Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a Monte Carlo approach, applied to the Fukushima nuclear accident

Ngoc Bao Tran Le, Irene Korsakissok, Vivien Mallet, Raphaël Périllat, Anne Mathieu

ATMOSPHERIC ENVIRONMENT: X 10 (2021) 100112

Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a Monte Carlo approach, applied to the Fukushima nuclear accident

Ngoc Bao Tran LE [a,b,c,*], Irène Korsakissok [a], Vivien Mallet [b], Raphaël Périllat [a,d], Anne Mathieu [a]

[a] IRSN–French Institute of Radiation Protection and Nuclear Safety, 31 Avenue de La Division Leclerc, Fontenay-aux-Roses, 92260, France
[b] INRIA–French Institute for Research in Computer Science and Automation, 2 Rue Simone IFF, 75012, Paris, France
[c] Sorbonne Université, CNRS, LJLL, 4 Place Jussieu, 75005, Paris, France
[d] Phimeca, 18 Boulevard de Reuilly, 75012, Paris, France

ARTICLE INFO

Keywords:
Atmospheric dispersion
Uncertainty
Monte Carlo
Weather ensemble
Fukushima

ABSTRACT

In emergency cases, when nuclear accidental releases take place, numerical models, developed by French Institute of Radiation Protection and Nuclear Safety (IRSN), are used to forecast the atmospheric dispersion of radionuclides. These models compute the quantity of radionuclides in the atmosphere, their deposited amount on the ground, and the subsequent gamma dose rate. Their results are used to make recommendations to protect the population in case of nuclear accident. However, the simulations are subject to considerable uncertainties. These uncertainties originate from different sources: input variables (weather forecasting, source term), physical parameters used in the models (turbulent diffusion, scavenging coefficient, deposition velocity, etc.) and model approximations (representativeness and numerical errors).

This paper presents the propagation of input uncertainties through a Eulerian radionuclide transport model, *ℓdX*, applied to the Fukushima nuclear disaster. This uncertainty propagation involves perturbing the input variables and making numerous calls to the model. The perturbations should be broad enough to cover the possible range of variation of uncertain variables. Weather forecast ensembles are used to take into account meteorological uncertainties, and several source terms from the literature are included. The following step is to evaluate the spread of the outputs in order to draw insights about the subsequent uncertainties. In order to assess the quality of the ensemble of simulations, comparisons with radiological observations were carried out, using statistical indicators, both deterministic such as Root Mean Square Error (RMSE) or Figure of Merit in Space (FMS), and probabilistic indicators such as rank histograms, Brier score and Discrete Ranked Probability Scores (DRPS).

1. Introduction

1.1. Context

Atmospheric dispersion modeling is the numerical implementation of physical models that describe the evolution of air pollutants in the atmosphere. At French Institute of Radiation Protection and Nuclear Safety (IRSN), atmospheric dispersion models are developed and used to determine the consequences of an accidental release of radionuclides in the atmosphere. The results provide an estimate of the geographical spread of the potentially contaminated areas and the corresponding impact on human health and environment. These estimations can be used as a tool for decision making. On their basis, countermeasures may be proposed to protect the population, such as sheltering, evacuation, thyroid protection by using potassium stable iodine, restriction of food intake or water supply.

IRSN's modeling capabilities include a short-range Gaussian puff model, *pX*, used within a few tens of kilometers from the source, and the long-range model, *ℓdX*. The latter, used in this study, is a Eulerian transport model, based on the solution of a system of 3D advection-diffusion equations. *ℓdX* is an operational version of Polyphemus/Polair3D (Mallet et al., 2007). The two models were used during and

* Corresponding author. IRSN–French Institute of Radiation Protection and Nuclear Safety, 31 Avenue de La Division Leclerc, Fontenay-aux-Roses, 92260, France.
E-mail addresses: ngocbaotran_le@hotmail.fr (N.B.T. LE), irene.korsakissok@irsn.fr (I. Korsakissok), pro@vivienmallet.net (V. Mallet), perillat@phimea.com (R. Périllat), anne.mathieu@irsn.fr (A. Mathieu).

https://doi.org/10.1016/j.aeaoa.2021.100112
Received 5 October 2020; Received in revised form 13 April 2021; Accepted 22 April 2021
Available online 3 May 2021

after the Fukushima nuclear disaster in 2011. Atmospheric dispersion simulations were carried out with the *pX* model within 80 km of the source (Korsakissok et al., 2013) and over the Honshu island with the *ℓdX* model (Mathieu et al., 2012; Saunier et al., 2013). The pX model is typically used between 1 and 30 km from the source, and this may be extended up to 80 km if necessary (Korsakissok et al., 2013). The *ℓdX* model is used at regional or continental scale and the domain may extend up to a few thousands kilometers. Its spatial resolution depends on that of the input weather data. These studies provided a better understanding of the different contamination events, but also shed light on the uncertainties that still linger years after the accident.

Dealing with uncertain information is inherent to a crisis situation, where lack of knowledge and incorrect information (e.g., human errors) are key issues, especially in the early stage of an accident. There are *deep uncertainty*, as defined by French (2015); French, S. et al. (2020), that is, epistemic uncertainties that we know too little of to model or quantify them. As the knowledge of the damaged installation's state and sequence of events improves with time, the confidence in the simulation results should increase accordingly. However, as proven by the Fukushima case, uncertainties can still be large at a post-accident stage (i.e. when major releases are over), even when a large number of environmental observations are available. Besides, time is a key issue in the decision making process during an emergency. Currently, numerical simulations are carried out in a deterministic way. A single result is obtained by making conservative hypotheses in an attempt to encompass all sources of uncertainties. This practical approach has its limitations, because (a) it is very difficult to ensure that no possible contamination scenario has been left out, and (b) it may lead to unnecessary protective actions that can have large human and economic consequences. Therefore, a better understanding and quantification of all sources of uncertainties related to atmospheric dispersion simulations is crucial.

