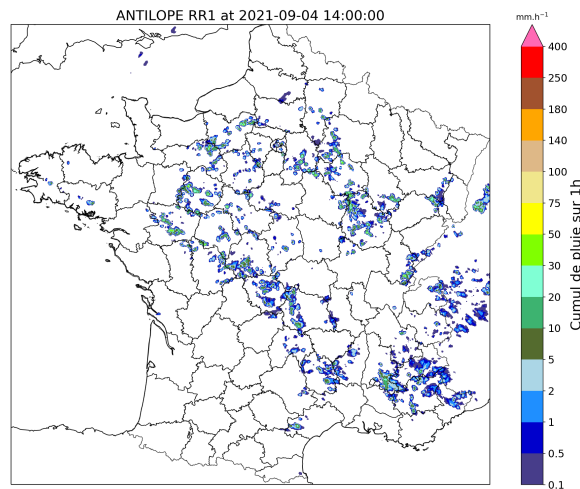


FIGURE 6.17 – Cumuls de pluie sur 1 heure observés (analyse de lame d'eau Antilope) le 04/09/2021 à 14:00 UTC.

La méthode KC16 employée utilise la même configuration que celle testée dans le chapitre 4 : les perturbations sont ajoutées aux tendances de température, d'humidité spécifique et des deux composantes horizontales du vent produites par le schéma de turbulence de Cuxart et al. (2000). Le même terme aléatoire est employé pour les 4 variables. Il s'agit d'un processus AR1 gaussien dont la corrélation temporelle est fixée à 10 minutes et la corrélation spatiale à 6.5 km, ce qui correspond à 5 fois la taille de la maille d'AROME (en accord avec les réglages de Kober et Craig 2016). Pour le schéma de convection stochastique, la condition de fermeture en flux de masse du schéma de Pergaud et al. (2009) est perturbée également avec un processus gaussien, avec une corrélation temporelle de 40 minutes comme dans les tests 1D, et une

corrélation spatiale d'environ 20 km, échelle à partir de laquelle l'hypothèse de quasi-équilibre pourrait être appliquée pour la convection peu profonde (Sakradzija et Klocke 2018).

Les simulations sont toutes initialisées avec l'analyse du 03/09/2021 03:00 UTC et on s'intéresse aux échéances d'environ +30 à +40 h couvrant la journée du lendemain, ceci afin de laisser le temps aux perturbations modèle de se développer.

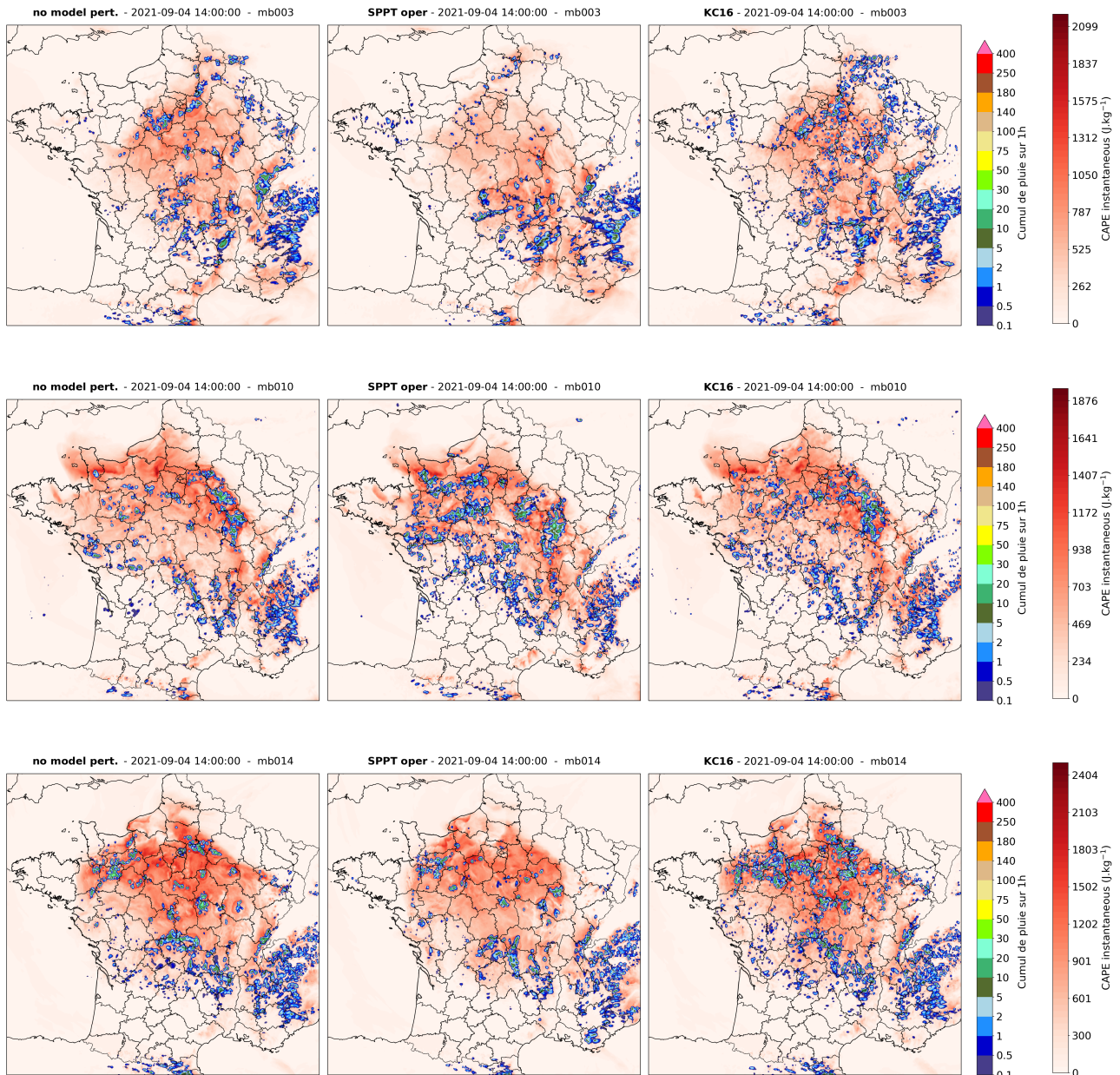


FIGURE 6.18 – Cartes de l'énergie convective disponible (CAPE) et des cumuls de précipitations sur 1 heure simulés pour le 04/09/2021 à 14:00 UTC par plusieurs membres de 3 expériences PEARO : un ensemble sans représentation de l'erreur modèle (colonne de gauche), l'ensemble utilisant la méthode opérationnelle SPPT (colonne du milieu), et un ensemble utilisant la méthode KC16 (colonne de droite). Les trois lignes correspondent respectivement aux membres 3, 10 et 14.

La figure 6.18 montre des cartes de cumuls de précipitations sur 1 heure obtenus dans certains membres des trois premières expériences. Dans tous les membres (y compris ceux non représentés sur la figure), on constate l'apparition de cellules convectives supplémentaires dans l'expérience "KC16" par rapport à celle sans perturbations modèle. Dans la majorité des membres, il y a également davantage de cellules convectives formées dans l'ensemble "KC16" que dans l'ensemble "SPPT" mais pour certains membres la différence est moins nette (voir par exemple le membre 10 : ligne du milieu sur la figure 6.18). Comme on peut le voir sur les cartes, les cellules supplémentaires dans l'ensemble "KC16" se forment dans les zones où la CAPE est non nulle. On peut donc supposer que les perturbations locales, introduites par la méthode KC16, sont à même d'initier de nouveaux nuages dans les zones d'instabilité convective, ce qui est l'un des objectifs de cette méthode. Les résultats de la figure 6.18 sont d'ailleurs cohérents avec ceux obtenus par Kober et Craig (2016) dans un cas de forçage synoptique faible (voir leurs figures 2 et 6).

Étant données les différences de corrélations temporelles et spatiales utilisées dans les méthodes SPPT et KC16 (6 h et 500 km pour l'une, 10 min et 6.5 km pour l'autre), on peut se demander si la tendance à générer davantage de cellules convectives avec KC16 peut simplement être due au fait que l'on utilise des perturbations plus localisées. Pour étudier cela, quelques expériences supplémentaires ont été réalisées en modifiant les corrélations spatio-temporelles des deux méthodes. Une comparaison des cumuls de précipitations sur 1 heure pour les expériences "SPPT", "KC16" et une expérience où les corrélations spatio-temporelles de la méthode SPPT ont été réduites à 6.5 km et 10 min est présentée sur la figure 6.19.

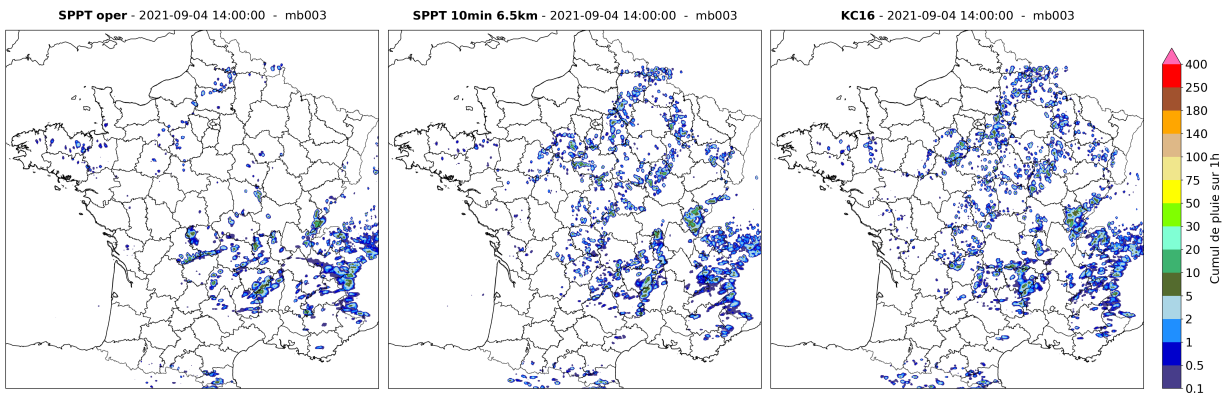


FIGURE 6.19 – Cartes de cumuls des précipitations sur 1 heure simulés pour le 04/09/2021 à 14:00 UTC par le membre 3 de trois expériences PEARO. De gauche à droite : un ensemble utilisant la méthode opérationnelle SPPT, un ensemble utilisant SPPT où les corrélations temporelles et spatiales ont été réglées à 10 min et 6.5 km respectivement, et un ensemble utilisant la méthode KC16.

On constate qu'en réduisant les corrélations spatio-temporelles de SPPT, la simulation devient beaucoup plus proche de celle de l'expérience "KC16", avec davantage de cellules convectives et une répartition spatiale similaire. Ceci est observé pour tous les membres, de façon plus ou moins marquée. Les membres de l'ensemble "KC16" semblent toujours simuler davantage de cellules convectives, et plus petites, mais la différence entre les deux expériences est globalement faible. Des expériences complémentaires où seule la corrélation temporelle ou bien la corrélation spatiale de la méthode SPPT est modifiée montrent que le principal effet sur la

multiplication des cellules convectives provient de la corrélation spatiale. Par ailleurs, régler la corrélation spatio-temporelle de la méthode KC16 à 500 km et 6 h modifie fortement les résultats des simulations et supprime l'effet de multiplication des cellules convectives observé dans les figures précédentes. Les courtes corrélations temporelles et spatiales utilisées dans la méthode KC16 semblent donc être déterminantes dans la capacité de cette méthode à initier davantage de cellules dans ce cas d'étude.

L'effet de la convection stochastique sur les précipitations est très similaire à celui de la méthode KC16 et n'est pas montré ici. Lorsque les deux méthodes sont utilisées en même temps, le nombre de cellules convectives créées augmente encore davantage.

La figure 6.20 présente l'évolution, durant la journée du 04/09/2021, des cumuls de précipitation sur 1 heure moyennés sur toute la France, simulés par les différents ensembles, ainsi que les observations correspondantes.

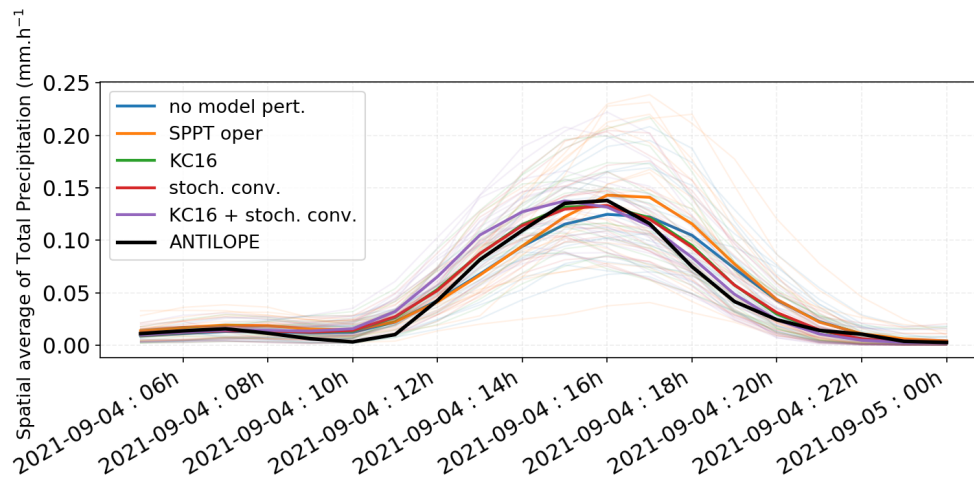


FIGURE 6.20 – Évolution temporelle du cumul de précipitation sur 1 heure moyenné sur toute la France durant la journée du 04/09/2019, pour différentes expériences PEARO. La courbe noire correspond aux observations (lame d'eau Antilope). Les courbes de couleur foncées correspondent à la moyenne de chaque ensemble, tandis que les courbes claires et plus fines représentent chaque membre des différents ensembles.

On constate que les ensembles utilisant la méthode KC16 ou la convection stochastique ne produisent pas nécessairement plus de pluie en moyenne (courbes vertes et rouges sur la figure 6.20, qui sont pratiquement superposées l'une à l'autre) mais que le timing est plus proche des observations que pour les autres ensembles. Les ensembles "SPPT" et sans perturbations modèle ont un pic de précipitation légèrement en retard par rapport aux observations. L'ensemble utilisant les deux méthodes KC16 et de convection stochastique semble au contraire produire des perturbations légèrement trop tôt. Il est donc possible de supposer que les deux méthodes POP de perturbation de la turbulence et de la convection favorisent l'initiation de cellules plus tôt dans la journée. Néanmoins, les différences sont assez faibles donc l'étude d'autres journées serait nécessaire pour le confirmer.

La figure 6.21 présente la dispersion calculée entre les membres des différents ensembles en chaque point de grille du domaine, puis moyennée sur tout le domaine, pour une des compo-

santes horizontales du vent proche de la surface (à 10 m). Pour tous les ensembles, on constate

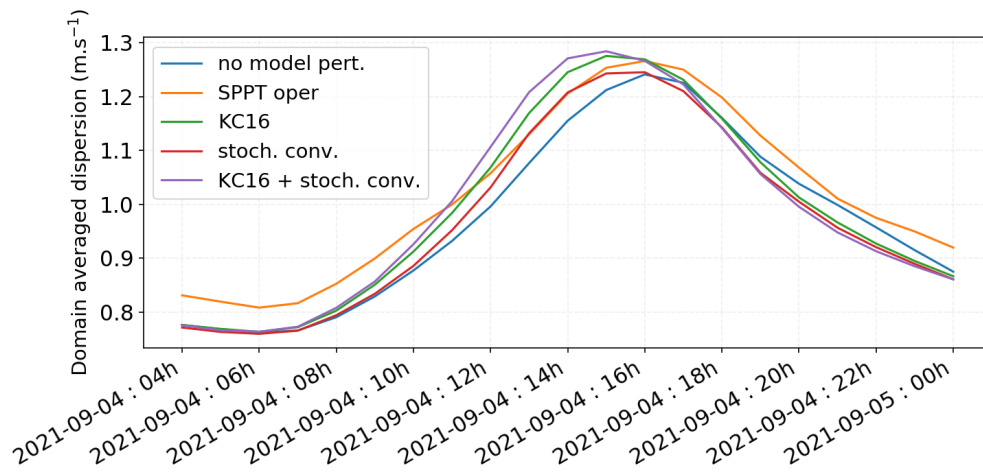


FIGURE 6.21 – Évolution temporelle de l'écart-type entre membres pour le vent zonal à 10 m, moyenné sur toute la France, durant la journée du 04/09/2019, pour différentes expériences PEARO.

une augmentation de la dispersion durant la journée et une rediminution le soir, en lien avec l'activité convective et l'augmentation de la turbulence en journée lorsque la couche limite se réchauffe. On constate que le cycle diurne est davantage marqué pour les ensembles utilisant la méthodes KC16 (et dans une moindre mesure la convection stochastique), ce qui peut s'expliquer par le fait que les perturbations de la turbulence sont contrôlées par la TKE, qui suit elle-même un cycle diurne marqué pour ce cas d'étude. De 04:00 à 06:00 UTC, on voit ainsi que la dispersion de ces ensembles est la même que celle de l'ensemble sans perturbations modèle, tandis qu'en milieu de journée, elle devient supérieure à celle de l'ensemble "SPPT". Les mêmes observations peuvent être faites pour la température et l'humidité à 2 m.

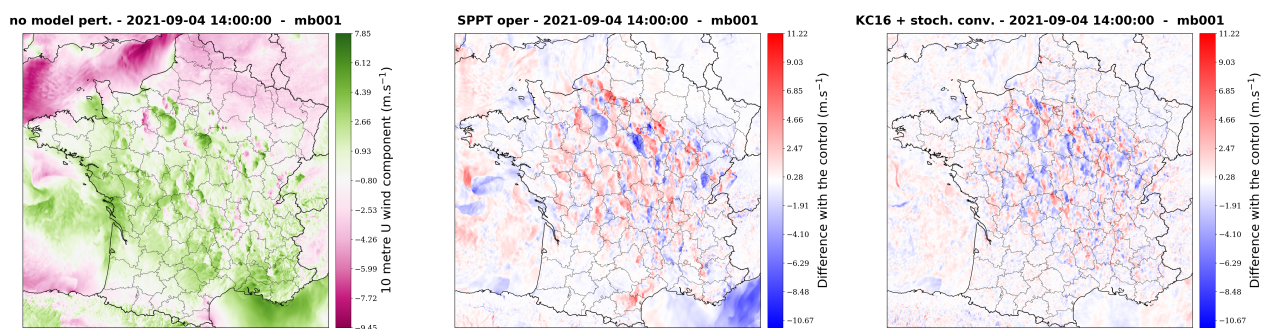


FIGURE 6.22 – Carte de vent zonal à 10 m simulé par le membre 1 de trois expériences PEARO pour le 04/09/2021 à 14:00 UTC. La carte de gauche correspond à l'expérience sans perturbations modèle. Les deux autres cartes présentent la différence par rapport à cette référence, pour l'expérience SPPT et celle utilisant la méthode KC16 et la convection stochastique. La même échelle de couleur est utilisée pour ces deux dernières cartes.

Néanmoins, même si les méthodes KC16 et de convection stochastique peuvent produire à l'échelle d'un point de grille une dispersion entre les membres parfois comparable à SPPT,

les perturbations sont très localisées et ne semblent pas à même d'avoir un impact à grande échelle. On voit sur les cartes de la figure 6.22 qu'en effet les différences de valeur de vent zonal à 10 m avec l'ensemble sans perturbation modèle sont plus localisées avec les méthodes KC16 et convection stochastique qu'avec la méthode SPPT. De façon générale, les membres des ensembles utilisant la méthode KC16 et/ou la convection stochastique restent très proches des membres de l'ensemble sans perturbations modèle. Il semblerait donc que les perturbations très locales introduites par KC16 et la convection stochastique ne parviennent pas à s'amplifier durant les simulations de ce cas d'étude.

6.2 Évaluation de trois ensembles sur le mois de septembre 2021

Pour prolonger l'étude de cas précédente et étudier davantage l'impact des méthodes KC16 et de convection stochastique sur le comportement du système PEARO, trois expériences sont lancées sur un mois complet (septembre 2021). Les trois expériences utilisent les mêmes perturbations initiales et de bord, seule la représentation de l'erreur modèle varie :

- un ensemble utilisant la méthode opérationnelle SPPT,
- un ensemble utilisant la variante iSPPT,
- un ensemble utilisant la méthode KC16 et la convection stochastique, et la méthode iSPPT sur les schémas de paramétrisation microphysique et radiatif.

Comparer un ensemble "SPPT" avec un ensemble utilisant seulement les perturbations de KC16 et la convection stochastique ne paraît pas équitable, puisque les perturbations SPPT affectent tous les schémas de paramétrisation. Pour la troisième expérience, on utilise donc la méthode iSPPT, qui permet d'appliquer des perturbations indépendamment à chaque paramétrisation, pour perturber en plus les schémas microphysique et radiatif.

Les résultats de ces expériences sont étudiés au travers des scores suivants.

Diagramme de fiabilité Ce diagramme présente l'adéquation entre les probabilités d'occurrence d'un évènement prédites par un ensemble et les fréquences correspondantes auxquelles l'évènement a été observé. Pour cela, l'intervalle des probabilités est séparé en plusieurs classes. Pour chaque point où la probabilité (calculée comme le rapport entre le nombre de membres prévoyant l'évènement et le nombre total de membres dans l'ensemble) se trouve dans une classe donnée, on regarde si l'évènement a eu lieu ou non selon les observations. Ceci permet alors de calculer la fréquence d'occurrence de l'évènement, que l'on associe à la classe de probabilité concernée. Un ensemble est fiable lorsque les probabilités prévues correspondent aux fréquences observées, soit lorsque la courbe du diagramme se situe sur la diagonale.

Courbe ROC (Relative Operating Characteristics) La courbe ROC est utilisée pour évaluer la capacité d'un ensemble à prédire un évènement binaire. Elle relie, pour cet évènement, le taux de fausses alarmes et le taux de détection. Le taux de fausses alarmes est défini comme le rapport entre le nombre de fois où l'évènement a été prédit à tort et le nombre total de fois où l'évènement a été prédit. Le taux de détection est défini comme le rapport entre le nombre de fois où l'évènement a été correctement prédit et le nombre total de fois où l'évènement s'est réalisé. Pour réaliser une prédiction binaire à l'aide d'un ensemble, il est nécessaire de fixer un seuil

de détection, c'est-à-dire une probabilité à partir de laquelle on considère que l'évènement est prédit par l'ensemble. Les différents points formant une courbe ROC correspondent à différents seuils de détection. Lorsque les points se trouvent sur la diagonale $y = x$, il y a autant de détections que de fausses alarmes, donc l'ensemble n'a pas d'utilité pour prédire l'évènement. Plus les points se rapprochent du coin supérieur gauche du graphe, plus l'ensemble est utile (forts taux de détection pour faibles taux de fausses alarmes).

Score de Brier Le score de Brier est l'équivalent d'une erreur quadratique moyenne, appliquée aux probabilités. Si l'on définit p_i la probabilité d'un certain évènement prévue par un ensemble et o_i l'observation associée alors le score est donné par :

$$BS = \frac{1}{N} \sum_i (p_i - o_i)^2,$$

avec N le nombre de couples observation/prévision. L'évènement à prévoir doit être binaire, de sorte que o_i vaut 0 ou 1. La probabilité p est elle comprise entre 0 et 1. Comme pour une erreur quadratique moyenne, plus ce score est proche de zéro, meilleures sont les prévisions. Ce score peut par ailleurs être décomposé en plusieurs composantes (Murphy 1973) reliées à l'incertitude de l'évènement, la fiabilité définie précédemment et la résolution, c'est-à-dire à quel point les probabilités prédites s'éloignent de leur moyenne climatologique.

Les scores présentés dans les figures 6.23, 6.24 et 6.25 concernent les précipitations. Ils sont établis en utilisant comme observations la lame d'eau de fusion de données radar et de stations de mesure au sol (lame d'eau Antilope). Deux évènements sont considérés : les cumuls de précipitation sur 1 heure dépassent 1 mm ou bien 5 mm. Étudier de plus forts cumuls n'est pas possible dans notre cas car avec seulement un mois de simulation, le nombre d'évènements serait trop faible.

Le constat général d'après les trois figures est que les scores des trois ensembles sont pratiquement identiques. Utiliser iSPPT plutôt que SPPT pour représenter l'erreur modèle n'aurait donc pas d'impact sur la qualité de la prévision probabiliste des pluies pour ce mois de septembre 2021. De même, remplacer les perturbations multiplicatives iSPPT des schémas de turbulence et de convection peu profonde par les perturbations de KC16 et de la convection stochastique n'affecte pas les scores. Ceci est un peu surprenant. Il faudrait probablement étendre la période de test, explorer d'autres moments de l'année et calculer les scores sur d'autres variables (vent, température, humidité, etc) pour approfondir ces résultats.

CHAPITRE 6. ÉVALUATION DES MÉTHODES DE PERTURBATION DANS DES SIMULATIONS 3D DU MODÈLE AROME

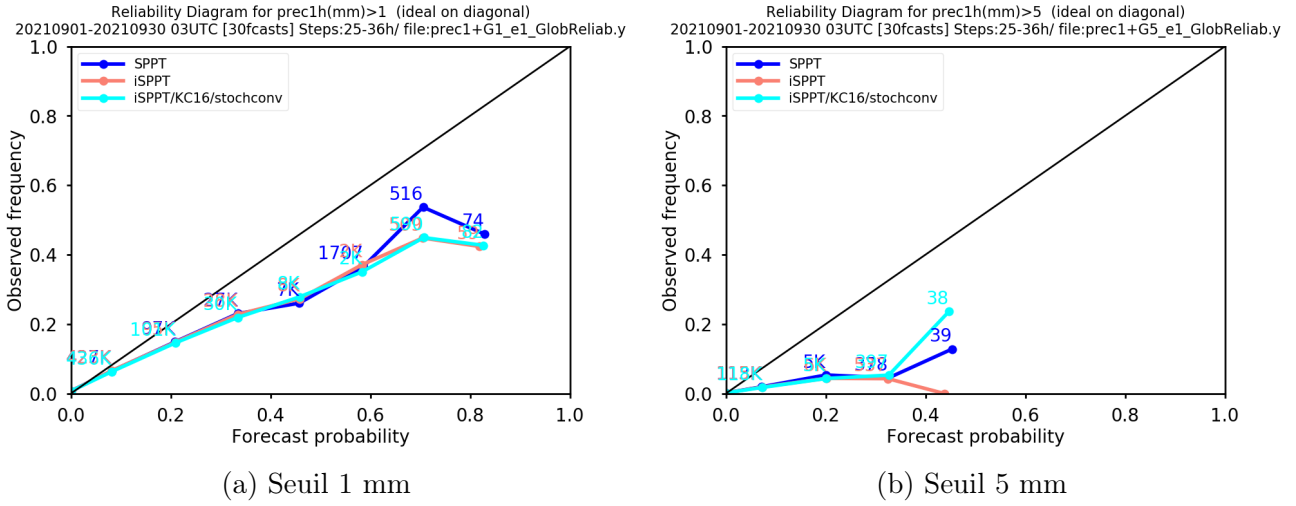


FIGURE 6.23 – Diagrammes de fiabilité

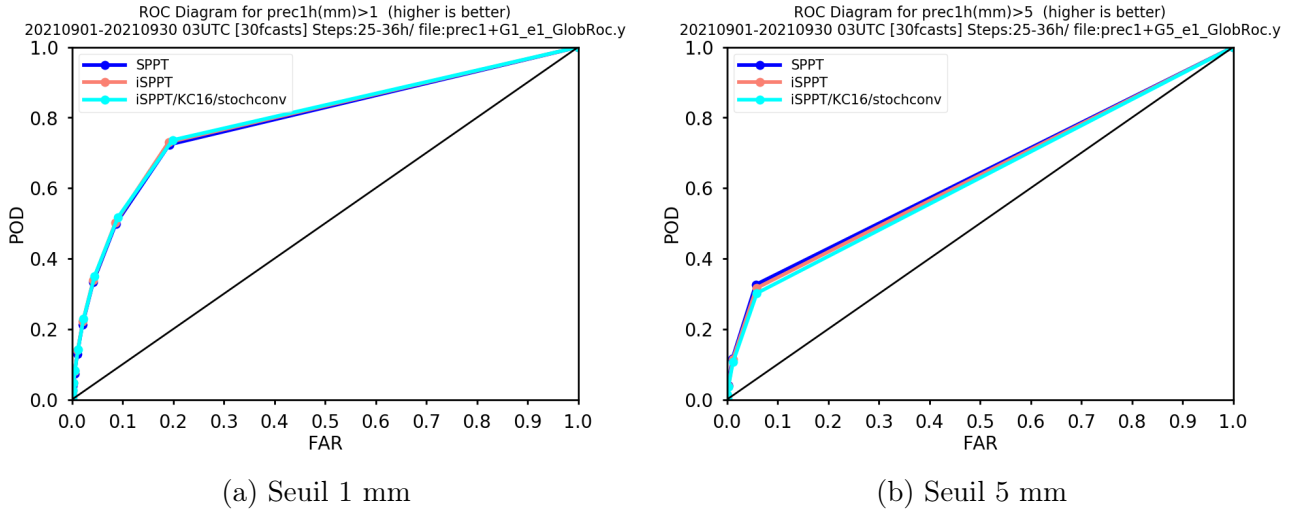


FIGURE 6.24 – Courbes ROC

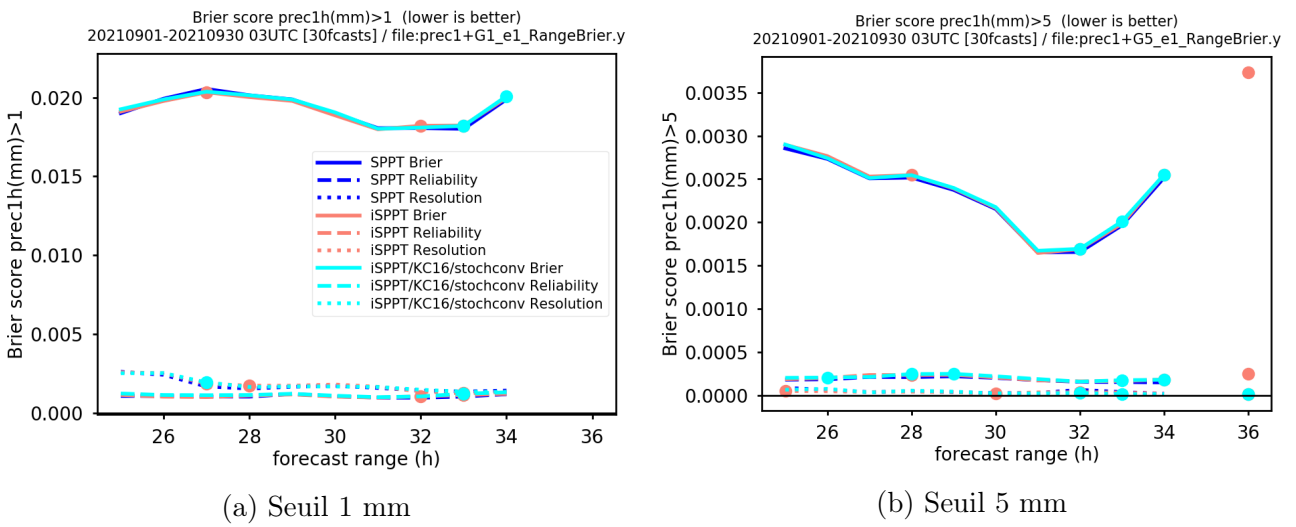


FIGURE 6.25 – Scores de Brier

Conclusion

Les modèles numériques de l’atmosphère sont de nos jours des outils indispensables à la prévision météorologique. Comme tout modèle, ils sont une simplification de la réalité. Pour utiliser au mieux leurs résultats, il est essentiel de caractériser les erreurs qu’ils commettent, mais il s’agit d’une tâche difficile car elles sont très dépendantes du temps, du lieu et des phénomènes modélisés. Depuis les années 1990, de nombreux centres météorologiques utilisent des *systèmes de prévision d’ensemble* pour tenter de fournir une estimation de l’incertitude de prévision. Le principe est de produire plusieurs scénarios équiprobables d’une situation à prévoir, généralement en introduisant des perturbations dans le système de prévision qui sont censées représenter toutes les sources d’incertitude pouvant mener à des erreurs. Les deux sources d’incertitude majeures pour la prévision météorologique sont les erreurs des conditions initiales (et de bord), dues aux données incomplètes d’observation de l’atmosphère, et les erreurs des modèles eux-mêmes, liées à leurs approximations.

Dans le domaine de la représentation des erreurs modèle, les méthodes de perturbations stochastiques sont fréquemment utilisées et ont démontré leur intérêt pour améliorer la performance des ensembles (e.g. Berner et al. 2009 ; Palmer et al. 2009). Une partie de ces techniques se base sur une représentation relativement sommaire de l’incertitude en utilisant un même terme aléatoire pour représenter tout un ensemble d’erreurs de natures différentes. Bien qu’elles soient peu coûteuses et efficaces pour produire de la dispersion entre les scénarios – ou membres – d’un ensemble, l’inconvénient de ces méthodes est que l’interprétation physique de leurs perturbations est parfois difficile et que celles-ci ne sont pas toujours adaptées à la situation météorologique. Ceci explique qu’actuellement, de nombreuses recherches se tournent vers des méthodes de perturbation plus précises ayant pour objectif de relier plus clairement les perturbations aux processus physiques à l’origine d’incertitude dans les modèles (Leutbecher et al. 2017).

Un des problèmes de longue date concernant les ensembles est le manque de dispersion : l’état réel de l’atmosphère se trouve trop souvent en dehors de l’intervalle prévu par les différents membres, ce qui indique que les méthodes actuelles ne permettent pas de représenter correctement toutes les sources d’incertitude. En particulier, il n’existe pas de véritable consensus sur la manière de représenter les erreurs modèle. Cette thèse s’est donc inscrite dans la problématique assez générale de savoir comment améliorer la représentation de l’incertitude modèle dans un système de prévision d’ensemble (le système PEARO de Météo-France). Plus précisément, l’objectif principal de la thèse était d’évaluer si des méthodes de perturbation plus élaborées du point de vue de leur interprétation physique peuvent avoir un intérêt en pratique pour améliorer la représentation de l’incertitude et les performances du système de prévision d’ensemble.

Pour cela, différentes méthodes de perturbation ont été testées dans chacun des schémas de

paramétrisation du modèle AROME. Un cadre de travail simplifié a été choisi pour caractériser l'impact de chaque perturbation sur le modèle, avant de procéder à des expériences 3D. Des simulations uni-colonne de plusieurs cas idéalisés (ARMCu, BOMEX, FIRE et LANFEX) ont ainsi été produites et perturbées, en s'appuyant éventuellement sur des simulations à haute résolution (LES) des mêmes cas d'étude. On présente dans la suite le résumé des travaux effectués, les principaux résultats, ainsi que leurs limites. Des perspectives sont également discutées.

Une première partie de la thèse a consisté à implémenter et tester deux méthodes de perturbation des processus (POP) dédiées à la turbulence et à la convection peu profonde. Leur caractéristique essentielle est de chercher à représenter une incertitude intrinsèque à la paramétrisation de ces deux processus physiques : les fluctuations dues à l'occurrence aléatoire et en nombre limité des structures sous-maille turbulentes et convectives que sont les tourbillons et les ascendances. Si ces fluctuations, qui s'apparentent à un bruit d'échantillonnage, peuvent être négligées lorsque la maille d'un modèle est grande, leur impact augmente à mesure que la résolution horizontale augmente. Avec une résolution de 1.3 km, le modèle AROME se trouve déjà dans la "zone grise" de la turbulence et de la convection (Honnert et al. 2011, 2016), où les hypothèses des paramétrisations de ces deux phénomènes commencent à être inadaptées, ce qui suggère l'intérêt de telles perturbations. L'originalité des deux méthodes repose sur la modélisation des fluctuations, à partir d'arguments physiques sur le comportement des ascendances et des tourbillons, pour construire leurs perturbations. Ces dernières ont donc des caractéristiques en terme d'amplitude, de corrélations spatiales et temporelles qui découlent directement de la modélisation – contrairement à d'autres méthodes stochastiques où elles sont fixées de façon arbitraire.

