A Single-Column Comparison of Model-Error Representations for Ensemble Prediction

François Bouttier, Axelle Fleury, Thierry Bergot, Sébastien Riette

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.

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 kilometric 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 that 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 in 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 pertubations (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 detraiment
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

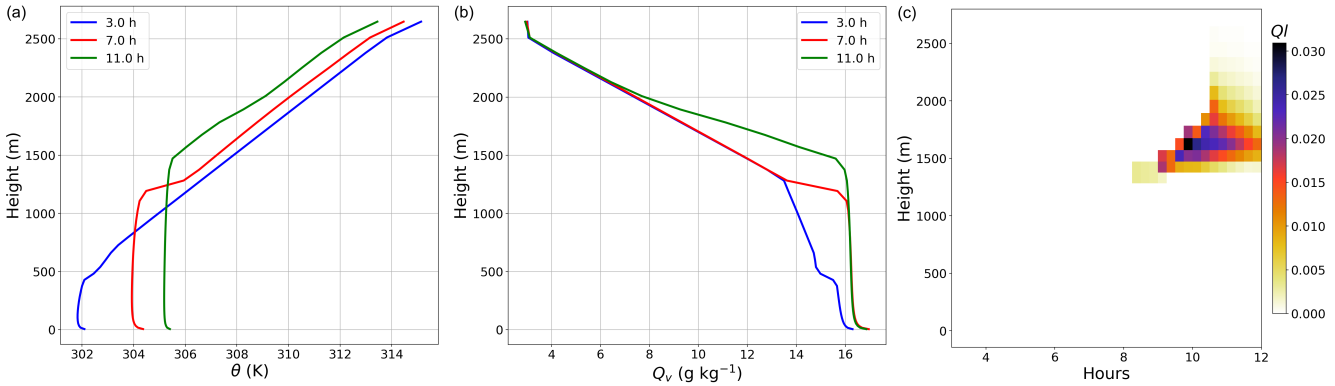

Figure 1: In the ARMCU case, evolution of the AROME vertical profiles of (a) θ and (b) Q_v in the lowest 2500 m after 3, 7, and 11 hours of simulation. (c) time–height distribution of Q_l.

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-

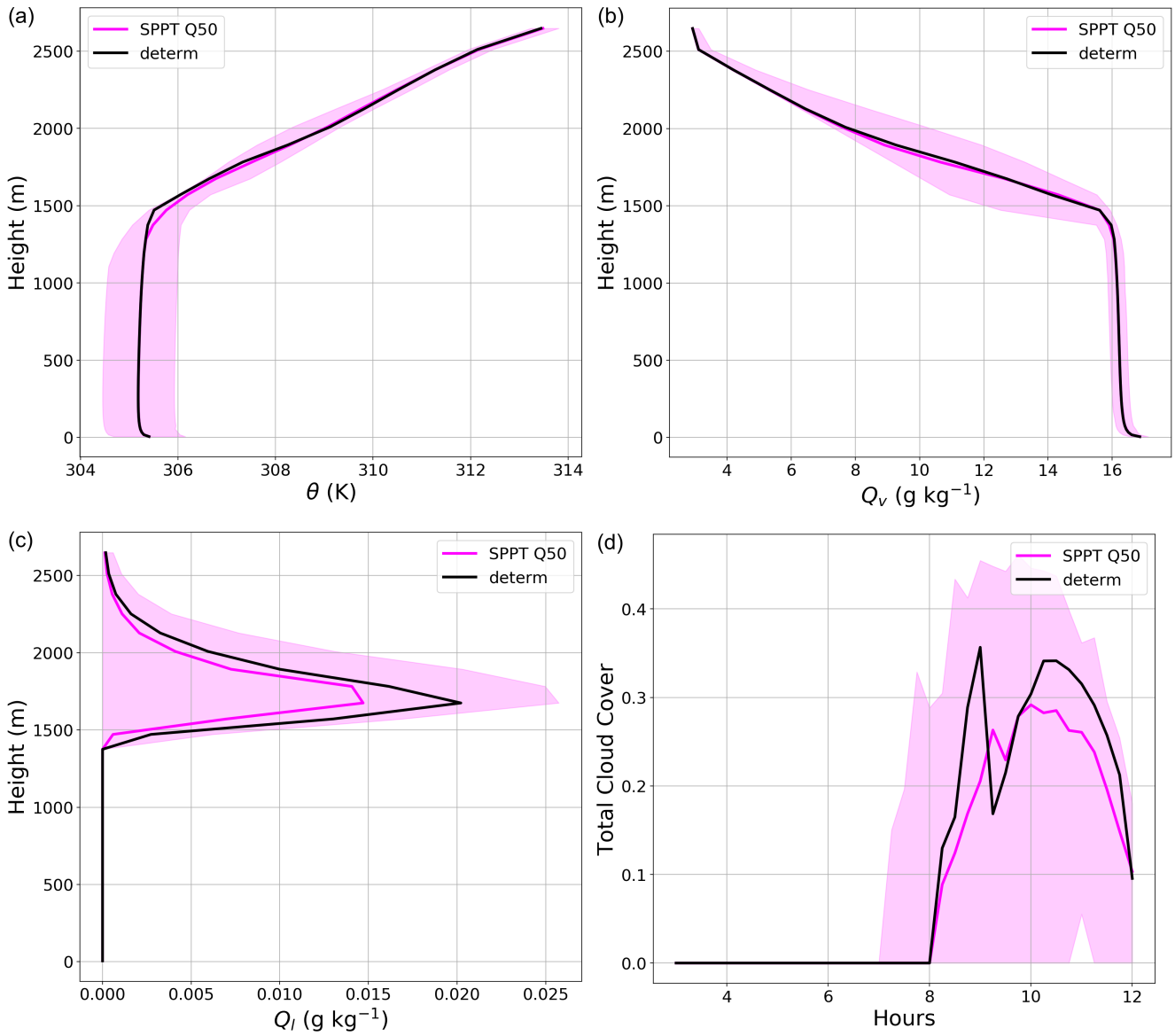

Figure 2: Comparison between the ARMCU SPPT ensemble (magenta) and the deterministic simulation (black): vertical profiles after 11 hours for (a) θ, (b) Q_v, (c) Q_l, and (d) time-series of total cloud cover fraction. The ensemble distribution is represented by its median (dark magenta curves) and the Q10–Q90 percentile interval (shaded).