1.2. Uncertainty analysis

The stages of uncertainty analysis of numerical models are, following Girard et al. (2014); Armand et al. (2014):

1. Identify the main sources of uncertainty and the parameters that significantly represent them,
2. For each parameter, construct a probability distribution that represents our *a priori* knowledge on this variable,
3. Determine the output variables that are sensitive to these input variables,
4. Propagate the uncertainties through the model,
5. Use the observations to improve our knowledge on these parameters (calibration).

This paper deals with stages 1–4 for the Fukushima case, as detailed in the following sections.

1.2.1. Input uncertainties (stages 1 and 2)

Uncertainties in atmospheric dispersion simulations primarily stem from input data, namely meteorological fields and source term. The former comprise two or three-dimensional fields varying in time for several variables such as wind, rain, or atmospheric stability. The latter consist in the time-varying release rate for several emitted species, as well as emission location and release height (including possible buoyancy effects).

As far as meteorological fields are concerned, some weather situations may lead to larger uncertainties than others. For instance, the timing and location of rain events, particularly convective showers, are difficult to forecast with accuracy. In the case of low wind speed, the wind direction is not well established and uncertainties can be very large. Therefore, meteorological uncertainties depend on the weather situation as well as on the forecast lead (i.e. how far in the future the weather forecast is used). They are also related to the meteorological model characteristics, especially its spatial and temporal resolution. In the case of the Fukushima accident, several studies highlighted the uncertainty of the meteorological inputs, and their effect on deposition patterns (Draxler et al., 2015; Arnold et al., 2015; Leadbetter et al., 2015).

The *a priori* knowledge of a source term, in the pre or early-release stage, is derived from the available knowledge of the state of the reactor and on-site operators' actions. In that case, uncertainties can be huge but very dependent on the accidental scenario. When environmental observations are available, *a posteriori* source terms can be constructed by taking into account these measurements. In the case of the Fukushima accident, all source terms in the literature were derived from environmental measurements, but there is still a large variability in these estimations, both on the total emitted quantities for the different radionuclides (a factor 2 or more) and the emission rate as a function of time (Mathieu et al., 2012, 2017).

Aside from these two sources of uncertainties, physical processes such as dry deposition, wet scavenging, or turbulent diffusion, are represented with physical parameterizations that are also subject to errors. Finally, the dispersion model in itself leads to biases due to numerical and mathematical errors. Bedwell et al. (2018) and Wellings et al. (2018) conducted a literature review of the range of variation for the most common physical parameters used in atmospheric dispersion models. Concerning the Fukushima case, the sensitivity to wet deposition schemes was investigated by Arnold et al. (2015); Quérel et al. (2015); Leadbetter et al. (2015).

A summary of these input uncertainties and how they can be modelled is provided in Leadbetter et al. (2020).

1.2.2. Sensitivity and uncertainty propagation (stages 3 and 4)

The aforementioned studies conducted after the Fukushima accident are all *local* sensitivity analyses, devoted to investigate the influence of one particular kind of variable on a given output. While these studies allow analyzing in detail the effect of a particular physical process, they do not provide a global view on the joint effect of all uncertain variables together. A *global* sensitivity analysis based on Polair3D was carried out by Girard et al. (2014, 2016) on the Fukushima case. These papers studied the relative influence of a set of uncertain inputs on several outputs. They concluded that the highest uncertainties in nuclear dispersion simulations for the Fukushima accident were due to the weather forecasts and the source term, and quantified the part of output variance related to each input variable. These studies were carried out by using large ranges of variation in an attempt to broadly encompass the input uncertainties, without trying to determine realistic input distributions. In particular, constant, homogeneous perturbations were applied to the meteorological fields, while meteorological uncertainties are inter-dependent and highly heterogeneous. To take these uncertainties into account, some studies added complex perturbations in meteorological inputs (Girard et al., 2020). Another way to represent these uncertainties with spatial and temporal variations is using meteorological ensembles, constructed by weather offices to be representative of weather forecast uncertainties. They have been used as input for dispersion models in the case of Fukushima (Sørensen et al., 2016; Kajino et al., 2019) as well as in hypothetical nuclear accident scenarios (Sørensen et al., 2016; Korsakissok et al., 2018, 2019, 2020; De Meutter et al., 2016). These studies accounted for meteorological uncertainties only, or, in some cases, both meteorological and source term uncertainties in a systematic approach (that is, all combinations of meteorology and source terms were simulated), which implies a significant computational burden.

To properly conduct an uncertainty analysis, prior input distributions have to be determined for all uncertain input parameters; these distributions are then sampled and propagated within the atmospheric dispersion model, typically with a Monte Carlo (MC) method. This approach, combined with the use of weather forecast ensembles, was

carried out with the short-range model *pX* on the Fukushima case and compared to the systematic, cross-simulations approach mentioned earlier (Périllat et al., 2016). Simulations were compared to radiological observations of dose rate and airborne deposition measurements collected in Japan.

1.3. Objectives of the study

The aim of this study is twofold: (a) to investigate the use of a coarse-resolution weather forecast ensemble for propagation of meteorological uncertainty, and (b) to propagate the uncertainties in the Fukushima case at continental scale (up to a few hundreds of km), using the *ℓdX* model.

Périllat et al. (2016) showed that using a high-resolution meteorological ensemble may not be significantly beneficial for our purpose, and a coarse-resolution, operational ensemble may be sufficient to encompass most uncertainties. The first part of this paper aims at analyzing more thoroughly an operational weather forecast ensemble provided by the European Center for Medium-range Weather Forecasts (ECMWF), as could be used in case of an emergency (Section 2). In particular, the effects of grid resolution (Section 2.2.1) are investigated.