Pour la turbulence comme pour la convection peu profonde, les deux méthodes sont tirées de résultats publiés dans la littérature, et ont été adaptées grâce aux données de simulations LES. Le schéma PSP (Physically based Stochastic Perturbation ; Kober et Craig 2016 ; Hirt et al. 2019) est employé pour perturber les tendances du schéma de turbulence d'AROME, grâce à une information sur la variance sous-maille des variables de vent, d'humidité et de température. Un schéma inspiré des travaux de Sakradzija et al. (2015), mais grandement simplifié à l'aide des LES, est utilisé pour perturber la condition de fermeture en flux de masse du schéma de convection peu profonde d'AROME. La comparaison avec la variabilité observée dans les champs LES (auxquels on applique une procédure de coarse-graining) montre que les ensembles de simulations uni-colonne produits avec ces deux méthodes ont une dispersion de température et d'humidité trop faible, en particulier au sein des nuages. En revanche, la dispersion du vent est comparable à celle des LES, essentiellement grâce aux perturbations de la turbulence. La perturbation de la convection a quant à elle un impact intéressant sur le contenu en eau nuageuse. Telles qu'implémentées dans AROME, ces deux méthodes atteignent donc partiellement leurs objectifs.

Les simulations LES ont été employées pour explorer la possibilité de perturber un troisième schéma de paramétrisation : l'ajustement à la saturation qui traite des processus microphysiques "rapides" de formation de gouttelettes et de glace nuageuses. La relation de Chaboureaud et Bechtold (2002) (CB02), diagnostiquant le contenu en hydrométéores non précipitants dans une maille, a été perturbée en s'appuyant sur les données issues du coarse-graining des simulations LES du cas de stratocumulus FIRE. Son impact est limité sur les simulations uni-colonne de FIRE, en revanche le cas de brouillard LANFEX y est plus sensible. Une limite importante de ces premiers tests est que l'analyse des LES n'a pas pris en compte le fait que dans AROME, le schéma de convection peu profonde contribue également à la formation d'eau ou de glace

nuageuse, et n'a donc pas pu établir clairement la part d'incertitude liée à la relation de CB02. La question peut se poser de savoir si, pour représenter la variabilité d'eau nuageuse observée dans les LES, la perturbation du schéma de convection peu profonde d'AROME est ou non suffisante. L'étude de simulations LES d'un cas non convectif (comme par exemple LANFEX) pourrait apporter des informations supplémentaires sur l'incertitude liée à la relation CB02.

Une deuxième partie des travaux de thèse a porté sur les schémas microphysique et radiatif. Pour ces schémas, une méthode de perturbation de deux paramètres réglant la distribution de taille des gouttelettes nuageuses (N_0 , la concentration en nombre de gouttelettes et ν , un paramètre de forme) a été testée. Les simulations des cas LANFEX et FIRE se sont révélées sensibles à ces deux paramètres. Leur impact non linéaire sur le modèle crée en revanche un biais dans les ensembles de simulations uni-colonne, ce qui a motivé la recherche de distributions de valeurs pour ces deux paramètres calibrées à l'aide de la résolution d'un problème inverse. Ensuite, en s'inspirant des travaux de Jahangir et al. (2021) montrant que N_0 et ν interviennent également dans le schéma radiatif au travers du calcul du rayon effectif des gouttelettes nuageuses, la méthode de perturbation a pu être étendue à la paramétrisation des flux radiatifs SW et LW. Les tests sur des simulations uni-colonne montrent que la perturbation des deux paramètres dans le schéma radiatif a un impact essentiellement sur les flux SW, et que la dispersion des ensembles est globalement plus faible que lorsque les perturbations concernent le schéma microphysique.

Dans une dernière partie, quelques simulations ont été réalisées avec la version 3D du modèle AROME de façon à étendre les travaux à un cadre plus opérationnel. Un test de sensibilité aux paramètres N_0 et ν d'une situation de brouillard sur la France en janvier 2020 a été réalisé et montre que la réponse du modèle aux changements de valeur de ces deux paramètres est cohérente avec les observations faites en 1D. L'application des deux méthodes POP de perturbation de la turbulence et de la convection peu profonde dans le système PEARO a également été testée, mais les résultats sont très préliminaires. Pour une journée de septembre 2021 avec un faible forçage synoptique et des conditions favorables au développement de la convection, il semble que les deux méthodes produisent une multiplication des cellules convectives. En revanche, les simulations réalisées sur un mois complet ne semblent pas indiquer d'impact significatif sur trois types de scores.

Limites et perspectives des travaux

Certaines limites sont à souligner, tout d'abord dans l'implémentation des méthodes POP de perturbation de la turbulence et de la convection peu profonde. Premièrement, la version du schéma de convection stochastique employée dans cette thèse est beaucoup moins réaliste que celle de Sakradzija et al. (2015) : la population aléatoire de nuages est réduite à une perturbation du flux de masse à la surface, et l'effet de mémoire dû au temps de vie des nuages est remplacé par une corrélation temporelle du terme aléatoire. Par ailleurs, dans les tests 3D, on ne respecte pas la logique du schéma initial qui est de perturber une condition de fermeture valide à une plus grande échelle que celle de la maille en moyennant le flux de masse sur plusieurs mailles adjacentes. Dans les tests 3D, le terme aléatoire a une corrélation spatiale mais les perturbations sont appliquées au flux de masse de chaque colonne sans effectuer de moyenne avec des mailles voisines. Il serait donc souhaitable d'ajouter cette possibilité dans de futures expériences.

Par ailleurs, le terme aléatoire perturbant le flux de masse repose sur des distributions calculées dans les LES de cas d'étude idéalisés, ce qui pose la question de sa généralisation

à des situations réelles. La similarité de la forme (gaussienne) et des caractéristiques de la distribution (rapport $\sigma_{\text{LES}}/\mu_{\text{LES}}$) pour chacun des trois cas étudiés (ARMCu, BOMEX et FIRE) laisse espérer qu’une formulation unifiée puisse être trouvée, mais l’examen d’autres situations météorologiques et de cas réels serait nécessaire pour le confirmer. Par ailleurs, contrairement à la méthode PSP et au schéma de Sakradzija et al. (2015), la perturbation de la convection peu profonde appliquée dans cette thèse n’est pas prévue pour s’adapter à la résolution du modèle, ce qui impliquerait de recommencer l’étude si la résolution du modèle change. De ce point de vue, le schéma de Sakradzija et al. (2015) et Sakradzija et Klocke (2018) est beaucoup plus complet. Le problème de son coût numérique prohibitif pourrait être contourné en utilisant le système d’équations stochastiques proposé par Machulskaya et Seifert (2019) pour mimer ce schéma. Cette approche est actuellement testée dans le modèle allemand ICON (Ahlgriem et al. 2022).

L’implémentation de la méthode de perturbation PSP est elle très proche de celle employée par Kober et Craig (2016) et Hirt et al. (2019). Certains points pourraient néanmoins être discutés, comme l’utilisation d’un même terme aléatoire (d’amplitude différente mais de signe identique) pour toutes les variables perturbées.

La méthode de perturbation de paramètres développée dans la deuxième partie de la thèse s’appuie sur la résolution d’un problème inverse. La méthode de Butler et al. (2018) a permis de régler les perturbations de N_0 et ν de façon à obtenir des ensembles non biaisés, ce qui pourrait constituer une piste intéressante pour les méthodes de perturbation de paramètres qui font parfois face aux problèmes d’apparition de biais dans les ensembles (Frogner et al. 2022; Bouttier et al. 2022 présenté dans l’annexe C). Néanmoins, le problème inverse a été ici défini pour un cas relativement simple de brouillard radiatif, et seulement deux paramètres. La possibilité de généraliser cette approche à davantage de paramètres et pour la simulation de cas réels pose question. Premièrement, on peut se demander si pour une utilisation opérationnelle, la résolution devrait être faite avec une configuration 3D du modèle au lieu d’une version uni-colonne. Utiliser un modèle 3D complet semble prohibitif du point de vue du coût numérique, même pour un problème à deux paramètres comme celui étudié dans la thèse (environ 2000 simulations effectuées), mais l’utilisation d’une version hydrostatique de moindre coût, comme dans Wimmer (2021), pourrait être une option. Dans ce cas, la question de la mesure utilisée pour estimer la distribution des paramètres se pose aussi. Avec les simulations uni-colonne, utiliser le contenu en eau nuageuse à +18 h à 100 m suffit à réduire le biais sur tout le profil du cas LANFEX. Pour un modèle 3D, la définition de la mesure semble moins évidente, étant donné que le biais des ensembles peut varier par exemple selon la position géographique. Avant d’envisager la résolution du problème inverse avec des simulations 3D, il pourrait donc être intéressant d’étudier d’abord la performance d’un ensemble 3D utilisant des distributions calibrées dans un cadre uni-colonne, possiblement sur différents cas idéalisés pour explorer plusieurs situations météorologiques.

Par ailleurs, inclure davantage de paramètres dans la résolution du problème inverse augmenterait là aussi le coût de calcul. Des algorithmes de type Monte Carlo par chaînes de Markov (MCMC), non utilisés dans ces travaux de thèse, s’avèreraient sûrement indispensables pour explorer efficacement la distribution *a posteriori* des paramètres et limiter le nombre de simulations requises.

Ces travaux de thèse ont été marqués par le décalage pouvant exister entre la justification physique d’une perturbation et son intérêt pratique. Ainsi, les méthodes POP de perturbation de la turbulence et de la convection, qui s’appuient sur des arguments physiques intéressants,

ont une efficacité parfois limitée à générer de la dispersion entre les membres d'un ensemble. Les méthodes POP représentent une source d'erreur "intrinsèque", c'est-à-dire qui existe du fait même de l'utilisation de schémas de paramétrisation, et l'approche stochastique pour ce type d'erreur d'échantillonnage semble justifiée et inévitable, tant que les modèles n'ont pas une résolution suffisante pour représenter explicitement ces processus et s'affranchir des paramétrisations. Néanmoins, il s'agit peut-être d'une source d'incertitude de modélisation qui n'est pas forcément dominante, et une large partie des erreurs pourrait plutôt provenir de diverses approximations et réglages faits dans les modèles. De ce point de vue, les méthodes de perturbation de paramètres semblent plus efficaces, tout en permettant de continuer à cibler précisément l'incertitude. Néanmoins, leurs perturbations sont plus empiriques que celles des méthodes POP : les valeurs des paramètres sont perturbées sans tenir compte de la situation météorologique, et avec des corrélations temporelles et spatiales souvent arbitraires.

La perturbation de paramètres pourrait d'une certaine façon être considérée comme un moyen de corriger les déficiences de certaines paramétrisations. Par exemple, les schémas microphysiques à 1 moment fixent la valeur de la concentration en nombre de gouttelettes bien que l'on sache que celle-ci varie en temps et en espace. Perturber le paramètre N_0 d'un schéma microphysique à 1 moment permet de relâcher cette hypothèse irréaliste, mais en appliquant des perturbations arbitraires, sans lien avec les conditions météorologiques. Une meilleure correction de cette source d'incertitude serait d'utiliser un schéma microphysique à 2 moments où N_0 serait une variable pronostique. La question est ainsi parfois posée de l'utilité du développement des méthodes de perturbation des modèles pour la prévision d'ensemble par rapport à l'amélioration des paramétrisations physiques. On peut argumenter à cela que les processus paramétrisés dans un modèle sont nombreux et complexes, interagissent ensemble et que de ce fait, améliorer les paramétrisations physiques est une tâche de longue haleine. Tant que des erreurs proviennent de ces schémas du fait de leurs approximations, l'utilisation et le développement de techniques de perturbation pour la prévision d'ensemble se justifient. Celles-ci sont néanmoins vouées à évoluer au gré de l'amélioration des modèles, comme cela est souligné dans la synthèse du groupe de travail 1 du workshop sur l'incertitude modèle du CEPMMT de mai 2022 (Plant et al. 2022). Au fur et à mesure que la résolution des modèles augmente et que des phénomènes autrefois paramétrés deviennent explicitement résolus, ou bien au fur et à mesure que les schémas de paramétrisation deviennent plus réalistes, certaines sources d'incertitude initialement importantes peuvent disparaître et laisser la place à d'autres incertitudes qui jusqu'alors n'étaient pas majoritaires. Certaines perturbations peuvent alors ne plus se justifier, tandis que le besoin de représenter d'autres sources d'erreur émerge. De façon générale, il semble qu'une coopération étroite entre le développement des paramétrisations physiques déterministes et celui des perturbations stochastiques soit désirable (Plant et al. 2022). Ceci est particulièrement marquant pour les méthodes de perturbation de type POP ou de perturbation des paramètres qui sont très dépendantes de la construction des schémas de paramétrisation physiques et de leurs hypothèses.

Dans cette thèse, on s'est attaché à étudier l'impact de différentes méthodes de perturbation sur la dispersion d'un ensemble, tout en cherchant à conserver un biais le plus faible possible par rapport à la référence du modèle déterministe non perturbé. En effet, l'objectif de la thèse n'était pas d'améliorer le modèle déterministe et ses paramétrisations physiques mais de chercher quelles perturbations peuvent avoir un impact significatif sur des simulations de diverses situations météorologiques. La littérature montre néanmoins que, au travers des non-linéarités des modèles, les perturbations stochastiques peuvent avoir un effet sur la climatologie d'un mo-

dèle et parfois contribuer à son amélioration (Berner et al. 2017 ; Sakradzija et Klocke 2018). Le parti pris de cette thèse sur ce point n'est donc pas la seule approche possible, et il pourrait être envisagé de régler les méthodes de perturbation en se basant davantage sur des observations que sur le modèle déterministe de référence.

Une des perspectives immédiates de ce travail de thèse serait d'étendre les tests 3D à toutes les perturbations testées dans le cadre uni-colonne, c'est-à-dire en incluant les perturbations de la relation CB02 ainsi que les perturbations des paramètres N_0 et ν . Par ailleurs, comme on l'a souligné dans ce manuscrit, les différentes méthodes proposées ciblent des sources d'incertitude particulières et ne sauraient donc seules représenter toutes les erreurs modèle. Elles doivent donc être envisagées en combinaison avec d'autres perturbations. Par exemple, elles pourraient être complétées par la perturbation des 8 paramètres d'AROME identifiés par Wimmer et al. (2022). Il serait néanmoins nécessaire de veiller à ce que les perturbations ne soient pas redondantes : ainsi, la perturbation du paramètre XCMF, intervenant dans le calcul de la condition de fermeture en flux de masse du schéma de convection peu profonde, ne serait probablement pas nécessaire si cette condition de fermeture était déjà perturbée par la méthode de convection stochastique. De même, la perturbation du paramètre RSWINHF affectant le calcul de l'épaisseur optique dans le rayonnement SW pourrait être redondante avec la perturbation des paramètres N_0 et ν . Par ailleurs, la perturbation de paramètres supplémentaires concernant la distribution de taille des hydrométéores pourrait être envisagée : le paramètre α de la distribution des gouttelettes nuageuses, mais également les paramètres réglant la distribution des gouttes de pluie ou des hydrométéores glacés (Thompson et al. 2021). Il faudrait pour cela réaliser de nouvelles études de sensibilité, car ces paramètres ne font pas partie de ceux sélectionnés dans l'étude de Wimmer et al. (2022).

La combinaison avec des méthodes multiplicatives de type SPPT ou iSPPT pourrait être envisagée mais semble plus délicate car ces dernières sont censées représenter tout un ensemble d'erreurs qui ne sont pas précisément identifiées. Il n'est donc pas garanti que les différentes perturbations représentent des sources d'incertitude distinctes. Néanmoins, perturber certains schémas avec une méthode multiplicative des tendances du type iSPPT, et d'autres schémas avec des méthodes POP ou de perturbation de paramètres semble envisageable. C'est ce qui a été fait dans cette thèse pour les premiers tests 3D de perturbation de la turbulence et de la convection : la méthode iSPPT a été appliquée aux schémas de paramétrisation microphysique et radiatif afin de compléter la représentation de l'incertitude modèle. Combiner une variante de iSPPT avec une méthode de perturbation de paramètres pour la turbulence a aussi récemment été examiné dans la littérature (Wastl et al. 2019a).

De façon générale, étant donnée la diversité des sources d'erreurs possibles dans un modèle, il semble que combiner des méthodes de perturbation différentes soit une piste intéressante pour améliorer la représentation de l'incertitude modèle pour la prévision d'ensemble. Cela nécessite néanmoins de réfléchir aux coûts de maintenance impliqués par une multiplication des méthodes, ainsi qu'à l'augmentation du nombre de paramètres réglables (amplitude, corrélations des termes aléatoires), qui pourrait éventuellement mener à des problèmes de sur-ajustement. Par ailleurs, toutes les méthodes présentées dans cette thèse n'ont pas une efficacité similaire pour générer de la dispersion, et cela peut varier notamment selon la situation météorologique. Il faut garder à l'esprit que certaines d'entre elles représentent une incertitude de petite échelle, n'ayant pas la capacité d'affecter significativement les prévisions pour des modèles à aire limitée tels qu'AROME, souvent très dépendants des conditions de bord et donc des performances des

systemes de prévision d'ensemble globaux. Néanmoins, utiliser des méthodes de perturbation justifiées physiquement présente l'avantage de traiter avec plus de précision les sources d'incertitude, d'avoir des perturbations mieux reliées aux processus physiques et donc une meilleure interprétabilité. Ceci pourrait être particulièrement important dans la compréhension du comportement d'un ensemble lors d'études de situations météorologiques précises.

Annexe A

Calcul de la vitesse terminale de chute des gouttelettes nuageuses

On rappelle l'équation de la vitesse terminale "moyenne" de chute des gouttelettes :

$$V_T = \frac{1}{\rho_d r_c} \int_0^\infty \frac{\pi}{6} D^3 \rho_l V(D) N(D) dD,$$

ainsi que l'équation de la vitesse de chute d'une goutte de diamètre D , donnée par la formule de Pruppacher et Klett (1978) :

$$V(D) = \frac{2 \left(\frac{D}{2}\right)^2 g (\rho_l - \rho_d)}{9\eta_a} \left(1 + 1.26\lambda_{a0} \left(\frac{P_0}{P}\right) \left(\frac{T}{T_0}\right) \frac{2}{D} \right) \quad (\text{A.1})$$

D'après cette équation et la formule du spectre de taille des gouttelettes (équation 2.24) on a :

$$\begin{aligned} V_T &= \frac{1}{\rho_d r_c} \int_0^\infty \frac{\pi}{6} D^3 \rho_l \frac{2 \left(\frac{D}{2}\right)^2 g (\rho_l - \rho_d)}{9\eta_a} \left(1 + 1.26\lambda_{a0} \left(\frac{P_0}{P}\right) \left(\frac{T}{T_0}\right) \frac{2}{D} \right) N_0 \frac{\alpha}{\Gamma(\nu)} \lambda^{\alpha\nu} D^{\alpha\nu-1} e^{-(\lambda D)^\alpha} dD \\ &= A \int_0^\infty D^3 \times \frac{D^2}{2} \left(1 + 1.26\lambda_{a0} \left(\frac{P_0}{P}\right) \left(\frac{T}{T_0}\right) \frac{2}{D} \right) D^{\alpha\nu-1} e^{-(\lambda D)^\alpha} dD \end{aligned} \quad (\text{A.2})$$

$$\text{avec } A = \frac{1}{\rho_d r_c} \frac{\pi}{6} \rho_l \frac{g (\rho_l - \rho_d)}{9\eta_a} N_0 \frac{\alpha}{\Gamma(\nu)} \lambda^{\alpha\nu}.$$

On peut alors diviser l'intégrale en deux termes :

$$V_T = A \left(\int_0^\infty \frac{1}{2} D^{5+\alpha\nu-1} e^{-(\lambda D)^\alpha} dD + \int_0^\infty B \times D^{4+\alpha\nu-1} e^{-(\lambda D)^\alpha} dD \right)$$

$$\text{avec } B = 1.26\lambda_{a0} \left(\frac{P_0}{P}\right) \left(\frac{T}{T_0}\right).$$

On note les deux intégrales I_1 et I_2 respectivement. De plus, on se place dans le cas $\alpha = 1$ qui est la valeur par défaut de ce paramètre sur continent dans AROME. On a :

$$\begin{aligned} I_1 &= \int_0^\infty \frac{1}{2} D^{5+\nu-1} e^{-\lambda D} dD \\ I_2 &= \int_0^\infty B \times D^{4+\nu-1} e^{-\lambda D} dD \end{aligned}$$

En appliquant $(5 + \nu - 1)$ fois une intégration par partie pour I_1 et $(4 + \nu - 1)$ fois pour I_2 on obtient :

$$\begin{aligned} I_1 &= \frac{1}{2}(-1)^{5+\nu-1}(5 + \nu - 1)! \left(-\frac{1}{\lambda}\right)^{5+\nu-1} \times \left(0 + \frac{1}{\lambda}\right) \\ &= \frac{1}{2}(5 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{5+\nu} \\ I_2 &= B(-1)^{4+\nu-1}(4 + \nu - 1)! \left(-\frac{1}{\lambda}\right)^{4+\nu-1} \times \left(0 + \frac{1}{\lambda}\right) \\ &= B(4 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{4+\nu} \end{aligned}$$

On en déduit alors la vitesse terminale V_T .

$$\begin{aligned} V_T &= A(I_1 + I_2) \\ &= A \left(\frac{1}{2}(5 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{5+\nu} + B(4 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{4+\nu} \right) \\ &= A(5 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{5+\nu} \frac{1}{2} \left(1 + 2B \frac{\lambda}{5 + \nu - 1} \right) \\ &= \frac{1}{\rho_d r_c} \frac{\pi}{6} \rho_l \frac{g(\rho_l - \rho_d)}{9\eta_a} N_0 \frac{1}{\Gamma(\nu)} \lambda^\nu (5 + \nu - 1)! \left(\frac{1}{\lambda}\right)^{5+\nu} \frac{1}{2} \left(1 + 2B \frac{\lambda}{5 + \nu - 1} \right) \\ &= \left(\frac{\pi}{6} \rho_l \frac{N_0}{\rho_d r_c} \frac{\Gamma(\nu + 3)}{\Gamma(\nu)} \right) \frac{(5 + \nu - 1)!}{\Gamma(\nu + 3)} \frac{g(\rho_l - \rho_d)}{9\eta_a} \lambda^\nu \left(\frac{1}{\lambda}\right)^{5+\nu} \frac{1}{2} \left(1 + 2B \frac{\lambda}{5 + \nu - 1} \right) \end{aligned}$$

On rappelle que λ est défini (pour $\alpha = 1$) par :

$$\lambda = \left(\frac{\pi}{6} \rho_l \frac{\Gamma(\nu + 3)}{\Gamma(\nu)} \frac{N_0}{\rho_d r_c} \right)^{1/3}$$

Donc on a :

$$\begin{aligned} V_T &= \lambda^3 \frac{\Gamma(5 + \nu)}{\Gamma(\nu + 3)} \frac{g(\rho_l - \rho_d)}{9\eta_a} \lambda^\nu \left(\frac{1}{\lambda}\right)^{5+\nu} \frac{1}{2} \left(1 + 2B \frac{\lambda}{5 + \nu - 1} \right) \\ &= \frac{\Gamma(\nu + 5)}{\Gamma(\nu + 3)} \frac{g(\rho_l - \rho_d)}{9\eta_a} \lambda^{3+\nu-5-\nu} \frac{1}{2} \left(1 + 2B \frac{\lambda}{5 + \nu - 1} \right) \end{aligned}$$

Soit

$$V_T = \frac{\Gamma(\nu + 5)}{\Gamma(\nu + 3)} \frac{g(\rho_l - \rho_d)}{9\eta_a} \lambda^{-2} \frac{1}{2} \left(1 + 1.26\lambda_{a0} \left(\frac{P_0}{P}\right) \left(\frac{T}{T_0}\right) \frac{2\lambda}{5 + \nu - 1} \right) \quad (\text{A.3})$$

L'équation A.3 est une fonction décroissante du paramètre de forme ν . Elle est représentée sur la figure A.1, pour $N_0 = 50$ gouttes.cm⁻³, $r_c = 10^{-4}$ kg.kg⁻¹, $T_0 = 293.15$ K, $T = 270$ K, $P_0 = 1013.25$ hPa, $P = 1013$ hPa.

Ceci peut sembler contre-intuitif puisque lorsque ν augmente, le diamètre moyen des gouttes augmente. Néanmoins, la figure 5.6 montre aussi que la queue de distribution à droite diminue : il y a moins de très grosses gouttes. Or les grosses gouttes ont plus d'impact que les petites sur la vitesse moyenne, car leur vitesse de chute est plus élevée, ce qu'illustre la figure A.2, tracée d'après l'équation A.1. Ainsi, la vitesse moyenne de chute des gouttes diminue lorsque ν augmente.

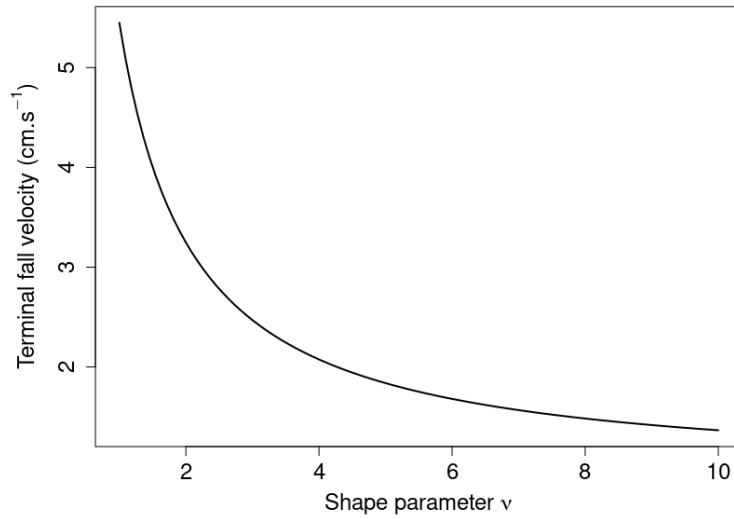


FIGURE A.1 – Vitesse terminale "moyenne" de chute des gouttelettes selon la valeur du paramètre de forme ν .

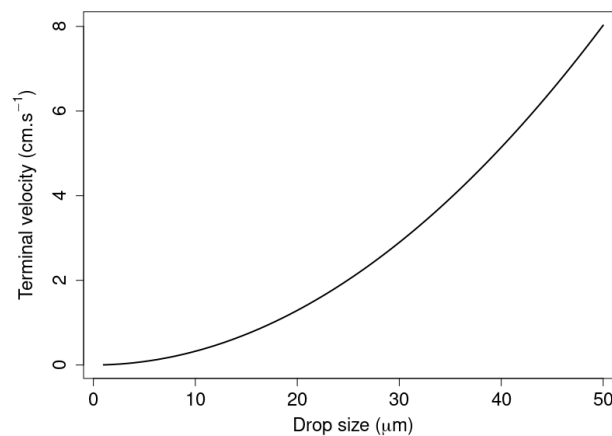


FIGURE A.2 – Vitesse de chute d'une gouttelette nuageuse en fonction de son diamètre.

Annexe B

Calcul du rayon effectif à partir de la distribution de taille des gouttelettes nuageuses

La densité de probabilité d'une variable aléatoire réelle X suivant une loi gamma généralisée est

$$f(x) = \frac{\alpha}{\Gamma(\nu)} \lambda^{\alpha\nu} x^{\alpha\nu-1} e^{-(\lambda x)^\alpha},$$

où α , ν et λ sont des constantes et où $\Gamma(\cdot)$ désigne la fonction gamma.

Le moment d'ordre m de X est

$$E(X^m) = \int_{X \in \mathbb{R}} x^m f(x) dx = \frac{1}{\lambda^m} \frac{\Gamma(\nu + m/\alpha)}{\Gamma(\nu)}. \quad (\text{B.1})$$

Dans le schéma microphysique ICE3 d'AROME, il est fait l'hypothèse que la distribution de taille des gouttelettes d'eau nuageuses est de type gamma généralisée :

$$n(r) = N_0 \frac{\alpha}{\Gamma(\nu)} \lambda^{\alpha\nu} r^{\alpha\nu-1} e^{-(\lambda r)^\alpha},$$

où r est le rayon des gouttelettes et N_0 est le nombre total de gouttelettes nuageuses par unité de volume.

Le rayon effectif est défini comme étant le rapport entre les troisième et deuxième moments de la distribution de taille des gouttelettes (Stephens 1978) :

$$r_{\text{eff}} = \frac{\int_0^\infty r^3 n(r) dr}{\int_0^\infty r^2 n(r) dr}$$

D'après l'équation B.1 on a donc

$$\begin{aligned} r_{\text{eff}} &= \frac{1}{\lambda^3} \frac{\Gamma(\nu + 3/\alpha)}{\Gamma(\nu)} \lambda^2 \frac{\Gamma(\nu)}{\Gamma(\nu + 2/\alpha)} \\ &= \frac{1}{\lambda} \frac{\Gamma(\nu + 3/\alpha)}{\Gamma(\nu + 2/\alpha)} \end{aligned} \quad (\text{B.2})$$

Le coefficient λ est réglé dans AROME de façon à respecter le contenu en eau nuageuse présent dans la maille. Si ρ_d est la masse volumique de l'air sec, ρ_l la masse volumique de l'eau liquide et r_c le rapport de mélange de l'eau liquide nuageuse dans la maille à un niveau vertical donné alors on doit avoir

$$\begin{aligned}\rho_d r_c &= \int_0^\infty \frac{4}{3} \pi r^3 \rho_l n(r) dr \\ &= \frac{4}{3} \pi \rho_l N_0 \int_0^\infty r^3 \frac{\alpha}{\Gamma(\nu)} \lambda^{\alpha\nu} r^{\alpha\nu-1} e^{-(\lambda r)^\alpha} dr.\end{aligned}$$

D'après la relation B.1 :

$$\rho_d r_c = \frac{4}{3} \pi \rho_l N_0 \frac{1}{\lambda^3} \frac{\Gamma(\nu + 3/\alpha)}{\Gamma(\nu)},$$

d'où

$$\lambda = \left(\frac{4\pi}{3} \rho_l \frac{N_0}{r_c \rho_d} \frac{\Gamma(\nu + 3/\alpha)}{\Gamma(\nu)} \right)^{1/3} \quad (\text{B.3})$$

Donc en remplaçant λ dans l'équation B.2 par B.3 on obtient :

$$\begin{aligned}r_{\text{eff}} &= \left(\frac{3r_c \rho_d}{4\pi \rho_l N_0} \frac{\Gamma(\nu)}{\Gamma(\nu + 3/\alpha)} \right)^{1/3} \frac{\Gamma(\nu + 3/\alpha)}{\Gamma(\nu + 2/\alpha)} \\ &= \left(\frac{3r_c \rho_d}{4\pi \rho_l N_0} \frac{\Gamma(\nu)}{\Gamma(\nu + 3/\alpha)} \frac{\Gamma(\nu + 3/\alpha)^3}{\Gamma(\nu + 2/\alpha)^3} \right)^{1/3} \\ &= \left(\frac{3r_c \rho_d}{4\pi \rho_l N_0} \frac{\Gamma(\nu) \Gamma(\nu + 3/\alpha)^2}{\Gamma(\nu + 2/\alpha)^3} \right)^{1/3}\end{aligned} \quad (\text{B.4})$$

Avec l'équation B.4 on retrouve bien une expression similaire à celle de Martin et al. (1994) :

$$r_{\text{eff}} = \left(\frac{3L}{4\pi \rho_l N_0 k} \right)^{1/3}.$$

L est le contenu en eau nuageuse défini dans Martin et al. (1994) par $L = \frac{4}{3} \pi \rho_l \int_0^\infty r^3 n(r) dr$ donc égal à $\rho_d r_c$ dans notre cas.

Les deux expressions sont identiques en définissant le coefficient k de la façon suivante :

$$k = \frac{\Gamma(\nu + 2/\alpha)^3}{\Gamma(\nu) \Gamma(\nu + 3/\alpha)^2} \quad (\text{B.5})$$

Si $\alpha = 1$, l'expression se réduit à

$$\begin{aligned}k &= \frac{\Gamma(\nu + 2)^3}{\Gamma(\nu) \Gamma(\nu + 3)^2} = \frac{\Gamma(\nu + 2)^3}{\Gamma(\nu) (\nu + 2)^2 \Gamma(\nu + 2)^2} \\ &= \frac{\Gamma(\nu + 2)}{\Gamma(\nu) (\nu + 2)^2} = \frac{(\nu + 1)\nu}{(\nu + 2)^2} = \frac{(\nu^2 + \nu)}{(\nu + 2)^2},\end{aligned}$$

qui est bien l'expression utilisée par Jahangir et al. (2021).

Annexe C

Comparaison de trois approches de représentation de l'erreur modèle dans des simulations uni-colonne de couche limite

Les pages suivantes sont un article intitulé "A Single-Column Comparison of Model Error Representations for Ensemble Prediction", publié dans la revue *Boundary-Layer Meteorology*. Cet article étudie l'impact de trois méthodes de perturbation sur des simulations uni-colonne de trois cas idéalisés de couche limite. Les méthodes de perturbation visent toutes à représenter l'incertitude provenant des paramétrisations physiques du modèle. Il s'agit de la méthode SPPT, sa variante iSPPT et d'une méthode de perturbation de paramètres, désignée "RP" dans l'article.

A Single-Column Comparison of Model Error Representations for Ensemble Prediction

François Bouttier, Axelle Fleury, Thierry Bergot, Sébastien Riette

10 January 2022

affiliation: CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

corresponding author: François Bouttier, CNRM/GMME/PRECIP Météo-France 42 Av. Coriolis F-31057 Toulouse cedex, France. Email: francois.bouttier@meteo.fr

Orcid identifier: François Bouttier, 0000-0001-6148-4510.

Funding information: Météo-France and CNRS.

This is an author's version of a peer-reviewed article. It is hereby distributed under Creative Commons Attribution Licence CC-BY-NC as of 25 July 2022, in accordance with French law regarding Government funded research (loi du 7 octobre 2016 pour une République Numérique).

It is available :

- in the free HAL repository at <https://hal.archives-ouvertes.fr/hal-03806323>
- as a Boundary Layer Meteorology journal publication typeset by the Editor at the following DOI (submitted on 4 May 2021, accepted on 6 Dec 2021, published online on 24 January 2022). pay-walled version: <https://www.doi.org/10.1007/s10546-021-00682-6>, free shareable version: <https://rdcu.be/cFD1G>

Cite as: Bouttier, F., A. Fleury, T. Bergot and S. Riette, 2022: A Single-Column Comparison of Model Error Representations for Ensemble Prediction. *Bound. Layer Meteor.*, **183**, 167-197. doi:10.1007/s10546-021-00682-6

Abstract

Various perturbation schemes have been proposed for representing model error in convection-permitting ensemble prediction. Their evaluation usually relies on time-averaged ensemble prediction statistics and on complex case studies. In this work, their detailed physical behaviour is studied in order to understand their differences, and to help their optimization. A process-level intercomparison framework is used to investigate the widely used SPPT (stochastic perturbations of physics tendencies), independent SPPT (iSPPT), and random parameters (RP) model perturbation schemes. Ensemble predictions with the single-column version of the Arome numerical weather prediction model are evaluated on three different boundary-layer regimes: cumulus convection, stratocumulus-topped boundary layer, and radiation fog. The iSPPT scheme is found to produce more dispersion than the SPPT scheme, particularly when several physics parametrizations are in near equilibrium. It also appears to be more numerically stable near the surface. The RP scheme perturbations are structurally very different from the other schemes, particularly regarding cloud structure. The iSPPT and RP ensembles have very different sensitivities to the atmospheric conditions, which suggests that intercomparisons of ensemble model error schemes should carefully account for situation dependency. Substantial forecast biases are produced by the RP scheme with respect to the unperturbed model. These results suggest that the iSPPT scheme can bring major improvements over the SPPT approach with minimal effort, that there is some complementarity between the iSPPT and RP approaches, but that implementing RP-type schemes in operational applications may require some careful tuning to avoid creating forecast biases.

keywords: atmospheric ensemble prediction, model error representation, numerical weather prediction, single-column model, stochastic perturbations.