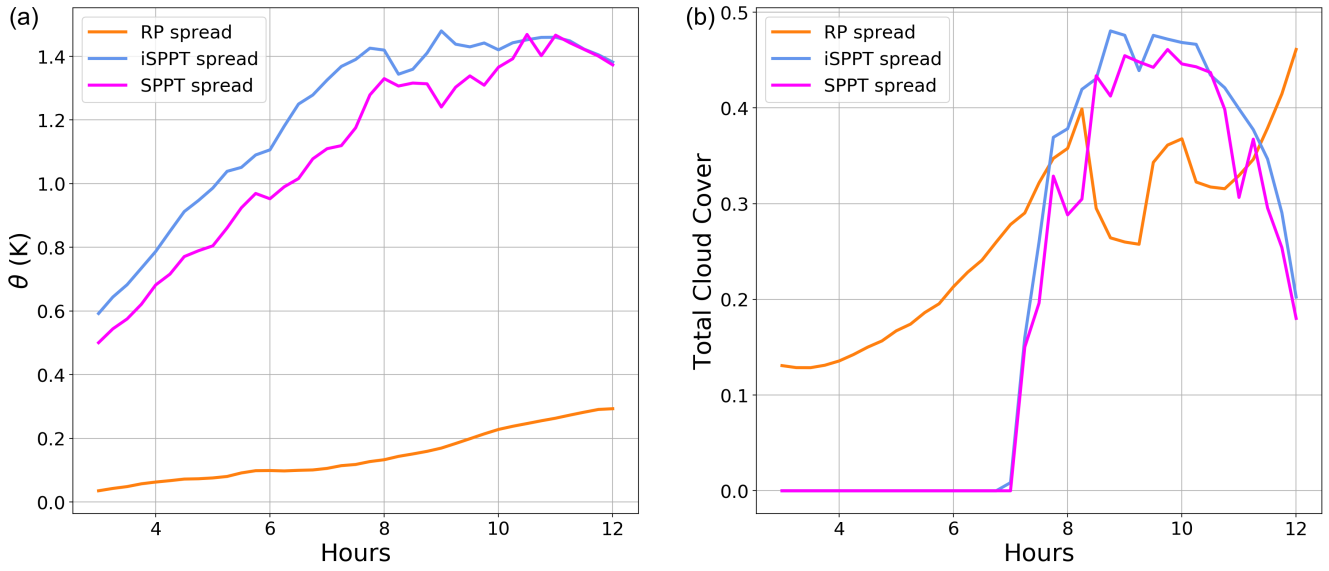

Figure 3: Comparison of ensemble spread for (a) θ at the lowest model level and (b) total cloud cover between the ARMCU SPPT, iSPPT (independent SPPT) and RP ensembles.