The second part of this paper (Section 3) presents MC simulations for the Fukushima disaster with *ℓdX* model. In this study, we used all members of the weather ensemble, six *a posteriori* source terms from the literature, and random perturbations on the other inputs. The simulations are compared to radiological observations of activity concentration in air, gamma dose rate and airborne deposition measurements collected in Japan. The principal objectives are the validation of the *a priori* distributions of the uncertain inputs and the analysis of sensitive outputs.

2. Analysis of a meteorological ensemble

2.1. Description of ensemble weather forecasts

2.1.1. Principle of ensemble forecast

Numerical weather prediction (NWP) models are used to predict the evolution of the state of the atmosphere. The first step is to evaluate the current state of the atmosphere (called the analysis). To achieve this goal, millions of observations are available everyday. The various observations and *a priori* simulated estimations are combined to make a consistent 3-D multivariate representation of the atmosphere, a process called data assimilation. The resulting analysis is subject to uncertainties coming from the observations partial coverage, observational errors as well as errors introduced by the data assimilation process itself. The analysis is used as initial condition to the subsequent forecasts made by NWP models, which rely on a discretization of the mathematical equations governing the physical processes occurring in the atmosphere. Therefore, further uncertainties are introduced in the forecasts by the physical and numerical approximations made by the model. Boundary conditions, such as interactions of the atmosphere with the land or ocean, are imperfectly taken into account, leading to an additional source of uncertainties. These three sources of uncertainty or error (initial conditions, modeling discrepancy and boundary conditions) can all be reasons for a forecast to differ from what actually occurs. Due to the chaotic nature of the atmosphere, small initial errors can grow during the forecast and lead to substantial changes in large-scale patterns. The rate at which these inevitable errors grow vary from day to day (it is flow-dependent). Sometimes, small errors in the initial conditions may result in dramatically different outcomes, while on other occasions, the forecasts will be more or less the same regardless of the slight perturbations in the initial conditions.

An ensemble is a set of forecasts run from different initial conditions to account for initial uncertainties, and possibly including NWP model errors as well. The ensemble of forecasts provides a range of possible future scenarios describing how the atmospheric flow may evolve, in accordance with our knowledge of its initial state and of the model capabilities. The ECMWF-ENS operational ensemble prediction system (EPS), as described in Leutbecher and Lang (2014), comprises fifty simulations representing slightly different starting conditions, plus one "control" member (without perturbation). Each simulation is called a member. The initial perturbations are generated from an ensemble data assimilation (EDA) process Buizza et al. (2008). It consists in adjusting the recent ensemble forecast (providing a "first guess") to available observations using 4D-var techniques. Ideally, the EDA would be sufficient to represent the initial uncertainties. As this step does not provide enough spread, this approach is completed by another one called Singular Vector–SV (Leutbecher and Lang, 2014). This mathematical technique generates perturbations to an initial state that will grow the fastest over a given time interval (here, 48 h). Using these perturbations together with the EDA helps to compensate for having a relatively small number of ensemble members, by ensuring to produce fast-growing forecast spread.

In addition, the ECMWF-ENS forecast takes account of model uncertainties. Physical parametrizations are used to represent the effect of processes that are unresolved (or partially resolved) by the model. In particular, subgrid-scale phenomena, occurring at a scale smaller than that of a grid box, are parametrized. Uncertainties arise from the parametrized processes due to two aspects:

- Using simplified representations of processes not fully resolved: the impact on the resolved flow is represented via some bulk/averaged quantity (e.g., surface drag, convection rates, phase transitions, radiation transfers),
- Representing processes that are poorly constrained due to lack of observational coverage, or not fully understood (e.g., vertical cloud-overlap, composition of the atmosphere, non-orographic drag).

ECMWF uses a combination of two methods: Stochastically Perturbed Parametrization Tendencies (SPPT) and Stochastic Kinetic Energy Backscatter scheme (SKEB) to account for these different aspects of the model uncertainty Palmer et al. (2009). The SPPT scheme introduces a random element to the parametrized variables, in order to account for the subgrid-scale uncertainty. The SKEB scheme accounts for energy transfers between subgrid scales and larger scales. In the ensemble construction, each initial (perturbed) state is then used with a different version of the numerical model, modified by random perturbations from SPPT and SKEB schemes. This inclusion of model uncertainties is made in order to generate a larger ensemble's spread, in an attempt to make the spread more representative of the error for variables used in weather forecasting. In practice, there are limitations to these schemes in terms of physical consistency of the forecast and partial representation of the model uncertainties. A better representation of model uncertainties is currently under development at ECMWF (Ollinaho et al., 2017).

2.1.2. Application of ensemble weather forecast to the Fukushima study

The ECMWF-ENS operational ensemble used in this study was retrieved from the ECMWF archive in 2017. It contains 50 members with 0.25° of horizontal resolution and 3-h time step, the domain spans 40 × 40 grid boxes that cover a great part of Honshu island. The vertical resolution of the data is coarse, with only five pressure levels to cover the boundary layer height: 50,000, 70,000, 85,000, 92,500, 100,000 Pa. For dispersion purposes, the meteorological variables are then interpolated on a finer grid adopted to our needs (see section 3.2). The ECMWF operational ensemble forecast is representative of the kind of data that could be retrieved and used on a relatively short term for emergency response. Ensembles of finer resolution have been made operational for weather forecast at national level as computer efficiency increases, but for a given computational burden, there is a balance between the respective benefit of increasing resolution compared to using more ensemble members to represent uncertainty (Raynaud and Bouttier, 2017). The huge amount of data for high-resolution ensembles currently

poses a challenge for downloading, processing, using and storing the data and results. On the other hand, previous studies have shown that coarse-resolution ensembles may be sufficient to provide a first, broad view of the main uncertainties related to dispersion (Périllat et al., 2016). One of the main objectives of this study is to further investigate to what extent an ensemble of this spatial and temporal resolution may be appropriate to account for uncertainties in the first stage of emergency response.