1 Introduction

Ensemble prediction is a standard technique used in numerical weather prediction centres to diagnose forecast uncertainty (Palmer, 2001). In this approach, forecast error probability distributions are simulated by perturbing known error sources, which includes initial, boundary condition error, and numerical model uncertainty. The perturbations are applied to a set of model runs, called ensemble members. Their output is a discrete sample of the distribution of possible forecasts, given the information available at forecast initiation time. Ensemble distributions can be presented as diagnostics of forecast uncertainties, such as the probability of occurrence of weather events (Fundel et al., 2019). Ensemble forecasts can be sensitive to their representation of model errors, for which several approaches have been proposed.

The focus of this work is on ensemble prediction in convection permitting models, i.e., models with a horizontal resolution of the order of a few kilometres and over a few days. On these scales, modelling error arises because of limitations to the numerical model resolution and approximations in the evolution equations. This uncertainty source can be represented by including several models in an ensemble prediction: some weather centres have built ensemble prediction systems from models developed by multiple institutes, which is the multi-model method (e.g., Hagedorn et al., 2005; Garcia-Moya et al., 2011). Alternatively, a single-model framework can be used with multiple dynamical cores and/or physical parametrizations, which is the mixed physics method (e.g., Clark et al., 2008; Duda et al., 2014; Gallus et al., 2019). These approaches are useful, but can be complex to use operationally because they require scientific and technical manpower to maintain model software developed by various teams. The implications of using members of unequal quality and varying forecast bias are not yet fully understood.

Because numerical weather prediction models tend to share many modelling principles and simplifying assumptions, it is not necessarily clear that the dispersion of a multimodel or mixed physics ensemble represents all significant sources of modelling errors.

Given these complications, weather prediction centres increasingly focus on model error representation by perturbing a single deterministic model. Some have constructed model perturbations through analysis of known physical processes, such as the upscale energy transfer of subgrid errors in the stochastic kinetic energy backscatter (SKEB) approach (Shutts, 2005; Berner et al., 2015), or partly resolved fluctuations near the model grid resolution from turbulent eddy mixing (Kober and Craig, 2016) or the triggering of shallow convection (Sakradzija and Klocke, 2018). Other methods treat the atmospheric model timestep as a black box with empirical perturbations applied either to its input settings (the parameter perturbation methods), or to its output (the stochastic tendency perturbation methods).

Stochastic tendency perturbation methods are widely used in the community (Buizza et al., 1999; Hou et al., 2006; Berner et al., 2015, among others). In this work we study the SPPT (stochastic perturbations of physics tendencies) approach of Bouttier et al. (2012), which is related to the ECMWF (European centre for medium-range weather forecasts) approach of Palmer et al. (2009). The SPPT scheme applies bulk perturbations to physics tendencies using random multiplicative noise. It is currently used at Météo-France in the AROME-France-EPS (applications of research to operations at mesoscale) ensemble prediction system and in other centres. Its design can be partly justified using idealized studies (e.g., Shutts and Pallarès, 2014), its success mostly comes from its ease of implementation. Its impact on forecast quality has been documented in various settings, including convection-permitting ensembles (Bouttier et al., 2012; Romine et al., 2014; Zhang, 2018; Jankov et al., 2019), global ensembles, and seasonal prediction systems.

More recent studies have shown that the SPPT scheme can be improved by relaxing some of its design assumptions, such as the correlation of perturbations applied to different physical parametrizations. This variant of the SPPT scheme, called 'independent SPPT' or iSPPT (Christensen et al., 2017), allows tuning the perturbation characteristics separately for different parametrizations. It has been proposed that the iSPPT scheme benefits from improving the interaction between physical processes. Other generalizations of the SPPT scheme have been suggested, such as decorrelating the perturbations between model variables (Christensen et al., 2017; Wastl et al., 2019a).

Parameter perturbation methods have taken various forms, including fixed parameters for each ensemble member (e.g., Bowler et al., 2008; Hacker et al., 2011), time-evolving parameters (Baker et al., 2014; McCabe et al., 2016), and parameters that simultaneously vary in space and time (Ollinaho et al., 2017; Lang et al., 2021). We collectively refer to these methods as RP since their key ingredient is the random sampling of model tunings. RP methods are a form of mixed-physics ensemble perturbations, because they assume that the model uncertainty can be explored by changing a small set of model constants; this is clearly justified for parameters tuned by imperfect calibration approaches. Random parameters can usually be linked to precise processes, and their perturbations act through parametrizations designed by physics experts: it suggests that the RP perturbations should be physically sound. By design, RP methods cannot represent physical uncertainties outside the subspace represented by the parametrization equations, which is a limitation.

Both SPPT and RP-type perturbations methods have their pros and cons, and the community is investing much effort in evaluating them, to understand their complementarities, and ultimately to combine them (Berner et al., 2015; Jankov et al., 2019; Lupo et al., 2020). Similar questions arise in the regional, global, and climate modelling communities. A fundamental problem of evaluating model perturbation

schemes is that model errors are difficult to diagnose. To date, most validation studies have relied either on probabilistic scores in long ensemble experiments, or on case studies with complex numerical models. In either approach, in three-dimensional experiments the impact of model perturbations interacts with other ensemble perturbations (initial condition, large-scale forcing, surface conditions) or between multiple processes. Thus, it is usually difficult to link the behaviour of ensembles to specific ingredients of the model perturbation scheme. The impact of model perturbations may be wrongly interpreted because they interact with other sources of ensemble dispersion that have their own flaws, and because they mix different atmospheric conditions in which the impact of model errors may not be comparable.

Model error schemes can also be investigated using coarse-graining studies (e.g., [Shutts and Pallarès, 2014](#); [Christensen, 2020](#)), which compare high- and low-resolution models. Such studies can be very useful, but complex and numerically expensive to perform. They tend to only address some aspects of model error because their results are intrinsically statistical. They usually assume that the higher-resolution model used is perfect. To date, coarse-graining studies have tended to focus on large-scale aspects of deep tropical convection and its variability, which is critical for seasonal and climate predictions (e.g., [Davini et al., 2017](#); [Christensen, 2020](#)). In convection-permitting mid-latitude short-range predictions, users are concerned with other weather processes, too, such as cloud and fog processes and their connection to boundary-layer physics.

In this work we evaluate ensemble model perturbations at the kilometeric scale, by intercomparing perturbation schemes in an idealized single-column framework. A state of the art numerical weather prediction model is used. The field experiment atmospheric cases used recognized in the meteorological community as representative of key modelling issues. Although the single-column geometry lacks a simulation of three-dimensional effects, it allows precise investigation of interactions between physical processes, because the parametrizations used act on the model columns. This gives full access to model variables and physics tendencies involved in the propagation of ensemble perturbations, from the model error schemes to the forecasted atmospheric profiles. The aim is to fully understand the mechanisms and physical implications of different model error perturbation approaches.

The central question of this work is : what are the physical implications of injecting model perturbations in ensemble prediction, and more specifically, how do the SPPT, iSPPT, and RP schemes compare in different weather conditions, can we identify basic issues for their use in ensemble predictions. There is no verification truth in this work, so we do not expect to conclude that one scheme is better than another, apart from the possible identification of unphysical behaviour. The goal of this work is to inform further improvement to these schemes, in particular the future design of hybrid schemes that would combine the SPPT, iSPPT, and RP schemes: previous studies ([Berner et al., 2015](#); [Ollinaho et al., 2017](#)) have shown that these schemes have very different impacts on ensemble performance. Their results suggest that they address different aspects of model error, so they could probably be combined in a constructive way, but the method to define such a combination is not yet clear and arguably requires a deep understanding of the behaviour of each scheme.

This paper is organized as follows: the numerical model and the ensemble perturbation schemes are explained in Sect. 2. Sections 3, 4, and 5 describe the impact of ensemble perturbations in three different atmospheric regimes, namely, the ARMCU (atmospheric radiation measurement cumulus), FIRE (first ISLSCP regional experiment) and LANFEX (local and nonlocal fog field experiment) cases, which are described in their respective sections. The results are discussed before the conclusion in the final sections.

2 Experimental design

2.1 *The single column model*

The prediction model used is a single-column model (SCM) version of the AROME numerical weather prediction system (Seity et al., 2011), with 1.3-km horizontal resolution. The focus is on the diurnal evolution of boundary-layer processes, with maximum forecast ranges between 12 and 30 h depending on the test case. The full AROME model uses the physical parametrizations listed below, which we shall identify in the text by the following nicknames:

turbulence is the Cuxart et al. (2000) one-dimensional (vertical) turbulent mixing parametrization, which relies on a prognostic TKE (subgrid turbulence kinetic energy) field to predict the mixing coefficients;

convection is the subgrid shallow convection parametrization of Pergaud et al. (2009), which models the effect of subgrid updrafts and downdrafts through a pseudo-historical mass-flux parametrization. It is designed to represent the net effect of dry thermals and non-precipitating clouds.

microphysics is the ICE3 one-moment bulk microphysics parametrization based on Cohard et al. (2000) which models the evolution of 5 prognostic condensed water species (cloud water and ice, rain, snow and graupel);

radiation is a simplified version of the ECMWF radiation parametrization (Fouquart and Bonnel, 1980; Mlawer, 1997), with climatological aerosols settings;

surface is the prognostic surface parametrization documented in Masson et al. (2013).

Each test case uses model configurations simplified as follows:

- the ARMCU case does not use the radiation, microphysics or surface parametrizations, surface effects are represented by prescribed surface fluxes of latent and specific heat
- the FIRE and LANFEX cases use all parametrizations, but the surface temperature is prescribed.

The cases were selected because they are known to be dominated by vertical interactions between physical processes in the boundary layer (references are provided in the depiction of each case in the following sections).

A more precise documentation of the AROME physics is provided in Termonia et al. (2018). We use the same 90-level vertical grid as in the Météo-France operational system in 2021: its lowest computational levels are at heights 5, 16, 32, 51 m above ground, its vertical resolution is approximately 10, 30, 80, 120 m at respective heights of 5, 100, 1000, 2000 m, respectively. Unless otherwise mentioned, ‘wind’ denotes the horizontal wind vector and its speed is U , near-surface variations of dry potential temperature θ are identified with temperature variations, relative humidity is computed with respect to liquid water saturation, and Q_v and Q_l are the water vapour and liquid water mixing ratios, respectively.

2.2 *The perturbation schemes*

Three model perturbation algorithms, called perturbation schemes, are studied in this paper. Each one defines an 80-member ensemble prediction that uses the same initial, surface, and boundary conditions, so that ensemble dispersion is completely determined by the model perturbation scheme. The reference simulation is a single deterministic (i.e., unperturbed) simulation. Each ensemble is named according to the model perturbation scheme used: SPPT, iSPPT or RP, as defined below. In the following, the ensembles are primarily compared using the median (indicated by Q50 in the plots) and a measure of ensemble dispersion, called spread, which is the interval between the 10% and 90% ensemble percentiles (indicated as Q10–Q90 in the plot legends).

This intercomparison study will also consider the bias, which is the difference between the median of two ensembles, or between an ensemble median and the deterministic simulation: as will be discussed in the following, a non-negligible bias (relative to the ensemble spread) is a sign of a systematic, physically significant change in ensemble behaviour.

The SPPT (stochastic perturbations of physical tendencies) scheme aims to represent model uncertainties linked to atmospheric physics, by assuming that tendency errors are proportional to their absolute amplitude. The SPPT scheme works by multiplying the total physics tendencies of horizontal wind, T , and Q_v by space- and time-correlated random numbers. Here, the SPPT scheme uses the configuration of the Météo-France AROME-France-EPS ensemble prediction system (Bouttier et al., 2012), with a random number decorrelation time scale of 6 h and a standard deviation of 0.3, except that the so-called boundary-layer tapering (a damping of the perturbations near the model surface) is switched off in order to study the SPPT impact at low levels. The impact of the tapering will be checked using the ARMCU case (Sect. 3).

The iSPPT scheme is a generalization of SPPT that allows different physical processes to be perturbed by different random numbers, following the idea that different physical parametrizations of the model may not produce errors at the same times and with the same amplitudes. The iSPPT scheme applies uncorrelated SPPT-like perturbations to the output tendencies from each parametrization. The iSPPT scheme has been shown to alleviate some weaknesses of SPPT, such as its tendency to overperturb clear-sky radiative tendencies (Christensen et al., 2017), and to reduce numerical stability problems in the boundary layer (Wastl et al., 2019a). Lupo et al. (2020) have compared different iSPPT versions to understand the processes that create ensemble spread in high-precipitation events.

Here, the iSPPT perturbations have the same characteristics as in SPPT (including the perturbation standard deviation), but they are applied to each active atmospheric parametrization. The parametrizations considered in this work are: radiation, turbulence, convection, and microphysics. Physical parametrizations in the AROME timestep are called in parallel and they compute tendencies as a function of the same model state. An exception is the saturation adjustment operator, which is performed at the beginning of the timestep: it is not perturbed in this work because the other parametrizations are designed to work on an adjusted model state. Perturbing the saturation adjustment in the iSPPT scheme can be important for predicting heavy precipitation (Lupo et al., 2020), this issue will be the topic of a future study. Because the random numbers used for each parametrization are mutually uncorrelated, the total tendency perturbation variance from iSPPT at each timestep is the sum of perturbation variances from each perturbed parametrization. We shall demonstrate that, due to interactions between processes during model integration, the iSPPT ensemble spread is not a simple sum of the spreads from each parametrization.

The RP perturbation scheme was originally used with constant parameter perturbations (Hacker et al.,

Table 1: Parameters perturbed in the RP (random parameters) scheme: AROME software variable name, (operational) values used when RP is not active, uniform distribution perturbation interval in the RP scheme, name of relevant parametrization, and summarized physical role of the parameter.

variable name	unperturbed value	perturbation interval	parametrization and impacted process
XCMF	0.065	[0.05, 0.08]	convection: updraft mass flux closure
XABUO	1	[0.8, 1.2]	convection: updraft buoyancy
XBDETR	0	[0, 0.4]	convection: updraft detrainment
XENTR_DRY	0.55	[0.4, 0.7]	convection: entrainment into cloud environment
XCED	0.85	[0.7 – 0.85]	turbulence: TKE dissipation
XCTP	4.65	[3.25 – 4.65]	turbulence: pressure impact on (T, Q_v) correlation
RSWINHF	1	[0.6, 1]	radiation: shortwave cloud inhomogeneity
RLWINHF	1	[0.6, 1]	radiation: longwave cloud inhomogeneity
VSIGQSAT	0.02	[0, 0.1]	microphysics: subgrid condensation
RCRIAUTI	2×10^{-4}	$[2 \times 10^{-5}, 2.3 \times 10^{-4}]$	microphysics: ice autoconversion
RCRIAUTC	10×10^{-4}	$[4 \times 10^{-4}, 10 \times 10^{-4}]$	microphysics: liquid water autoconversion

2011). Although it has been demonstrated that time-varying perturbations can be preferable to constant parameter perturbations (e.g., Christensen et al., 2015; McCabe et al., 2016; Christensen et al., 2017; Ollinaho et al., 2017), we keep parameter perturbations constant during model integration because the short timespan of the simulations performed here do not warrant the complexity of evolving parameters. We perturb 11 physics-tuning parameters using independent, uniformly distributed random numbers. The parameters and their perturbation intervals are documented in Table 1. Our RP settings have been chosen to be reasonably consistent with the ones already published in related work (Bowler et al., 2008; Gebhardt et al., 2011; Hacker et al., 2011; Baker et al., 2014; McCabe et al., 2016; Ollinaho et al., 2017; Wastl et al., 2019b) that is generally based on suggestions by physics experts.

3 Results on the ARMCU cumulus case

The ARMCU case (Brown et al., 2002) is characterized by the diurnal development of a convective boundary layer (CBL) forced by surface heating over a land surface. In the afternoon, the top of the CBL reaches saturation, leading to the formation of cumulus clouds that dissipate later in the evening. The model is forced by evolving lateral boundary conditions for temperature, humidity, steady geostrophic wind forcing, and by prescribed time-dependent surface fluxes of sensible and latent heat with a Bowen ratio of about 0.3. As explained in Brown et al. (2002) the simulation of this case is not very sensitive to the lateral boundary conditions, because the boundary-layer evolution is driven by the surface buoyancy fluxes. The only active physical parametrizations are the turbulence and convection parametrizations. This model set-up does not include as many processes as in the other cases (FIRE and LANFEX), because it does not have radiation or microphysics and the surface fluxes are prescribed. Thus, its response to ensemble perturbations may be different from a model where more processes are allowed to interact.

With this model set-up, the atmosphere is initially dry and stably stratified, and a well-mixed CBL rises during morning up to about 1000 m, after which it is topped by a cumulus-like cloud that grows up to 300 m thick until it dissipates 5 h later. The simulation is summarized in Fig. 1 by the evolution of vertical profiles of θ , Q_v , and Q_l . The daytime warming of the CBL is driven by a mix of turbulence and convection that somewhat balance each other, except in the upper CBL where convection is the dominant process. There are some Q_l oscillations (Fig. 1c) near the cloud top during early stages of the cloud development; this is an artifact of the limited vertical resolution of the model. Nevertheless, the AROME SCM simulation is quite consistent with published observational and modelling studies of the case (e.g., Brown et al., 2002).

3.1 Impact of the SPPT scheme on ARMCU

The ensemble dispersion produced by the SPPT perturbations is represented in Figs. 2 and 3. The dispersion of U (not plotted) is very small. Temperature spread (Fig. 2a and 3a) is about 1 K, evenly distributed in the CBL, and negatively correlated with a boundary-layer height spread of about 200 m at midday. The Q_v spread (Figs. 2b) is 0.4 g kg^{-1} in the CBL, and larger (2 g kg^{-1}) in the cloud layer. Temperature and humidity distributions are centred around their deterministic values in the CBL, but SPPT creates a moist Q_v bias with respect to the deterministic model, which is visible above 2000 m in Fig. 2b. This bias is linked to an asymmetry in the SPPT cloud perturbations as illustrated by Q_l profiles (Fig. 2c): SPPT strongly perturbs the cloud-top height, whereas cloud-base perturbations are much smaller. The diagnosed cloud-cover fraction (Fig. 2d) has large spread and some members dissipate the

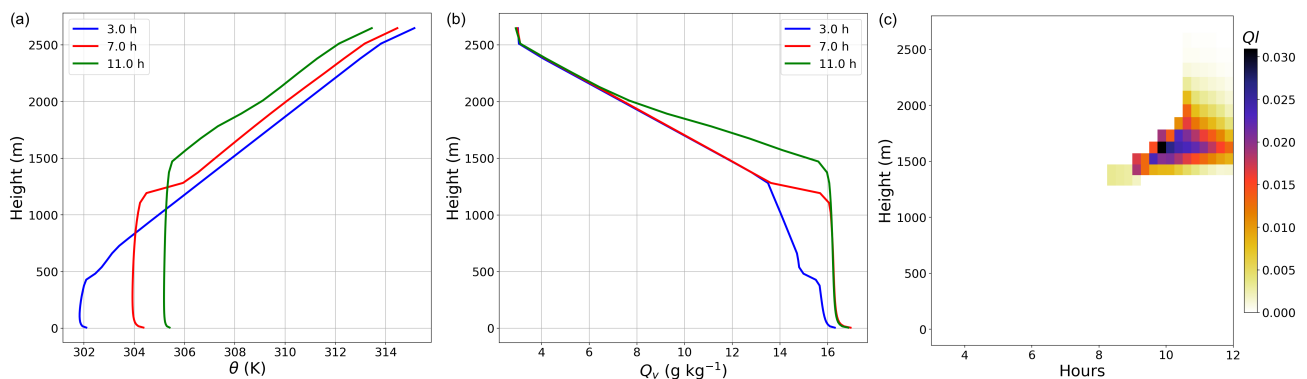


Figure 1: In the ARMCU case, evolution of the AROME vertical profiles of (a) θ and (b) Q_v in the lowest 2500 m after 3, 7, and 11 hours of simulation. (c) time–height distribution of Q_l .

cloud. The cloud onset time has a spread of about 2 h. The TKE spread (Fig. 4a) increases linearly with height from surface to cloud base. Its amplitude relative to the deterministic run (not shown) is about 20% in the CBL and 100% at the CBL top and inside the cloud. The vertical SPPT instantaneous tendency profiles exhibit numerical instabilities near the ground, as will be illustrated in the next section.

In summary, the SPPT ensemble dispersion for humidity is largest at the CBL top and in the cloud, which leads to large cloud perturbations, and spread is smaller near the ground. The situation is reversed for temperature which has its maximum dispersion inside the CBL.

The physical tendencies (Fig. 4b–f) illustrate how SPPT produces ensemble dispersion. Only temperature tendencies are shown, noting that they are highly coupled with humidity tendencies in the cloud layer through latent heat from condensation and evaporation. Turbulence transfers the surface heating to the low-level air, and convection transfers it to the top of the boundary layer. Near the surface, convection and turbulence tendencies are in near equilibrium, so the total instantaneous tendency is relatively small (about 5 K per day, Fig. 4c) and the SPPT scheme does not perturb it much. This aspect of the SPPT ensemble prediction is unrealistic, because diurnal boundary-layer development can be sensitive to surface fluxes, which are often imperfectly known (Barthlott and Kalthoff, 2011). Our ARMCU experimental set-up only perturbs atmospheric tendencies, not the surface fluxes. Representing their uncertainty would require perturbing surface fields directly or through soil moisture as in Bouttier et al. (2016). Figure 4c,d shows numerical instabilities triggered by SPPT, which will be further addressed in the next section.

In the upper part of the CBL and in the cloud (above 1000 m in Fig. 4), the convection parametrization is not balanced by turbulence, so that SPPT creates significant instantaneous tendency spread of both temperature and humidity, which enhances model state spread (Fig. 4c). Convective tendencies are maximum at the cloud top, as shown by the spike in Fig. 4f : there, moistening and cooling occur through detrainment of the cloud updraft. Perturbations to this process lead to ensemble dispersion of Q_v , Q_l , and cloud-top height (Fig. 2b and c).

An additional experiment was carried out with the same set-up, except that the tapering (damping of the perturbations near the model surface) of the SPPT scheme was switched on, as in the operational AROME-France-EPS system. Doing so reduces the spread by about half (not shown), in terms of the near-surface variables. It also suppresses the numerical instabilities mentioned above. The tapering does not change the other aspects of ensemble behaviour mentioned in this study.

In conclusion on this case, most of the SPPT-induced spread comes from convective tendency per-

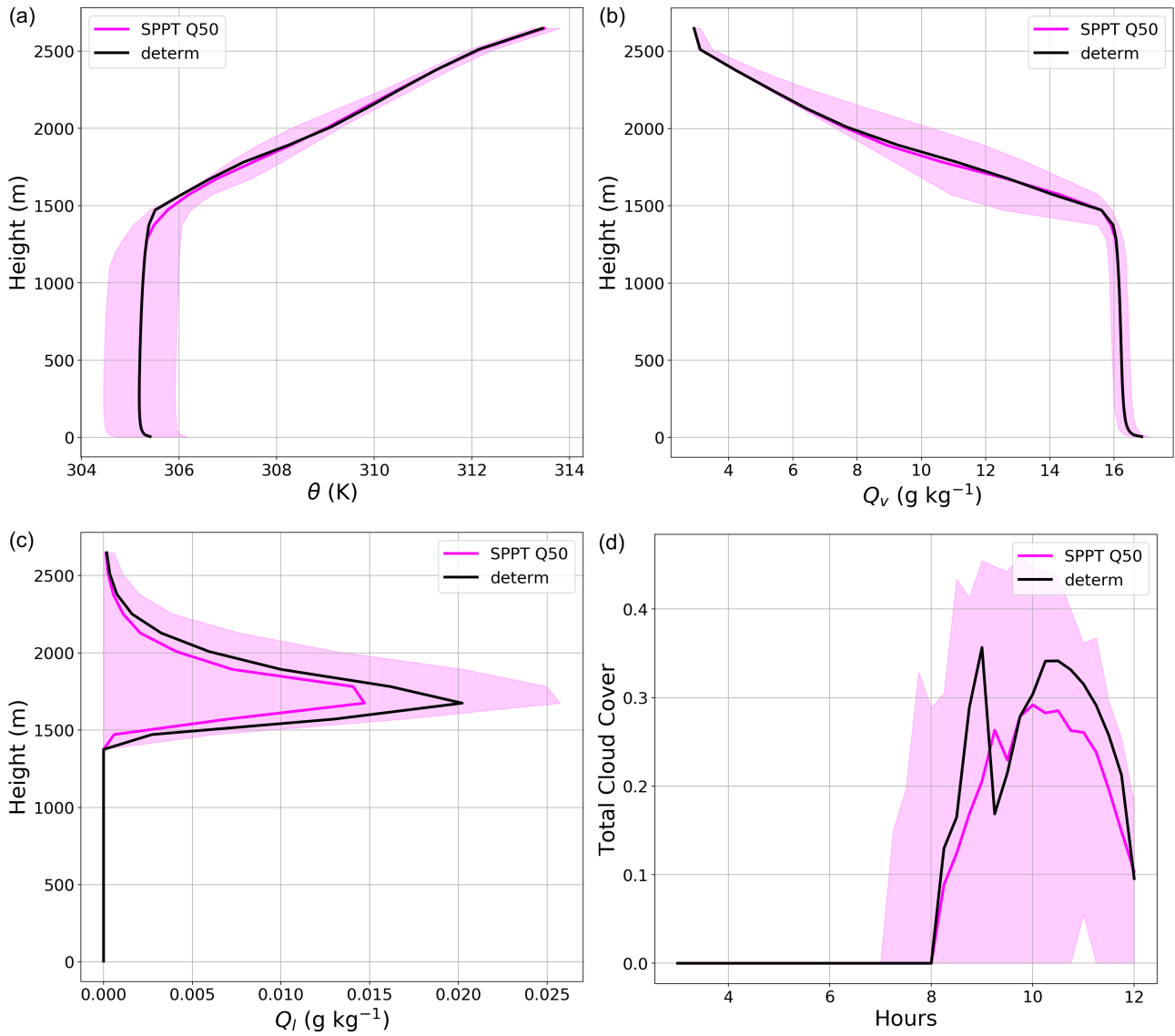


Figure 2: Comparison between the ARMCU SPPT ensemble (magenta) and the deterministic simulation (black): vertical profiles after 11 hours for (a) θ , (b) Q_v , (c) Q_l , and (d) time-series of total cloud cover fraction. The ensemble distribution is represented by its median (dark magenta curves) and the Q10–Q90 percentile interval (shaded).

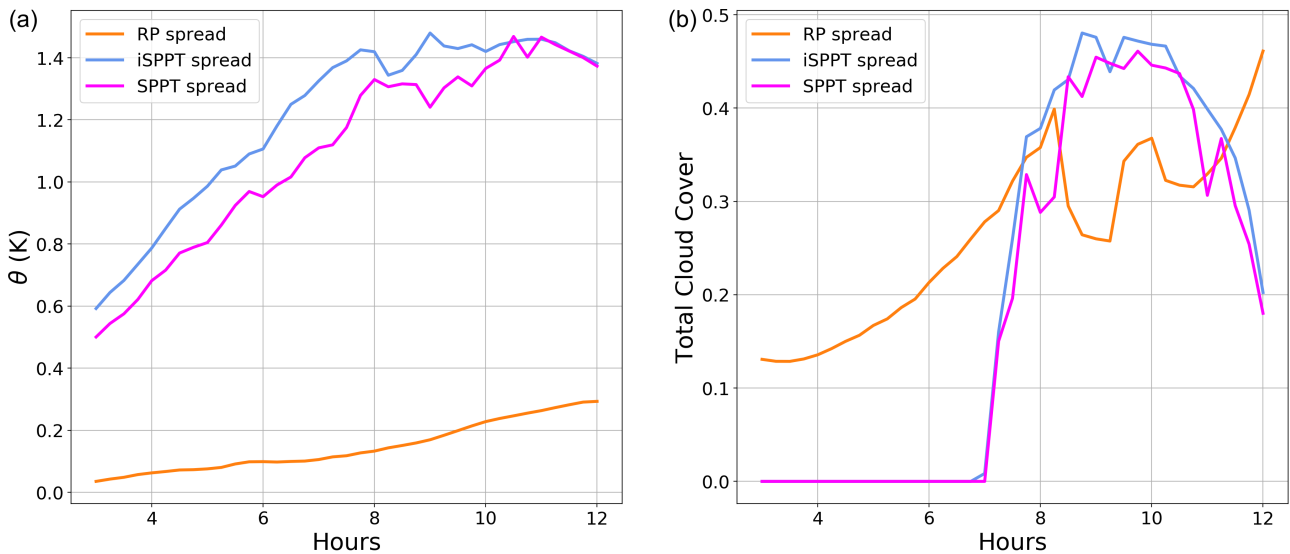


Figure 3: Comparison of ensemble spread for (a) θ at the lowest model level and (b) total cloud cover between the ARMCU SPPT, iSPPT (independent SPPT) and RP ensembles.

turbations at the top of the boundary layer. Near the ground, ensemble perturbations are relatively small because (1) surface fluxes are not perturbed, (2) SPPT is not very active because turbulence is in quasi-equilibrium with shallow convection, and (3) perturbations are smoothed out by turbulent mixing in the boundary layer. The main meteorological impacts are some spread of near-surface temperature, cumulus onset time, cover and cloud top height.

3.2 Impact of the iSPPT scheme on ARMCU

The SPPT and iSPPT ensembles have similar characteristics, with a temperature spread that is 20% larger with iSPPT at the lowest levels (Figs. 3a and 4c). There, large tendencies occur, so perturbing them independently in iSPPT disturbs the model temperature much more than perturbing their nearly zero sum in SPPT. Higher up in the CBL, the temperature, humidity and wind spreads are not much increased by iSPPT perturbations, because the near-surface perturbation enhancement is smoothed out by the strong vertical mixing.

The cloud liquid water, cloud-top height (not shown), and cloud cover (Fig. 3b) spreads grow by 10%, driven by the convection parametrization, which directly links the near-surface temperature with the cloud structure through the updraft model equations (Pergaud et al., 2009). The iSPPT scheme does not add spread to the timing of the cloud cover. It strongly increases the TKE spread in the lower CBL (Fig. 4a), where the balance between turbulence and convection is disturbed by applying different random numbers to convective and turbulent tendencies. Changes to turbulence tendency spread (Fig. 4c) are reflected in the TKE (Fig. 4a).

The interplay between the parametrizations of turbulence and convection is further illustrated by the tendencies in Fig. 4: turbulence is most active under 300 m height (Fig. 4e), with similar spread in the SPPT and iSPPT ensembles. The total spread is smaller than the turbulent and convective tendency spreads, because there is some compensation between both parametrizations.

The SPPT scheme creates numerical noise that can be seen in the tendency wiggles at low levels, particularly between 8 and 9 h of simulation time (Fig. 4a, b). The iSPPT tendencies are much smoother. SPPT is known to cause stability problems that can lead to numerical blow-ups in various models, which is why it is usually deactivated at the lowest model levels (e.g., Palmer et al., 2009; Wastl et al., 2019a). Here, iSPPT appears to be more numerically stable than SPPT, which confirms the interpretations of Christensen et al. (2017) and Wastl et al. (2019a) that the design of iSPPT perturbations allows finer interactions than SPPT between the parametrizations. Inside the AROME model timestep, the parametrizations are mostly called in parallel and do not interact through model state changes. Nevertheless, some interaction between turbulence and convection can occur inside the timestep because convection contributes to the TKE, which is then used by the turbulence parametrization. Thus, stabilization of the low-level numerical noise by iSPPT necessarily occurs either through TKE adjustments inside the timestep, or through more subtle changes to the model vertical profiles between timesteps.

Above 1400 m, the iSPPT behaviour is very similar to SPPT because only one parametrization (convection) is active, so that both algorithms are nearly equivalent in terms of model state spread. Some residual impact of iSPPT comes from the TKE changes at lower levels.

From a forecasting point of view, replacing SPPT by iSPPT in ARMCU slightly increases the ensemble spread of temperature and cloud variables. The largest impact is the increased spread of TKE in the lower CBL, which may be significant for predicting gusts.

3.3 Impact of RP (random parameters) on ARMCU

The replacement of iSPPT by RP ensemble perturbations is now checked on the ARMCU case. The temperature spread (Figs. 3a, 5a, 5b) is reduced tenfold at most levels, except inside the cloud. The RP scheme introduces a cold bias with respect to the other ensembles in the CBL: its Q10–Q90 interval is on a single side of the deterministic simulation (Fig. 5a), unlike the iSPPT (and SPPT) ensembles. The Q_v spread is reduced too (Fig. 5c, d) and the RP ensemble members are moister than the deterministic run. The vertical cloud distribution is strongly changed by the RP scheme: while the SPPT and iSPPT ensembles had more spread in terms of cloud-top than cloud-base height, this is the reverse with RP (Fig. 5f). RP perturbations create a 15% positive bias in cloud cover (Fig. 5e), and a very large spread in terms of cloud timing since half of the members form a cumulus more than 3 h before the deterministic run (as indicated by the RP ensemble in Fig. 5e). Some RP ensemble members even form some cloud in the early morning, when cloud formation is very unlikely according to the simulation set-up: the air is dry and the boundary layer is very shallow. The large cloud dispersion in the RP ensemble is not caused by changes to the mixed layer height, but by the production of cloud liquid water in many members, at humidities far below saturation. In the AROME model, cloud formation is diagnosed from relative humidity by a cloud parametrization that accounts for subgrid fluctuations of cloud formation using a probability distribution function. The width of this distribution is perturbed in our RP approach: widening the cloud parametrization distribution allows cloud to form much more easily than in the deterministic run. In a separate test (not shown), the cloud parametrization perturbations in RP have been switched off, which reduces the spread of cloud liquid water and cloud cover. In a realistic ensemble prediction set-up, the cloud parametrization should definitely be perturbed because it is a known source of modelling uncertainty. The large sensitivity to this parameter means that its perturbation distribution should be carefully tuned to avoid unphysical ensemble behaviour.

Model tendencies indicate that the RP perturbations are more confined to the CBL top than with the

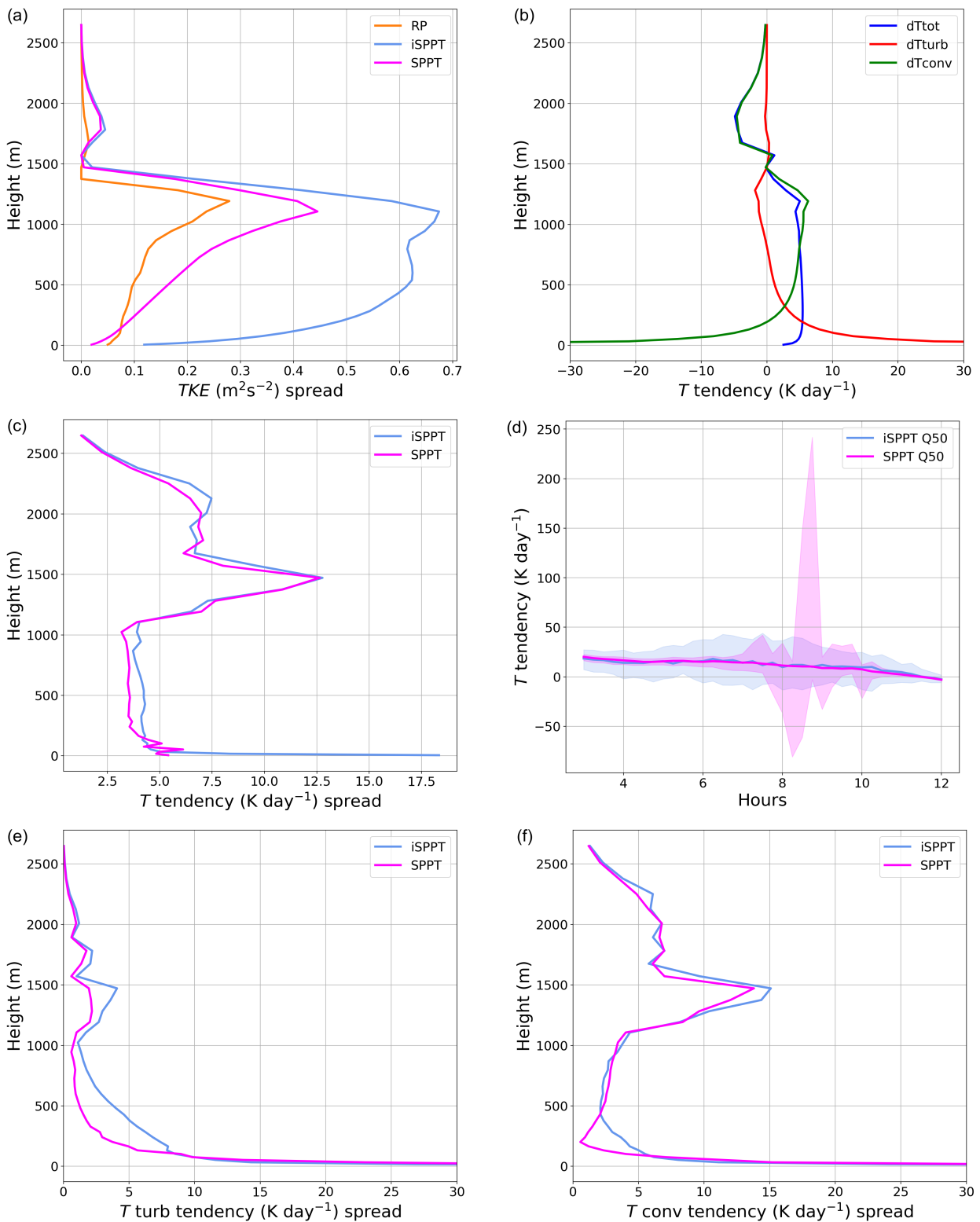


Figure 4: Vertical profiles of (a) TKE and (b, c, e, f) T tendencies and their spread after 11 hours, in the ARMCU case. The ensembles are as indicated in each panel, except (b) which provides for reference the T budget in the deterministic run. Panels (c, e, f) show the vertical profiles of the ensemble spread for each term of the T budget: dT_{tot} , dT_{turb} , and dT_{conv} stand for total, turbulent, and convective T tendencies, respectively. (d) evolution of the ensemble distributions of total T tendency at the lowest model level. Shading depicts the ensemble spread.