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

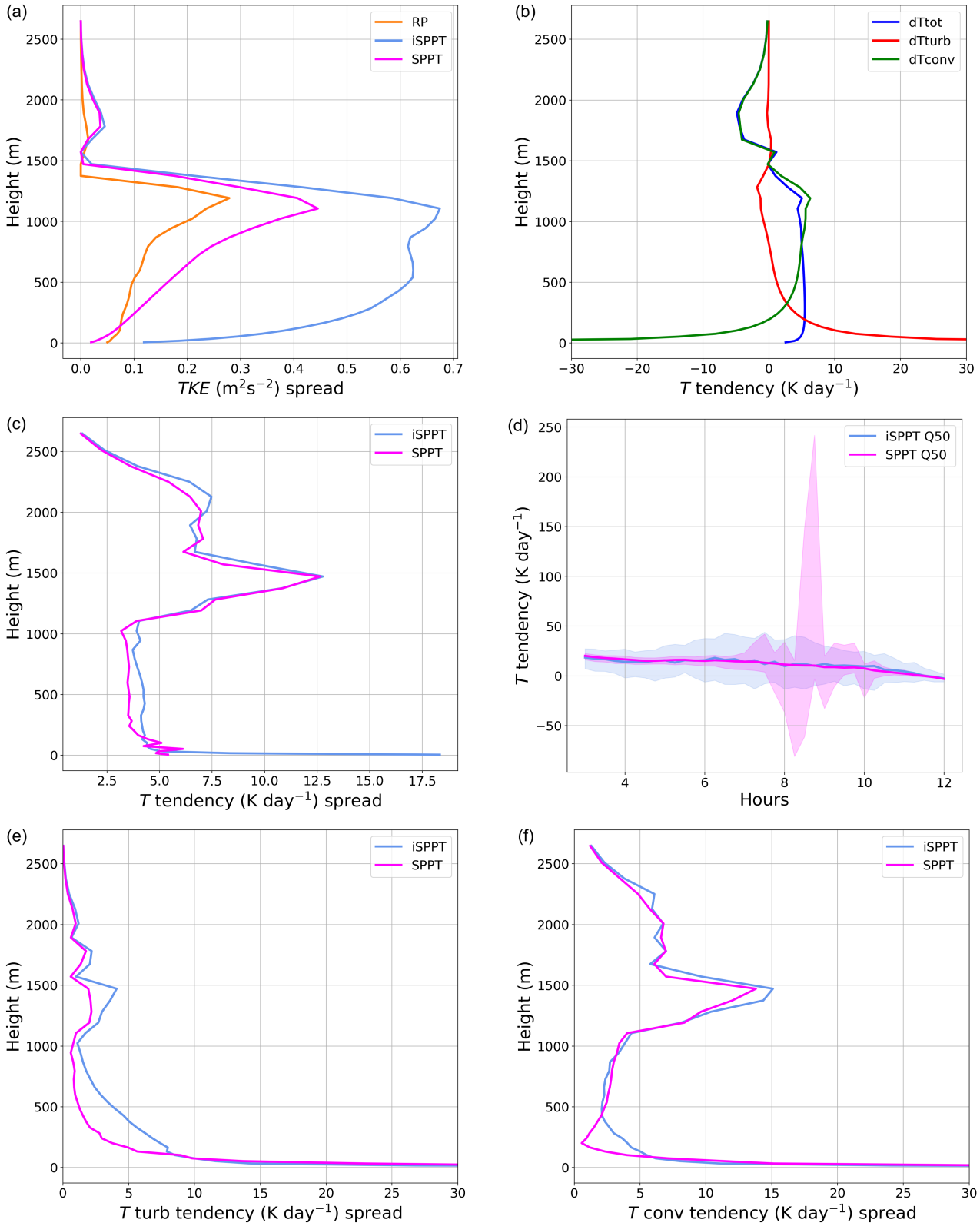

Figure 4: Vertical profiles of (a) TKE and (b, c, e, f) T tendencies and their spread after 11 hours, in the ARMCU case. The ensembles are as indicated in each panel, except (b) which provides for reference the T budget in the deterministic run. Panels (c, e, f) show the vertical profiles of the ensemble spread for each term of the T budget: dTtot, dTturb, and dTconv 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.

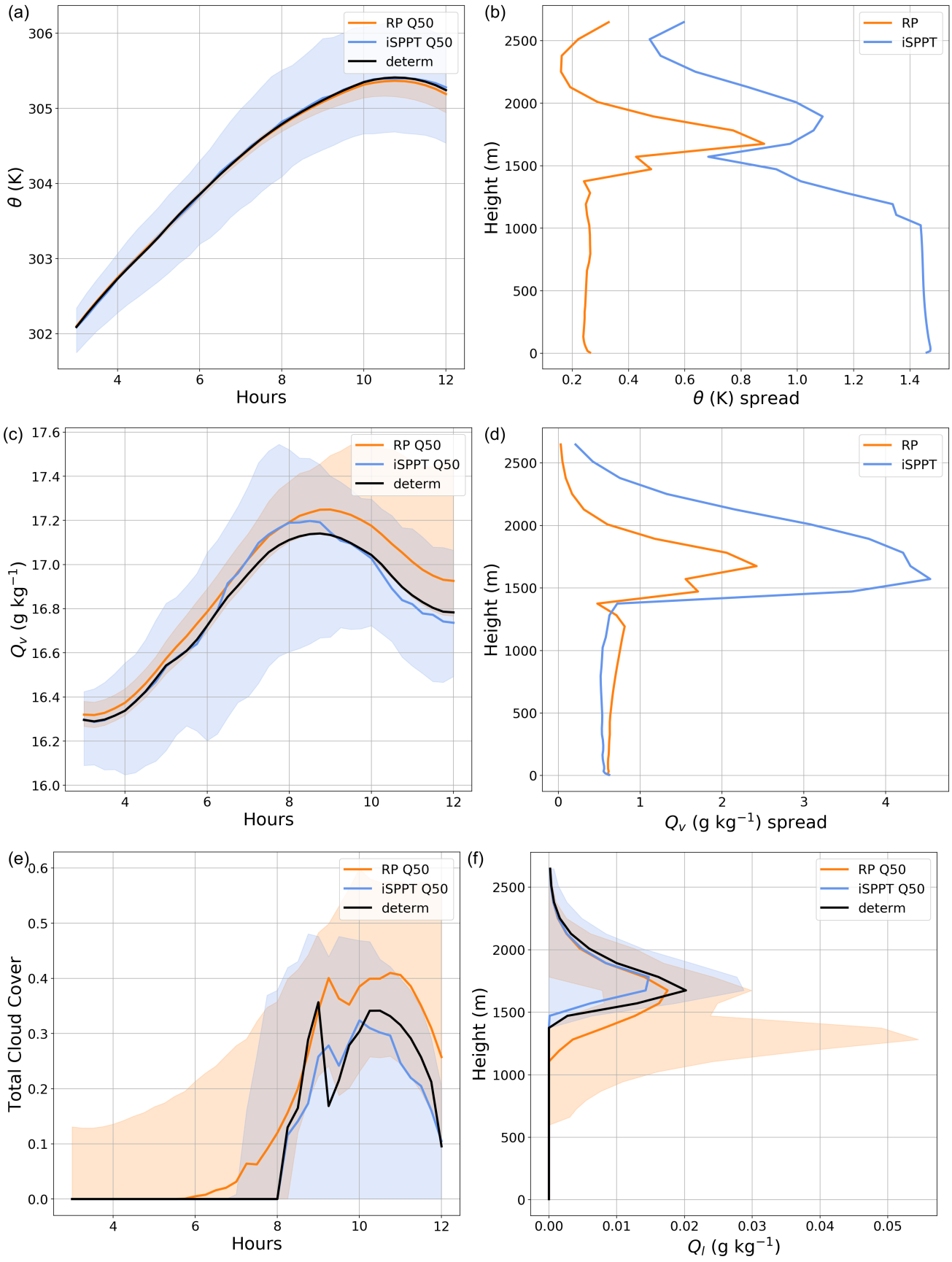

Figure 5: iSPPT (independent SPPT, blue) and RP (random parameters, orange) ensemble distributions for (a,b) θ, (c,d) Q_v, and (e,f) cloud cover and Q_l, respectively. The thick lines indicate the ensemble median and the shading shows the ensemble spread. The deterministic run is shown in black in panels (a, c, e, f). The vertical profiles are extracted 11 hours into the simulation.