Another concern is that such an ensemble is designed to capture the spread of meteorological variables on synoptic scales, while the domain and scales in this study are much smaller. Weather forecast ensembles, however, are also used in practice to predict local variables within the boundary layer, such as rain showers or wind at 10-m level. It is acknowledged that, generally, these ensembles are under-dispersive when compared to boundary layer variables, since they fail to take into account all sources of uncertainties at this scale. Before using them for dispersion within the boundary layer, a first step is to perform a model-to-data comparison to such local-scale variables in order to have a better view of this under-dispersion (section 2.2).

A third challenge in this study is posed by the length of the accident (three weeks). Usually, weather forecasts are made up to about five days. When the accidental scenario spans a longer time period, it raises the issue of how to properly combine different ensemble forecasts made from different initial times to cover the whole simulation period. To deal with this issue, this study uses a meteorological ensemble made of forecasts that started from 12 h to 36 h before the current time. For day D at 00:00, the ensemble of forecasts was simulated starting from analyzed states (i.e., initial states merging simulations and observations) for 12:00 of the day $D-1$. We make use of the subsequent forecasts until the end of day D. For day $D+1$, we switch to the better forecasts that originate from day D at 12:00. The perturbations of the initial state are independent from one day to the other. This means that the simulations of a member M of the day D and D + 1 do not come from the same perturbation which may lead to assimilation jumps when updating the ensemble member. Fig. 1 illustrates the cycle, using forecasts with a time step of 3 h. The period covered by the meteorological ensemble constructed that way is 11–30 March 2011.

This process leaves clear room for improvement and research on methods to insure a better temporal continuity in the ensemble members for long simulation periods. This ensemble, however, is representative of what could be constructed in a real emergency case, in the sense that 24 h is the typical forecast period that would be used in order to anticipate the consequences of potential releases. The Fukushima accident is a succession of independent releases coming from different units and

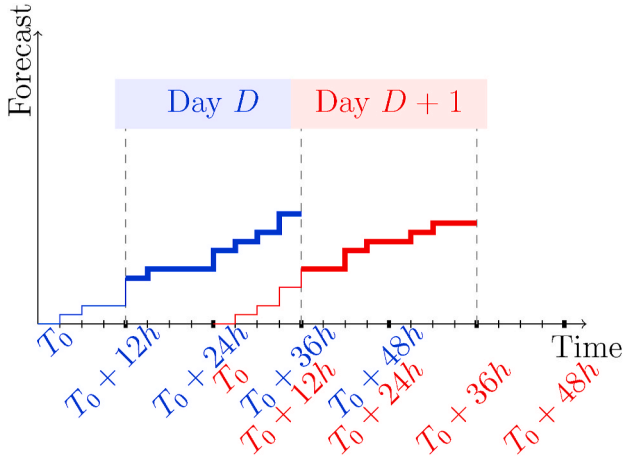


Fig. 1. Illustration of which meteorological forecasts are used (in bold) during the day. The forecasts originate from 12:00 on one day and are used the next day only.

installation events (Mathieu et al., 2018). It is probable that this kind of situation would be dealt with by repeatedly using 24-h forecasts used to infer or update recommendations for the protection of the population. The ensemble constructed here is made in this perspective. The first 12 h of the forecast are not used in order to let the perturbations grow, then the next 24 h are used by the simulation.

2.2. Model-to-data comparison of the meteorological ensemble

The Automated Meteorological Data Acquisition System (AMeDAS) in Japan provides observations of wind direction and velocity, precipitation and/or temperature on more than a thousand measurement stations, from which 653 are within our simulation domain. Stations within the same grid cell are averaged for comparison purposes, and there are 354 cells containing at least one meteorological measurement point. The comparisons are made from March 12^{th} to March 30^{th} 2011; some observations are given every 10 minutes, others every 1 hour. The meteorological time step is 3 h but the model output is saved every 1 h in order to correspond to the observation time resolution. Therefore, our comparisons are made on 354 points and 150 time steps. Finally, among the selected points and time steps, the observed values of wind and temperature are unavailable about 20% of the time. Indeed, many stations were out of order during the first days following the tsunami that triggered the Fukushima accident.

2.2.1. Subgrid-scale variability

While ground observations are representative of one given location, a forecast gives the average of variables over a three-dimensional grid cell. In addition, the horizontal resolution of this ensemble is 0.25°, i. e about 22 km × 28 km at the center of the simulation domain. A cell can be spread over different topographies. Therefore, the comparison to observations from a given monitoring stations can be biased. For example, the temperature predicted at Fujisan is much higher than the observation; but the model-to-data difference is constant in time, about 20°C. This phenomenon can be explained as follow:

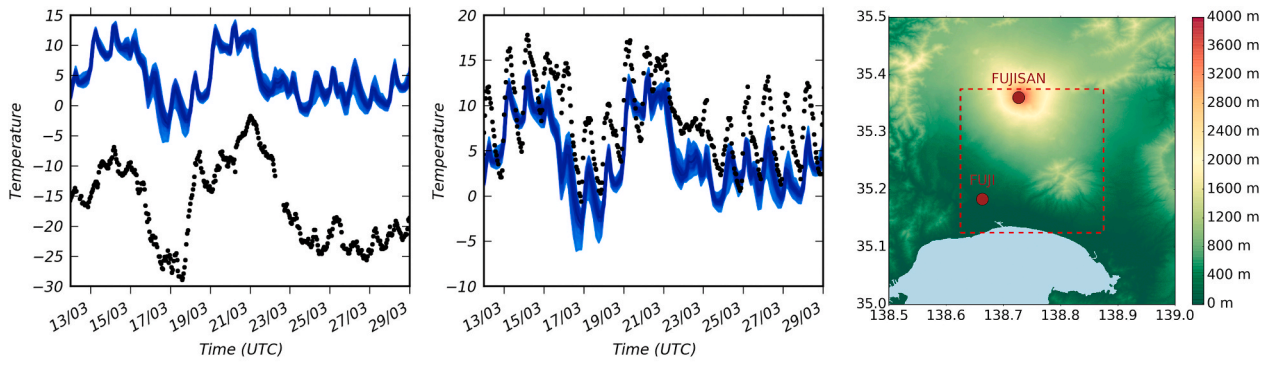
- The altitude of Fujisan station is about 3775 m above sea level. The temperature observed at this station is obviously lower than that at a lower altitude;
- With a large resolution, the cell containing the Fujisan station has a large variation in altitude. Weather forecast takes into account this variation and compute the average across the cell, Fig. 2.

Stations like Fujisan create a bias on the evaluation of weather ensemble. If we compare the forecast interpolated at each station to the measurements, the bias can be important. To circumvent this issue, it is preferable to compare the simulations to averages of all stations in each grid cell. However, this approach is limited by the small number of stations in one grid cell.

2.2.2. Probabilistic indicators for ensemble evaluation

There are several output products for ensemble weather forecast systems, such as meteograms, which display the time evolution of the distribution of meteorological parameters from the ensemble at a given location, as illustrated by Fig. 2. To evaluate the quality of an ensemble by comparison to observations, it is necessary to derive probabilistic scores that will provide a view on the capability of the ensemble to correctly represent the uncertainty for a given meteorological variable. At ECMWF, the overall performance of the forecasts is evaluated by using a set of headline scores which summarize different aspects of the forecast. Commonly used verification metrics are the rank histogram, spread-skill diagrams, Brier score (BS) and Discrete Ranked Probability Score (DRPS) (Hudson and Ebert, 2017).

While the ECMWF products are routinely verified against observations using these scores (Haiden et al., 2019), the variables of interest are those of synoptic scale, as discussed in Section 2.1, whilst we focus



(a) Fujisan (b) Fuji (c) Topographical map

Fig. 2. Temperature comparisons at two stations in the same grid cell (Fujisan and Fuji), and topographical map of this cell. The weather forecast ensemble, which contains 50 members, corresponds to the blue band and the black dots are measurement values. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

here on boundary layer variables, which are much more difficult to represent but are of prime interest for dispersion. Besides, routine verification is often made for a particular forecast lead time, whereas in this case different lead times are used (from 12 to 36 h after analysis time), to better represent what will actually be used for dispersion. Therefore, the verification conducted here is not representative of the ensemble quality for weather forecast, but rather, of how it may be suitable for use for atmospheric dispersion purposes. Fig. 3 gives a first view of three variables: rain, temperature and wind speed of a cell which contains weather and radiological stations. The first map in this figure represents the location of the cell in Japanese territory.

2.2.2.1. Rank histogram. Rank histograms are a good way to measure how well the ensemble is representative of the observed uncertainty. The rank of an observation is determined by counting how many members of the ensemble are below this observation. A "perfect" ensemble would have a flat histogram (see appendix A.1).

Fig. 4 shows the rank histograms for precipitation, temperature, wind speed and wind direction of the ensemble compared with all measurement stations. For the wind speed, we removed the values below $1 \mathrm{ms}^{-1}$ because the measurements of low wind speeds are highly uncertain. The ensemble does not spread out enough, hence its rank histograms show high external bars. There is a tendency to overestimate the wind speed (high left bar). This bias is partially explained by all the coastal cells that contain a large proportion of sea area, where the wind speed is generally higher than over land. The histogram for precipitation shows, with a high bar on the right, that the weather ensemble underestimates the observations. The chosen threshold was $0.5 \mathrm{mm\,h}^{-1}$, the rain gauge threshold, in order to avoid a bias due to the number of occurrences with no rain, where all members and observations are equal to zero. Therefore, only periods where either an observation or a member of the ensemble features a rain episode are taken into account. This leads to a limited number of observations for the construction of the rank diagram. There are only 9.7% rain observations whose values are higher than $0.5 \mathrm{\,mm}^{-1}$. As discussed in Section 2.1, the operational ensemble forecasts are not designed to capture the full uncertainty of boundary layer parameters. This error is still larger due to the spatial and temporal scales of the ensemble. To accurately represent precipitation, the spatial and temporal scales are crucial. The subgrid effect described in Section 2.2.1 can lead to significant bias, due to the discrepancy between the cell average and the observation on station. Besides, the 3-h time step may lead to a delay in the beginning or end of the rain episode, since rainfall rates are interpolated between two time steps. These representativeness errors are not well taken into account by the ensemble approach, and are the main limitation of this dataset.

2.2.2.2. BS and DRPS. The BS is an aggregated measure of discrepancy between the observed outcome and a prediction of threshold exceedance–the average squared error difference. In other words, it is a cost function that measures the accuracy of probabilistic predictions. Thus, the lower the BS is, the more accurate is the prediction. In its most common formulation, the best and worst possible BSs are 0 and 1 respectively.

Assume that we have n binary outcomes $o_1, \ldots, o_n \in \{0, 1\}$ that we want to predict. Let our predictions be denoted by p_i, probability that ith outcome is 1, for $i = 1, \ldots, n$. The BS for these predictions is given by the following formula

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2$$

In order to evaluate weather forecasts, the BS can be computed for a sequence of m thresholds of observed variables. Let us take an example in computing BS of rain forecast, we consider $H = (h_1, \ldots, h_m)$, the vector which contains rain thresholds in mm/h. The BS is computed with each h_j for all $j = 1, \ldots, m$

$$o_i = \begin{cases} 1 & \text{if } o_i > h_j \\ 0 & \text{otherwise} \end{cases}$$

and $p_i = P(i\text{th outcome} > h_j)$.