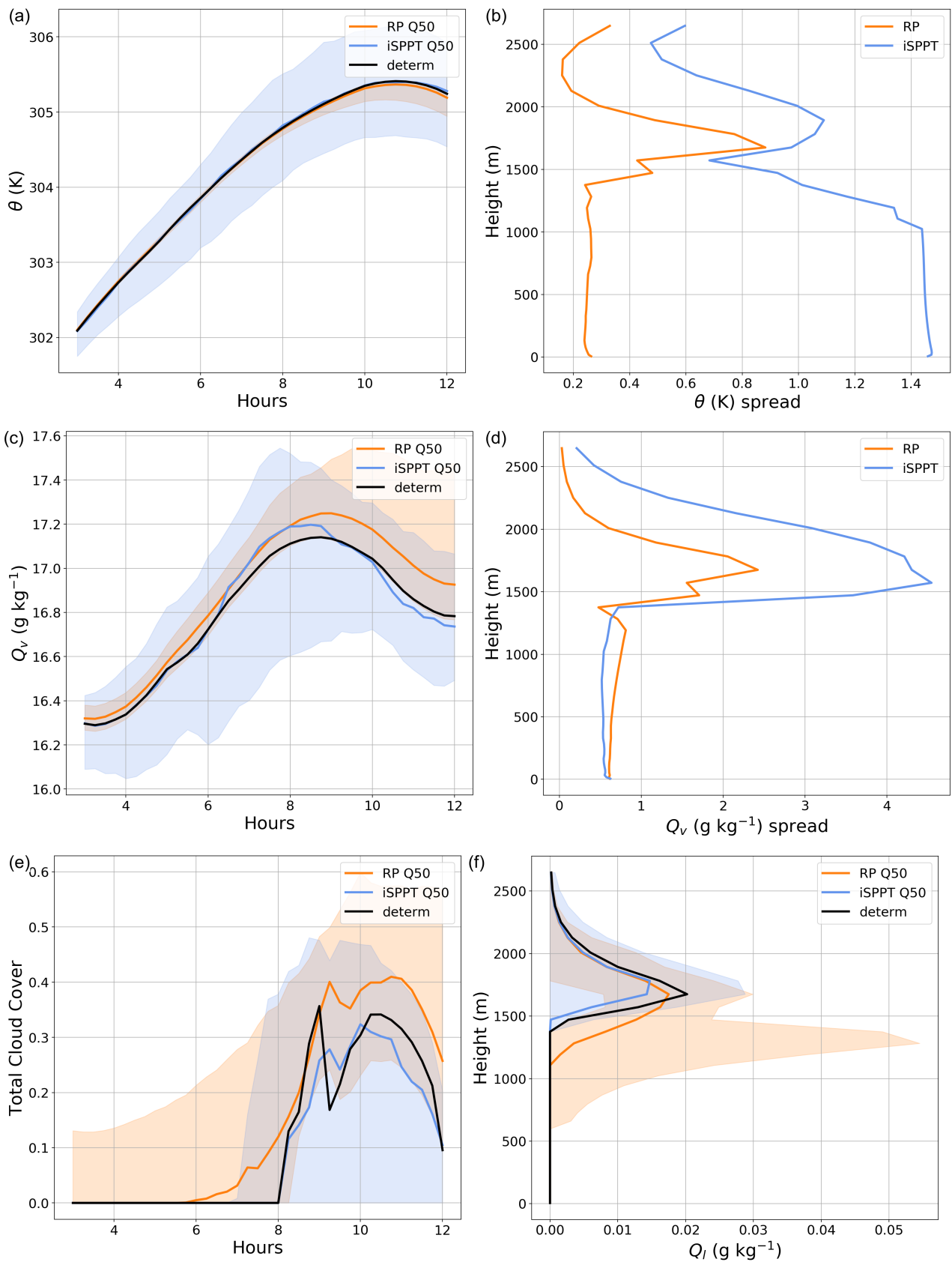


Figure 5: iSPPT (independent SPPT, blue) and RP (random parameters, orange) ensemble distributions for (a,b) θ , (c,d) Q_v , and (e,f) cloud cover and Q_i , respectively. The thick lines indicate the ensemble median and the shading shows the ensemble spread. The deterministic run is shown in black in panels (a, c, e, f). The vertical profiles are extracted 11 hours into the simulation.

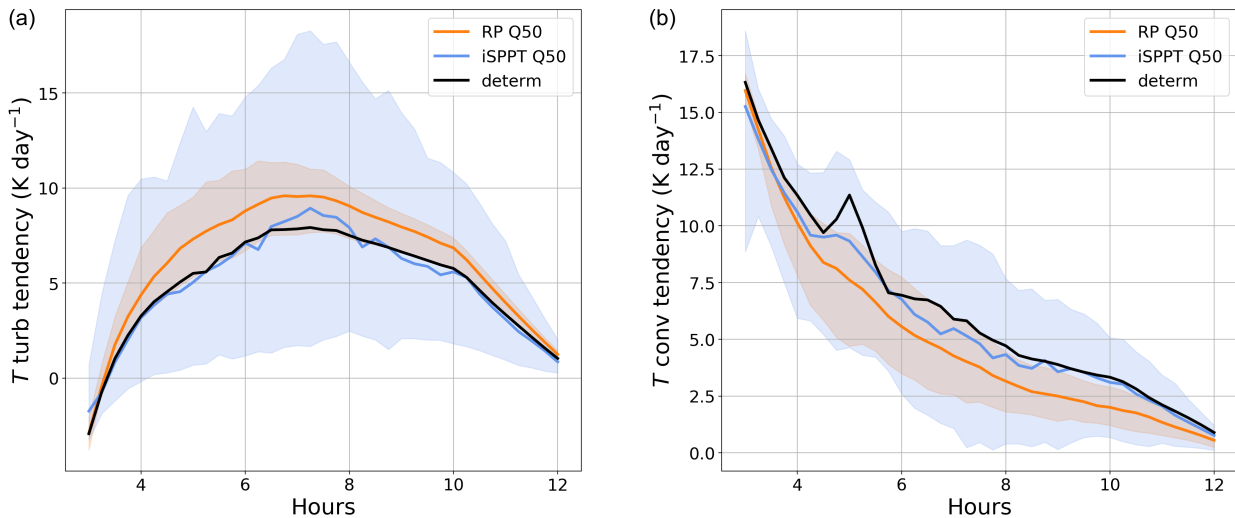


Figure 6: Time evolutions of the RP and iSPPT ensemble distributions of instantaneous temperature tendencies from (a) turbulence and (b) convection at 270 m, i.e., in the lower CBL.

SPPT and iSPPT schemes (not shown). The RP scheme creates some spread to turbulent and convective tendencies near the ground, but they tend to compensate each other. This is illustrated at level 270 m in Fig. 6, which shows that while the instantaneous iSPPT tendency perturbations are nearly centred around the deterministic values, RP perturbations are strongly biased: the RP scheme tends to enhance turbulence at the expense of convection, which probably explains why it does not create as much cloud top spread as iSPPT.

From a forecasting point of view, the random-parameters-based ARMCU ensemble has less dispersion than the ensembles based on SPPT and iSPPT, with the exception of the cloud-related parameters, which have excessive spread that could probably be solved by fine-tuning the cloud parametrization parameter perturbations. RP ensemble members are heavily biased with respect to the deterministic model. RP ensemble perturbations are nearly systematic changes to the model behaviour, we have shown that they result from biases in the physics tendencies. One can suggest two explanations for these biases : (1) the parameter perturbations intervals are asymmetric about their reference values, and (2) the model response to some parameter perturbations is very non-linear. Better designed parameter distributions could possibly alleviate the RP issues, indeed [Ollinaho et al. \(2017\)](#) have demonstrated that the behaviour of a large-scale ensemble can be sensitive to the distributions used in RP. Hence, the RP scheme may require careful tuning before implementation in operational ensembles: acceptable distributions for all perturbed parameters have to be defined, then they have to be updated as forecast model physics evolve. By comparison, the SPPT and iSPPT ensembles are nearly centred on the deterministic model they are derived from. The ensemble spread created by the iSPPT and RP schemes have very different structures, as shown in this case on the cloud representation, so it seems that both schemes produce complementary forms of ensemble perturbations and that they could usefully be combined.

4 Results on the FIRE stratocumulus case

The FIRE case of 14 July 1987 (Duynkerke et al., 2004; DeRoode and Wang, 2007) is characterized by a stratocumulus-topped, well-mixed boundary layer over sea. The cloud top is at about 600 m height, where radiative cooling is in near balance with cloud condensation under a strong inversion of about 8 K. The atmosphere is stably stratified and dry above the inversion. The conditions are nearly stationary, with a weak diurnal cycle: during the day, solar heating slightly warms the CBL, which raises the cloud base. At night, the cloud top cools down, which encourages moisture transport from the surface to the upper part of the cloud; as a result, cloud water content increases, which occasionally triggers drizzle.

In the AROME SCM, this case is implemented using all model parametrizations with a sea surface of fixed temperature. There is a time-independent lateral boundary forcing of geostrophic wind, temperature, and humidity advection, as well as a forcing on vertical velocity to create subsidence. The diurnal cycle is forced by the downward solar and infrared fluxes computed from the local time by the radiative parametrization. The simulation runs over 36 h, dropping the first hour for spinup, after which the model exhibits a nearly periodic diurnal behaviour, so that time series plots are truncated to 30 h.

The FIRE simulation is illustrated by Fig. 7, which shows that the inversion separates a moist CBL from a drier and stably stratified free atmosphere. The CBL is weakly stable because the shallow convection parametrization is more active than the turbulent mixing, except very close to the surface (less than 150 m) where turbulence is triggered by the surface conditions (the sea surface is slightly warmer than the air above). The wind direction is nearly independent of height, so the wind distribution is well summarized by the vertical U speed profiles. The cloud-base height has a diurnal cycle similar to observations (see Fig. 5 in Duynkerke et al. (2004)), but the observed variations in cloud-top height are poorly simulated because of insufficient vertical resolution (which is 50 m at 500 m).

The strongest temperature tendencies (see below) occur near the cloud top: they are the radiative cooling (100 K per day), followed by the latent heat release by condensation (in the microphysics), and the evaporative cooling from convective updraft detrainment. At lower levels, the temperature tendencies are smaller, the main ones being the cooling by drizzle evaporation in the lower parts of the cloud, some convection, and the turbulent heating from the surface at the lowest model levels.

The notable water vapour tendencies (not shown) are: some moistening just above the cloud top by turbulent vertical mixing, a humidity source at the surface, and a water vapour sink by microphysics (i.e., condensation) near the cloud top, which is nearly balanced by the updraft evaporation from the convection parametrization. The wind tendencies are weak, with some surface friction and a downward momentum flux from the convection parametrization.

This case can be summarized as a near equilibrium between three processes: a surface humidity and heat source, a radiative heat sink at the cloud top which drives explicit humidity condensation and subgrid convective detrainment, and an upward transport of heat and moisture by convection inside the mixed layer. The cloud-top processes occur over just two model layers, due to the coarse vertical resolution of the model, so that the cloud-top inversion shows up as a discontinuity in model profiles (Fig. 7).

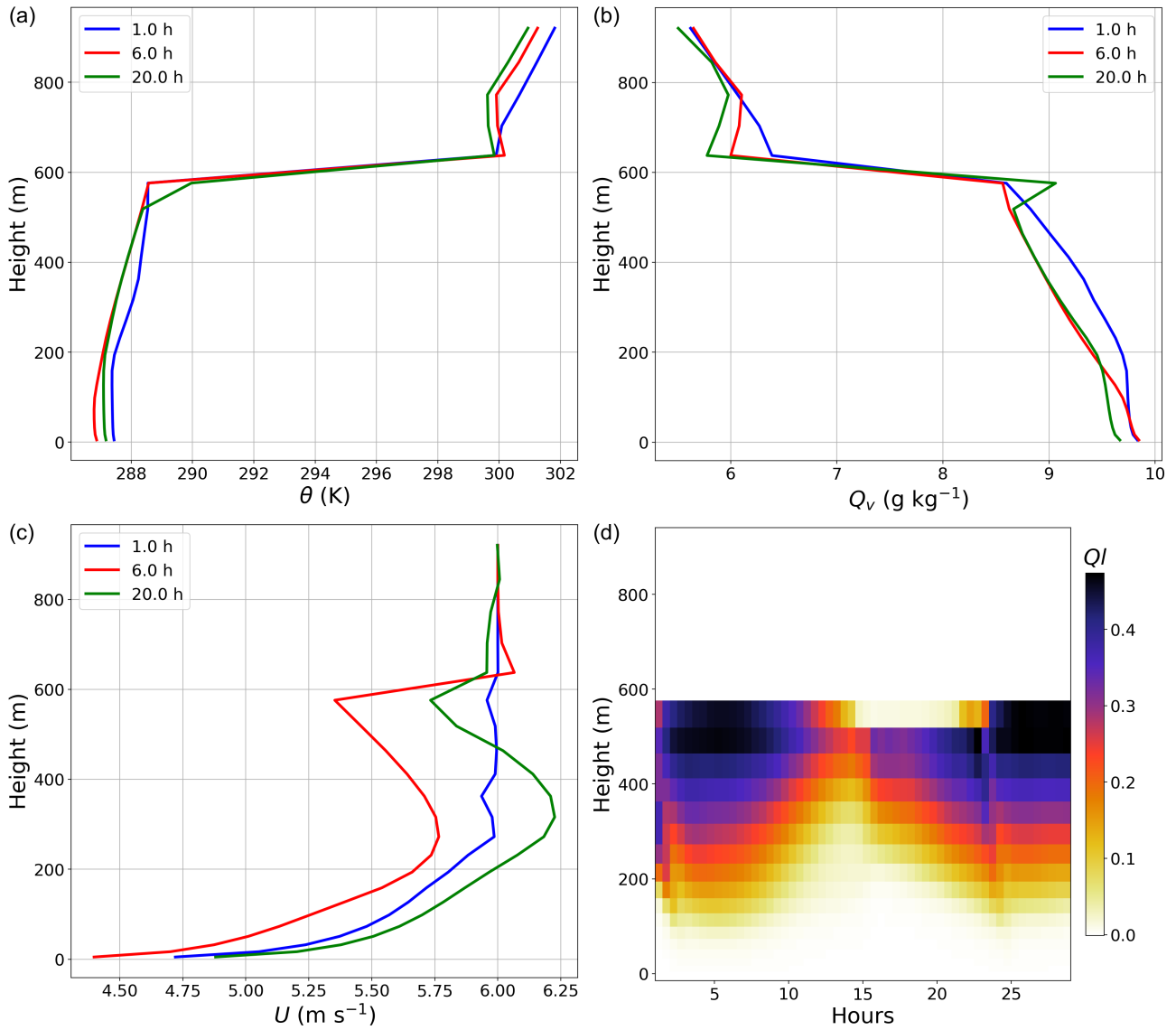


Figure 7: Sample vertical profiles of (a) θ , (b) Q_v , and (c) U after 1, 6, and 20 hours of simulation in the FIRE case. (d) time–height evolution of Q_l .

4.1 *Impact of the SPPT scheme on FIRE*

The ensemble dispersion from the SPPT perturbations (Fig. 8) is relatively weak: the U spread is less than 0.2 m s^{-1} near the surface, caused by perturbations to the vertical momentum flux linked to wind shear. The spreads of θ and Q_v are very small except near the inversion (Fig. 10). The dispersions of cloud liquid water content in the cloud layer, LWP (liquid water path) (Fig. 10), TKE and surface fluxes (not shown) are about 10% of their deterministic values.

From a forecasting point of view, the dispersions of near-surface wind speed, temperature, and humidity are rather small, so SPPT has little impact for practical purposes. Despite the spread of cloud liquid water content there is hardly any ensemble dispersion in cloud base, top height, or cloud cover.

The instantaneous tendencies are illustrated using temperature in Fig 9: the tendency spread is dominated by radiation and microphysics at the cloud top, there is very little spread in convection and turbulence. The microphysics tendency spread is nearly 100% of its deterministic value, which is linked to small fluctuations of the cloud-top height. As in ARMCU, there is a lot of compensation between the parametrizations, which acts to dampen the perturbations of the prognostic variables.

The conclusion is that the SPPT perturbations only have a small impact on the FIRE case, because (a) it is a quasi-stationary atmospheric state, so there are not much physics tendencies to perturb, and (b) the model evolution is dynamically stable, in the sense that tendency perturbations that are injected by SPPT (mainly at the cloud top) are damped. Thus, the SPPT ensemble members never depart much from the deterministic run, except at the cloud-top level, where spread is mostly caused by small ($< 100\text{m}$) displacements of sharp vertical gradients. An exception is Q_l : its spread is created by large tendency perturbations at the cloud top, and it propagates evenly down the cloud layer. Nevertheless, since the cloud is relatively thick, these fluctuations in Q_l and cloud-top height are not large enough to cause significant T and Q_v spread in the CBL.

4.2 *Impact of the iSPPT scheme on FIRE*

In terms of model prognostic variables, the iSPPT ensemble behaviour is qualitatively similar to the SPPT ensemble, but the ensemble spread (Fig. 8) is much larger, except for temperature (Fig. 8b) at low levels. Temperature spread is increased aloft. The iSPPT scheme does not change the cloud structure much, but it increases the dispersion of liquid water near the cloud top (Fig. 8d), by randomly shifting the cloud-top position. The instantaneous temperature tendency profiles (Fig. 9) show that replacing SPPT by iSPPT has a large impact near the cloud top, where the tendencies are large. The tendency spread of the dominant processes (radiation and microphysics) is increased by iSPPT, leading to a marked increase in total tendency spread (Fig. 9b), despite the fact the spreads are not additive: some compensation occurs between the tendencies from various physical parametrizations.

4.3 *Impact of the RP scheme on FIRE*

The impact of the RP scheme is stronger in FIRE than in the ARMCU case. The model variable spread (Figs. 8 and 10) is similar to the iSPPT scheme and even larger at cloud-top level, where a peak of spread is produced by small fluctuations of the inversion height. The RP scheme lowers the inversion (Fig. 10a,b). It brings more dispersion in the cloud water content and cloud-base height (Fig. 10c),

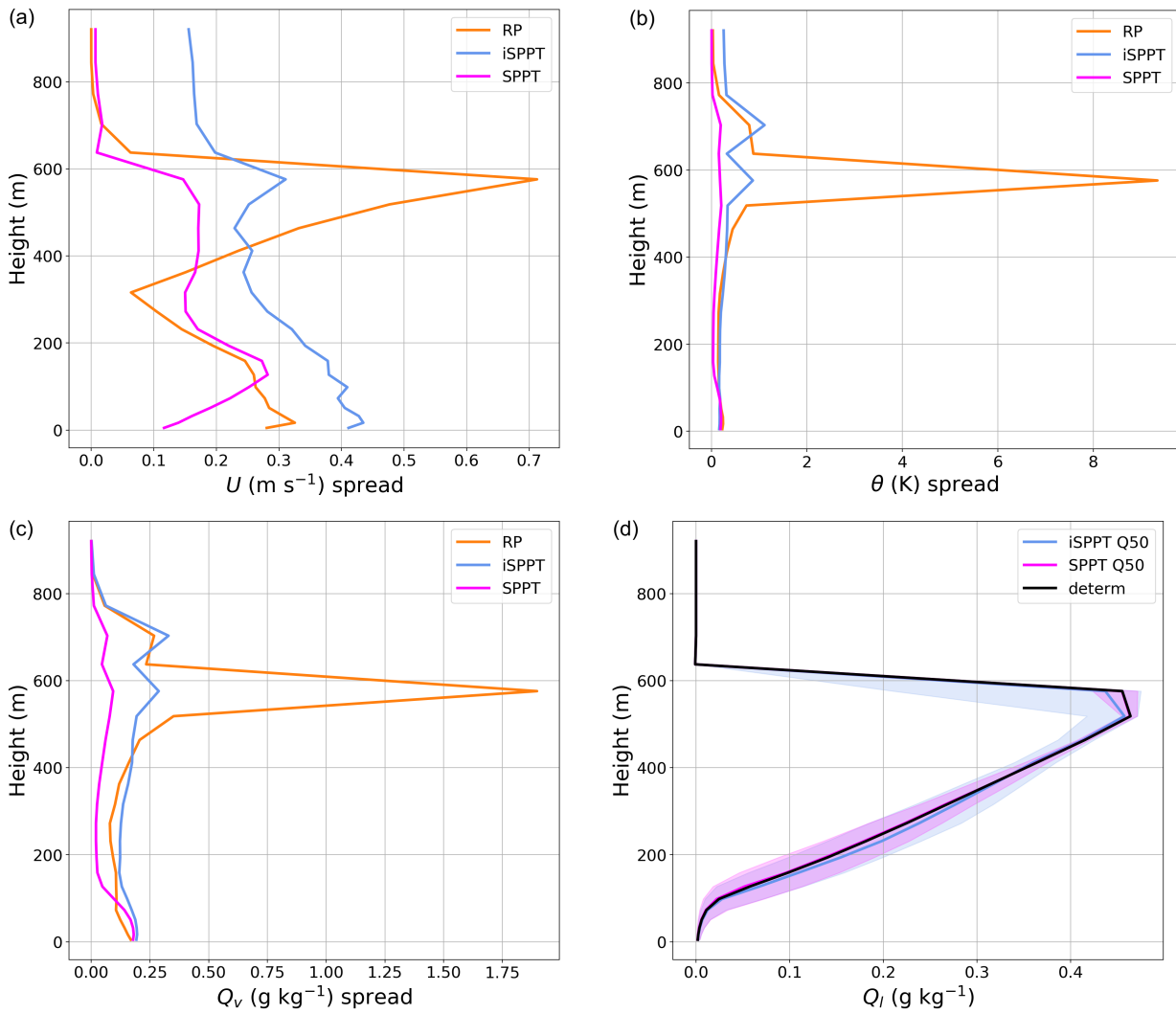


Figure 8: Vertical profile after 6 hours in the FIRE SPPT and iSPPT ensembles for (a) U spread, (b) θ spread, (c) Q_v spread, (d) Q_l distribution. The RP ensemble spread is also shown in panels (a), (b), (c).

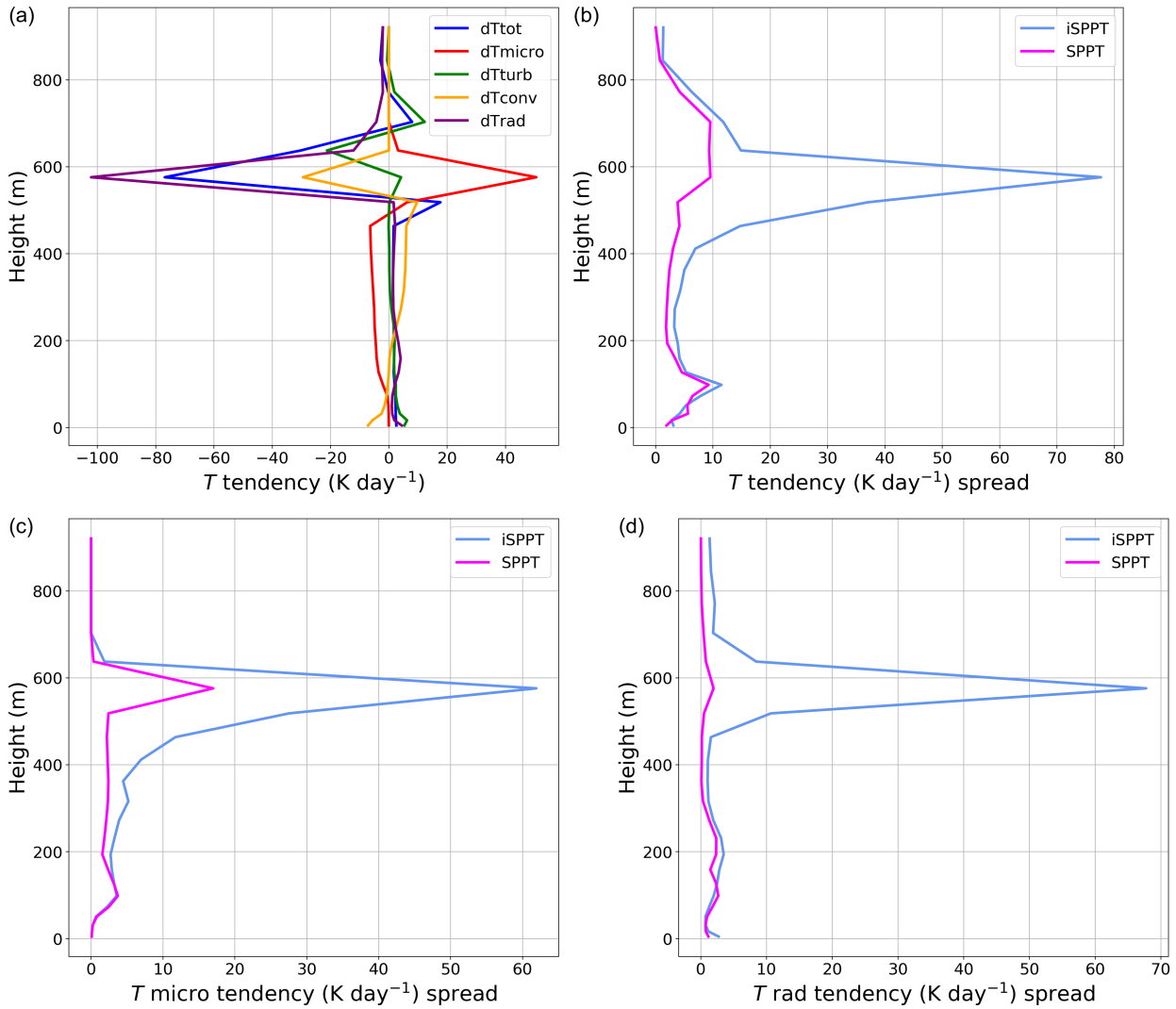


Figure 9: Vertical distributions of the instantaneous temperature tendencies after 6 hours in FIRE: (a) tendency budget in the deterministic run (dTtot, dTmicro, dTturb, dTconv, and dTrad stand for total, microphysics, convection, and radiation tendency, respectively), comparison of the SPPT and iSPPT ensemble spreads in terms of the (b) total, (c) microphysics, and (d) radiation tendency.

leading to a threefold increase in LWP spread. Thus, the RP schemes strongly perturb the vertical cloud structure. This could be desirable in the sense that stratocumulus cloud profiles are often uncertain in numerical weather prediction, but it comes at the price of strong biases with respect to the other ensembles, clearly visible in Fig. 10. While the SPPT and iSPPT ensemble dispersions are nearly centred around the deterministic run, RP-induced biases significantly affect most variables: near the ground, the RP ensemble has a slow (0.3 m s^{-1}) wind speed bias (not shown), it is moister (5% relative humidity) and colder (0.5 K) than the deterministic model, and the total cloud cover is higher with a near doubling of LWP (Fig. 10d), which reduces solar radiation and evaporation at the surface. The cloud base is much lower in the RP ensemble. These biases are also present, but much smaller, in the iSPPT ensemble.

In many aspects, the unperturbed simulation is far outside the Q10–Q90 interquartile range of the RP ensemble. It means that the RP ensemble cannot be claimed to be a representation of uncertainty around the unperturbed model simulation, as is commonly assumed by forecasters when using ensembles. RP model tendencies (not shown) indicate that the perturbation amplitude in the RP ensemble is generally similar to the iSPPT ensemble, with some location changes, for instance the RP scheme produces more tendency spread near the cloud top and the surface, but less at the cloud base.

4.4 Summary of FIRE case

In this case, the SPPT scheme only slightly perturbs the active processes. There is little forecast spread and any model state deviation from the deterministic simulation appears to be damped. The only significant dispersion in terms of point forecasts comes from small shifts of large vertical gradients at the cloud-top inversion, and in the wind shear zone inside the boundary layer. The iSPPT scheme is much more effective than SPPT at creating ensemble spread, by changing the relative amplitude of the various processes at play: this perturbs the equilibrium state itself. The RP scheme achieves a higher level of ensemble spread through comparable mechanisms, but it creates biases as large as the spread itself. In other words, RP simulates uncertainties of a model that has a quite different climate from the unperturbed model. As in the ARMCU case, it would be interesting to clarify if RP-induced biases are due to intrinsic model non-linearities or to asymmetries in the parameter perturbations. Model non-linearities do impact all ensemble perturbation techniques, as shown by the biases of iSPPT ensemble, but the biases are very large with RP. Hence, there is a difficulty with random-parameters-based ensembles for estimating the uncertainty of a deterministic prediction, which is the way most ensembles are used in current operational forecasting applications. Some improvement to the RP scheme might alleviate this problem. For instance, [Christensen et al. \(2015\)](#) have suggested that randomly varying parameters may alleviate the RP-induced biases, based on ten-day integrations of a lower-resolution model. It would be interesting to clarify if this result remains true at forecast ranges shorter than the decorrelation time scale of RP perturbations.

5 Results on the LANFEX fog case

The LANFEX IOP1 fog case of 24–25 November 2014 ([Boutle et al., 2018](#)) is now studied. This case is derived from the LANFEX field experiment ([Price et al., 2018](#)). Our model set-up closely follows the specification of the intercomparison study in [Boutle et al. \(2021\)](#), with a cloud droplet number concentration of 100 cm^{-3} . It is characterized by a stably stratified, nearly saturated surface layer with light wind (U is of the order of 1 m s^{-1}), over a slowly cooling surface. During the night, radiation fog forms

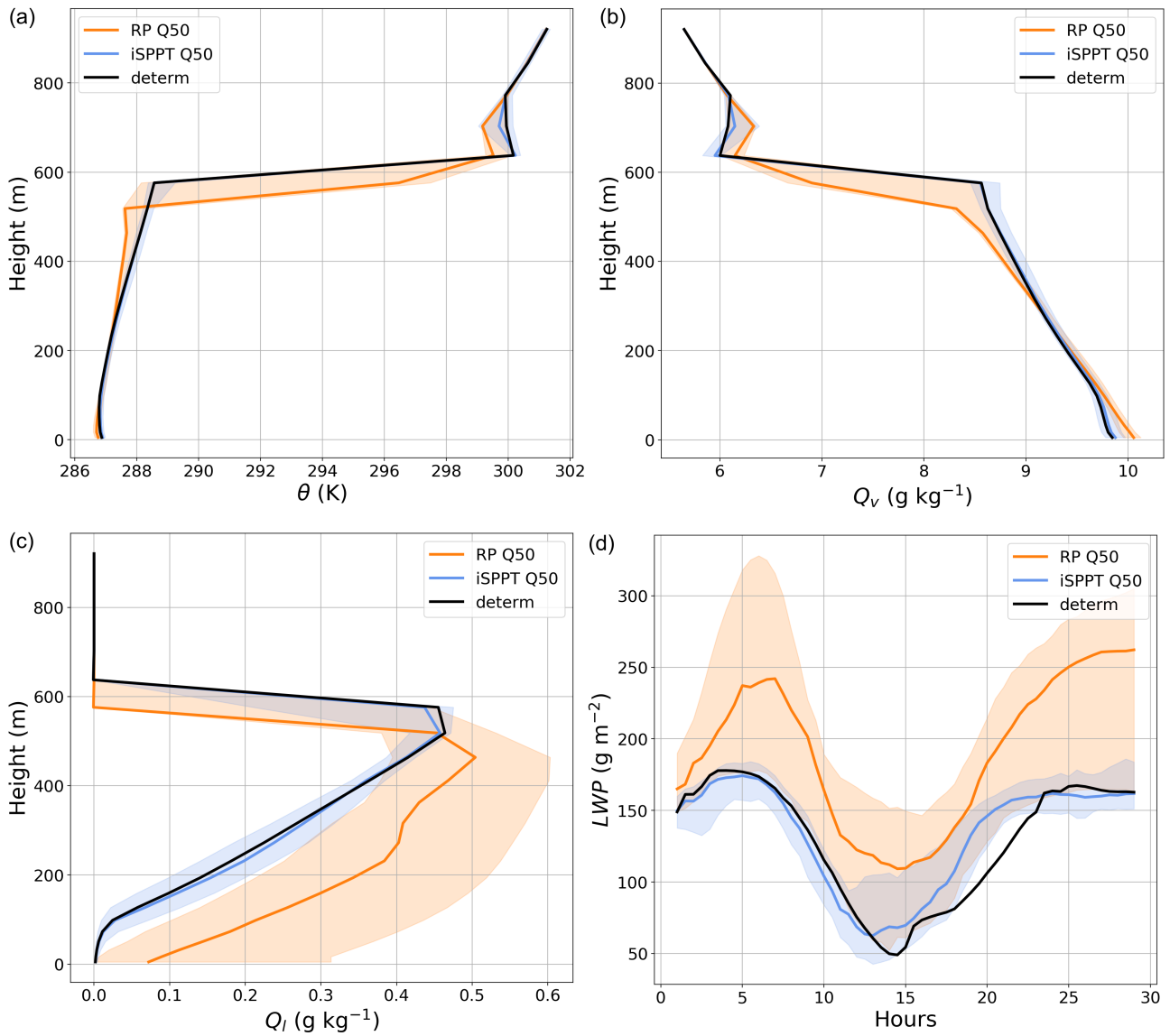


Figure 10: iSPPT and RP ensemble spread on the FIRE case, in terms of 6-hour vertical profiles of (a) θ , (b) Q_v , (c) Q_i , and (d) LWP timeseries.

and grows until sunrise (15 hours into the simulation), then the fog begins to lift. For this simulation, no lateral boundary forcing is prescribed, the atmospheric evolution is entirely driven by the radiative forcing and the surface condition. The simulation stops before noon because the experiment set-up is not designed to represent fog dissipation during daytime.

The deterministic LANFEX simulation is summarized in Fig. 11: after 30 minutes, the lowest model level saturates and fog forms as indicated by the appearance of non-zero Q_l at the lowest model level in Fig. 11d. Then, the mixed fog layer grows almost linearly with time under an inversion of nearly 10 K. This evolution is mainly driven by radiative cooling tendencies of about 80 K per day in the uppermost 50 m of fog. The cooling is mitigated by turbulent mixing across the inversion and latent heat release from water condensation, which generates a foggy layer with some drizzle. In this regime, the model behaviour is similar to the FIRE case, except that the model convection parametrization is not active until sunrise, because of a lack of upward heat fluxes at the surface.