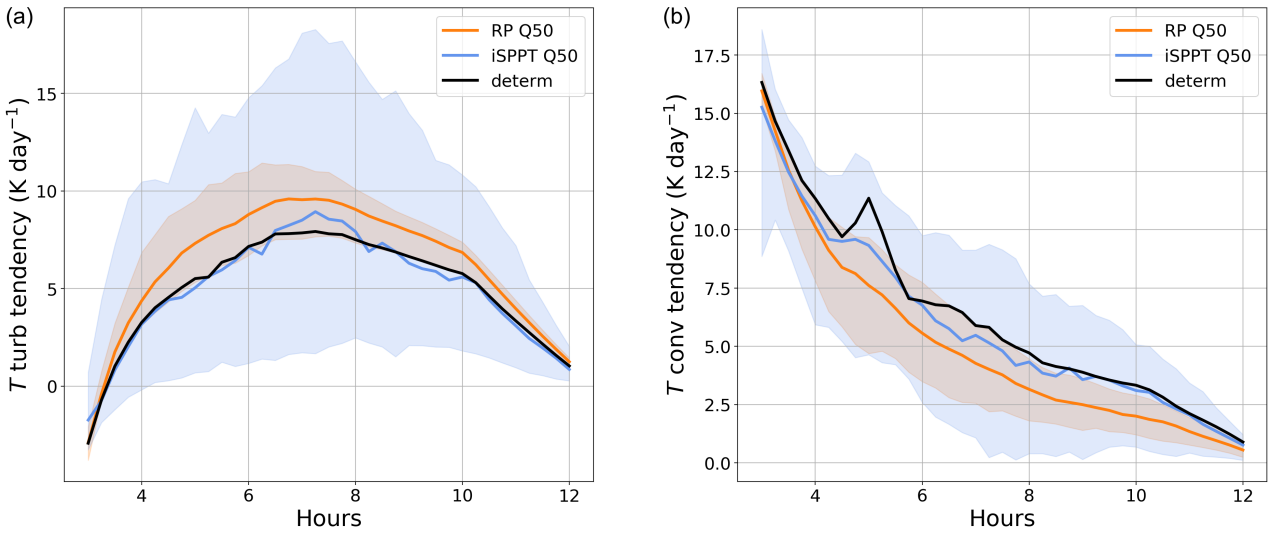

Figure 6: Time evolutions of the RP and iSPPT ensemble distributions of instantaneous temperature tendencies from (a) turbulence and (b) convection at 270 m, i.e., in the lower CBL.

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).

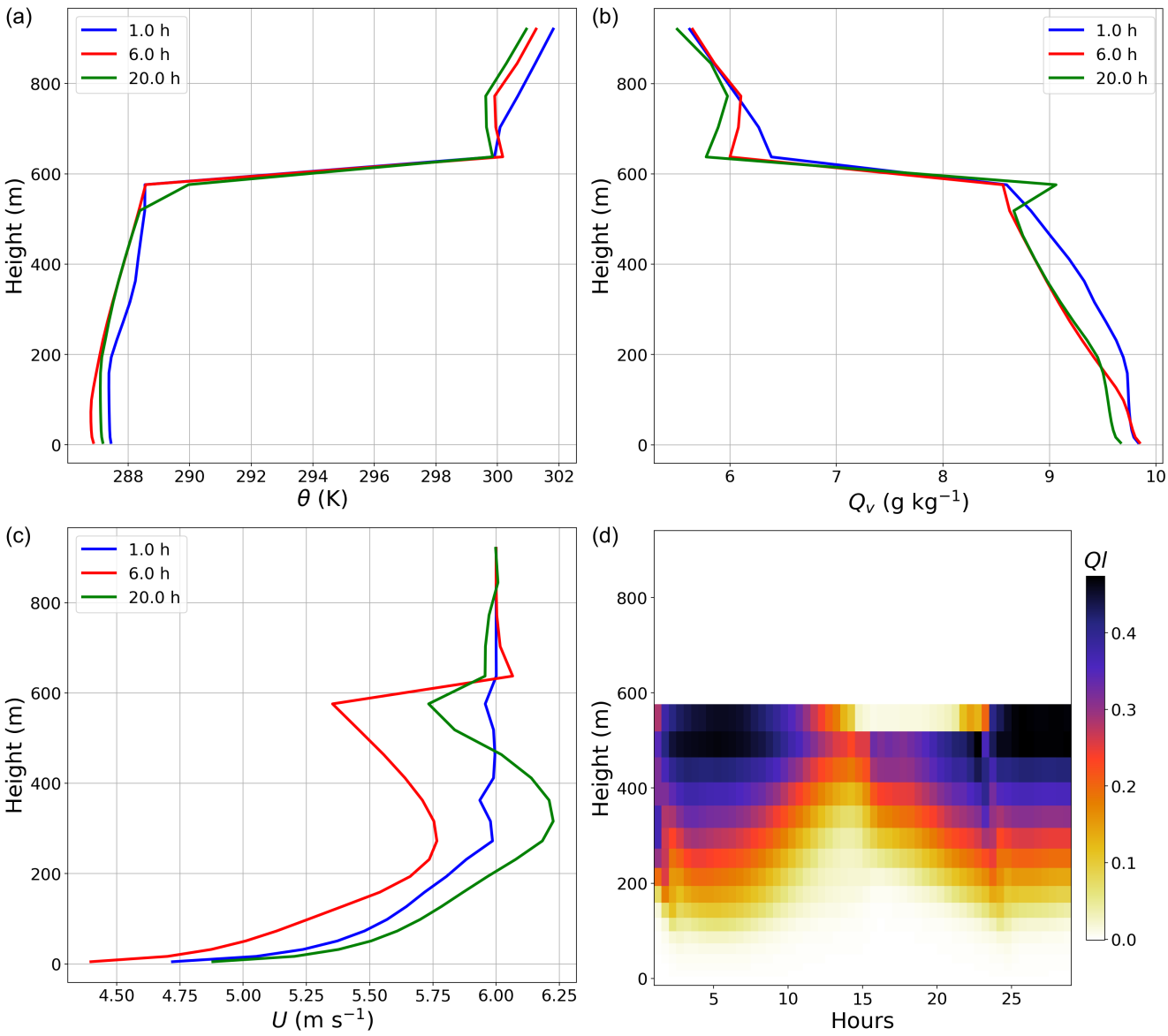

Figure 7: Sample vertical profiles of (a) θ, (b) Q_v, and (c) U after 1, 6, and 20 hours of simulation in the FIRE case. (d) time–height evolution of Q_l.