Fig. 5 shows the Brier Score as a function of threshold for rain, wind and temperature. The BS as a function of the threshold should converge toward zero for large enough threshold values, where both simulated and observed values are always below the threshold (thus being in perfect agreement regarding the threshold exceedance). Before that, the BS curve may be always decreasing (as for the rain in Fig. 5 (a)) or increase and reach a maximum before decreasing (as for the wind speed in Fig. 5c). The choice of the range of variation of the variable of interest is of particular importance for the result interpretation. Here, the ranges of variation are $[0.5, 10]$ mm/hr for rain (to be consistent with the rain gauges' accuracy), $[-20, 20]$ °C for temperature (in accordance with wintertime temperatures which may be locally very low, as shown by 2, and $[0.5, 30]$ m/s for wind speed. The highest Brier Score obtained is around 0.5 for wind speed and temperature, and 0.4 for rain. The highest BS, which correspond to the largest discrepancies, are obtained for low threshold values. For instance, in the case of rain, the prediction of threshold exceedance is less accurate when model or observations have zero values (meaning that there are a lot of false-positive or false-negative occurrences).

To aggregate the information provided by these figures, a single score called CRPS may be computed. It is obtained by integrating the BS on all the thresholds of the variable of interest and dividing by the range of variation of the variable. When the range of variation is discrete, the

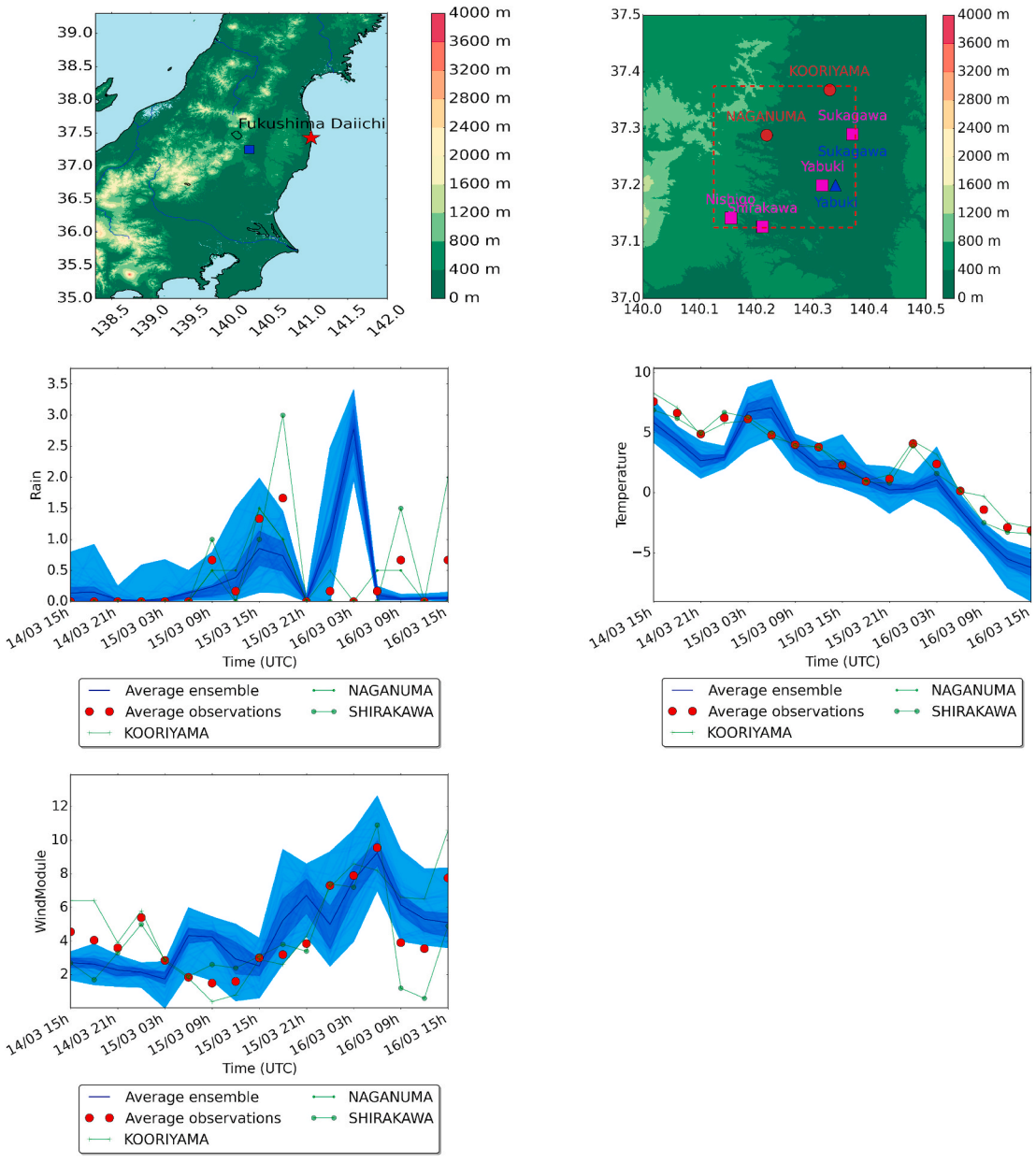


Fig. 3. Rain [mm], temperature [◦C] and wind speed [m/s] in a grid cell on March 14–16. The map on the right is a zoom on this cell. The red points are the meteorological stations, the pink points are the concentration stations and the blue points are the dose-rate stations. Note that Shirakawa is both a meteorological station and a dose-rate station. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

area below the curve is not computed by an integral, but as a sum over "bins" that discretize the x-abscissa. The associated score is called "discrete rank probability score" (DRPS). Again, a good prediction will have a low CRPS or DRPS. However, it should be noted that these scores carry the same unit as the variable of interest. So the CRPS (or DRPS) of different variables, as shown in Table 1, should not be compared.