Fifteen hours into the simulation, the sun rises and warms up the bottom layers, which triggers shallow convection. At this stage the liquid water content of the lowest fog layers begin to decrease, while the fog top continues to rise. The θ profiles (Fig. 11a) show how the night-time near-surface cooling stops a few hours after fog onset because of turbulent mixing between the ground and the inversion, the mixed layer is heated by condensation at the top of the fog and by entrainment from the warmer air aloft. Inside the fog, Q_v (Fig. 11b) closely follows the temperature evolution because it is driven by variations of saturation humidity. The cloud liquid water Q_l (Fig. 11c, d) is nearly constant at about 0.2 g kg^{-1} throughout the fog layer. It is regulated by a slow sedimentation flux (0.03 mm h^{-1} , not shown) and some occasional drizzle events. Drizzle happens 4 and 9 h into the simulation when the condensate concentration temporarily reaches 0.5 g kg^{-1} , which is a threshold for conversion from cloud particles to larger raindrops in the model microphysics parametrization.

After sunrise, shallow convection kicks in and begins to pump humidity aloft, which decreases low-level cloud liquid water. This evolution is normally associated with increasing horizontal visibility.

5.1 Impact of the SPPT and iSPPT perturbations on LANFEX

The spread of the SPPT ensemble (Fig. 12) is negligible for practical purposes except for some small wind dispersion. This situation lasts until sunrise, when diurnal heating triggers shallow convection associated with increasing spread of temperature and humidity. From a forecasting point of view, the SPPT scheme does not significantly perturb the prediction of this radiative fog case.

The iSPPT ensemble, on the other hand, produces more spread of wind, temperature, and humidity (Fig. 12a, c, e) when fog begins to form, after which the ensemble spread remains nearly constant. Spread remains high at the cloud-top level (Fig. 13c). Q_l (Fig. 12b) has an approximately 15% relative spread in the fog layer throughout the fog life cycle. It implies a large dispersion in the fog top height (Fig. 12d), and in LWP growth over time (Fig. 12f), which is in line with the model-related LWP uncertainties described by [Boutle et al. \(2021\)](#). (2021).

The instantaneous tendency budgets (Fig 13a) indicate that fog evolution results from a balance between radiation and microphysics (i.e., condensation). The iSPPT scheme changes the relative strength of these processes. This mechanism produces more spread than SPPT: the radiative cooling has little spread in the lower fog layer, but iSPPT perturbs the radiation and condensation above, which further perturbs temperature through latent heat release, and the Q_l/Q_v ratio. At the fog top, iSPPT radiation

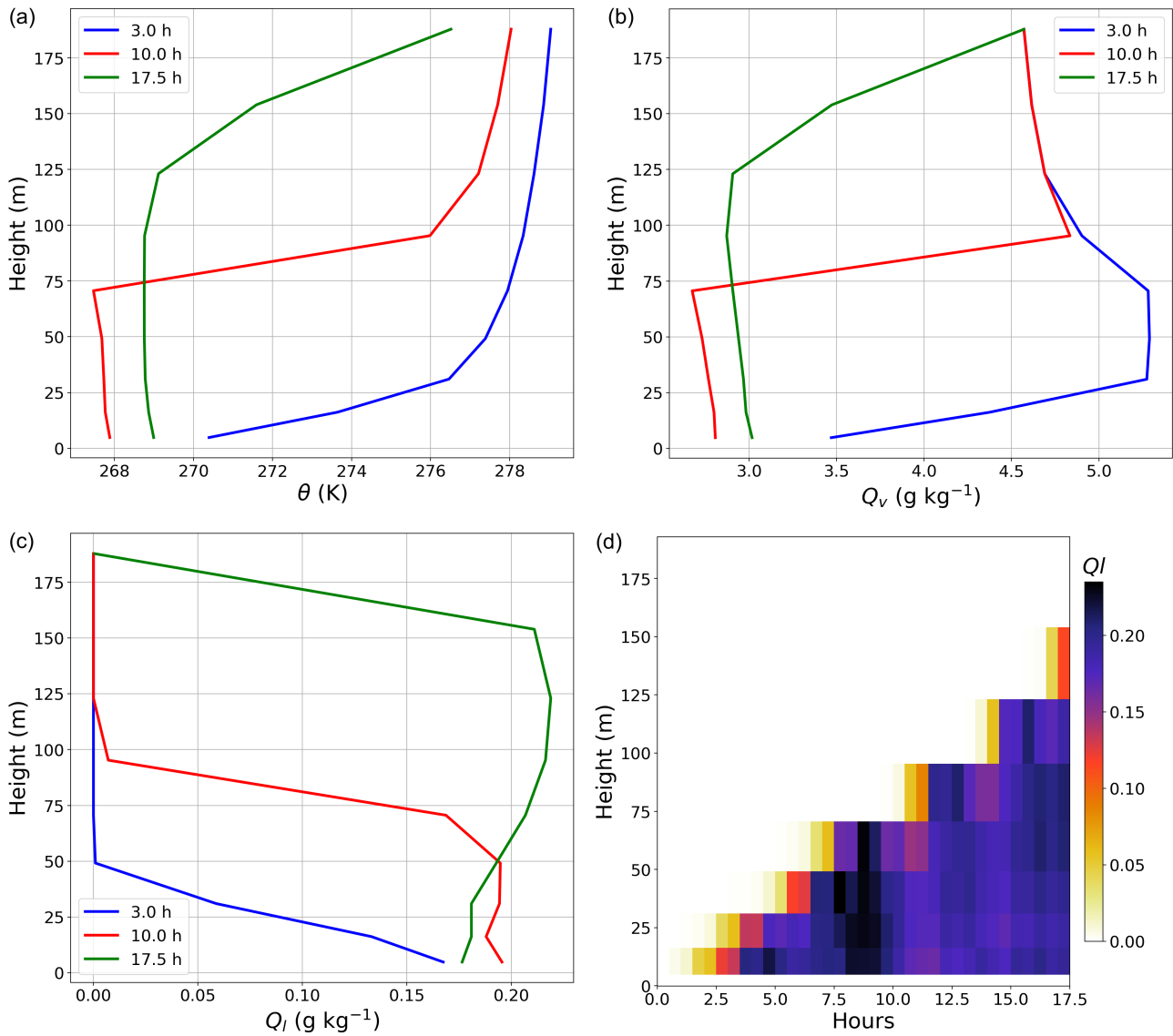


Figure 11: Deterministic LANFEX simulation vertical profiles for (a) θ , (b) Q_v , (c) Q_l after 3 hours (shortly after fog formation), 10 hours (mature fog), and 17.5 hours (2.5 hours after sunrise), (d) time–height evolution of Q_l .

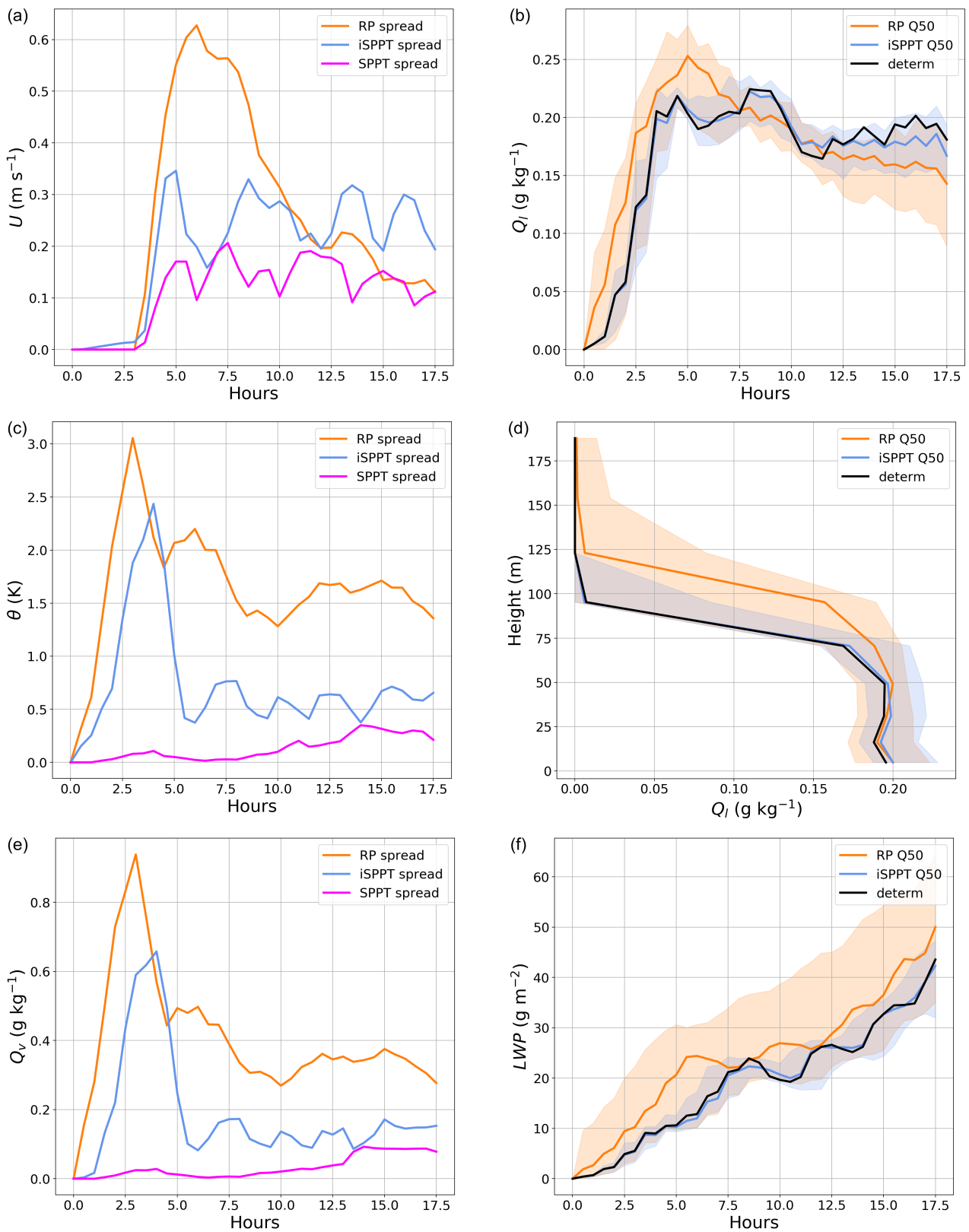


Figure 12: left column: LANFEX ensemble spread evolution at the second lowest model level with the SPPT, iSPPT and RP schemes, for (a) U , (c) θ , (e) Q_v . Right column: evolution of the median and spread of Q_l depicted by its (b) evolution at the second lowest model level, (d) vertical profile after 10 hours, and (f) LWP evolution.

perturbations are much more active than with SPPT, so that fog thickness has more spread (Fig. 12d, f). This is significant from a forecasting point of view, since changing the total water content and thickness of the fog layer can be expected to affect fog dissipation later, after the ending time of our simulations. The spread in Q_l and LWP growth rates suggests that iSPPT produces a dispersion of about 1 h for fog onset and 3 h for fog dispersion. This is consistent with uncertainty estimates on fog onset times from large-eddy simulations (LES) on this case (L. Ducongé, personal communication). Our experimental set-up does not allow much spread in the fog onset process, because it occurs very early into the simulation. A study of fog onset sensitivity would require simulations that start earlier, allowing ensemble perturbations to grow before water saturation is reached.

5.2 Impact of the RP scheme on LANFEX

We now compare the impact of the RP perturbations with iSPPT in this fog case. The RP schemes produces much more spread than iSPPT (Figs. 12 and 13), with a near doubling for most variables. The instantaneous tendency and time–height plots (Fig. 13) show that RP strongly perturbs the radiation tendencies and the fog thickness (about 50% of total thickness after 10 h). It implies some large spread on the fog timing, with about 2 h of spread on fog onset time, and more than 6 h on fog thickness reaching 100 m. This ensemble spread is much larger than the prediction uncertainties suggested in the intercomparison of [Boutle et al. \(2021\)](#), according to available results (Boutle, personal communication).

As in the other cases, ensemble biases are a concern, with many output variables of the RP ensemble exhibiting one-sided perturbations. The Q_l biases are significant (Fig. 12b, d, f): if the deterministic run is assumed to be unbiased, then the RP ensemble fog forms too early, it is too thick with a too large LWP, and it starts to dissipate too early. This would have a large impact on visibility forecasts.

The conclusion on the LANFEX fog case is that the iSPPT and RP schemes are both much more effective than SPPT at creating ensemble dispersion; the dispersion of the RP ensemble is the largest. Unlike the iSPPT ensemble, the RP ensemble is heavily biased with respect to the deterministic simulation.

6 Summary and discussion

This study has provided a detailed comparison of mechanisms by which the SPPT, iSPPT, and RP schemes produce dispersion in ensemble prediction, in several distinct atmospheric regimes.

In all cases, the iSPPT perturbation structure is qualitatively similar to the SPPT one, with a generally larger amplitude. This amplification is strongest in conditions of near equilibrium between several compensating processes, as represented by the model physical parametrizations: in the ARMCU case, the main equilibrium is between turbulence and convection in the mixed layer. In the FIRE and LANFEX case, it is between radiation and condensation near the cloud top. Most of the ensemble dispersion is injected into the model at these levels by the SPPT and iSPPT perturbations. The perturbations propagate downwards to the rest of the boundary layer, with some damping and a redistribution between variables that is highly case-dependent. The damping is most pronounced near the inversions, and the perturbations do not carry enough energy to significantly perturb the inversion heights, although ensemble spread is locally large because of small vertical displacements of the atmospheric profiles in areas of sharp gradients.

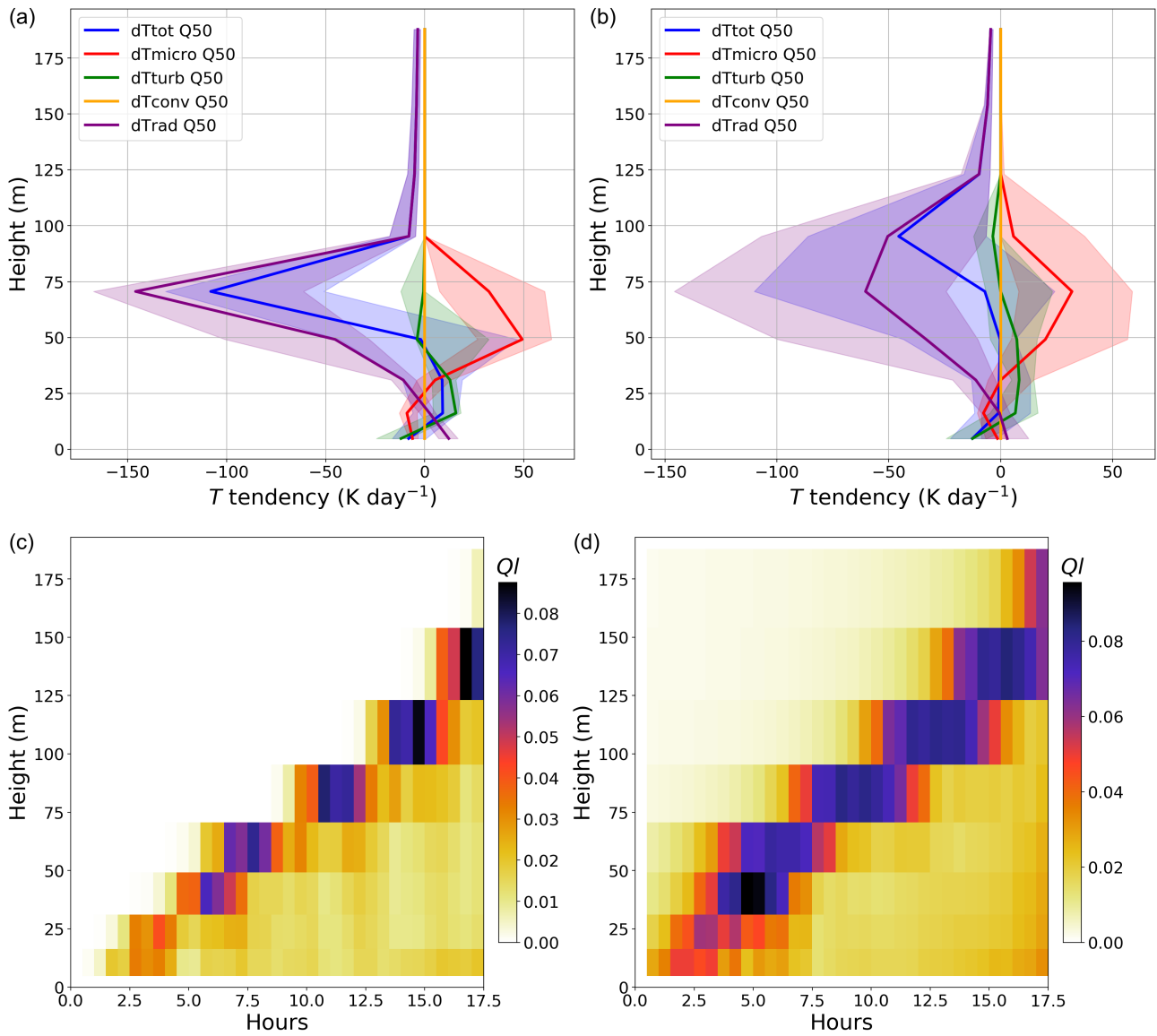


Figure 13: (a) (b) vertical profiles of the instantaneous temperature tendencies in the LANFEX ensemble after 8 hours, in the (a) iSPPT and (b) RP ensembles, with the same conventions as in Fig. 9a (i.e., the thick lines show the median tendencies, the shaded areas depict the Q10–Q90 ensemble quantiles). (c) (d) time–height distributions of the ensemble standard deviation of Q_l in the iSPPT and RP ensembles, respectively.

The spread of low-level weather parameters results from a complex damping of the ensemble spread aloft, and in some cases the spread is very small, which suggests that routinely available near-surface observations may not be appropriate for evaluating the impact of model error schemes: ensemble spread appears to be much more sensitive to these perturbations near the boundary-layer top and near the cloud top. The cloud profiles are quite sensitive to the SPPT and iSPPT perturbations, with large cloud-top spread in ARMCU, and cloud-base spread in FIRE. The differences between the iSPPT and SPPT ensembles can be understood through the study of model tendencies: iSPPT brings more ensemble dispersion than SPPT because it perturbs the relative strengths of different model processes, although some of the implied additional spread is lost by internal compensation mechanisms during model integration. The ARMCU case confirms a practically important result of [Christensen et al. \(2017\)](#) and [Wastl et al. \(2019a\)](#): iSPPT appears numerically more stable than the SPPT scheme, to the extent that it seems usable down to the model surface, unlike SPPT which frequently requires some low-level tapering to avoid model crashes.

The effect of RP perturbations is quite different from the other approaches: the RP ensemble spread has a different case dependency, it is smaller in ARMCU, and larger in FIRE and LANFEX, although the same settings are used in all cases. It means that the RP approach is more effective at perturbing some processes than others, so its impact will depend on the type of boundary layer and cloud. Our version of RP produces particularly large cloud perturbations, which is linked to cloud parametrization perturbations, as well as (in the FIRE and LANFEX cases) to the link between cloud variables and radiation. RP perturbations produce large model biases, in the sense that the differences between the RP ensemble members and the deterministic simulation are almost one-sided, and the RP ensemble median is clearly different from the deterministic simulation. Most of the RP ensemble spread comes from the creation of biases. The iSPPT ensembles exhibit some bias too, but with a much lower amplitude. Other set-ups of the RP approach may produce different results. It is possible that RP-induced biases could be reduced by varying the random parameters, as noticed by [Christensen et al. \(2015\)](#) at much longer forecast ranges (10 days) than the time scale of random parameters perturbations (about 6 h). This result is not guaranteed to hold at shorter forecast ranges (one day or less) which presumably offer less potential for compensation between successive realizations of the parameter perturbations.

The presence of large biases in our presumably naive RP configuration is worrying for the use of this approach in operational ensembles. It suggests that forecast biases should be carefully checked when designing RP ensemble perturbations. For instance, one could attempt to tune the random perturbation distributions to minimize the ensemble bias. Further experimentation would be needed to understand if the RP ensemble bias comes from asymmetry in the distributions of the random parameters, or from non-linearity of the model response to parameter perturbations. The bias could probably be reduced by narrowing the distribution of the perturbations, but the purpose of having a model perturbation approach would be defeated if the RP-induced perturbation amplitudes became negligibly small in the process. Hence, tuning the random parameters distributions may alleviate the bias issue that is highlighted in our experiments, but adding this constraint on top of the need to optimized ensemble spread seems to worsen the complexity and manpower costs of the RP approach already noted by [Ollinaho et al. \(2017\)](#).

The bias issue of the RP ensembles leads to a more general question: is it a problem if an ensemble is biased with respect to the deterministic system it is derived from? The answer seems to depend on the intended application. Ensembles are often used to assess the uncertainty of an unperturbed deterministic forecast, e.g., for decision-making in an forecasting environment, or to generate an optimal analysis using an ensemble data assimilation procedure: if the physical behaviour of the ensemble members is too different from the unperturbed runs, they probably sample inconsistent forecast errors, which is an issue worth investigating. Conversely, in a long-range, seasonal, or climate simulation setting, deterministic

runs are hardly used, the valuable information is in the ensemble forecasts, so changing their biases using RP (or any other perturbation scheme) might actually be beneficial if it brings the ensemble climate closer to the truth, as reported in, e.g., [Ollinaho et al. \(2017\)](#).

All model error schemes used here have changed some aspects of the cloud simulation, in particular its vertical extent, but the perturbations in cloud timing were relatively limited, and no member had no cloud. There was relatively little perturbation to the vertical extent of the mixed layer, and to the timing of its evolution. The largest perturbations of model prognostic variables have been found in atmospheric parameters that are sensitive to the accumulation of physics tendencies during model integration, i.e., afternoon convection in ARMCU and mature fog in LANFEX. The ensemble perturbations from all tested schemes have remained negligible in the free atmosphere where physical parametrizations are hardly active, and near the surface because the perturbations injected into the upper mixed layer and into the cloud tend to be heavily diluted by vertical mixing.

These limitations are unavoidable because those perturbation schemes are all constrained to act inside the atmospheric physical parametrizations of the model: by design, they are unable to directly perturb atmospheric features related to the model dynamics, to the initial state, the horizontal advection, the surface fluxes, or atmospheric processes that are not represented in this model, e.g., aerosols or chemistry. Our perturbation approaches do represent some of these processes in an indirect and limited way: perturbing turbulence near the surface can be a proxy for perturbing surface fluxes, and perturbing the microphysics tendencies in fog can emulate uncertainties related to aerosols. In a model setting that includes realistic interaction between a land surface and the atmospheric column, one would expect a strong feedback between surface fluxes and the evolution of cumulus and fog during daytime. This should enhance the spread of near-surface weather variables, even if there is no ensemble perturbation in the surface model.

This study has shown that ensemble perturbations created by model perturbation schemes have structures and amplitudes that are sensitive to the weather situation and to the design of the perturbation scheme. Tuning these schemes, using, e.g., observations or higher resolution models as reference, will require consideration of ensemble behaviour in various weather regimes. Current ensemble testing approaches generally ignore this aspect and use either average scores over long periods and large domains, or case studies with limited weather type diversity. Realistic regional ensemble systems have a much more complex set-up than this study: they use three-dimensional atmospheric models and a mixture of initial, lateral boundary, and surface perturbations. Modifications of model error representations in such ensembles will interact in a complex way with these other perturbation sources, and in general it will not be possible to isolate the impact of modifying these model error schemes, because as we have shown, their impact on observable variables such as near-surface variables or cloudiness can be very small and/or indirect. It is recommended that future evaluations of model error schemes in three-dimensional ensembles take into account weather types, vertical levels and variables that are directly sensitive to these schemes.

Field experiment data and instrumented observation sites may help to better understand and design model perturbation schemes, through precise comparison between vertical profile observations and meteorologically consistent numerical simulations, such as the single column runs used here, or more sophisticated numerical models including large-eddy simulations. Doing so would allow a better understanding of model error representation issues than the impact studies such as the one presented here, by indicating what is physically right or wrong in model perturbation schemes.

This study has used a very simplified modelling framework, so it can only be used to understand specific aspects of the ensemble perturbation, namely, those that depend on vertical interactions in the boundary layer over a few hours. The impact of horizontal forcings, including horizontal advections,

which are prescribed in our set-up, would require a three-dimensional modelling framework to be properly investigated. The panel of cases used is not representative of all weather situations; in particular it would be useful to include a stable boundary layer case and some precipitating cases, as well as a more realistic interaction with the surface.

The perturbation schemes used could be improved regarding the conservation of water, and the random perturbation distributions in the iSPPT and RP schemes. The iSPPT perturbations could be applied deeper inside the physical parametrizations, e.g., at different stages inside the microphysics and radiation, they could also allow for some decorrelation between variables perturbations, following the pSPPT (partial SPPT) scheme proposal (Wastl et al 2019a). As discussed above, the perturbations in the RP scheme could be made time-dependent, and include correlations between parameters (Christensen et al., 2015) in a follow-up paper. Some physically based stochastic perturbation approaches (Kober and Craig, 2016) should be included in the set of tested options, as they have rather sound physical justifications, and they will generate significant ensemble dispersion in specific weather conditions. Other important sources of ensemble spread are missing in this experimentation and should be evaluated: perturbations to the initial, surface, and lateral boundary conditions in particular may have a strong impact in our framework, possibly larger than the model error perturbations. It should be noted that our study is too simplified to conclude that one scheme is superior to another, except from clearly undesirable features like numerical instabilities or strong biases. This can only be regarded as a comparison between several approaches and weather regimes, designed to better understand the behaviour of ensembles.

The impact of model perturbation schemes on water conservation has been mentioned as a concern in several previous studies (Bouttier et al., 2012; Christensen et al., 2015). Examination of total column water vapour (TCWV) biases and dispersion (not shown) in our experiments indicates that the used cases only provide trivial conclusions on this issue. Namely, in the ARMCU case, the simulated clouds are nonprecipitating and they only condense a small fraction of the atmospheric water vapour; the surface and lateral water fluxes are prescribed, so that the TCWV spread is tiny, i.e., much smaller than its mean value. None of the three tested schemes produced significant TCWV bias with respect to the deterministic run. In the FIRE and LANFEX cases, variations of TCWV are driven by condensation processes, so the TCWV biases are nearly opposite to the Q_t and LWP biases already discussed in Sects. 4 and 5 (see Figures 10 and 12). It seems that a study of the humidity biases reported by other authors would require a more complex experimental set-up, with significant precipitation and interaction with the surface fluxes of humidity. The approach of Lupo et al. (2020), for instance, seems more appropriate, because it focuses on deep convective events with large rainrates.

7 Conclusion and outlook

In conclusion, we have shown in three different cases that both iSPPT and RP approaches produce more ensemble spread than the SPPT scheme thanks to their representation of model uncertainties, and we have identified the model processes responsible for the main variations in ensemble dispersion. The iSPPT scheme can be regarded as superior to SPPT because it is physically more general and it is numerically more stable. The RP scheme can create at least as much ensemble spread as iSPPT, but its impact is structurally different, which suggests that there is some complementarity between the iSPPT and RP schemes. Both perturbation schemes create some ensemble bias. The random-parameters-induced biases represent such a large part of the ensemble dispersion that they should be carefully checked in ensemble systems that use this approach. A risk is that when tuning model error schemes in real-size ensembles, if

some ensemble perturbation schemes are used to compensate for biases in other aspects of the forecasting system, the maintenance and further development of such a system will become cumbersome and difficult to physically base.

The differences between iSPPT and RP perturbation structures suggest that a hybrid perturbation scheme could combine the iSPPT and RP schemes as a unified model error approach for ensemble prediction, taking care to avoid redundancies between both schemes. Such an approach would maximize the generality of the model errors that are represented in an ensemble.

This study has also shown that the behaviour of model error schemes is very situation-dependent. Hence, the tuning of model error schemes would benefit from targeting specific weather regimes, and variables that are sensitive to these schemes. Otherwise, there is a risk that model perturbation schemes will be used to compensate for deficiencies in other aspects of the ensemble generation, e.g., lack or excess of spread in initial conditions, surface or large-scale forcings. This issue will be tackled in a future study using weather regime-dependent ensemble verification, and observations of vertical profiles in the boundary layer, in order to improve future model error representations in ensemble prediction.

Acknowledgements

This study was funded by the French Government through Météo-France, CNRS and the University of Toulouse. It received the IPSIPE research grant from the LEFE/MANU programme of CNRS/INSU. The LANFEX case set-up used data that was kindly provided by I. Boutle and J. Price. The manuscript was improved by the comments and suggestions of two reviewers.

References

- Baker LH, Rudd AC, Migliorini S, Bannister RN (2014) Representation of model error in a convective-scale ensemble prediction system. *Nonlin Process Geophys* 21:19–39 doi:10.5194/npg-21-19-2014
- Barthlott C, Kalthoff N (2011) A numerical sensitivity study on the impact of soil moisture on convection-related parameters and convective precipitation over complex terrain. *J Atmos Sci* 68:2971–2987 doi:10.1175/JAS-D-11-027.1
- Berner J, Fossell KR, Ha SY, Hacker JP, Snyder C (2015) Increasing the skill of probabilistic forecasts: understanding performance improvements from model-error representations. *Mon Weather Rev* 153:1295–1320 doi:10.1175/MWR-D-14-00091.1
- Boutle I, Price J, Kudzotsa I, Kokkola H, Romakkaniemi S (2018) Aerosol-fog interaction and the transition to well-mixed radiation fog, *Atmos Chem Phys* 18:7827–7840, doi:10.5194/acp-18-7827-2018
- Boutle I, Angevine W, Bao JW, Bergot T, Bhattacharya R, Bott A, Ducongé L, Forbes R, Goecke T, Grell E, Hill A, Igel A, Kudzotsa I, Lac C, Maronga B, Romakkaniemi S, Schmidli J, Schwenkel J, Steenveld GJ, Vié B (2021) Demistify: an LES and SCM intercomparison of radiation fog. *Atmos Chem Phys preprint* doi:10.5194/acp-2021-832

- Bouttier F, Vié B, Nuissier O, Raynaud L (2012) Impact of stochastic physics in a convection-permitting ensemble. *Mon Weather Rev* 140:3706–3721. doi:10.1175/MWR-D-12-00031.1
- Bouttier F, Raynaud L, Nuissier O, Ménétrier B, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Q J R Meteorol Soc* 142:390–403. doi:10.1002/qj.2622
- Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE (2008) The MOGREPS short-range ensemble prediction system. *Q J R Meteorol Soc* 134:703–722. doi:10.1002/qj.234
- Brown AR, Cederwall RT, Chlond A, Duynkerke PG, Golaz JC, Khairoutdinov M, Lewellen DC, Lock AP, MacVean MK, Moeng CH, Neggers RAJ, Siebesma AP, Stevens B (2002) Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q J R Meteorol Soc* 128:1075–1093. doi:10.1256/003590002320373210
- Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q J R Meteorol Soc* 125:2887–2908. doi:10.1002/qj.49712556006
- Christensen HM, Moroz IM, Palmer TN (2015) Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *J Atmos Sci* 72:2525–2544. doi:10.1175/JAS-D-14-0250.s1
- Christensen, HM, Lock SJ, Moroz IM, Palmer TN (2017) Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Q J R Meteorol Soc* 143:2168–2181. doi:10.1002/qj.3075
- Christensen HM (2020) Constraining stochastic parametrisation schemes using high-resolution simulations. *Q J R Meteorol Soc* 146:938–962. doi:10.1002/qj.3717
- Clark AJ, Gallus WA, Chen TC (2008) Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon Weather Rev* 136:2140–2156. doi:10.1175/2007MWR2029.1
- Cohard JM, Pinty JP (2000a): A comprehensive two-moment warm microphysical bulk scheme. I: Description and tests. *Q J R Meteorol Soc* 126:1815–1842. doi:10.1256/smsqj.56613
- Cuxart J, Bougeault P, Redelsperger JL (2000) A turbulence scheme allowing for mesoscale and large-eddy simulations. *Q J R Meteorol Soc* 126:1–30. doi:10.1002/qj.49712656202
- Davini P, von Hardenberg J, Corti S, Christensen H, Juricke S, Subramanian A, Watson P, Weisheimer A and Palmer TN (2017) Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geosci Model Dev* 10:1383–1402. doi:10.5194/gmd-10-1383-2017
- De Roode S, Wang Q (2007) Do stratocumulus clouds detrain ? FIRE I data revisited. *Boundary-Layer Meteorol* 122:479–491. doi:10.1007/s10546-006-9113-1
- Duda JD, Wang X, Kong F, Xue M (2014) Using varied microphysics to account for uncertainty in warm-Season QPF in a convection-allowing ensemble. *Mon Weather Rev* 142:2198–2219. doi:10.1175/MWR-D-13-00297.1

- Duynkerke PG, de Roode SR, van Zanten MC, Calvo J, Cuxart J, Cheinet S, Chlond A, Grenier H, Jonker PJ, M Köhler, Lenderink G, Lewellen D, Lappen C, Lock AP, Moeng C, Muller F, Olmeda D, Piriou JM, Sanchez E, Sednev I (2004) Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case. *Q J R Meteorol Soc* 130:3269–3296. doi:10.1256/qj.03.139
- Fundel VJ, Fleischhut N, Herzog SM, Goeber M, Hagedorn R (2019) Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Q J R Meteorol Soc* 145:210–231. doi:10.1002/qj.3482
- Fouquart Y, Bonnel B (1980) Computations of solar heating of the earth's atmosphere: A new parameterization. *Beitr Phys Atmos* 53:35–62.
- Gallus jr W, Wolff J, Gotway JH, Harrold M, Blank M, Beck J (2019) The impacts of using mixed physics in the community leveraged unified ensemble. *Weather Forecast* 34:849–867 doi:10.1175/WAF-D-18-0197.1
- Garcia-Moya JA, Callado A, Escribá P, Santos C, Santos-Munoz D, Simarro, J (2011) Predictability of short-range forecasting: a multimodel approach. *Tellus A* 63:550-563 doi:10.1111/j.1600-0870.2010.00506.x
- Gebhardt C, Theis SE, Paulat M, Ben Bouallègue Z (2011) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos Res* 100:168–177 doi:10.1016/j.atmosres.2010.12.008
- Hacker JP, Snyder C, Ha SY, Pocerlich (2011) Linear and non-linear response to parameter variations in a mesoscale model. *Tellus A* 63:429–444 doi:10.1111/j.1600-0870.2010.00505.x
- Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. I. Basic concept. *Tellus A* 57:219–233. doi:10.3402/tellusa.v57i3.14657
- Hou D, Toth Z, Zhu Y (2006) A stochastic parameterization scheme within the NCEP global ensemble forecast system. 2006 AMS workshop proceedings, section 4.5, 5pp. https://ams.confex.com/ams/Annual2006/techprogram/paper_101401.htm (last accessed: 20 July 2021)
- Kober K, Craig G. (2016) Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J Atmos Sci* 73:2893–2911. doi:10.1175/JAS-D-15-0144.1
- Jankov I, Beck J, Wolff J, Harrold M, Olson J, Smirnova T, Alexander C (2019) Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon Weather Rev* 147:153–173. doi:10.1175/MWR-D-18-0092.1
- Lang S, Lock SJ, Leutbencher M, Bechtold P, Forbes R (2021) Revision of the stochastically perturbed parameterisations model uncertainty scheme in the integrated forecasting system. *Q J R Meteorol Soc* 147:1364–1381. doi:10.1002/qj.3978
- Lupo K, Torn R, Yang SC (2020) Evaluation of stochastic perturbed parameterization tendencies on convective-permitting ensemble forecasts of heavy rainfall events in New York and Taiwan. *Weather Forecast* 35:5–24, doi:10.1175/WAF-D-19-0064.1