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 (< 100m) 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),

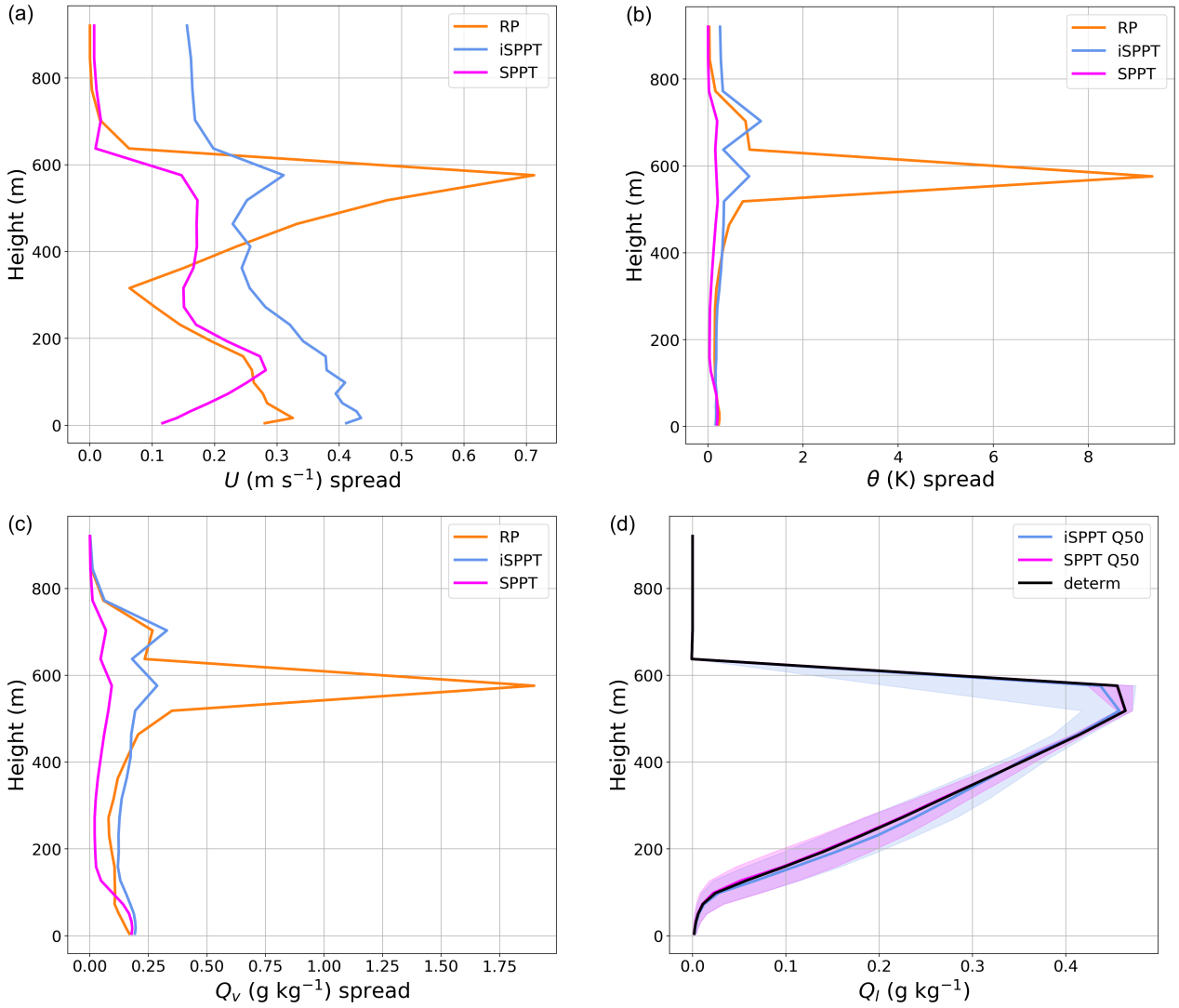

Figure 8: Vertical profile after 6 hours in the FIRE SPPT and iSPPT ensembles for (a) U spread, (b) θ spread, (c) Q_v spread, (d) Q_l distribution. The RP ensemble spread is also shown in panels (a), (b), (c).