The ECMWF ensemble could be used in case of an emergency situation, although it may not represent very well the boundary-layer meteorological variables and should be considered with caution. The goal of the next sections is to propagate these uncertainties, along with others, through an atmospheric dispersion model and determine whether this might nonetheless provide a first assessment of uncertainties for operational purposes. A possibility is that the uncertainties on plume trajectories may accumulate over time and/or space, and this, combined with source term and dispersion model uncertainties, might compensate for the limitations of the meteorological ensemble itself.

3. Monte Carlo simulations

3.1. ℓdX description

ℓdX is a Eulerian transport model that solves advection-diffusion equations over a fixed 3D-grid. It is derived from Polyphemus/Polair3D which is an air quality modeling system. IRSN modified this model to meet the requirements and specificities of a nuclear emergency. The advection-diffusion equation of this model for a given radionuclide, noted by index r, is written as follows:

$$\frac{\partial c_r}{\partial t} + div(\boldsymbol{w} c_r) = div\left(\rho \boldsymbol{K} \nabla \frac{c_r}{\rho}\right) - \boldsymbol{F}\boldsymbol{c} + E_r - \Lambda_r c_r \quad (1)$$

where c_r is the concentration in the air of radionuclide r, $\boldsymbol{c}$ the vector of concentrations of all radionuclides, $\boldsymbol{F}$ the matrix of decay coefficients for all radionuclides, $\boldsymbol{w} = (w_u, w_v, w_z)$ the wind velocity vector, ρ the air

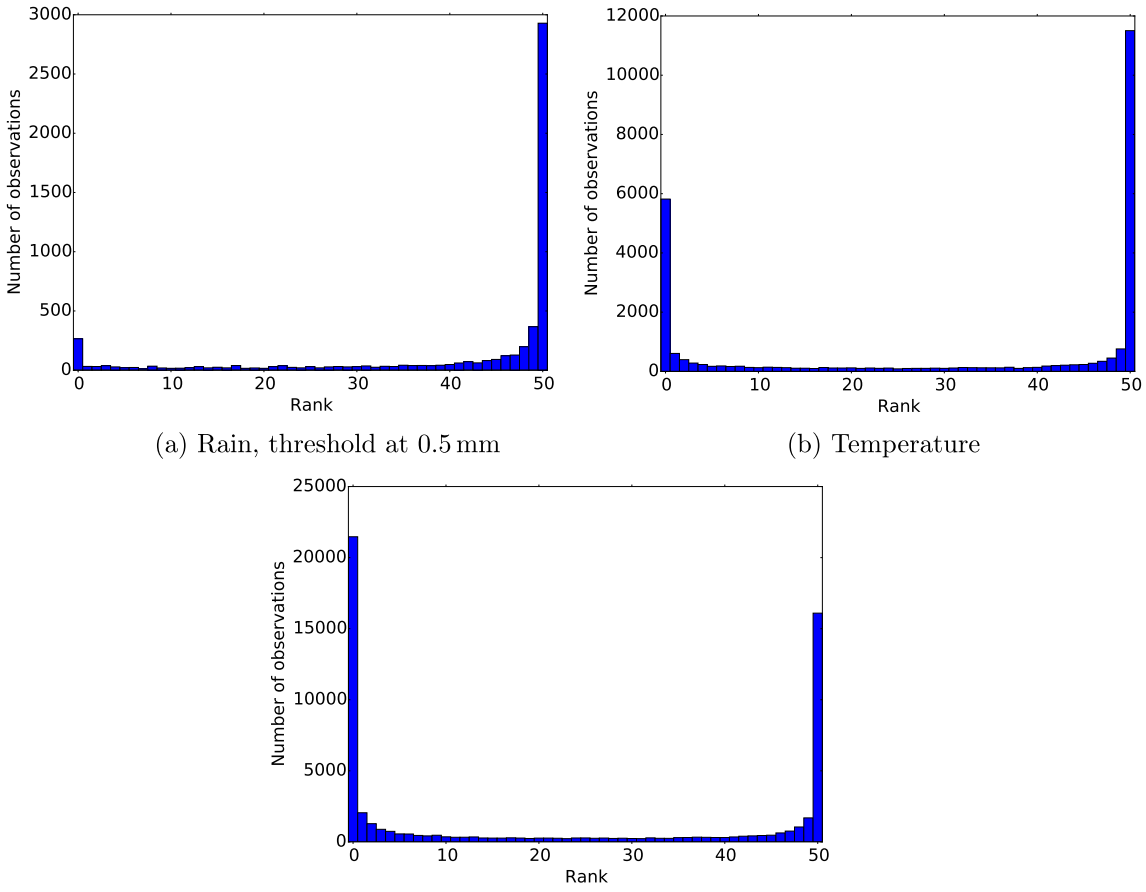


Fig. 4. Rank histograms of ensembles.

density, $\boldsymbol{K}$ the turbulent diffusion matrix, assumed to be a diagonal matrix with diagonal (K_u, K_v, K_z), E_r the emission term and Λ_r the scavenging coefficient for the given radionuclide. The coefficient of horizontal diffusion is computed following a probability density in Table 2. The coefficient of vertical diffusion is computed by using Troen-Mahrt parametrization within the boundary layer and the Louis formula above (Louis, 1979; Troen and Mahrt, 1986). The equation is solved using first-order operation splitting, with diffusion integrated after advection.