- Masson V, Le Moigne P, Martin E, Faroux S, Alias A, Alkama R, Belamari S, Barbu A, Boone A, Bouysse F, Brousseau P, Brun E, Calvet JC, Carrer D, Decharme B, Delire C, Donier S, Essaouini K, Gibelin AL, Giordani H, Habets F, Jidane M, Kerdraon G, Kourzeneva E, Lafaysse M, Lafont S, Lebeaupin Brossier C, Lemonsu A, Mahfouf JF, Marguinaud P, Mokhtari M, Morin S, Pigeon G, Salgado R, Seity Y, Taillefer F, Tanguy G, Tulet P, Vincendon B, Vionnet V, Voldoire, A (2013) The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, *Geosci Model Dev* 6:929–960 doi:10.5194/gmd-6-929-2013
- McCabe A, Swinbank R, Tennant W, Lock A (2016) Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. *Q J R Meteorol Soc* 142:2897–2910. doi:10.1002/qj.2876
- Mlawer EJ, Taubman SJ, Brown P, Iacono MJ, Clough, SA (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave *J Geophys Res* 102:16663–16682. doi:10.1029/97JD00237
- Ollinaho P, Lock SJ, Leutbecher M, Bechtold P, Beljaars A, Bozzo A, Forbes RM, Haiden T, Hogan RJ, Sandu I (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Q J R Meteorol Soc* 143:408–422. doi:10.1002/qj.2931
- Palmer, TN (2001) A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q J R Meteorol Soc* 127:279–304. doi:10.1002/qj.49712757202
- Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts G, Steinheimer M, Weisheimer A (2009) Stochastic parametrization and model uncertainty. ECMWF Res. Dept. Tech Memo no.598, 44pp. available from www.ecmwf.int doi:10.21957/ps8gbwbdv
- Pergaud J, Masson V, Malardel S, Couvreux F (2009) A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-Layer Meteorol* 132:83–106 doi:10.1007/s10546-009-9388-0
- Price J, Lane S, Boutle I, Smith D, Bergot T, Lac C, Ducongé L, McGregor J, Kerr-Munslow A, Pickering M, Clark R (2018) LANFEX: a field and modelling study to improve our understanding and forecasting of radiation fog. *Bull Am Meteorol Soc* 99:2061–2077 doi:10.1175/BAMS-D-16-0299.1
- Romine GS, Schwartz CS, Berner J, Fossell KR, Snyder C, Anderson JL, Weisman ML, (2014) Representing forecast error in a convection-permitting ensemble system. *Mon Weather Rev* 142:4519–4541 doi:10.1175/MWR-D-14-00100.1
- Sakradzija M, Klocke D, (2018) Physically constrained stochastic shallow convection in realistic kilometer-scale simulations *J Adv Model Earth Syst* 10:2755–2776 doi:10.1029/2018MS001358
- Shutts GJ (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q J R Meteorol Soc* 131:3079–3102 doi:10.1256/qj.04.106
- Shutts GJ, Pallarès AC (2014) Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Philos Trans R Soc A* 372:20130284 doi:10.1098/rsta.2013.0284

- Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, Lac C, Masson V (2011) The AROME-France convective scale operational model. *Mon Weather Rev* 139:976–99 doi:10.1175/2010MWR3425.1
- Termonia P, Fischer C, Bazile E, Bouyssel F, Brožková R, Bénard P, Bochenek B, Degrauwe D, Derková M, El Khatib R, Hamdi R, Mašek J, Pottier P, Pristov N, Seity Y, Smolíková, P., Španiel O, Tudor M, Wang Y, Wittmann C, Joly A (2018) The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1 *Geosci Model Dev* 11:257–281 doi:10.5194/gmd-11-257-2018
- Wastl C, Wang Y, Atencia A, Wittmann C (2019a) Independent perturbations for physics parametrization tendencies in a convection-permitting ensemble (pSPPT). *Geosci Model Dev* 12:261–273 doi:10.5194/gmd-12-261-2019
- Wastl C, Wang Y, Atencia A, Wittmann C (2019b) A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. *Mon Weather Rev* 147:2217–2230 doi:10.1175/MWR-D-18-0415.1
- Zhang X (2018) Application of a convection-permitting ensemble prediction system to quantitative precipitation forecasts over southern China: Preliminary results during SCMREX. *Q J R Meteorol Soc* 144:2842–2862 doi:10.1002/qj.3411

Bibliographie

- Ahlgrimm, M., Sakradzija, M., Machulskaya, E., Stiller, O., Seifert, A. & De Lozar, A. (2022). *A stochastic shallow convection scheme for operational use – impact on ensemble spread* [ECMWF Workshop on model uncertainty]. Récupérée 24 août 2022, à partir de https://events.ecmwf.int/event/290/contributions/2972/attachments/1815/3283/Model-WS_Ahlgrimm.pdf
- Albrecht, B. A., Randall, D. A. & Nicholls, S. (1988). Observations of marine stratocumulus clouds during FIRE. *Bulletin of the American Meteorological Society*, 69(6), 618-626. [https://doi.org/10.1175/1520-0477\(1988\)069<0618:OOMSCD>2.0.CO;2](https://doi.org/10.1175/1520-0477(1988)069<0618:OOMSCD>2.0.CO;2)
- Angevine, W. M., Olson, J., Kenyon, J., Gustafson, W. I., Endo, S., Suselj, K. & Turner, D. D. (2018). Shallow cumulus in WRF parameterizations evaluated against LASSO large-eddy simulations. *Monthly Weather Review*, 146(12), 4303-4322. <https://doi.org/10.1175/MWR-D-18-0115.1>
- Arakawa, A. & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *Journal of the Atmospheric Sciences*, 31(3), 674-701. [https://doi.org/10.1175/1520-0469\(1974\)031<0674:IOACCE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2)
- Arnold, H. M. (2013). *Stochastic parametrisation and model uncertainty* (thèse de doct.). University of Oxford. Oxford, UK.
- Arribas, A., Robertson, K. B. & Mylne, K. R. (2005). Test of a poor man's ensemble prediction system for short-range probability forecasting. *Monthly Weather Review*, 133(7), 1825-1839. <https://doi.org/10.1175/MWR2911.1>
- Aster, R. C., Borchers, B. & Thurber, C. H. (2018). *Parameter estimation and inverse problems*. Elsevier.
- Baker, L., Rudd, A., Migliorini, S. & Bannister, R. (2014). Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, 21(1), 19-39. <https://doi.org/10.5194/npg-21-19-2014>
- Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. & Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the COSMO model : Description and sensitivities. *Monthly Weather Review*, 139(12), 3887-3905. <https://doi.org/10.1175/MWR-D-10-05013.1>
- Barker, H. W., Cole, J., Morcrette, J.-J., Pincus, R., Räisänen, P., von Salzen, K. & Vaillancourt, P. A. (2008). The Monte Carlo independent column approximation : An assessment using several global atmospheric models. *Quarterly Journal of the Royal Meteorological Society*, 134(635), 1463-1478. <https://doi.org/10.1002/qj.303>
- Barker, H. W., Stephens, G. L. & Fu, Q. (1999). The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Quarterly Journal of the Royal Meteorological Society*, 125(558), 2127-2152. <https://doi.org/10.1002/qj.49712555810>
- Barthlott, C. & Kalthoff, N. (2011). A numerical sensitivity study on the impact of soil moisture on convection-related parameters and convective precipitation over complex terrain.

- Journal of the atmospheric sciences*, 68(12), 2971-2987. <https://doi.org/10.1175/JAS-D-11-027.1>
- Batté, L. & Doblas-Reyes, F. J. (2015). Stochastic atmospheric perturbations in the EC-Earth3 global coupled model : Impact of SPPT on seasonal forecast quality. *Climate Dynamics*, 45(11), 3419-3439. <https://doi.org/10.1007/s00382-015-2548-7>
- Bechtold, P., Bazile, E., Guichard, F., Mascart, P. & Richard, E. (2001). A mass-flux convection scheme for regional and global models. *Quarterly Journal of the Royal Meteorological Society*, 127(573), 869-886. <https://doi.org/10.1002/qj.49712757309>
- Bechtold, P., Cuijpers, J. W. M., Mascart, P. & Trouilhet, P. (1995). Modeling of trade wind cumuli with a low-order turbulence model : Toward a unified description of Cu and Se clouds in meteorological models. *Journal of the atmospheric sciences*, 52(4), 455-463. [https://doi.org/10.1175/1520-0469\(1995\)052<0455:MOTWCW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0455:MOTWCW>2.0.CO;2)
- Bechtold, P., Semane, N., Lopez, P., Chaboureaud, J.-P., Beljaars, A. & Bormann, N. (2014). Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, 71(2), 734-753. <https://doi.org/10.1175/JAS-D-13-0163.1>
- Bengtsson, L. (2022). *On cellular automata in stochastic convective parameterizations over the years - feats and defeats* [ECMWF Workshop on model uncertainty]. Récupérée 24 août 2022, à partir de https://events.ecmwf.int/event/290/contributions/2963/attachments/1828/3310/Model-WS_Bengtsson.pdf
- Bengtsson, L., Dias, J., Tulich, S., Gehne, M. & Bao, J.-W. (2021). A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System. *Journal of Advances in Modeling Earth Systems*, 13(1), e2020MS002260. <https://doi.org/10.1029/2020MS002260>
- Bengtsson, L., Steinheimer, M., Bechtold, P. & Geleyn, J.-F. (2013). A stochastic parameterization for deep convection using cellular automata. *Quarterly Journal of the Royal Meteorological Society*, 139(675), 1533-1543. <https://doi.org/10.1002/qj.2108>
- Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Camara, A., Christensen, H. M., Colangeli, M., Coleman, D. R. B., Crommelin, D., Dolaptchiev, S. I., Franzke, C. L. E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., ... Yano, J.-I. (2017). Stochastic parameterization : Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3), 565-588. <https://doi.org/10.1175/BAMS-D-15-00268.1>
- Berner, J., Fossell, K., Ha, S.-Y., Hacker, J. & Snyder, C. (2015). Increasing the skill of probabilistic forecasts : Understanding performance improvements from model-error representations. *Monthly Weather Review*, 143(4), 1295-1320. <https://doi.org/10.1175/MWR-D-14-00091.1>
- Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A. & Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system : Stochastic versus multiphysics representations. *Monthly Weather Review*, 139(6), 1972-1995. <https://doi.org/10.1175/2010MWR3595.1>
- Berner, J., Shutts, G. J., Leutbecher, M. & Palmer, T. N. (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences*, 66(3), 603-626. <https://doi.org/10.1175/2008JAS2677.1>
- Berre, L., Pannekoucke, O., Desroziers, G., Stefanescu, S. E., Chapnik, B. & Raynaud, L. (2007). A variational assimilation ensemble and the spatial filtering of its error covariances : Increase of sample size by local spatial averaging. *Proceedings of the ECMWF Workshop on Flow-Dependent Aspects of Data Assimilation*, 151-168. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/node/8172>

- Bougeault, P. (1981). Modeling the trade-wind cumulus boundary layer. Part I : Testing the ensemble cloud relations against numerical data. *Journal of Atmospheric Sciences*, 38(11), 2414-2428. [https://doi.org/10.1175/1520-0469\(1981\)038<2414:MTTWCB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<2414:MTTWCB>2.0.CO;2)
- Bougeault, P. (1982). Cloud-ensemble relations based on the gamma probability distribution for the higher-order models of the planetary boundary layer. *Journal of Atmospheric Sciences*, 39(12), 2691-2700. [https://doi.org/10.1175/1520-0469\(1982\)039<2691:CERBOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<2691:CERBOT>2.0.CO;2)
- Bougeault, P. & Lacarrère, P. (1989). Parameterization of orography-induced turbulence in a mesobeta-scale model. *Monthly weather review*, 117(8), 1872-1890. [https://doi.org/10.1175/1520-0493\(1989\)117<1872:POOITI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1872:POOITI>2.0.CO;2)
- Bouteloup, Y., Seity, Y. & Bazile, E. (2011). Description of the sedimentation scheme used operationally in all Météo-France NWP models. *Tellus A : Dynamic Meteorology and Oceanography*, 63(2), 300-311. <https://doi.org/10.1111/j.1600-0870.2010.00484.x>
- Boutle, I., Angevine, W., Bao, J.-W., Bergot, T., Bhattacharya, R., Bott, A., Ducongé, L., Forbes, R., Goecke, T., Grell, E., Hill, A., Igel, A., Kudzotsa, I., Lac, C., Maronga, B., Romakkaniemi, S., Schmidli, J., Schwenkel, J., Steeneveld, G.-J. & Vié, B. (2022). Demistify : a large-eddy simulation (LES) and single-column model (SCM) intercomparison of radiation fog. *Atmospheric Chemistry and Physics*, 22(1), 319-333. <https://doi.org/10.5194/acp-22-319-2022>
- Boutle, I., Price, J. D., Kudzotsa, I., Kokkola, H. & Romakkaniemi, S. (2018). Aerosol-fog interaction and the transition to well-mixed radiation fog. *Atmospheric Chemistry and Physics*, 18(11), 7827-7840. <https://doi.org/10.5194/acp-18-7827-2018>
- Bouttier, F. & Raynaud, L. (2018). Clustering and selection of boundary conditions for limited-area ensemble prediction. *Quarterly Journal of the Royal Meteorological Society*, 144(717), 2381-2391. <https://doi.org/10.1002/qj.3304>
- Bouttier, F., Raynaud, L., Nuissier, O. & Ménétrier, B. (2015). Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quarterly Journal of the Royal Meteorological Society*, 142, 390-403. <https://doi.org/10.1002/qj.2622>
- Bouttier, F., Vié, B., Nuissier, O. & Raynaud, L. (2012). Impact of stochastic physics in a convection-permitting ensemble. *Monthly Weather Review*, 140(11), 3706-3721. <https://doi.org/10.1175/MWR-D-12-00031.1>
- Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. & Beare, S. E. (2008). The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134(632), 703-722. <https://doi.org/10.1002/qj.234>
- Brient, F., Couvreur, F., Villefranque, N., Rio, C. & Honnert, R. (2019). Object-oriented identification of coherent structures in large eddy simulations : Importance of downdrafts in stratocumulus. *Geophysical Research Letters*, 46(5), 2854-2864. <https://doi.org/10.1029/2018GL081499>
- Bright, D. R. & Mullen, S. L. (2002). Short-range ensemble forecasts of precipitation during the southwest monsoon. *Weather and Forecasting*, 17(5), 1080-1100. [https://doi.org/10.1175/1520-0434\(2002\)017<1080:SREFOP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<1080:SREFOP>2.0.CO;2)
- Brousseau, P., Seity, Y., Ricard, D. & Léger, J. (2016). Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142(699), 2231-2243. <https://doi.org/10.1002/qj.2822>
- Brown, A. R. (1996). Evaluation of parametrization schemes for the convective boundary layer using large-eddy simulation results. *Boundary-Layer Meteorology*, 81(2), 167-200. <https://doi.org/10.1007/BF00119064>

- Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C., Khairoutdinov, M., Lewellen, D. C., Lock, A. P., MacVean, M. K., Moeng, C.-H. et al. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1075-1093. <https://doi.org/10.1256/003590002320373210>
- Brown, B. G. & Murphy, A. H. (1996). Improving forecasting performance by combining forecasts : The example of road-surface temperature forecasts. *Meteorological Applications*, 3(3), 257-265. <https://doi.org/10.1002/met.5060030307>
- Bubnová, R., Hello, G., Bénard, P. & Geleyn, J.-F. (1995). Integration of the fully elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/Aladin NWP system. *Monthly weather review*, 123(2), 515-535. [https://doi.org/10.1175/1520-0493\(1995\)123<0515:IOTFEE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<0515:IOTFEE>2.0.CO;2)
- Buizza, R., Leutbecher, M. & Isaksen, L. (2008). Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 2051-2066. <https://doi.org/10.1002/qj.346>
- Buizza, R., Miller, M. & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560), 2887-2908. <https://doi.org/10.1002/qj.49712556006>
- Burnet, F., Lac, C., Martinet, P., Fourrié, N., Haeffelin, M., Delanoë, J., Price, J., Barrau, S., Canut, G., Cayez, G., Dabas, A., Denjean, C., Dupont, J.-C., Honnert, R., Mahfouf, J.-F., Montmerle, T., Roberts, G., Seity, Y. & Vié, B. (2020). The SOuth west FOGs 3D experiment for processes study (SOFOG3D) project. *EGU General Assembly 2020*. <https://doi.org/10.5194/egusphere-egu2020-17836>
- Butler, T., Jakeman, J. & Wildey, T. (2018). Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems. *SIAM Journal on Scientific Computing*, 40(2), A984-A1011. <https://doi.org/10.1137/16M1087229>
- Candille, G. (2009). The multiensemble approach : The NAEFS example. *Monthly Weather Review*, 137(5), 1655-1665. <https://doi.org/10.1175/2008MWR2682.1>
- Caniaux, G., Redelsperger, J.-L. & Lafore, J.-P. (1994). A numerical study of the stratiform region of a fast-moving squall line. Part I : General description and water and heat budgets. *Journal of the atmospheric sciences*, 51(14), 2046-2074. [https://doi.org/10.1175/1520-0469\(1994\)051<2046:ANSOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<2046:ANSOTS>2.0.CO;2)
- Chaboureaud, J.-P. & Bechtold, P. (2002). A simple cloud parameterization derived from cloud resolving model data : Diagnostic and prognostic applications. *Journal of the atmospheric sciences*, 59(15), 2362-2372. [https://doi.org/10.1175/1520-0469\(2002\)059<2362:ASCPDF>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<2362:ASCPDF>2.0.CO;2)
- Charron, M., Pellerin, G., Spacek, L., Houtekamer, P., Gagnon, N., Mitchell, H. L. & Michelin, L. (2010). Toward random sampling of model error in the Canadian ensemble prediction system. *Monthly Weather Review*, 138(5), 1877-1901. <https://doi.org/10.1175/2009MWR3187.1>
- Chen, X., Bryan, G. H., Hazelton, A., Marks, F. D. & Fitzpatrick, P. (2022). Evaluation and Improvement of a TKE-Based Eddy-Diffusivity Mass-Flux (EDMF) Planetary Boundary Layer Scheme in Hurricane Conditions. *Weather and Forecasting*, 37(6), 935-951. <https://doi.org/10.1175/WAF-D-21-0168.1>
- Chlond, A., Müller, F. & Sednev, I. (2004). Numerical simulation of the diurnal cycle of marine stratocumulus during FIRE—An LES and SCM modelling study. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3297-3321. <https://doi.org/10.1256/qj.03.128>

- Christensen, H. M. (2020). Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society*, 146(727), 938-962. <https://doi.org/10.1002/qj.3717>
- Christensen, H. M., Lock, S.-J., Moroz, I. M. & Palmer, T. N. (2017). Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society*, 143(706), 2168-2181. <https://doi.org/10.1002/qj.3075>
- Christensen, H. M., Moroz, I. M. & Palmer, T. N. (2015). Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *Journal of the Atmospheric Sciences*, 72(6), 2525-2544. <https://doi.org/10.1175/JAS-D-14-0250.1>
- Clark, P., Halliwell, C. & Flack, D. (2021). A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part I : Formulation and Evaluation in a Convection-Permitting Model. *Journal of the Atmospheric Sciences*, 78(3), 727-746. <https://doi.org/10.1175/JAS-D-19-0291.1>
- Cohen, B. G. & Craig, G. C. (2006). Fluctuations in an equilibrium convective ensemble. Part II : Numerical experiments. *Journal of the atmospheric sciences*, 63(8), 2005-2015. <https://doi.org/10.1175/JAS3710.1>
- Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Rotunno, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T. & Tarpley, J. D. (2003). Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). *Journal of Geophysical Research : Atmospheres*, 108(D22). <https://doi.org/10.1029/2002JD003316>
- Courtier, P., Freyrier, C., Geleyn, J.-F., Rabier, F. & Rochas, M. (1991). The Arpege project at Météo France. *Seminar on Numerical Methods in Atmospheric Models, 9-13 September 1991*, 193-232. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/node/8798>
- Couvreur, F., Hourdin, F. & Rio, C. (2010). Resolved versus parametrized boundary-layer plumes. Part I : A parametrization-oriented conditional sampling in large-eddy simulations. *Boundary-layer meteorology*, 134(3), 441-458. <https://doi.org/10.1007/s10546-009-9456-5>
- Couvreur, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Briant, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q. & Xu, W. (2021). Process-based climate model development harnessing machine learning : I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, 13(3), e2020MS002217. <https://doi.org/10.1029/2020MS002217>
- Craig, G. C. & Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. Part I : Theoretical formulation. *Journal of the atmospheric sciences*, 63(8), 1996-2004. <https://doi.org/10.1175/JAS3709.1>
- Cuijpers, J. & Bechtold, P. (1995). A simple parameterization of cloud water related variables for use in boundary layer models. *Journal of Atmospheric Sciences*, 52(13), 2486-2490. [https://doi.org/10.1175/1520-0469\(1995\)052<2486:ASPOCW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<2486:ASPOCW>2.0.CO;2)
- Cuxart, J., Bougeault, P. & Redelsperger, C. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, 126(562), 1-30. <https://doi.org/10.1002/qj.49712656202>
- Deardorff, J. W. (1972a). Theoretical expression for the countergradient vertical heat flux. *Journal of Geophysical Research*, 77(30), 5900-5904. <https://doi.org/10.1029/JC077i030p05900>

- Deardorff, J. W. (1972b). Numerical investigation of neutral and unstable planetary boundary layers. *Journal of Atmospheric Sciences*, 29(1), 91-115. [https://doi.org/10.1175/1520-0469\(1972\)029<0091:NIONAU>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<0091:NIONAU>2.0.CO;2)
- de Rooy, W. C., Siebesma, P., Baas, P., Lenderink, G., De Roode, S. R., De Vries, H., Van Meijgaard, E., Meirink, J. F., Tijm, S. & Van't Veen, B. (2022). Model development in practice : a comprehensive update to the boundary layer schemes in HARMONIE-AROME cycle 40. *Geoscientific Model Development*, 15(4), 1513-1543. <https://doi.org/10.5194/gmd-15-1513-2022>
- Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P. & Cébron, P. (2015). PEARP, the Météo-France short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1671-1685. <https://doi.org/10.1002/qj.2469>
- Desroziers, G., Berre, L., Chabot, V. & Chapnik, B. (2009). A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137(10), 3420-3436. <https://doi.org/10.1175/2009MWR2778.1>
- Dombrowski-Etchevers, I., Sanchez, I. & Seity, Y. (2018). *La visibilité : nouveau paramètre issu des modèles de prévision météorologique* [Ateliers de Modélisation de l'Atmosphère - Février 2018]. Récupérée 1 octobre 2022, à partir de http://www.meteo.fr/cic/meetings/2018/AMA/presentations/AMA_1402/13-%5C%20presentation_AMAs2018_Visi.pdf
- Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J. & Vogel, G. (2011). *A description of the non-hydrostatic regional COSMO model, Part II : physical parameterization* (rapp. tech.). Récupérée 22 juillet 2022, à partir de <http://www.cosmo-model.org/content/model/documentation/core/>
- Dorrestijn, J., Crommelin, D. T., Siebesma, A. P. & Jonker, H. J. (2013). Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theoretical and Computational Fluid Dynamics*, 27(1-2), 133-148. <https://doi.org/10.1007/s00162-012-0281-y>
- Draper, C. S. (2021). Accounting for Land Model Uncertainty in Numerical Weather Prediction Ensemble Systems : Toward Ensemble-Based Coupled Land-Atmosphere Data Assimilation. *Journal of Hydrometeorology*, 22(8), 2089-2104. <https://doi.org/10.1175/JHM-D-21-0016.1>
- Duda, J. D., Wang, X., Kong, F. & Xue, M. (2014). Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Monthly Weather Review*, 142(6), 2198-2219. <https://doi.org/10.1175/MWR-D-13-00297.1>
- Duynkerke, P. G., de Roode, S. R., van Zanten, M. C., Calvo, J., Cuxart, J., Cheinet, S., Chlond, A., Grenier, H., Jonker, P. J., Köhler, M. et al. (2004). Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3269-3296. <https://doi.org/10.1256/qj.03.139>
- Ebert, E. E. & Curry, J. A. (1992). A parameterization of ice cloud optical properties for climate models. *Journal of Geophysical Research : Atmospheres*, 97(D4), 3831-3836. <https://doi.org/10.1029/91JD02472>
- Eckel, F. A. & Mass, C. F. (2005). Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, 20(3), 328-350. <https://doi.org/10.1175/WAF843.1>
- Ehrendorfer, M. (1994). The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I : Theory. *Monthly Weather Review*, 122(4), 703-713. [https://doi.org/10.1175/1520-0493\(1994\)122<0703:TLEAIP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0703:TLEAIP>2.0.CO;2)

- Emanuel, K. A. (1991). A scheme for representing cumulus convection in large-scale models. *Journal of the atmospheric sciences*, 48(21), 2313-2329. [https://doi.org/10.1175/1520-0469\(1991\)048<2313:ASFRCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1991)048<2313:ASFRCC>2.0.CO;2)
- Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus*, 21(6), 739-759. <https://doi.org/10.3402/tellusa.v21i6.10143>
- Errico, R. & Baumhefner, D. (1987). Predictability experiments using a high-resolution limited-area model. *Monthly Weather Review*, 115(2), 488-504. [https://doi.org/10.1175/1520-0493\(1987\)115<0488:PEUAHR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0488:PEUAHR>2.0.CO;2)
- Evans, R., Harrison, M., Graham, R. & Mylne, K. (2000). Joint medium-range ensembles from the Met. Office and ECMWF systems. *Monthly weather review*, 128(9), 3104-3127. [https://doi.org/10.1175/1520-0493\(2000\)128<3104:JMREFT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3104:JMREFT>2.0.CO;2)
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143-10162. <https://doi.org/10.1029/94JC00572>
- Fleming, R. J. (1971). On stochastic dynamic prediction : I. The energetics of uncertainty and the question of closure. *Monthly Weather Review*, 99(11), 851-872. [https://doi.org/10.1175/1520-0493\(1971\)099<0851:OSDP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1971)099<0851:OSDP>2.3.CO;2)
- Fouquart, Y. & Bonnel, F. (1980). Computations of solar heating of the Earth's atmosphere : A new parameterization. *Beiträge zur Physik der Atmosphäre*, 53, 35-62.
- Fresnay, S., Hally, A., Garnaud, C., Richard, E. & Lambert, D. (2012). Heavy precipitation events in the Mediterranean : sensitivity to cloud physics parameterisation uncertainties. *Natural Hazards and Earth System Sciences*, 12(8), 2671-2688. <https://doi.org/10.5194/nhess-12-2671-2012>
- Fritsch, J. M., Hilliker, J., Ross, J. & Vislocky, R. L. (2000). Model consensus. *Weather and forecasting*, 15(5), 571-582. [https://doi.org/10.1175/1520-0434\(2000\)015<0571:MC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0571:MC>2.0.CO;2)
- Frogner, I.-L., Andrae, U., Ollinaho, P., Hally, A., Härmäläinen, K., Kauhanen, J., Ivarsson, K.-I. & Yazgi, D. (2022). Model Uncertainty Representation in a Convection-Permitting Ensemble—SPP and SPPT in HarmonEPS. *Monthly Weather Review*, 150(4), 775-795. <https://doi.org/10.1175/MWR-D-21-0099.1>
- Frogner, I.-L. & Iversen, T. (2002). High-resolution limited-area ensemble predictions based on low-resolution targeted singular vectors. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1321-1341. <https://doi.org/10.1256/003590002320373319>
- Fu, R., Del Genio, A. D. & Rossow, W. B. (1994). Influence of Ocean Surface Conditions on Atmospheric Vertical Thermodynamic Structure and Deep Convection. *Journal of Climate*, 7(7), 1092-1108. [https://doi.org/10.1175/1520-0442\(1994\)007<1092:IOOSCO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1092:IOOSCO>2.0.CO;2)
- Fujita, T., Stensrud, D. J. & Dowell, D. C. (2007). Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Monthly weather review*, 135(5), 1846-1868. <https://doi.org/10.1175/MWR3391.1>
- Gage, K. S. (1979). Evidence for a $k^{-5/3}$ law inertial range in mesoscale two-dimensional turbulence. *Journal of Atmospheric Sciences*, 36(10), 1950-1954. [https://doi.org/10.1175/1520-0469\(1979\)036<1950:EFALIR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<1950:EFALIR>2.0.CO;2)
- Gallus, J., William A, Wolff, J., Gotway, J. H., Harrold, M., Blank, L. & Beck, J. (2019). The impacts of using mixed physics in the Community Leveraged Unified Ensemble. *Weather and Forecasting*, 34(4), 849-867. <https://doi.org/10.1175/WAF-D-18-0197.1>

- Garcia, J. R. & Mellado, J. P. (2014). The two-layer structure of the entrainment zone in the convective boundary layer. *Journal of the Atmospheric Sciences*, 71(6), 1935-1955. <https://doi.org/10.1175/JAS-D-13-0148.1>
- García-Moya, J.-A., Callado, A., Escribà, P., Santos, C., Santos-Muñoz, D. & Simarro, J. (2011). Predictability of short-range forecasting : a multimodel approach. *Tellus A : Dynamic Meteorology and Oceanography*, 63(3), 550-563. <https://doi.org/10.1111/j.1600-0870.2010.00506.x>
- Gebhardt, C., Theis, S. E., Paulat, M. & Bouallègue, Z. B. (2011). Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, 100(2-3), 168-177. <https://doi.org/10.1016/j.atmosres.2010.12.008>
- Geleyn, J.-F., Catry, B., Bouteloup, Y. & Brožková, R. (2008). A statistical approach for sedimentation inside a microphysical precipitation scheme. *Tellus A : Dynamic Meteorology and Oceanography*, 60(4), 649-662. <https://doi.org/10.1111/j.1600-0870.2007.00323.x>
- Gerard, L., Piriou, J.-M., Brožková, R., Geleyn, J.-F. & Banciu, D. (2009). Cloud and precipitation parameterization in a meso-gamma-scale operational weather prediction model. *Monthly Weather Review*, 137(11), 3960-3977. <https://doi.org/10.1175/2009MWR2750.1>
- Gilmore, M. S., Straka, J. M. & Rasmussen, E. N. (2004). Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Monthly weather review*, 132(11), 2610-2627. <https://doi.org/10.1175/MWR2810.1>
- Gleeson, T. A. (1970). Statistical-dynamical predictions. *Journal of Applied Meteorology and Climatology*, 9(3), 333-344. [https://doi.org/10.1175/1520-0450\(1970\)009<0333:SDP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0333:SDP>2.0.CO;2)
- Golaz, J.-C., Larson, V. E. & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I : Method and model description. *Journal of the atmospheric sciences*, 59(24), 3540-3551. [https://doi.org/10.1175/1520-0469\(2002\)059<3540:APBMFB>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2)
- Grabowski, W. W. (2003). Impact of cloud microphysics on convective-radiative quasi equilibrium revealed by cloud-resolving convection parameterization. *Journal of Climate*, 16(21), 3463-3475. [https://doi.org/10.1175/1520-0442\(2003\)016<3463:IOCMOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3463:IOCMOC>2.0.CO;2)
- Grabowski, W. W., Morrison, H., Shima, S.-I., Abade, G. C., Dziekan, P. & Pawlowska, H. (2019). Modeling of cloud microphysics : Can we do better? *Bulletin of the American Meteorological Society*, 100(4), 655-672. <https://doi.org/10.1175/BAMS-D-18-0005.1>
- Grabowski, W. W. & Wang, L.-P. (2013). Growth of cloud droplets in a turbulent environment. *Annual review of fluid mechanics*, 45(1), 293-324. <https://doi.org/10.1146/annurev-fluid-011212-140750>
- Grimit, E. P. & Mass, C. F. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting*, 17(2), 192-205. [https://doi.org/10.1175/1520-0434\(2002\)017<0192:IROAMS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0192:IROAMS>2.0.CO;2)
- Hacker, J. P., Snyder, C., Ha, S.-Y. & Pocerlich, M. (2011a). Linear and non-linear response to parameter variations in a mesoscale model. *Tellus A : Dynamic Meteorology and Oceanography*, 63(3), 429-444. <https://doi.org/10.1111/j.1600-0870.2010.00505.x>
- Hacker, J. P., Ha, S.-Y., Snyder, C., Berner, J., Eckel, F., Kuchera, E., Pocerlich, M., Rugg, S., Schramm, J. & Wang, X. (2011b). The US Air Force Weather Agency's mesoscale ensemble : scientific description and performance results. *Tellus A : Dynamic Meteorology and Oceanography*, 63(3), 625-641. <https://doi.org/10.1111/j.1600-0870.2010.00497.x>

- Hagedorn, R., Doblas-Reyes, F. J. & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting — I. Basic concept. *Tellus A : Dynamic Meteorology and Oceanography*, 57(3), 219-233. <https://doi.org/10.3402/tellusa.v57i3.14657>
- Hall, W. D. (1980). A detailed microphysical model within a two-dimensional dynamic framework : Model description and preliminary results. *Journal of Atmospheric Sciences*, 37(11), 2486-2507. [https://doi.org/10.1175/1520-0469\(1980\)037<2486:ADMMWA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1980)037<2486:ADMMWA>2.0.CO;2)
- Han, J., Witek, M. L., Teixeira, J., Sun, R., Pan, H.-L., Fletcher, J. K. & Bretherton, C. S. (2016). Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Weather and Forecasting*, 31(1), 341-352. <https://doi.org/10.1175/WAF-D-15-0053.1>
- Harrison, M. S. J., Palmer, T. N., Richardson, D. S. & Buizza, R. (1999). Analysis and model dependencies in medium-range ensembles : Two transplant case-studies. *Quarterly Journal of the Royal Meteorological Society*, 125(559), 2487-2515. <https://doi.org/10.1002/qj.49712555908>
- Hirt, M., Rasp, S., Blahak, U. & Craig, G. C. (2019). Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Monthly Weather Review*, 147(11), 3917-3934. <https://doi.org/10.1175/MWR-D-19-0060.1>
- Hohenegger, C., Walser, A., Langhans, W. & Schär, C. (2008). Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quarterly Journal of the Royal Meteorological Society*, 134(633), 889-904. <https://doi.org/10.1002/qj.252>
- Holland, J. Z. & Rasmusson, E. M. (1973). Measurements of the atmospheric mass, energy, and momentum budgets over a 500-kilometer square of tropical ocean. *Monthly Weather Review*, 101(1), 44-55. [https://doi.org/10.1175/1520-0493\(1973\)101<0044:MOTAME>2.3.CO;2](https://doi.org/10.1175/1520-0493(1973)101<0044:MOTAME>2.3.CO;2)
- Hollingsworth, A. (1980). An experiment in Monte Carlo forecasting. *Proceedings of the ECMWF Workshop on Stochastic Dynamic Forecasting*, 65-85. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/node/9945>
- Honnert, R., Couvreux, F., Masson, V. & Lancz, D. (2016). Sampling the structure of convective turbulence and implications for grey-zone parametrizations. *Boundary-layer meteorology*, 160(1), 133-156. <https://doi.org/10.1007/s10546-016-0130-4>
- Honnert, R., Masson, V. & Couvreux, F. (2011). A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometeric scale. *Journal of the Atmospheric Sciences*, 68(12), 3112-3131. <https://doi.org/10.1175/JAS-D-11-061.1>
- Hou, D., Kalnay, E. & Droegemeier, K. K. (2001). Objective verification of the SAMEX'98 ensemble forecasts. *Monthly Weather Review*, 129(1), 73-91. [https://doi.org/10.1175/1520-0493\(2001\)129<0073:OVOTSE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2)
- Hourdin, F., Couvreux, F. & Menut, L. (2002). Parameterization of the dry convective boundary layer based on a mass flux representation of thermals. *Journal of the atmospheric sciences*, 59(6), 1105-1123. [https://doi.org/10.1175/1520-0469\(2002\)059<1105:POTDCB>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1105:POTDCB>2.0.CO;2)
- Houtekamer, P. L. (1995). The construction of optimal perturbations. *Monthly weather review*, 123(9), 2888-2898. [https://doi.org/10.1175/1520-0493\(1995\)123<2888:TCOOP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2888:TCOOP>2.0.CO;2)