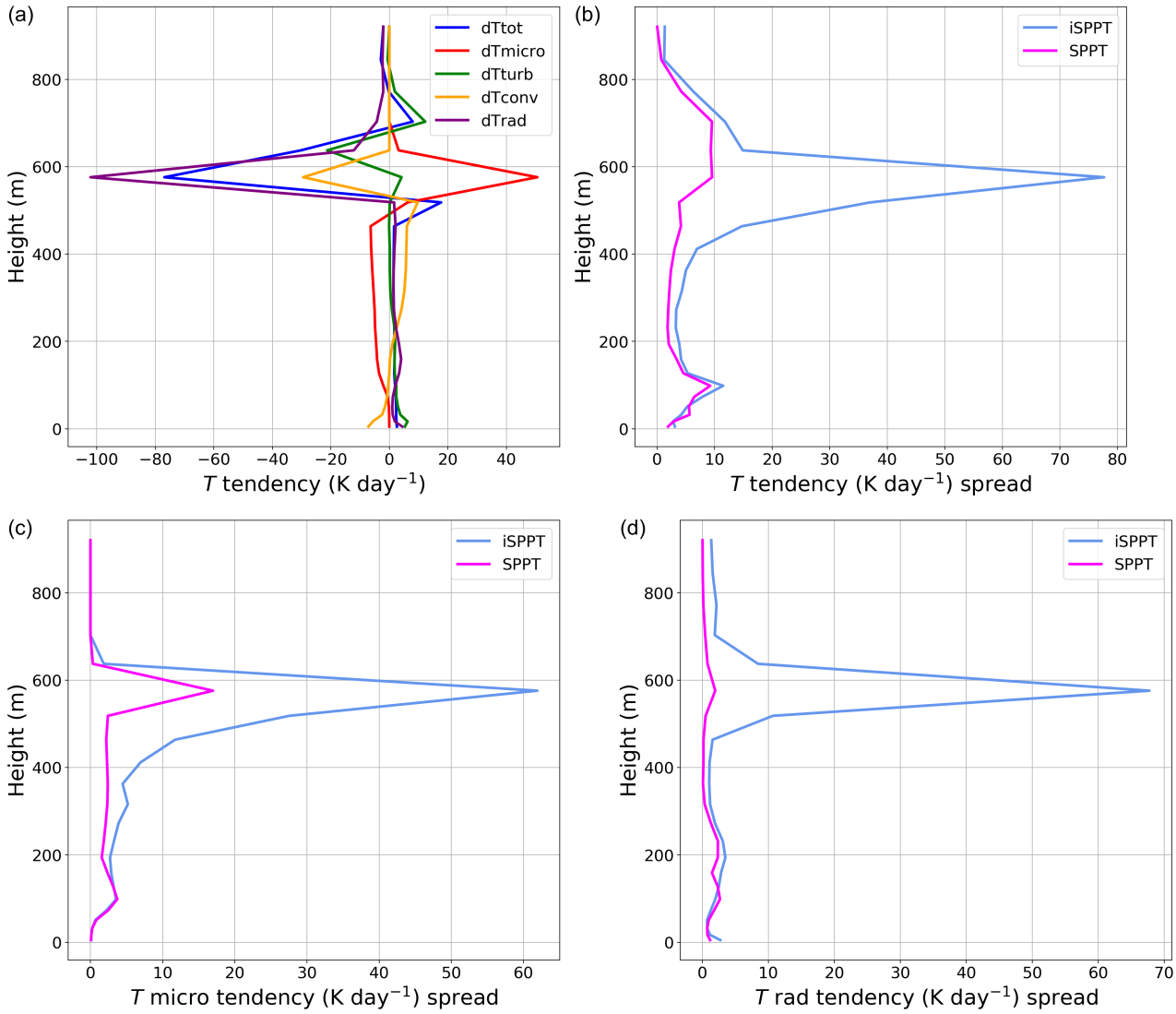

Figure 9: Vertical distributions of the instantaneous temperature tendencies after 6 hours in FIRE: (a) tendency budget in the deterministic run (dTtot, dTmicro, dTturb, dTconv, and dTrad stand for total, microphysics, convection, and radiation tendency, respectively), comparison of the SPPT and iSPPT 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 perturbs 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 interquantile 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

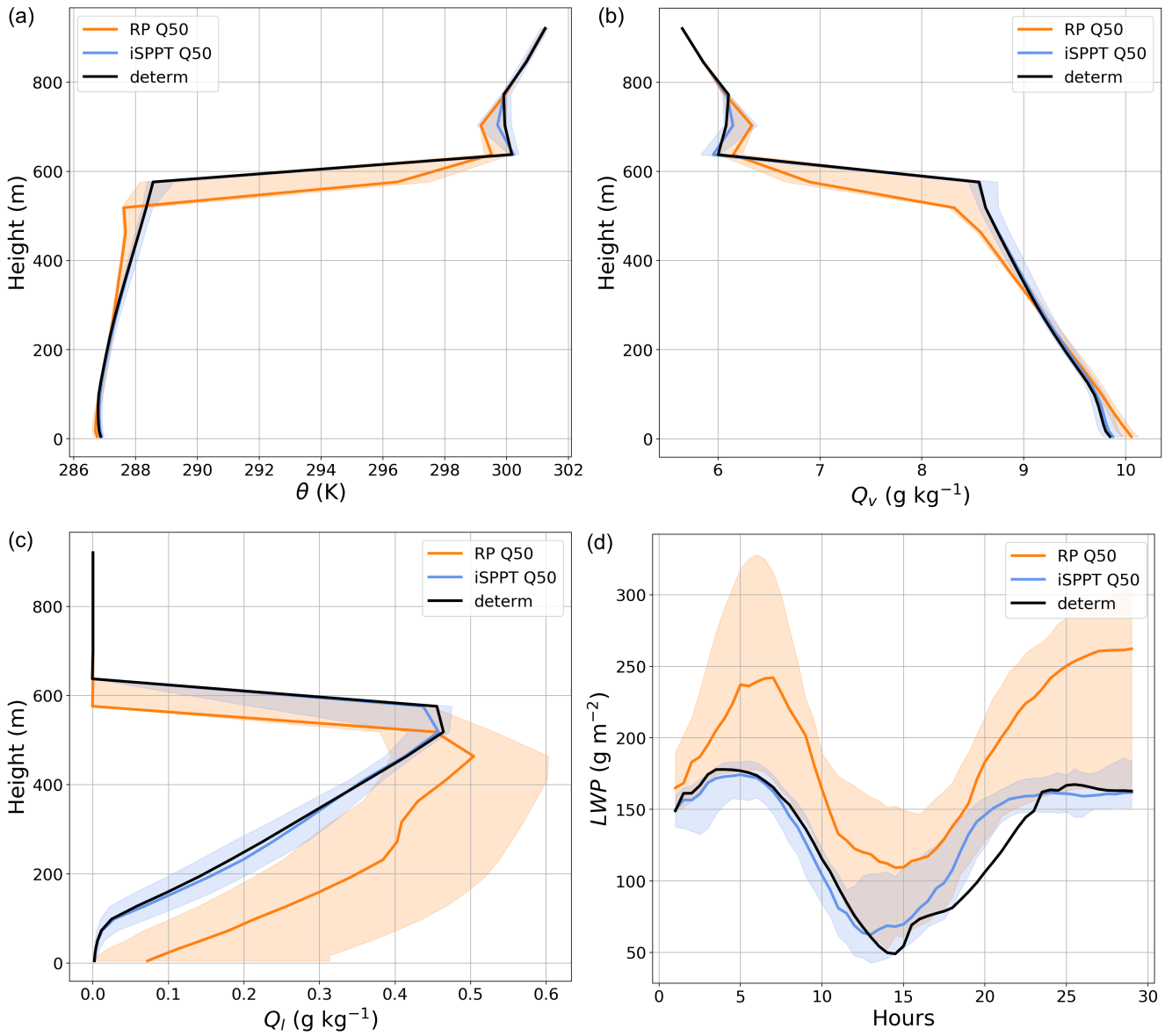

Figure 10: iSPPT and RP ensemble spread on the FIRE case, in terms of 6-hour vertical profiles of (a) θ, (b) Q_v, (c) Q_l, and (d) LWP timeseries.