- Houtekamer, P. L., Lefaiivre, L., Derome, J., Ritchie, H. & Mitchell, H. L. (1996). A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124(6), 1225-1242. [https://doi.org/10.1175/1520-0493\(1996\)124<1225:ASSATE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2)
- Houtekamer, P. L. & Mitchell, H. L. (2005). Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3269-3289. <https://doi.org/10.1256/qj.05.135>
- Igel, A. L. & van den Heever, S. C. (2017a). The Importance of the Shape of Cloud Droplet Size Distributions in Shallow Cumulus Clouds. Part I : Bin Microphysics Simulations. *Journal of the Atmospheric Sciences*, 74(1), 249-258. <https://doi.org/10.1175/JAS-D-15-0382.1>
- Igel, A. L. & van den Heever, S. C. (2017b). The importance of the shape of cloud droplet size distributions in shallow cumulus clouds. Part II : Bulk microphysics simulations. *Journal of the Atmospheric Sciences*, 74(1), 259-273. <https://doi.org/10.1175/JAS-D-15-0383.1>
- Isaksen, L., Haseler, J., Buizza, R. & Leutbecher, M. (2010). The new Ensemble of Data Assimilations. *ECMWF Newsletter*, (123), 17-21. <https://doi.org/10.21957/vzbtkbf8>
- Iversen, T., Deckmyn, A., Santos, C., Sattler, K., Bremnes, J. B., Feddersen, H. & Frogner, I.-L. (2011). Evaluation of ‘GLAMEPS’—a proposed multimodel EPS for short range forecasting. *Tellus A : Dynamic Meteorology and Oceanography*, 63(3), 513-530. <https://doi.org/10.1111/j.1600-0870.2010.00507.x>
- Jahangir, E., Libois, Q., Couvreur, F., Vié, B. & Saint-Martin, D. (2021). Uncertainty of SW cloud radiative effect in atmospheric models due to the parameterization of liquid cloud optical properties. *Journal of Advances in Modeling Earth Systems*, 13(12), e2021MS002742. <https://doi.org/10.1029/2021MS002742>
- Jam, A., Hourdin, F., Rio, C. & Couvreur, F. (2013). Resolved versus parametrized boundary-layer plumes. Part III : Derivation of a statistical scheme for cumulus clouds. *Boundary-layer meteorology*, 147(3), 421-441. <https://doi.org/10.1007/s10546-012-9789-3>
- Jankov, I., Beck, J., Wolff, J., Harrold, M., Olson, J. B., Smirnova, T., Alexander, C. & Berner, J. (2019). Stochastically perturbed parameterizations in an HRRR-based ensemble. *Monthly Weather Review*, 147(1), 153-173. <https://doi.org/10.1175/MWR-D-18-0092.1>
- Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., Smirnova, T. G., Benjamin, S. G. & Brown, J. M. (2017). A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Monthly Weather Review*, 145(4), 1161-1179. <https://doi.org/10.1175/MWR-D-16-0160.1>
- Järvinen, H., Laine, M., Solonen, A. & Haario, H. (2012). Ensemble prediction and parameter estimation system : The concept. *Quarterly journal of the royal meteorological society*, 138(663), 281-288. <https://doi.org/10.1002/qj.923>
- Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A. & Haario, H. (2010). Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmospheric Chemistry and Physics*, 10(20), 9993-10002. <https://doi.org/10.5194/acp-10-9993-2010>
- Jeffery, C. A., Reisner, J. M. & Andrejczuk, M. (2007). Another look at stochastic condensation for subgrid cloud modeling : Adiabatic evolution and effects. *Journal of the atmospheric sciences*, 64(11), 3949-3969. <https://doi.org/10.1175/2006JAS2147.1>
- Kain, J. S. & Fritsch, J. M. (1990). A one-dimensional entraining/detraining plume model and its application in convective parameterization. *Journal of the Atmospheric Sciences*, 47(23), 2784-2802. [https://doi.org/10.1175/1520-0469\(1990\)047<2784:AODEPM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2)
- Kalina, E. A., Jankov, I., Alcott, T., Olson, J., Beck, J., Berner, J., Dowell, D. & Alexander, C. (2021). A progress report on the development of the High-Resolution Rapid Refresh

- ensemble. *Weather and Forecasting*, 36(3), 791-804. <https://doi.org/10.1175/WAF-D-20-0098.1>
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1), 35-45. <https://doi.org/10.1115/1.3662552>
- Kessler, E. (1969). *On the distribution and continuity of water substance in atmospheric circulations*. Springer. https://doi.org/10.1007/978-1-935704-36-2_1
- Khvorostyanov, V. I. & Curry, J. A. (1999a). Toward the theory of stochastic condensation in clouds. Part I : A general kinetic equation. *Journal of the atmospheric sciences*, 56(23), 3985-3996. [https://doi.org/10.1175/1520-0469\(1999\)056<3985:TTTOSC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3985:TTTOSC>2.0.CO;2)
- Khvorostyanov, V. I. & Curry, J. A. (1999b). Toward the theory of stochastic condensation in clouds. Part II : analytical solutions of the gamma-distribution type. *Journal of the atmospheric sciences*, 56(23), 3997-4013. [https://doi.org/10.1175/1520-0469\(1999\)056<3997:TTTOSC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3997:TTTOSC>2.0.CO;2)
- Klasa, C., Arpagaus, M., Walser, A. & Wernli, H. (2018). An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. *Quarterly Journal of the Royal Meteorological Society*, 144(712), 744-764. <https://doi.org/10.1002/qj.3245>
- Kober, K. & Craig, G. C. (2016). Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences*, 73(7), 2893-2911. <https://doi.org/10.1175/JAS-D-15-0144.1>
- Kolmogorov, A. N. (1941). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Doklady Akademii Nauk*, 30, 9-13. Récupérée 7 septembre 2022, à partir de <https://www.jstor.org/stable/51980>
- Koschmieder, H. (1924). Theorie der horizontalen Sichtweite. *Beitrage zur Physik der freien Atmosphere*, 12, 33-53.
- Kosmopoulos, P. G., Kazadzis, S., Lagouvardos, K., Kotroni, V. & Bais, A. (2015). Solar energy prediction and verification using operational model forecasts and ground-based solar measurements. *Energy*, 93, 1918-1930. <https://doi.org/10.1016/j.energy.2015.10.054>
- Kunkel, B. A. (1984). Parameterization of droplet terminal velocity and extinction coefficient in fog models. *Journal of Applied Meteorology and Climatology*, 23(1), 34-41. [https://doi.org/10.1175/1520-0450\(1984\)023<0034:PODTVA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<0034:PODTVA>2.0.CO;2)
- Kuo, H.-L. (1974). Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *Journal of Atmospheric Sciences*, 31(5), 1232-1240. [https://doi.org/10.1175/1520-0469\(1974\)031<1232:FSOTPO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<1232:FSOTPO>2.0.CO;2)
- Lac, C., Chaboureaud, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., Leriche, M., Barthe, C., Aouizerats, B., Augros, C., Aumond, P., Auguste, F., Bechtold, P., Berthet, S., Bielli, S., Bosseur, F., Caumont, O., Cohard, J.-M., Colin, J., ... Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, 11(5), 1929-1969. <https://doi.org/10.5194/gmd-11-1929-2018>
- Lafore, J.-P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., Fischer, C., Hérelil, P., Mascart, P., Masson, V., Pinty, J.-P., Redelsperger, J.-L., Richard, E. & Vilà-Guerau de Arellano, J. (1998). The Meso-NH atmospheric simulation system. Part I : Adiabatic formulation and control simulations. *16(1)*, 90-109. <https://doi.org/10.1007/s00585-997-0090-6>
- Laine, M., Solonen, A., Haario, H. & Järvinen, H. (2012). Ensemble prediction and parameter estimation system : the method. *Quarterly Journal of the Royal Meteorological Society*, 138(663), 289-297. <https://doi.org/10.1002/qj.922>

- Lang, S. T. K., Lock, S.-J., Leutbecher, M., Bechtold, P. & Forbes, R. M. (2021). Revision of the stochastically perturbed parametrisations model uncertainty scheme in the integrated forecasting system. *Quarterly Journal of the Royal Meteorological Society*, 147(735), 1364-1381. <https://doi.org/10.1002/qj.3978>
- Lavaysse, C., Carrera, M., Bélair, S., Gagnon, N., Frenette, R., Charron, M. & Yau, M. K. (2013). Impact of surface parameter uncertainties within the Canadian regional ensemble prediction system. *Monthly weather review*, 141(5), 1506-1526. <https://doi.org/10.1175/MWR-D-11-00354.1>
- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B. & Slater, A. G. (2011). Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, 3(1). <https://doi.org/10.1029/2011MS00045>
- Le Moigne, P. (2012). *SURFEX scientific documentation* (rapp. tech.). Récupérée 22 juillet 2022, à partir de <https://www.umr-cnrm.fr/surfex/>
- Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly weather review*, 102(6), 409-418. [https://doi.org/10.1175/1520-0493\(1974\)102<0409:TSOMCF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2)
- Leoncini, G., Plant, R. S., Gray, S. L. & Clark, P. A. (2010). Perturbation growth at the convective scale for CSIP IOP18. *Quarterly Journal of the Royal Meteorological Society*, 136(648), 653-670. <https://doi.org/10.1002/qj.587>
- Leoncini, G., Plant, R. S., Gray, S. L. & Clark, P. A. (2013). Ensemble forecasts of a flood-producing storm : comparison of the influence of model-state perturbations and parameter modifications. *Quarterly Journal of the Royal Meteorological Society*, 139(670), 198-211. <https://doi.org/10.1002/qj.1951>
- Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H. M., Diamantakis, M., Dutra, E. et al. (2017). Stochastic representations of model uncertainties at ECMWF : State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2315-2339. <https://doi.org/10.1002/qj.3094>
- Lewellen, W. & Yoh, S. (1993). Binormal model of ensemble partial cloudiness. *Journal of the atmospheric sciences*, 50(9), 1228-1237. [https://doi.org/10.1175/1520-0469\(1993\)050<1228:BMOEPC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<1228:BMOEPC>2.0.CO;2)
- Li, R., Jin, J., Wang, S.-Y. & Gillies, R. R. (2015). Significant impacts of radiation physics in the Weather Research and Forecasting model on the precipitation and dynamics of the West African Monsoon. *Climate Dynamics*, 44(5), 1583-1594. <https://doi.org/10.1007/s00382-014-2294-2>
- Lilly, D. K. (1967). The Representation of Small-Scale Turbulence in Numerical Simulation Experiments. *IBM Scientific Computing Symposium on Environmental Sciences*, 195-210. Récupérée 5 septembre 2022, à partir de <http://courses.washington.edu/mengr544/handouts-12/lilly-67.pdf>
- Lilly, D. K. (1983). Stratified turbulence and the mesoscale variability of the atmosphere. *Journal of Atmospheric Sciences*, 40(3), 749-761. [https://doi.org/10.1175/1520-0469\(1983\)040<0749:STATMV>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<0749:STATMV>2.0.CO;2)
- Lock, S.-J., Lang, S. T. K., Leutbecher, M., Hogan, R. J. & Vitart, F. (2019). Treatment of model uncertainty from radiation by the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme and associated revisions in the ECMWF ensembles. *Quarterly Journal of the Royal Meteorological Society*, 145, 75-89. <https://doi.org/10.1002/qj.3570>

- Loftus, A. M., Cotton, W. R. & Carrió, G. G. (2014). A triple-moment hail bulk microphysics scheme. Part I : Description and initial evaluation. *Atmospheric research*, 149, 35-57. <https://doi.org/10.1016/j.atmosres.2014.05.013>
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2), 130-141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17(3), 321-333. <https://doi.org/10.3402/tellusa.v17i3.9076>
- Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus*, 21(3), 289-307. <https://doi.org/10.3402/tellusa.v21i3.10086>
- Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, 34(6), 505-513. <https://doi.org/10.3402/tellusa.v34i6.10836>
- Louis, J.-F. (1979). A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, 17(2), 187-202. <https://doi.org/10.1007/BF00117978>
- Lu, C., Liu, Y., Zhang, G. J., Wu, X., Endo, S., Cao, L., Li, Y. & Guo, X. (2016). Improving parameterization of entrainment rate for shallow convection with aircraft measurements and large-eddy simulation. *Journal of the Atmospheric Sciences*, 73(2), 761-773. <https://doi.org/10.1175/JAS-D-15-0050.1>
- Lupo, K. M., Torn, R. D. & Yang, S.-C. (2020). Evaluation of stochastic perturbed parameterization tendencies on convective-permitting ensemble forecasts of heavy rainfall events in New York and Taiwan. *Weather and Forecasting*, 35(1), 5-24. <https://doi.org/10.1175/WAF-D-19-0064.1>
- Machulskaya, E. & Seifert, A. (2019). Stochastic differential equations for the variability of atmospheric convection fluctuating around the equilibrium. *Journal of Advances in Modeling Earth Systems*, 11(8), 2708-2727. <https://doi.org/10.1029/2019MS001638>
- MacLeod, D. A., Cloke, H. L., Pappenberger, F. & Weisheimer, A. (2016). Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 79-90. <https://doi.org/https://doi.org/10.1002/qj.2631>
- Madec, G., Bell, M., Bourdallé-Badie, R., Chanut, J., Clementi, E., Coward, A., Drudi, M., Epicoco, I., Ethé, C., Iovino, D., Lea, D., Lévy, C., Martin, N., Masson, S., Mathiot, P., Mele, F., Mocavero, S., Moulin, A., Müeller, S., ... Storkey, D. (2022). *NEMO ocean engine* (Scientific Notes of Climate Modelling Center, 27, ISSN 1288-1619). Institut Pierre-Simon Laplace (IPSL). <https://doi.org/10.5281/zenodo.6334656>
- Malardel, S. (2008a). *MUSC : (Modèle Unifié, Simple Colonne) for Arpege-Aladin-Arome-Alaro-Hirlam-(IFS) (CY31T1 version)* (rapp. tech.). Météo-France. Récupérée 29 août 2022, à partir de http://www.umr-cnrm.fr/gmapdoc/IMG/pdf_DOC_1D_MODEL.pdf
- Malardel, S. (2008b). *SCUM or MUSC*. Récupérée 29 août 2022, à partir de http://www.umr-cnrm.fr/gmapdoc/IMG/pdf_scum.pdf
- Marshall, J. S. & Palmer, W. M. K. (1948). The distribution of raindrops with size. *Journal of Atmospheric Sciences*, 5(4), 165-166. [https://doi.org/10.1175/1520-0469\(1948\)005<0165:TDORWS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2)
- Marsigli, C., Boccanera, F., Montani, A. & Paccagnella, T. (2005). The COSMO-LEPS mesoscale ensemble system : validation of the methodology and verification. *Nonlinear Processes in Geophysics*, 12(4), 527-536. <https://doi.org/10.5194/npg-12-527-2005>
- Martin, G., Johnson, D. & Spice, A. (1994). The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *Journal of Atmospheric Sciences*,

- 51(13), 1823-1842. [https://doi.org/10.1175/1520-0469\(1994\)051<1823:TMAPOE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<1823:TMAPOE>2.0.CO;2)
- Mason, P. J. & Thomson, D. J. (1992). Stochastic backscatter in large-eddy simulations of boundary layers. *Journal of Fluid Mechanics*, 242, 51-78. <https://doi.org/10.1017/S0022112092002271>
- Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essauouini, K., Gibelin, A.-L., ... Voldoire, A. (2013). The SURFEXv7. 2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, 6(4), 929-960. <https://doi.org/10.5194/gmd-6-929-2013>
- McCabe, A., Swinbank, R., Tennant, W. & Lock, A. (2016). Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. *Quarterly Journal of the Royal Meteorological Society*, 142(700), 2897-2910. <https://doi.org/10.1002/qj.2876>
- McLay, J., Flatau, M., Reynolds, C., Cummings, J., Hogan, T. & Flatau, P. (2012). Inclusion of sea-surface temperature variation in the US Navy ensemble-transform global ensemble prediction system. *Journal of Geophysical Research : Atmospheres*, 117(D19). <https://doi.org/10.1029/2011JD016937>
- McTaggart-Cowan, R., Separovic, L., Aider, R., Charron, M., Desgagné, M., Houtekamer, P. L., Paquin-Ricard, D., Vaillancourt, P. A. & Zadra, A. (2022a). Using Stochastically Perturbed Parameterizations to Represent Model Uncertainty, Part I : Implementation and Parameter Sensitivity. *Monthly Weather Review*. <https://doi.org/10.1175/MWR-D-21-0315.1>
- McTaggart-Cowan, R., Separovic, L., Charron, M., Deng, X., Gagnon, N., Houtekamer, P. L. & Patoine, A. (2022b). Using Stochastically Perturbed Parameterizations to Represent Model Uncertainty, Part II : Comparison with Existing Techniques in an Operational Ensemble. *Monthly Weather Review*. <https://doi.org/10.1175/MWR-D-21-0316.1>
- Mellor, G. L. (1977). The Gaussian cloud model relations. *Journal of the Atmospheric Sciences*, 34(2), 356-358. [https://doi.org/10.1175/1520-0469\(1977\)034<0356:TGCMR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1977)034<0356:TGCMR>2.0.CO;2)
- Miles, N. L., Verlinde, J. & Clothiaux, E. E. (2000). Cloud droplet size distributions in low-level stratiform clouds. *Journal of the atmospheric sciences*, 57(2), 295-311. [https://doi.org/10.1175/1520-0469\(2000\)057<0295:CDSDIL>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<0295:CDSDIL>2.0.CO;2)
- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J. & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres : RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research : Atmospheres*, 102(D14), 16663-16682. <https://doi.org/10.1029/97JD00237>
- Moeng, C.-H. (1984). A large-eddy-simulation model for the study of planetary boundary-layer turbulence. *Journal of the Atmospheric Sciences*, 41(13), 2052-2062. [https://doi.org/10.1175/1520-0469\(1984\)041<2052:ALESMF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<2052:ALESMF>2.0.CO;2)
- Moin, P. & Mahesh, K. (1998). Direct numerical simulation : a tool in turbulence research. *Annual review of fluid mechanics*, 30, 539-578. <https://doi.org/10.1146/annurev.fluid.30.1.539>
- Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. & Paccagnella, T. (2001). A strategy for high-resolution ensemble prediction. I : Definition of representative members and global-model experiments. *Quarterly Journal of the Royal Meteorological Society*, 127(576), 2069-2094. <https://doi.org/https://doi.org/10.1002/qj.49712757612>

- Molteni, F., Buizza, R., Palmer, T. N. & Petroliagis, T. (1996). The ECMWF ensemble prediction system : Methodology and validation. *Quarterly journal of the royal meteorological society*, 122(529), 73-119. <https://doi.org/10.1002/qj.49712252905>
- Morcrette, J.-J. (2000). *Radiation transfer* (Meteorological Training Course Lecture Series). ECMWF. Récupérée 25 septembre 2022, à partir de <https://www.ecmwf.int/en/elibrary/16964-radiation-transfer>
- Morcrette, J.-J. & Fouquart, Y. (1986). The overlapping of cloud layers in shortwave radiation parameterizations. *Journal of the atmospheric sciences*, 43(4), 321-328. [https://doi.org/10.1175/1520-0469\(1986\)043<0321:TOOCLI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043<0321:TOOCLI>2.0.CO;2)
- Morrison, H. & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I : Scheme description and idealized tests. *Journal of the Atmospheric Sciences*, 72(1), 287-311. <https://doi.org/10.1175/JAS-D-14-0065.1>
- Mosegaard, K. & Tarantola, A. (2002). Probabilistic approach to inverse problems. *International Handbook of Earthquake and Engineering Seismology* (p. 237-265). Academic Press. Récupérée 1 octobre 2022, à partir de http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/Papers_PDF/InverseProblemHandbk.pdf
- Mureau, R., Molteni, F. & Palmer, T. N. (1991). Ensemble predictions using dynamically-conditioned perturbations. *Proceedings of ECMWF Workshop on New Developments in Predictability*, 143-174. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/node/11325>
- Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology and Climatology*, 12(4), 595-600. [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2)
- Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001), 768-772. <https://doi.org/10.1038/nature02771>
- Mylne, K. R., Evans, R. E. & Clark, R. T. (2002). Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. *Quarterly Journal of the Royal Meteorological Society*, 128(579), 361-384. <https://doi.org/10.1256/00359000260498923>
- Neggers, R. A. (2009). A dual mass flux framework for boundary layer convection. Part II : Clouds. *Journal of the Atmospheric Sciences*, 66(6), 1489-1506. <https://doi.org/10.1175/2008JAS2636.1>
- Nikitidou, E., Kazantzidis, A. & Salamalikis, V. (2014). The aerosol effect on direct normal irradiance in Europe under clear skies. *Renewable energy*, 68, 475-484. <https://doi.org/10.1016/j.renene.2014.02.034>
- Nuissier, O., Joly, B., Vié, B. & Ducrocq, V. (2012). Uncertainty of lateral boundary conditions in a convection-permitting ensemble : a strategy of selection for Mediterranean heavy precipitation events. *Natural Hazards and Earth System Sciences*, 12(10), 2993-3011. <https://doi.org/10.5194/nhess-12-2993-2012>
- Nutter, P., Stensrud, D. & Xue, M. (2004). Effects of coarsely resolved and temporally interpolated lateral boundary conditions on the dispersion of limited-area ensemble forecasts. *Monthly weather review*, 132(10), 2358-2377. [https://doi.org/10.1175/1520-0493\(2004\)132<2358:EOCRAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2358:EOCRAT>2.0.CO;2)
- Ollinaho, P., Bechtold, P., Leutbecher, M., Laine, M., Solonen, A., Haario, H. & Järvinen, H. (2013). Parameter variations in prediction skill optimization at ECMWF. *Nonlinear processes in geophysics*, 20(6), 1001-1010. <https://doi.org/10.5194/npg-20-1001-2013>

- Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J. & Sandu, I. (2017). Towards process-level representation of model uncertainties : stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 408-422. <https://doi.org/10.1002/qj.2931>
- Ooyama, K. (1971). A theory on parameterization of cumulus convection. *Journal of the Meteorological Society of Japan*, 49, 744-756. https://doi.org/10.2151/jmsj1965.49A.0_744
- Orth, R., Dutra, E. & Pappenberger, F. (2016). Improving weather predictability by including land surface model parameter uncertainty. *Monthly Weather Review*, 144(4), 1551-1569. <https://doi.org/10.1175/MWR-D-15-0283.1>
- Ovtchinnikov, M. & Kogan, Y. L. (2000). An investigation of ice production mechanisms in small cumuliform clouds using a 3D model with explicit microphysics. Part I : Model description. *Journal of the atmospheric sciences*, 57(18), 2989-3003. [https://doi.org/10.1175/1520-0469\(2000\)057<2989:AIOIPM>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<2989:AIOIPM>2.0.CO;2)
- Pailleux, J., Geleyn, J.-F. & Legrand, E. (2000). La prévision numérique du temps avec les modèles ARPÈGE et ALADIN-Bilan et perspectives. *La Météorologie*, (30), 32-60. <https://doi.org/10.4267/2042/36123>
- Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, F. J., Jung, T. & Leutbecher, M. (2005). Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, 33, 163-193. <https://doi.org/10.1146/annurev.earth.33.092203.122552>
- Palmer, T. N. (1999). *Predicting uncertainty in forecasts of weather and climate : ECMWF Technical Memoranda* (rapp. tech. N° 294). ECMWF. <https://doi.org/10.21957/2iy67frln>
- Palmer, T. N. (2001). A nonlinear dynamical perspective on model error : A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, 127(572), 279-304. <https://doi.org/10.1002/qj.49712757202>
- Palmer, T. N. (2019). The ECMWF ensemble prediction system : Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, 145, 12-24. <https://doi.org/10.1002/qj.3383>
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delécluse, P., Déqué, M., Diez, E., Doblas-Reyes, F. J., Feddersen, H. et al. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85(6), 853-872. <https://doi.org/10.1175/BAMS-85-6-853>
- Palmer, T. N., Barkmeijer, J., Buizza, R. & Petroliaigis, T. (1997). The ECMWF ensemble prediction system. *Meteorological Applications*, 4(4), 301-304. <https://doi.org/10.1017/S1350482797000649>
- Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M. & Weisheimer, A. (2009). *Stochastic parametrization and model uncertainty* (Technical Memorandum). ECMWF. Récupérée 1 octobre 2022, à partir de https://www2.physics.ox.ac.uk/sites/default/files/2011-08-15/techmemo598_stochphys_2009_pdf_50419.pdf
- Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. & Tribbia, J. (1993). Ensemble prediction. *Proceedings of ECMWF Seminar on Validation of models over Europe*, 1, 21-66. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/sites/default/files/elibrary/1992/11488-ensemble-prediction.pdf>

- Park, Y.-Y., Buizza, R. & Leutbecher, M. (2008). TIGGE : Preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134(637), 2029-2050. <https://doi.org/10.1002/qj.334>
- Perez, D. M., Martin, G. M. & Sanchez, J. C. (2014). Statistical condensation scheme : Vertical variation of the saturation and convolution pdf. *ALADIN-HIRLAM Newsletter*, (3), 36-44. Récupérée 3 octobre 2022, à partir de https://www.umr-cnrm.fr/aladin/IMG/pdf/aladin-hirlam-newsletter_no3_septembre2014.pdf
- Pergaud, J., Masson, V., Malardel, S. & Couvreur, F. (2009). A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-layer meteorology*, 132(1), 83. <https://doi.org/10.1007/s10546-009-9388-0>
- Perraud, E., Couvreur, F., Malardel, S., Lac, C., Masson, V. & Thouron, O. (2011). Evaluation of statistical distributions for the parametrization of subgrid boundary-layer clouds. *Boundary-layer meteorology*, 140(2), 263-294. <https://doi.org/10.1007/s10546-011-9607-3>
- Petch, J. C. (2006). Sensitivity studies of developing convection in a cloud-resolving model. *Quarterly Journal of the Royal Meteorological Society*, 132(615), 345-358. <https://doi.org/10.1256/qj.05.71>
- Pincus, R., Barker, H. W. & Morcrette, J.-J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research : Atmospheres*, 108(D13). <https://doi.org/10.1029/2002JD003322>
- Pinty, J.-P. & Jabouille, P. (1998). A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model : simulations of a squall line and of orographic precipitations. *Proceedings of the conference on cloud physics*, 217-220. Récupérée 20 septembre 2022, à partir de http://mesonh.aero.obs-mip.fr/mesonh/dir_publication/pinty_jabouille_ams_ccp1998.pdf
- Piriou, J.-M., Redelsperger, J.-L., Geleyn, J.-F., Lafore, J.-P. & Guichard, F. (2007). An approach for convective parameterization with memory : Separating microphysics and transport in grid-scale equations. *Journal of the Atmospheric Sciences*, 64(11), 4127-4139. <https://doi.org/10.1175/2007JAS2144.1>
- Pitcher, E. J. (1977). Application of stochastic dynamic prediction to real data. *Journal of Atmospheric Sciences*, 34(1), 3-21. [https://doi.org/10.1175/1520-0469\(1977\)034<0003:AOSDPT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1977)034<0003:AOSDPT>2.0.CO;2)
- Plant, R., Hally, A., Lock, S.-J., Ahlgrimm, M., Arpagus, M., Bauer, W., Cafaro, C., Chen, I., Fannon, J., Feddersen, H., Fleury, A., Frogner, I.-L., Gebhardt, C., Hamalainen, K., Hieronymus, M., Kauhanen, J., Kouhen, S., Kuntze, P., Lean, H., . . . Xu, Z. (2022). Summary and Recommendations from Working Group 1 : model uncertainty representations in convection-permitting / shorter lead-time / limited-area ensembles. *ECMWF Workshop on Model Uncertainty*. Récupérée 17 août 2022, à partir de https://events.ecmwf.int/event/290/attachments/1794/3437/WS2022_ModelUnc_WG1_summary.pdf
- Plant, R. S. & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, 65(1), 87-105. <https://doi.org/10.1175/2007JAS2263.1>
- Poku, C., Ross, A. N., Blyth, A. M., Hill, A. A. & Price, J. D. (2019). How important are aerosol-fog interactions for the successful modelling of nocturnal radiation fog? *Weather*, 74(7), 237-243. <https://doi.org/10.1002/wea.3503>
- Posselt, D. J. & Vukicevic, T. (2010). Robust characterization of model physics uncertainty for simulations of deep moist convection. *Monthly Weather Review*, 138(5), 1513-1535. <https://doi.org/10.1175/2009MWR3094.1>

- Price, J. D. (2019). On the formation and development of radiation fog : an observational study. *Boundary-Layer Meteorology*, 172(2), 167-197. <https://doi.org/10.1007/s10546-019-00444-5>
- Price, J. D., Lane, S., Boutle, I. A., Smith, D. K. E., Bergot, T., Lac, C., Duconge, L., McGregor, J., Kerr-Munslow, A., Pickering, M. et al. (2018). LANFEX : a field and modeling study to improve our understanding and forecasting of radiation fog. *Bulletin of the American Meteorological Society*, 99(10), 2061-2077. <https://doi.org/10.1175/BAMS-D-16-0299.1>
- Pruppacher, H. R. & Klett, J. D. (1978). *Microphysics of Clouds and Precipitation* (1^{re} éd.). Springer Science & Business Media. <https://doi.org/10.1007/978-94-009-9905-3>
- Ramanathan, V., Pitcher, E. J., Malone, R. C. & Blackmon, M. L. (1983). The response of a spectral general circulation model to refinements in radiative processes. *Journal of the Atmospheric Sciences*, 40(3), 605-630. [https://doi.org/10.1175/1520-0469\(1983\)040<0605:TROASG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<0605:TROASG>2.0.CO;2)
- Redelsperger, J.-L. & Sommeria, G. (1981). Méthode de représentation de la turbulence d'échelle inférieure à la maille pour un modèle tri-dimensionnel de convection nuageuse. *Boundary-Layer Meteorology*, 21(4), 509-530. <https://doi.org/10.1007/BF02033598>
- Reynolds, O. (1895). IV. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical transactions of the royal society of london. (a.)*, (186), 123-164. <https://doi.org/10.1098/rsta.1895.0004>
- Rieger, D., Steiner, A., Bachmann, V., Gasch, P., Förstner, J., Deetz, K., Vogel, B. & Vogel, H. (2017). Impact of the 4 April 2014 Saharan dust outbreak on the photovoltaic power generation in Germany. *Atmospheric Chemistry and Physics*, 17(21), 13391-13415. <https://doi.org/10.5194/acp-17-13391-2017>
- Rochetin, N., Couvreur, F., Grandpeix, J.-Y. & Rio, C. (2014). Deep convection triggering by boundary layer thermals. Part I : LES analysis and stochastic triggering formulation. *Journal of the Atmospheric Sciences*, 71(2), 496-514. <https://doi.org/10.1175/JAS-D-12-0336.1>
- Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Abdel-Lathif, A. Y., Bazile, E., Belamari, S., Blein, S., Bouniol, D., Bouteloup, Y., Cattiaux, J., Chauvin, F., Chevallier, M., Colin, J., Douville, H., Marquet, P., ... Sénési, S. (2020). The CNRM Global Atmosphere Model ARPEGE-Climat 6.3 : Description and Evaluation. *Journal of Advances in Modeling Earth Systems*, 12(7), e2020MS002075. <https://doi.org/10.1029/2020MS002075>
- Romine, G. S., Schwartz, C. S., Berner, J., Fossell, K. R., Snyder, C., Anderson, J. L. & Weisman, M. L. (2014). Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, 142(12), 4519-4541. <https://doi.org/10.1175/MWR-D-14-00100.1>
- Rontu, L. & Lindfors, A. V. (2018). Comparison of radiation parametrizations within the HARMONIE-AROME NWP model. *Advances in Science and Research*, 15, 81-90. <https://doi.org/10.5194/asr-15-81-2018>
- Ryerson, W. R. & Hacker, J. P. (2014). The potential for mesoscale visibility predictions with a multimodel ensemble. *Weather and Forecasting*, 29(3), 543-562. <https://doi.org/10.1175/WAF-D-13-00067.1>
- Sablon, J. (2020). *Estimation de l'impact des aérosols sur la production photovoltaïque en France* (Rapport de stage). ISAE-SUPAERO/CNRM. Récupérée 29 septembre 2022, à partir de https://www.umr-cnrm.fr/IMG/pdf/2020_cnrm_sablon_julien_rapport.pdf

- Saito, K., Kuroda, T., Kunii, M. & Kohno, N. (2010). Numerical simulation of Myanmar cyclone Nargis and the associated storm surge part II : ensemble prediction. *Journal of the Meteorological Society of Japan*, 88(3), 547-570. <https://doi.org/10.2151/jmsj.2010-316>
- Sakradzija, M. & Hohenegger, C. (2017). What determines the distribution of shallow convective mass flux through a cloud base? *Journal of the Atmospheric Sciences*, 74(8), 2615-2632. <https://doi.org/10.1175/JAS-D-16-0326.1>
- Sakradzija, M. & Klocke, D. (2018). Physically Constrained Stochastic Shallow Convection in Realistic Kilometer-Scale Simulations. *Journal of Advances in Modeling Earth Systems*, 10(11), 2755-2776. <https://doi.org/10.1029/2018MS001358>
- Sakradzija, M., Seifert, A. & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, 8(2), 786-812. <https://doi.org/10.1002/2016MS000634>
- Sakradzija, M., Seifert, A. & Heus, T. (2015). Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, 22, 65-85. <https://doi.org/10.5194/npg-22-65-2015>
- Schäfer, S., Köhler, M., Hogan, R., Ahlgrimm, M., Rieger, D., Seifert, A., Lozar, A. d. & Zängl, G. (2021). Uncertainties, stochastic and deterministic components in radiation modelling. *COSMO Uncertainty workshop*. Récupérée 25 septembre 2022, à partir de http://www.cosmo-model.org/content/tasks/workGroups/wg7/docs/Schaefer_StochasticRadiation.pdf
- Schwenkel, J. & Maronga, B. (2020). Towards a better representation of fog microphysics in Large-Eddy Simulations based on an embedded lagrangian cloud model. *Atmosphere*, 11(5), 466. <https://doi.org/10.3390/atmos11050466>
- Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. & Masson, V. (2011). The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3), 976-991. <https://doi.org/10.1175/2010MWR3425.1>
- Separovic, L., Charron, M., Erfani, A., Gagnon, N., Zadra, A. & Vaillancourt, P. (2016). *Model Error Representation in the Canadian Ensemble Prediction Systems* [ECMWF/WWRP Workshop on model uncertainty]. Récupérée 2 janvier 2023, à partir de <https://www.ecmwf.int/sites/default/files/elibrary/2016/16742-model-error-representation-canadian-ensemble-prediction-systems.pdf>
- Shima, S.-I., Kusano, K., Kawano, A., Sugiyama, T. & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation : A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, 135(642), 1307-1320. <https://doi.org/10.1002/qj.441>
- Shiogama, H., Watanabe, M., Ogura, T., Yokohata, T. & Kimoto, M. (2014). Multi-Parameter Multi-Physics Ensemble (MPMPE) : A new approach exploring the uncertainties of climate sensitivity. *Atmospheric Science Letters*, 15(2), 97-102. <https://doi.org/10.1002/asl2.472>
- Shonk, J. K. P. & Hogan, R. J. (2008). Tripleclouds : An efficient method for representing horizontal cloud inhomogeneity in 1D radiation schemes by using three regions at each height. *Journal of Climate*, 21(11), 2352-2370. <https://doi.org/10.1175/2007JCLI1940.1>
- Shonk, J. K. P. & Hogan, R. J. (2010). Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part II : The global effects. *Quarterly Journal of the Royal Meteorological Society*, 136(650), 1205-1215. <https://doi.org/10.1002/qj.646>