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

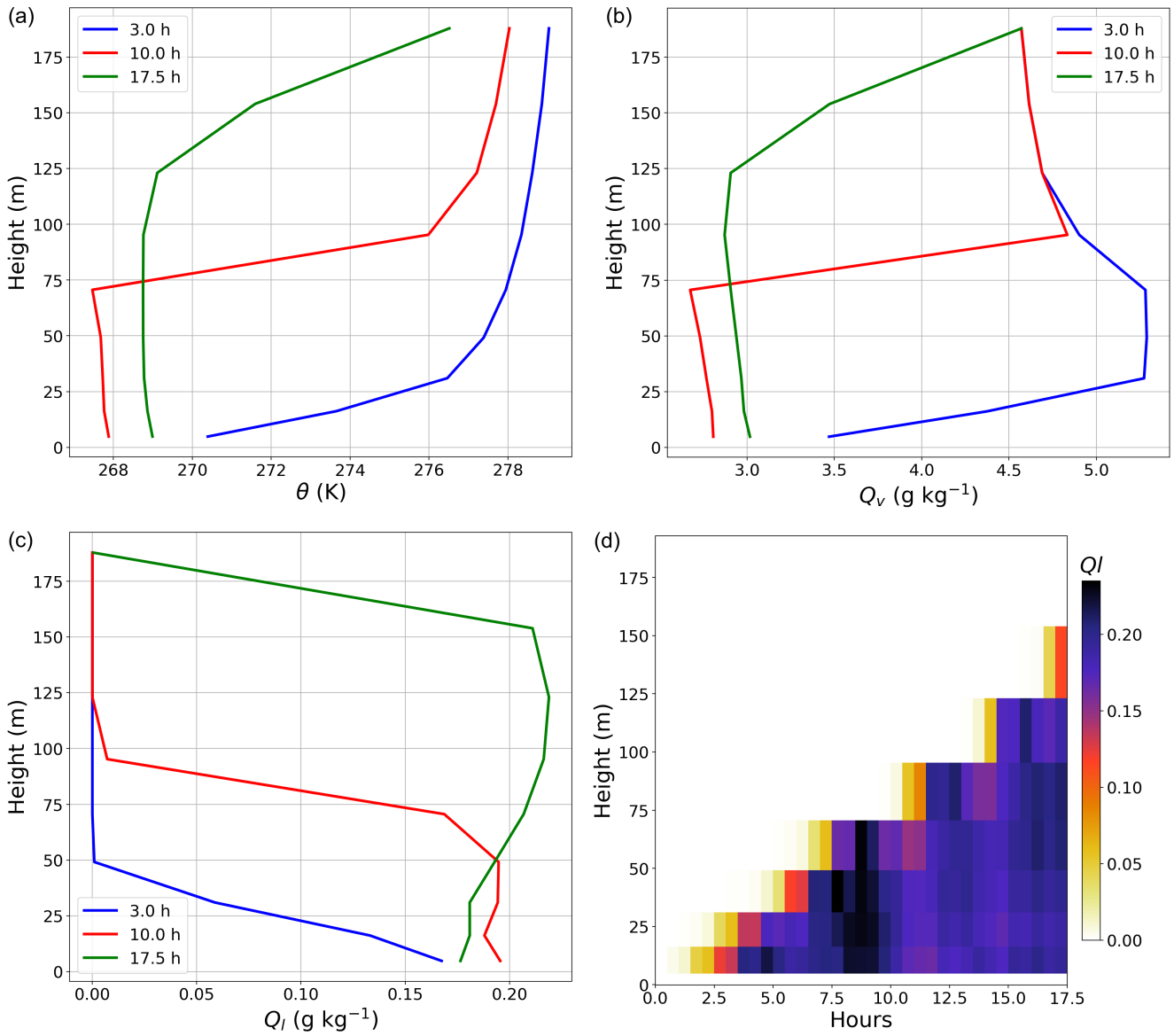

Figure 11: Deterministic LANFEX simulation vertical profiles for (a) θ, (b) Q_v, (c) Q_l after 3 hours (shortly after fog formation), 10 hours (mature fog), and 17.5 hours (2.5 hours after sunrise), (d) time–height evolution of Q_l.