- Shutts, G. J. & Palmer, T. N. (2007). Convective forcing fluctuations in a cloud-resolving model : Relevance to the stochastic parameterization problem. *Journal of climate*, 20(2), 187-202. <https://doi.org/10.1175/JCLI3954.1>
- Shutts, G. J. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131(612), 3079-3102. <https://doi.org/10.1256/qj.04.106>
- Shutts, G. J., Leutbecher, M., Weisheimer, A., Stockdale, T., Isaksen, L. & Bonavita, M. (2011). Representing model uncertainty : stochastic parametrizations at ECMWF. *ECMWF Newsletter*, (129), 19-24. <https://doi.org/10.21957/fbqmkhv7>
- Shutts, G. J. & Pallarès, A. C. (2014). Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models. *Philosophical Transactions of the Royal Society A*, 372(2018), 20130284. <https://doi.org/10.1098/rsta.2013.0284>
- Shutts, G. J. & Palmer, T. N. (2004). The use of high resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. *ECMWF/CLIVAR Workshop on Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO, 3-6 November 2003*, 83-102. Récupérée 1 octobre 2022, à partir de <https://www.ecmwf.int/node/12212>
- Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., Jiang, H., Khairoutdinov, M., Lewellen, D., Moeng, C.-H. et al. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, 60(10), 1201-1219. [https://doi.org/10.1175/1520-0469\(2003\)60<1201:ALESIS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)60<1201:ALESIS>2.0.CO;2)
- Siebesma, A. P. & Cuijpers, J. W. M. (1995). Evaluation of parametric assumptions for shallow cumulus convection. *Journal of Atmospheric Sciences*, 52(6), 650-666. [https://doi.org/10.1175/1520-0469\(1995\)052<0650:EOPAFS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<0650:EOPAFS>2.0.CO;2)
- Siebesma, A. P., Soares, P. M. & Teixeira, J. (2007). A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the atmospheric sciences*, 64(4), 1230-1248. <https://doi.org/10.1175/JAS3888.1>
- Slingo, A. (1990). Sensitivity of the Earth's radiation budget to changes in low clouds. *Nature*, 343(6253), 49-51. <https://doi.org/10.1038/343049a0>
- Smith, D. K. E., Renfrew, I. A., Price, J. D. & Dorling, S. R. (2018). Numerical modelling of the evolution of the boundary layer during a radiation fog event. *Weather*, 73(10), 310-316. <https://doi.org/10.1002/wea.3305>
- Smith, E. A. & Shi, L. (1992). Surface forcing of the infrared cooling profile over the Tibetan plateau. Part I : Influence of relative longwave radiative heating at high altitude. *Journal of the atmospheric sciences*, 49(10), 805-822. [https://doi.org/10.1175/1520-0469\(1992\)049<0805:SFOTIC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0805:SFOTIC>2.0.CO;2)
- Smith, R. N. B. (1990). A scheme for predicting layer clouds and their water content in a general circulation model. *Quarterly Journal of the Royal Meteorological Society*, 116(492), 435-460. <https://doi.org/10.1002/qj.49711649210>
- Soares, P., Miranda, P., Siebesma, A. P. & Teixeira, J. (2004). An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3365-3383. <https://doi.org/10.1256/qj.03.223>
- Sommeria, G. & Deardorff, J. W. (1977). Subgrid-scale condensation in models of nonprecipitating clouds. *Journal of Atmospheric Sciences*, 34(2), 344-355. [https://doi.org/10.1175/1520-0469\(1977\)034<0344:SSCIMO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1977)034<0344:SSCIMO>2.0.CO;2)

- Sommeria, G. (1976). Three-dimensional simulation of turbulent processes in an undisturbed trade wind boundary layer. *Journal of Atmospheric Sciences*, 33(2), 216-241. [https://doi.org/10.1175/1520-0469\(1976\)033<0216:TDSOTP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<0216:TDSOTP>2.0.CO;2)
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J. & Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024), 403-406. <https://doi.org/10.1038/nature03301>
- Steiner, J. T. (1973). A three-dimensional model of cumulus cloud development. *Journal of Atmospheric Sciences*, 30(3), 414-435. [https://doi.org/10.1175/1520-0469\(1973\)030<0414:ATDMOC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1973)030<0414:ATDMOC>2.0.CO;2)
- Stensrud, D. J., Bao, J.-W. & Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, 128(7), 2077-2107. [https://doi.org/10.1175/1520-0493\(2000\)128<2077:UICAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2)
- Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S. & Rogers, E. (1999). Using ensembles for short-range forecasting. *Monthly Weather Review*, 127(4), 433-446. [https://doi.org/10.1175/1520-0493\(1999\)127<0433:UEFSRF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0433:UEFSRF>2.0.CO;2)
- Stephens, G. L. (1978). Radiation Profiles in Extended Water Clouds. II : Parameterization Schemes. *Journal of Atmospheric Sciences*, 35(11), 2123-2132. [https://doi.org/10.1175/1520-0469\(1978\)035<2123:RPIEWC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<2123:RPIEWC>2.0.CO;2)
- Strømmen, K., Christensen, H. M., Berner, J. & Palmer, T. N. (2018). The impact of stochastic parametrisations on the representation of the Asian summer monsoon. *Climate dynamics*, 50(5), 2269-2282. <https://doi.org/10.1007/s00382-017-3749-z>
- Stull, R. B. (1988). *An introduction to boundary layer meteorology* (T. 13). Springer Science & Business Media. <https://doi.org/10.1007/978-94-009-3027-8>
- Sud, Y., Shukla, J. & Mintz, Y. (1988). Influence of land surface roughness on atmospheric circulation and precipitation : A sensitivity study with a general circulation model. *Journal of Applied Meteorology and Climatology*, 27(9), 1036-1054. [https://doi.org/10.1175/1520-0450\(1988\)027<1036:IOLSRO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1988)027<1036:IOLSRO>2.0.CO;2)
- Sušelj, K., Hogan, T. F. & Teixeira, J. (2014). Implementation of a stochastic eddy-diffusivity/mass-flux parameterization into the Navy Global Environmental Model. *Weather and forecasting*, 29(6), 1374-1390. <https://doi.org/10.1175/WAF-D-14-00043.1>
- Sušelj, K., Teixeira, J. & Chung, D. (2013). A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *Journal of the Atmospheric Sciences*, 70(7), 1929-1953. <https://doi.org/10.1175/JAS-D-12-0106.1>
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Tennant, W. (2016). *Impacts of a spatially varying random parameter scheme on cloud cover in MOGREPS* [ECMWF/WWRP]. Récupérée 9 août 2022, à partir de <https://www.ecmwf.int/sites/default/files/elibrary/2016/16301-workshop-model-uncertainty-poster-warren-tennant.pdf>
- Tennant, W. & Beare, S. (2014). New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1150-1160. <https://doi.org/10.1002/qj.2202>
- Tennant, W. & Jardak, M. (2020). Recent Developments in Global Ensemble Prediction at the Met Office. *Proceedings of the International Conference on Ensemble Methods in Modelling and Data Assimilation (EMMDA)*. Récupérée 1 octobre 2022, à partir de

- https://www.ncmrwf.gov.in/event/emmda/Presentations/WTennant_EMMDA_TS01_Day01.pdf
- Thompson, G., Berner, J., Frediani, M., Otkin, J. A. & Griffin, S. M. (2021). A Stochastic Parameter Perturbation Method to Represent Uncertainty in a Microphysics Scheme. *Monthly Weather Review*, 149(5), 1481-1497. <https://doi.org/10.1175/MWR-D-20-0077.1>
- Thompson, G. & Eidhammer, T. (2014). A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *Journal of the atmospheric sciences*, 71(10), 3636-3658. <https://doi.org/10.1175/JAS-D-13-0305.1>
- Thompson, P. D. (1985). Prediction of the probable errors of predictions. *Monthly weather review*, 113(2), 248-259. [https://doi.org/10.1175/1520-0493\(1985\)113<0248:POTPEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<0248:POTPEO>2.0.CO;2)
- Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather review*, 117(8), 1779-1800. [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2)
- Tompkins, A. M. (2002). A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *Journal of the atmospheric sciences*, 59(12), 1917-1942. [https://doi.org/10.1175/1520-0469\(2002\)059<1917:APPFTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1917:APPFTS>2.0.CO;2)
- Torn, R. D., Hakim, G. J. & Snyder, C. (2006). Boundary conditions for limited-area ensemble Kalman filters. *Monthly weather review*, 134(9), 2490-2502. <https://doi.org/10.1175/MWR3187.1>
- Toth, Z. & Kalnay, E. (1993). Ensemble Forecasting at NMC : The Generation of Perturbations. *Bulletin of the American Meteorological Society*, 74(12), 2317-2330. [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2)
- Toth, Z. & Kalnay, E. (1997). Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 125(12), 3297-3319. [https://doi.org/10.1175/1520-0493\(1997\)125<3297:EFANAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2)
- Tsiringakis, A., Contreras, S., de Rooy, W. & Barkmeijer, J. (2022). New convection and microphysical parameters for the SPP scheme of the HarmonEPS. *ACCORD All Staff Workshop*. Récupérée 20 septembre 2022, à partir de https://www.umr-cnrm.fr/accord/IMG/pdf/tsiringakis_accord_all_staff_workshop.pdf
- van Den Heever, S. C. & Cotton, W. R. (2004). The impact of hail size on simulated supercell storms. *Journal of the atmospheric sciences*, 61(13), 1596-1609. [https://doi.org/10.1175/1520-0469\(2004\)061<1596:TIOHSO>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1596:TIOHSO>2.0.CO;2)
- van Lier-Walqui, M., Morrison, H., Kumjian, M. R., Reimel, K. J., Prat, O. P., Lunderman, S. & Morzfeld, M. (2020). A Bayesian approach for statistical–physical bulk parameterization of rain microphysics. Part II : Idealized Markov chain Monte Carlo experiments. *Journal of the Atmospheric Sciences*, 77(3), 1043-1064. <https://doi.org/10.1175/JAS-D-19-0071.1>
- van Lier-Walqui, M., Vukicevic, T. & Posselt, D. J. (2012). Quantification of cloud microphysical parameterization uncertainty using radar reflectivity. *Monthly Weather Review*, 140(11), 3442-3466. <https://doi.org/10.1175/MWR-D-11-00216.1>
- van Weverberg, K., van Lipzig, N. P. & Delobbe, L. (2011). The impact of size distribution assumptions in a bulk one-moment microphysics scheme on simulated surface precipitation and storm dynamics during a low-topped supercell case in Belgium. *Monthly Weather Review*, 139(4), 1131-1147. <https://doi.org/10.1175/2010MWR3481.1>

- Vié, B., Pinty, J.-P., Berthet, S. & Leriche, M. (2016). LIMA (v1. 0) : A quasi two-moment microphysical scheme driven by a multimodal population of cloud condensation and ice freezing nuclei. *Geoscientific Model Development*, 9(2), 567-586. <https://doi.org/10.5194/gmd-9-567-2016>
- Vié, B., Nuissier, O. & Ducrocq, V. (2011). Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events : Uncertainty on initial conditions and lateral boundary conditions. *Monthly Weather Review*, 139(2), 403-423. <https://doi.org/10.1175/2010MWR3487.1>
- Vislocky, R. L. & Fritsch, J. M. (1995). Improved model output and statistics through model consensus. *Bulletin of the American Meteorological Society*, 76(7), 1157-1164. [https://doi.org/10.1175/1520-0477\(1995\)076<1157:IMOSFT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<1157:IMOSFT>2.0.CO;2)
- Vukicevic, T. & Errico, R. M. (1990). The influence of artificial and physical factors upon predictability estimates using a complex limited-area model. *Monthly weather review*, 118(7), 1460-1482. [https://doi.org/10.1175/1520-0493\(1990\)118<1460:TIOAAP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1460:TIOAAP>2.0.CO;2)
- Vukicevic, T. & Paegle, J. (1989). The influence of one-way interacting lateral boundary conditions upon predictability of flow in bounded numerical models. *Monthly weather review*, 117(2), 340-350. [https://doi.org/10.1175/1520-0493\(1989\)117<0340:TIOOWI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0340:TIOOWI>2.0.CO;2)
- Vukicevic, T. & Posselt, D. (2008). Analysis of the impact of model nonlinearities in inverse problem solving. *Journal of the atmospheric sciences*, 65(9), 2803-2823. <https://doi.org/10.1175/2008JAS2534.1>
- Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., ... Zerroukat, M. (2019). The Met Office Unified Model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geoscientific Model Development*, 12(5), 1909-1963. <https://doi.org/10.5194/gmd-12-1909-2019>
- Wang, L., Shen, X., Liu, J. & Wang, B. (2020). Model uncertainty representation for a convection-allowing ensemble prediction system based on CNOP-P. *Advances in Atmospheric Sciences*, 37(8), 817-831. <https://doi.org/10.1007/s00376-020-9262-z>
- Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S. et al. (2011). The Central European limited-area ensemble forecasting system : ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 137(655), 483-502. <https://doi.org/10.1002/qj.751>
- Wang, Y., Belluš, M., Weidle, F., Wittmann, C., Tang, J., Meier, F., Xia, F. & Keresturi, E. (2019). Impact of land surface stochastic physics in ALADIN-LAEF. *Quarterly Journal of the Royal Meteorological Society*, 145(724), 3333-3350. <https://doi.org/10.1002/qj.3623>
- Warner, T. T., Peterson, R. A. & Treadon, R. E. (1997). A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, 78(11), 2599-2618. [https://doi.org/10.1175/1520-0477\(1997\)078<2599:ATOLBC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2)
- Wastl, C., Wang, Y., Atencia, A. & Wittmann, C. (2019a). A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. *Monthly Weather Review*, 147(6), 2217-2230. <https://doi.org/10.1175/MWR-D-18-0415.1>

- Wastl, C., Wang, Y., Atencia, A. & Wittmann, C. (2019b). Independent perturbations for physics parametrization tendencies in a convection-permitting ensemble (pSPPT). *Geoscientific Model Development*, 12(1), 261-273. <https://doi.org/10.5194/gmd-12-261-2019>
- Wei, M., Toth, Z., Wobus, R. & Zhu, Y. (2008). Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A : Dynamic Meteorology and Oceanography*, 60(1), 62-79. <https://doi.org/10.1111/j.1600-0870.2007.00273.x>
- Weisheimer, A., Corti, S., Palmer, T. & Vitart, F. (2014). Addressing model error through atmospheric stochastic physical parametrizations : Impact on the coupled ECMWF seasonal forecasting system. *Philosophical Transactions of the Royal Society A*, 372(2018), 20130290. <https://doi.org/10.1098/rsta.2013.0290>
- Wimmer, M. (2021). *Représentation des erreurs de modélisation dans le système de prévision d'ensemble régional PEARO* (thèse de doct.). Université Paul Sabatier-Toulouse III.
- Wimmer, M., Raynaud, L., Descamps, L., Berre, L. & Seity, Y. (2022). Sensitivity analysis of the convective-scale AROME model to physical and dynamical parameters. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 920-942. <https://doi.org/10.1002/qj.4239>
- Wobus, R. L. & Kalnay, E. (1995). Three years of operational prediction of forecast skill at NMC. *Monthly weather review*, 123(7), 2132-2148. [https://doi.org/10.1175/1520-0493\(1995\)123<2132:TYOOPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2132:TYOOPO>2.0.CO;2)
- Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, 140(8), 2373-2423. <https://doi.org/10.1175/MWR-D-11-00121.1>
- Wyngaard, J. C. (2004). Toward numerical modeling in the “Terra Incognita”. *Journal of the atmospheric sciences*, 61(14), 1816-1826. [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TNMITT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2)
- Xu, Z., Chen, J., Jin, Z., Li, H. & Chen, F. (2020a). Assessment of the forecast skill of multi-physics and multistochastic methods within the GRAPES regional ensemble prediction system in the east Asian monsoon region. *Weather and Forecasting*, 35(3), 1145-1171. <https://doi.org/10.1175/WAF-D-19-0021.1>
- Xu, Z., Chen, J., Jin, Z., Li, H. & Chen, F. (2020b). Representing model uncertainty by multi-stochastic physics approaches in the GRAPES ensemble. *Advances in Atmospheric Sciences*, 37(4), 328-346. <https://doi.org/10.1007/s00376-020-9171-1>
- Xu, Z., Chen, J., Mu, M., Tao, L., Dai, G., Wang, J. & Ma, Y. (2022). A stochastic and non-linear representation of model uncertainty in a convective-scale ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 148(746), 2507-2531. <https://doi.org/10.1002/qj.4322>
- Yanai, M., Esbensen, S. & Chu, J.-H. (1973). Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *Journal of Atmospheric Sciences*, 30(4), 611-627. [https://doi.org/10.1175/1520-0469\(1973\)030<0611:DOBPOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1973)030<0611:DOBPOT>2.0.CO;2)
- Yokohata, T., Webb, M. J., Collins, M., Williams, K. D., Yoshimori, M., Hargreaves, J. C. & Annan, J. D. (2010). Structural similarities and differences in climate responses to CO2 increase between two perturbed physics ensembles. *Journal of Climate*, 23(6), 1392-1410. <https://doi.org/10.1175/2009JCLI2917.1>
- Yonehara, H. & Ujiie, M. (2011). A Stochastic physics scheme for model uncertainties in the JMA one-week Ensemble Prediction System. In C. WGNE (Éd.), *Research activities in atmospheric and oceanic modelling*. Récupérée 1 août 2022, à partir de <https://www.>

wcrp-climate.org/WGNE/BlueBook/2011/individual-articles/06_Yonehara_Hitoshi_06_Yonehara_Hitoshi_WEPS.pdf

- Yussouf, N. & Stensrud, D. J. (2012). Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble Kalman filter. *Monthly weather review*, 140(2), 562-586. <https://doi.org/10.1175/MWR-D-10-05074.1>
- Zängl, G., Reinert, D., Rípodas, P. & Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M : Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 563-579. <https://doi.org/10.1002/qj.2378>
- Zhou, B. & Du, J. (2010). Fog prediction from a multimodel mesoscale ensemble prediction system. *Weather and Forecasting*, 25(1), 303-322. <https://doi.org/10.1175/2009WAF2222289.1>
- Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., Sinsky, E., Kolczynski, W., Xue, X., Luo, Y., Peng, J., Yang, B., Tallapragada, V. & Pегion, P. (2022). The Development of the NCEP Global Ensemble Forecast System Version 12. *Weather and Forecasting*, 37(6), 1069-1084. <https://doi.org/10.1175/WAF-D-21-0112.1>

Table des figures

1	Cumuls de précipitations simulés et observés entre 00 UTC et 12 UTC le 14/09/2021 dans le sud de la France. De gauche à droite : prévisions AROME-France initialisées avec les analyses du 13/09/2021 12 UTC, du 14/09/2021 00 UTC et observations des pluies (analyse de lame d'eau Antilope).	1
1.1	Illustration de la discrétisation de l'atmosphère terrestre.	4
1.2	Illustration de trajectoires dans l'espace des phases. Les cercles bleus représentent plusieurs états initiaux distincts. Les trajectoires correspondent aux différents états par lesquels passe le système. Les flèches bleues illustrent les états atteints à un instant donné $t > 0$	6
1.3	Illustration des différentes méthodes de perturbations utilisées pour représenter les erreurs modèle en prévision d'ensemble. En orange sont indiquées les méthodes multimodèle/multiphysique décrites dans la section 1.4.1, en rouge sont représentées les méthodes de perturbation stochastique "globales", décrites dans la section 1.4.2, en violet les méthodes de perturbation des paramètres, décrites dans la section 1.4.3 et en vert les méthodes de perturbation des processus, décrites dans la section 1.4.4.	12
1.4	Illustration de la différence entre les approches multimodèle et perturbations stochastiques. Voir Palmer (1999).	14
1.5	Exemple de champ aléatoire ϵ gaussien utilisé dans l'ensemble global PEARP de Météo France.	15
1.6	Schéma simplifié des interactions entre les systèmes d'assimilation AEARP, AEARO et les systèmes de production PEARP et PEARO.	27
2.1	Représentation de nuages convectifs dans une maille du modèle (illustration de l'article d'Arakawa et Schubert 1974). La fraction de l'espace occupée par les nuages est supposée très petite.	38
2.2	Illustration du moyennage sur un grand domaine autour de la maille.	43
2.3	Schéma des échanges entre une plume convective et son environnement tiré de Pergaud et al. (2009).	45
2.4	Illustration des schémas de condensation "all or nothing" ou statistique.	47
2.5	Formules de Chaboureau et Bechtold (2002) de diagnostic de la fraction nuageuse et du contenu en eau moyen sur la maille d'après l'écart à la saturation moyen Q_1	48
2.6	Processus microphysiques modélisés par le schéma microphysique ICE3 (illustration de Pinty et Jabouille 1998). Dans AROME, le processus de formation et d'évaporation des gouttelettes liquides nuageuses (flèches "CND") est traité par l'ajustement à la saturation.	49

2.7	Principe du schéma de sédimentation de Bouteloup et al. (2011). Les ronds bleus représentent des gouttelettes nuageuses de différentes tailles, et les deux traits horizontaux délimitent deux couches verticales successives du modèle.	52
3.1	Évolution temporelle des profils verticaux du contenu en eau nuageuse, de la température potentielle, ainsi que des flux de chaleur à la surface pour le cas ARMCu.	62
3.2	Profils verticaux des tendances issues des schémas de paramétrisation d'AROME, moyennées sur l'intervalle +6-10 h pour le cas ARMCu.	62
3.3	Évolution temporelle des profils verticaux du contenu en eau nuageuse, de la température potentielle, ainsi que des flux de chaleur à la surface pour le cas BOMEX.	63
3.4	Profils verticaux des tendances issues des schémas de paramétrisation d'AROME, moyennées sur l'intervalle +2-23 h pour le cas BOMEX.	63
3.5	Évolution temporelle des profils verticaux d'eau nuageuse et de température potentielle pour le cas FIRE. Les barres verticales sur la figure a indiquent l'intervalle de temps couvert par la figure b.	64
3.6	Profils verticaux des tendances issues des schémas de paramétrisation d'AROME, moyennées sur l'intervalle +48-85 h pour le cas FIRE.	65
3.7	Évolution temporelle des profils verticaux d'eau liquide nuageuse et de température potentielle, et évolution du contenu en eau liquide intégré (LWP) pour le cas LANFEX. Les simulations démarrent à 17:00 UTC. Les échéances +6h, +12h et +18h correspondent à 23:00, 05:00 et 11:00 UTC respectivement.	66
3.8	Profils verticaux des tendances issues des schémas de paramétrisation d'AROME, moyennées sur l'intervalle +0-19 h pour le cas LANFEX.	66
3.9	Graphes hovmoller de l'eau liquide nuageuse moyennée sur tout le domaine des LES des cas ARMCu, BOMEX et FIRE.	68
4.1	Procédure de coarse-graining appliquée aux champs LES.	70
4.2	Coupes verticales de l'anomalie de concentration en traceur avec les contours des objets identifiés grâce à l'algorithme de N. Villefranque et F. Brient pour les cas d'étude ARMCu, BOMEX et FIRE.	71
4.3	Histogrammes des valeurs de r_{np} et s observées dans le domaine LES et distributions théoriques pour le cas FIRE, pour trois échéances (+6, +14, +20 h) et trois niveaux verticaux (400, 600, 800 m).	108
4.4	Histogrammes des valeurs de r_{np} et s observées dans le domaine LES et distributions théoriques pour le cas ARMCu, pour trois échéances (+5, +8, +11 h) et trois niveaux verticaux (environ 500, 1000 et 1500 m).	108
4.5	Correspondance entre Q_1 et \bar{r}_c/σ_s calculés dans les LES des cas ARMCu à +8 h, BOMEX à +10 h et FIRE à +14 h, avec la procédure de coarse-graining. Chaque point rouge correspond à un sous-domaine et à un niveau vertical.	109
4.6	Profils verticaux de l'eau nuageuse et de Q_1 à +14 h dans les simulations AROME-MUSC du cas FIRE. La courbe en pointillé rouge de la figure a correspond à la moyenne des 50 simulations AROME-MUSC et la zone en rouge correspond à l'intervalle entre les percentiles 10 et 90 de l'ensemble. La courbe noire de la figure a et la courbe de la figure b proviennent d'une simulation AROME-MUSC non perturbée.	110
4.7	Identique à la figure 4.6 pour le cas LANFEX à +18 h.	110

4.8	Identique à la figure 4.6a, mais pour un ensemble de simulations AROME-MUSC du cas FIRE où le paramètre VSIGQSAT est perturbé.	111
5.1	Profils verticaux de dispersion de température potentielle dans des ensembles de simulations AROME-MUSC des cas FIRE à +24 h et LANFEX à +18 h. La dispersion est mesurée ici comme l'intervalle entre les percentiles 10 et 90 des ensembles. Les méthodes de perturbation utilisées pour construire les ensembles sont la méthode SPPT (courbe rouge), la méthode de KC16 (courbe bleu foncée), la convection stochastique (courbe cyan) et la perturbation de la relation de CB02 (courbe verte).	113
5.2	Distributions de taille des gouttelettes pour trois valeurs de N_0	114
5.3	Profils verticaux de la vitesse de sédimentation.	114
5.4	Évolution temporelle du contenu en eau intégré verticalement (LWP).	114
5.5	Diagrammes hovmoller du contenu en eau nuageuse	114
5.6	Distributions de taille des gouttelettes pour trois valeurs de ν	116
5.7	Vitesse de sédimentation	116
5.8	Évolution temporelle du LWP.	116
5.9	Profils verticaux du contenu en eau nuageuse à +18 h pour quatre ensembles de simulations LANFEX faisant varier la valeur des paramètres N_0 et ν selon une distribution uniforme ou lognormale. La zone en rouge correspond à l'intervalle entre les percentiles 10 et 90 de l'ensemble. La courbe noire provient d'une simulation AROME-MUSC non perturbée.	117
5.10	De haut en bas et de gauche à droite : distribution des observations (f_Y), distribution jointe théorique ($g_{\Lambda,Y}$), la distribution jointe <i>a posteriori</i> (p) et la distribution marginale <i>a posteriori</i> (p_Λ).	121
5.11	Modélisation de la distribution <i>a posteriori</i> p_Λ par une loi lognormale au fil des itérations de la méthode de Vukicevic et Posselt (2008)	122
5.12	Évolution de la moyenne expérimentale de la distribution <i>a posteriori</i> p_Λ obtenue au fil des itérations.	122
5.13	Construction de la distribution $\pi_Y^{G(\text{prior})}$	124
5.14	Calcul de la distribution $\pi_\Lambda^{\text{post}}$	124
5.15	Distributions <i>a posteriori</i> $\pi_\Lambda^{\text{post}}$ obtenues avec la méthode de Butler et al. (2018).	125
5.16	Profils verticaux du contenu en eau nuageuse à +18h pour deux ensembles de simulations faisant varier la valeur des paramètres N_0 ou ν selon une loi lognormale réglée d'après les résultats de la résolution du problème inverse.	125
5.17	Estimation du modèle direct $G(\lambda)$ pour $\lambda = (N_0, \nu)$ à l'aide de 1995 simulations du cas LANFEX.	126
5.18	Distribution <i>a posteriori</i> $\pi_\Lambda^{\text{post}}(\lambda)$ pour $\lambda = (N_0, \nu)$	126
5.19	Échantillonnage de 10000 valeurs de la distribution <i>a posteriori</i> à l'aide d'une méthode de rejet.	126
5.20	Distribution de la mesure obtenue avec la distributions de paramètre <i>a priori</i> ("push-forward of the prior") et avec la distribution <i>a posteriori</i> ("push-forward of the posterior").	126

5.21	Profils verticaux du contenu en eau nuageuse à +18 h dans la simulation de référence LANFEX (courbe noire) et deux ensembles RPP de 50 simulations : en rouge, un ensemble où les valeurs de N_0 et ν sont tirées chacune dans une distribution lognormale centrée sur la valeur de référence du paramètre, et en bleu un ensemble où les valeurs de N_0 et ν proviennent de l'échantillonnage de la distribution <i>a posteriori</i> du problème inverse. Les courbes en pointillé correspondent à la moyenne de chaque ensemble, et les zones colorées représentent l'intervalle entre les percentiles 10 et 90.	127
5.22	Évolution du coefficient k en fonction de ν d'après l'équation 5.8 pour deux valeurs de α	130
5.23	Évolution temporelle du LWP dans les cas d'étude LANFEX et FIRE pour deux configurations différentes du schéma radiatif. Les courbes noires correspondent à la configuration originale du schéma de Martin et al. (1994), utilisée opérationnellement dans AROME, tandis que les courbes rouges en pointillé correspondent à la version modifiée de ce schéma, où les hypothèses concernant le spectre de taille des gouttelettes sont cohérentes avec celles du schéma microphysique.	131
5.24	Profils verticaux à +18h de plusieurs variables calculées par le schéma radiatif pour des simulations du cas LANFEX où la valeur de N_0 ou ν a été modifiée (dans le schéma radiatif uniquement).	131
5.25	Évolution temporelle de l'épaisseur optique du brouillard du cas LANFEX. Il s'agit de la somme des épaisseurs optiques calculées pour chaque niveau vertical du modèle. La courbe noire est la simulation de référence, et les courbes bleue et rouge correspondent à des simulations où la valeur de N_0 a été réglée à 10 et 300 gouttes/cm ³ respectivement.	132
5.26	Diagramme hovmoller des flux nets SW dans la simulation de référence du cas LANFEX.	132
5.27	Profil vertical à +18h de l'eau liquide nuageuse dans cinq simulations du cas LANFEX.	132
5.28	Diagramme hovmoller des flux nets SW dans la simulation de référence du cas FIRE.	133
5.29	Profil vertical à +14h de l'eau liquide nuageuse dans cinq simulations du cas FIRE.	133
5.30	Profils verticaux à +10h de la tendance de température du schéma radiatif, de la contribution des flux SW et des flux LW, pour des simulations du cas FIRE où la valeur de N_0 ou ν a été modifiée (dans le schéma radiatif uniquement).	133
5.31	Profils verticaux de la dispersion d'eau nuageuse dans différents ensembles de simulations AROME-MUSC du cas FIRE à +14 h. La dispersion est calculée comme l'intervalle entre les percentiles 10 et 90 des ensembles.	134
5.32	Profils verticaux d'eau nuageuse pour les quatre cas d'étude, et deux types d'ensembles : un construit avec la méthode SPPT (en rouge) et un construit avec les perturbations listées dans le tableau 5.2 (COMB, en bleu). Les courbes noires représentent les simulations de référence pour chaque cas. Les courbes en pointillé sont les moyennes des ensembles. Les zones colorées représentent l'intervalle entre les percentiles 10 et 90 des ensembles.	136
5.33	Profils verticaux de dispersion de température potentielle pour les quatre cas d'étude, et les deux types d'ensembles. La dispersion est calculée comme l'intervalle entre les percentiles 10 et 90 des ensembles.	137
5.34	Identique à la figure 5.33 mais pour le vent zonal.	137

5.35	Profils verticaux de la dispersion et du biais dans plusieurs ensembles de simulations AROME-MUSC du cas ARMCu à +9 h.	139
6.1	Contenu en eau liquide nuageuse à 10 mètres au-dessus du sol à différentes échéances d'une simulation AROME initialisée avec l'analyse du 05/01/2020 03:00 UTC. L'étoile sur la troisième carte correspond au point pour lequel le diagramme hovmoller de la figure 6.3 est tracé.	142
6.2	Direction et force du vent à 10 m à différentes échéances d'une simulation AROME initialisée avec l'analyse du 05/01/2020 03:00 UTC.	142
6.3	Diagramme hovmoller de l'eau liquide nuageuse durant la nuit du 5 au 6 janvier 2020 simulée par AROME au point indiqué sur la figure 6.1.	143
6.4	Diagnostic de visibilité à +25h dans l'expérience A, et différence par rapport à cette référence pour les expériences B et C.	145
6.5	Contenu en eau nuageuse à 10 m à +25h dans l'expérience A, et différence par rapport à cette référence pour les expériences B et C.	145
6.6	Profils verticaux du contenu en eau nuageuse à +25h sur cinq points du domaine. La courbe noire est l'expérience contrôle (Exp A), la courbe bleue est l'expérience B et la rouge est l'expérience C.	146
6.7	LWP à +25h dans l'expérience A, et différence par rapport à cette référence pour les expériences B et C.	146
6.8	Echéance à laquelle le brouillard apparaît dans les expériences A, B, C. Le calcul a été effectué entre les échéances +12 et +36 du modèle.	147
6.9	Echéance à laquelle le brouillard disparaît dans les expériences A, B, C. Le calcul a été effectué entre les échéances +12 et +36 du modèle.	147
6.10	Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés près de Bordeaux et Toulouse respectivement. La courbe noire est l'expérience contrôle (Exp A), la courbe bleue est l'expérience B et la rouge est l'expérience C.	148
6.11	Identique à la figure 6.10 pour les points 2, 3 et 4.	148
6.12	LWP à +25h dans l'expérience A, et différence par rapport à cette référence pour les expériences G et H.	149
6.13	Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés près de Bordeaux et Toulouse respectivement. La courbe noire est l'expérience contrôle (Exp D), la courbe bleue est l'expérience E et la rouge est l'expérience F.	150
6.14	LWP à +25h dans l'expérience D, et différence par rapport à cette référence pour les expériences I et J.	150
6.15	Évolution temporelle du contenu en eau nuageuse à 10 m aux points 1 et 5 (voir figure 6.5) situés aux environs de Bordeaux et Toulouse respectivement. La courbe noire est l'expérience contrôle (Exp D), la courbe bleu foncé est l'expérience I et la courbe bleu clair est l'expérience J.	151
6.16	Visualisation des données de simulation des modèles ARPEGE et AROME le 04/09/2021 à 12:00 UTC. (a) Géopotential à 500 hPa et température à 850 hPa simulés par le modèle global ARPEGE. (b) Température à 2 m et vent à 10 m simulés par le modèle AROME. Les cartes ont été extraites depuis http://intra.cnrm.meteo.fr/visumodeles (site interne).	152
6.17	Cumuls de pluie sur 1 heure observés (analyse de lame d'eau Antilope) le 04/09/2021 à 14:00 UTC.	152

6.18	Cartes de l'énergie convective disponible (CAPE) et des cumuls de précipitations sur 1 heure simulés pour le 04/09/2021 à 14:00 UTC par plusieurs membres de 3 expériences PEARO : un ensemble sans représentation de l'erreur modèle (colonne de gauche), l'ensemble utilisant la méthode opérationnelle SPPT (colonne du milieu), et un ensemble utilisant la méthode KC16 (colonne de droite). Les trois lignes correspondent respectivement aux membres 3, 10 et 14.	153
6.19	Cartes de cumuls des précipitations sur 1 heure simulés pour le 04/09/2021 à 14:00 UTC par le membre 3 de trois expériences PEARO. De gauche à droite : un ensemble utilisant la méthode opérationnelle SPPT, un ensemble utilisant SPPT où les corrélations temporelles et spatiales ont été réglées à 10 min et 6.5 km respectivement, et un ensemble utilisant la méthode KC16.	154
6.20	Évolution temporelle du cumul de précipitation sur 1 heure moyenné sur toute la France durant la journée du 04/09/2019, pour différentes expériences PEARO. La courbe noire correspond aux observations (lame d'eau Antilope). Les courbes de couleur foncées correspondent à la moyenne de chaque ensemble, tandis que les courbes claires et plus fines représentent chaque membre des différents ensembles.	155
6.21	Évolution temporelle de l'écart-type entre membres pour le vent zonal à 10 m, moyenné sur toute la France, durant la journée du 04/09/2019, pour différentes expériences PEARO.	156
6.22	Carte de vent zonal à 10 m simulé par le membre 1 de trois expériences PEARO pour le 04/09/2021 à 14:00 UTC. La carte de gauche correspond à l'expérience sans perturbations modèle. Les deux autres cartes présentent la différence par rapport à cette référence, pour l'expérience SPPT et celle utilisant la méthode KC16 et la convection stochastique. La même échelle de couleur est utilisée pour ces deux dernières cartes.	156
6.23	Diagrammes de fiabilité	159
6.24	Courbes ROC	159
6.25	Scores de Brier	159
A.1	Vitesse terminale "moyenne" de chute des gouttelettes selon la valeur du paramètre de forme ν	169
A.2	Vitesse de chute d'une gouttelette nuageuse en fonction de son diamètre.	169

Liste des tableaux

1.1	Liste des 8 paramètres d'AROME les plus influents identifiés par Wimmer et al. (2022).	20
1.2	Liste de méthodes POP affectant la paramétrisation de différents processus sous-maille. Les méthodes indiquées en gris ont été développées pour être utilisées dans un modèle déterministe, mais leur principe pourrait être utilisé pour générer un ensemble.	23
3.1	Réglages utilisés pour les simulations uni-colonnes des différents cas d'étude. . .	60
3.2	Caractéristiques des différentes LES utilisées.	67
5.1	Moyenne et écart-type expérimentaux des distributions <i>a posteriori</i> obtenues avec la méthode de Butler et al. (2018).	123
5.2	Caractéristiques des différentes perturbations utilisées. (*) les écart-types de 25 (ARMCu) et 50 (FIRE, BOMEX), et 0.5 (FIRE, BOMEX) et 2 (ARMCu) sont utilisés lorsque N_0 et ν sont tirés dans une distribution lognormale et non dans la distribution <i>a posteriori</i> du problème inverse.	135
6.1	Description des expériences 3D réalisées pour tester l'impact de ν et N_0 sur la simulation du brouillard par AROME. Les valeurs de N_0 sont données en gouttes/cm ³	144