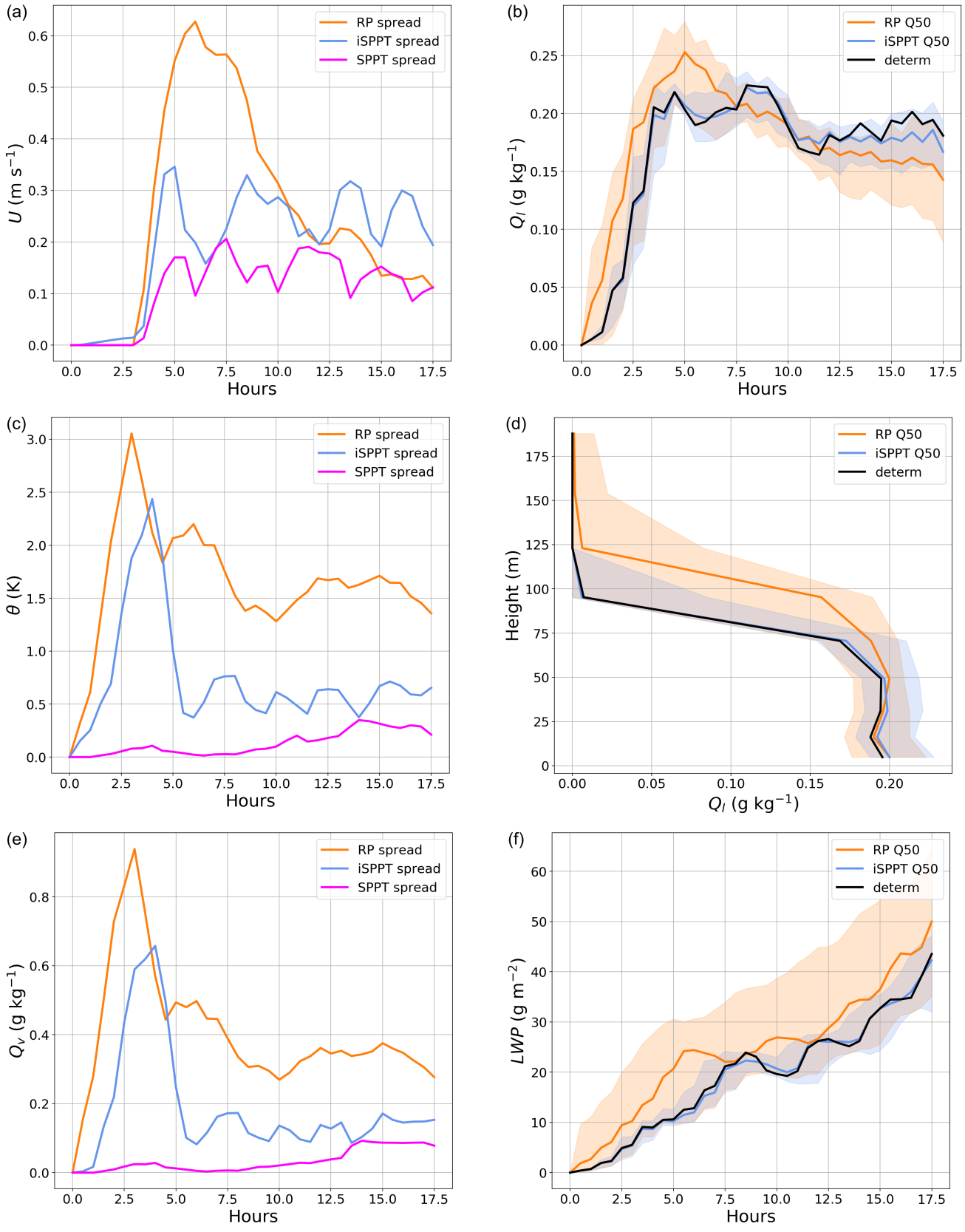

Figure 12: left column: LANFEX ensemble spread evolution at the second lowest model level with the SPPT, iSPPT and RP schemes, for (a) U, (c) θ, (e) Q_v. Right column: evolution of the median and spread of Q_l depicted by its (b) evolution at the second lowest model level, (d) vertical profile after 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.

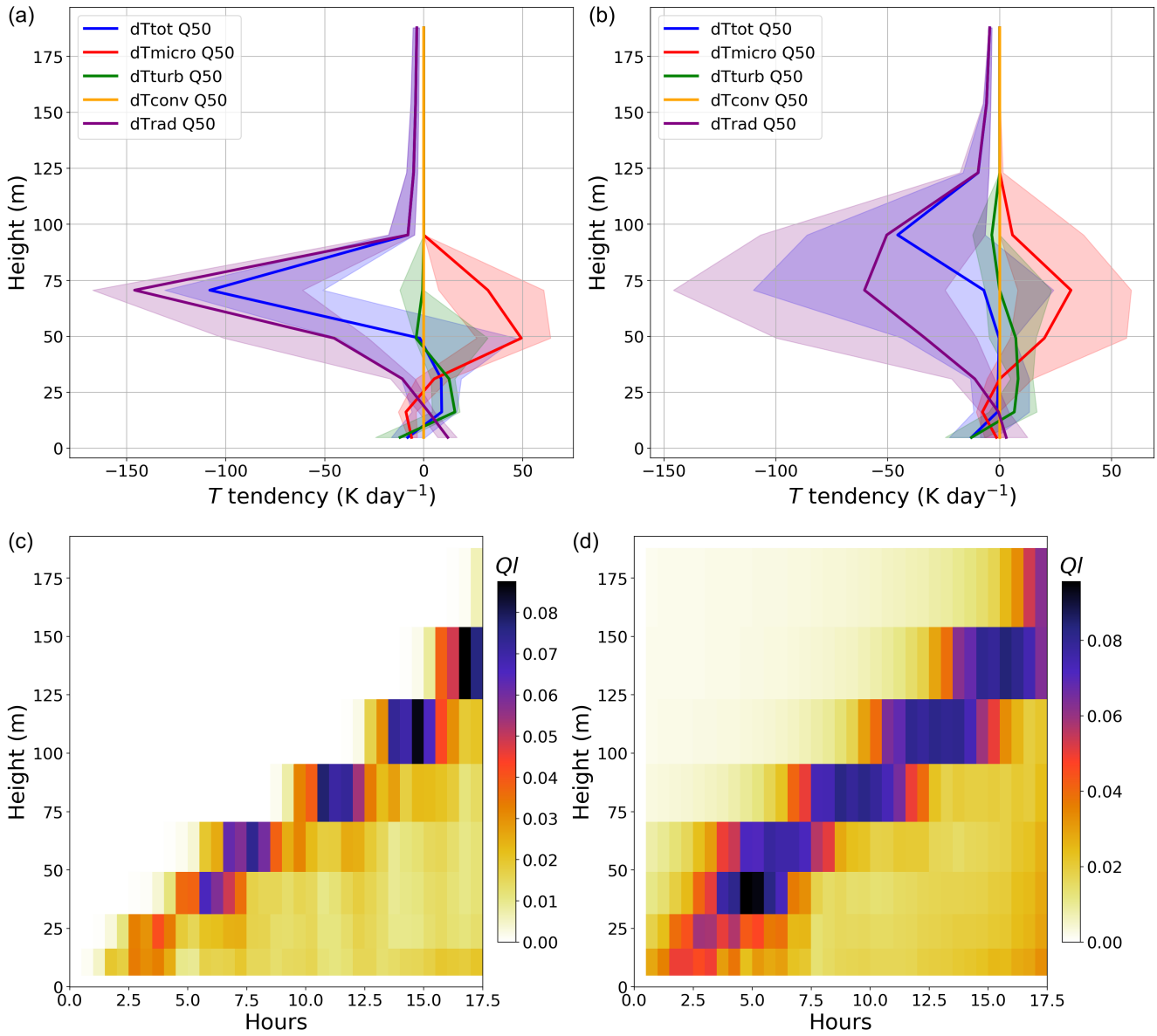

Figure 13: (a) (b) vertical profiles of the instantaneous temperature tendencies in the LANFEX ensemble after 8 hours, in the (a) iSPPT and (b) RP ensembles, with the same conventions as in Fig. 9a (i.e., the thick lines show the median tendencies, the shaded areas depict the Q10–Q90 ensemble quantiles). (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_l 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 appopriate, 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 ensembles 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, Pocernich (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 parametrisations 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, Bouyssel 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