ℓdX uses a weather forecast and a source term as inputs to calculate the concentration in the air and the amount of radionuclides deposited on the ground. Speed and amount of deposition depend on weather condition, land use coverage and radionuclide's physical properties. Given the concentration in the air and deposition on the ground for each radionuclide, the gamma dose rate in the atmosphere can be inferred. This makes it possible to estimate the radioactive consequences on human health and environment.

3.2. Simulation setup

Once all sources of uncertainties have been identified, we sample them according to given probability distributions. The MC ensemble contains 200 simulations and is carried out on ten parameters, detailed in Table 2. For each simulation, a meteorological member and a source term are randomly chosen in the weather ensemble and the six source terms, with a uniform distribution. For the other parameters, we have reused the same variations as proposed by Girard et al. (2014). The parameters are not all perturbed in the same way, some are added a random value, some are multiplied by a random value, and others are replaced by a random value. All parameters are perturbed independently. The six source terms used are from the literature. Almost ten years after the accident, no reliable source term has been produced by relying only on the reactor physics and on-site events. All source terms in the literature are based on more or less sophisticated inverse or reverse modeling techniques, which combine environmental observation data, meteorological and dispersion modeling. An account of the different source terms available in the literature and the various techniques and observation datasets used to obtain them is available in Mathieu et al. (2018). Here, six source terms have been retained. Some source terms were not selected, either because they did not significantly differ from another source term (to avoid introducing bias in the ensemble), because they were not deemed relevant due to the spatial and temporal scales of interest in the study, or because insufficient information was available. Although they are all *a posteriori* source terms based on observations, the variability in the overall released quantities is important. The total release of ^{137}Cs over the whole period ranges from 8 PBq to 20.6 PBq, and the differences are even larger when looking at particular release events, due to large differences in the release kinetics Mathieu et al. (2018). These six source terms present complete time series of releases for different radionuclides during the three weeks considered in the article. Fig. 6 illustrates the time evolution of the release rate for ^{137}Cs for the six source terms. It shows the large variability between the temporal evolution of the different release estimates, with peaks of various intensity corresponding to the different release events during the three weeks. In addition, the number of radionuclides considered also varies between one and eight depending on the source term.

As this variability is representative of the *a posteriori* uncertainty, additional perturbations have been applied on the released quantities

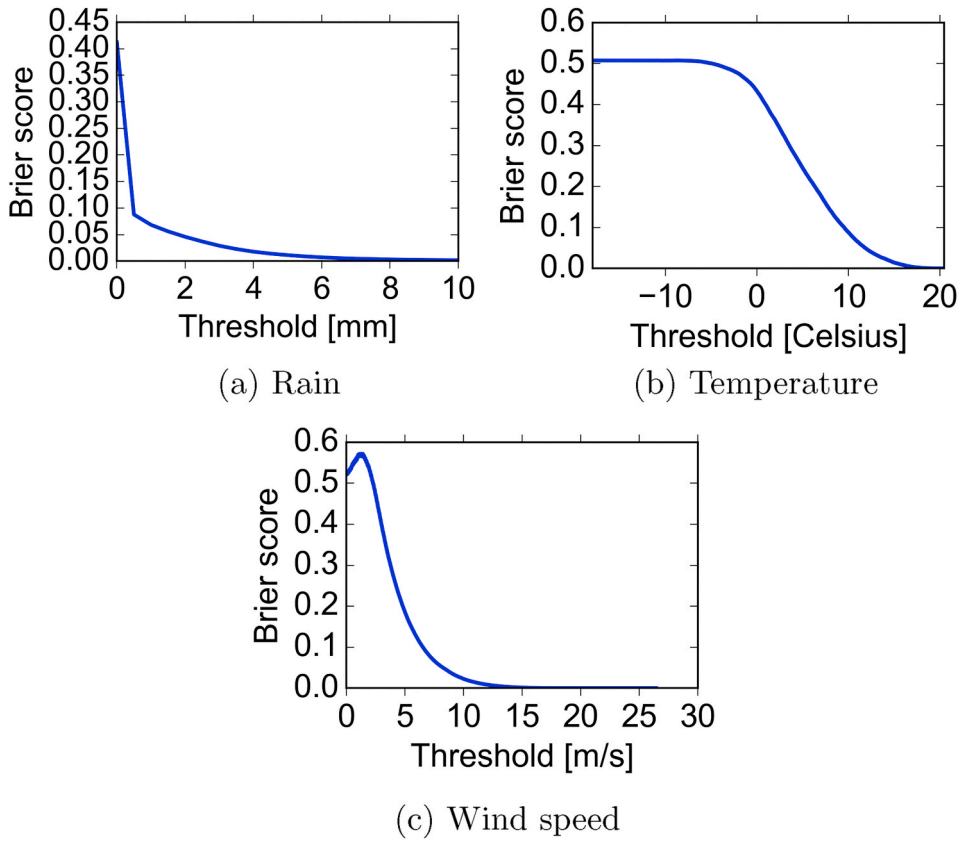


Fig. 5. BS of original weather forecast ensemble.

Table 1
DRPS of original weather forecast.

Variable	DRPS
Temperature [°C]	11.69
Wind speed [m/s]	2.57
Precipitation [mm/h]	0.42

Table 2
Variation space of ℓdX input variables.

Variable	Distribution/ Method	Variation space
ECMWF ensemble member	Discrete/replace	[member 1, …, member 50]
Source term	Discrete replace	[Katata et al. (2015); Terada et al. (2012); Mathieu et al. (2012); Saunier et al. (2013, 2016); Winiarek et al. (2012)]
Emission factor	Log-normal/ multiplication	[1/3, 3]
Source elevation [m]	Discrete/replace	[0–40, 40–160, 160–280, 280–400]
Emission delay [hours]	Truncated normal/ addition	$[-6, +6]$
Dry deposition velocity [m/s]	Uniform/replace	$[5.10^{-4}, 5.10^{-3}]$
Scavenging factor a[a] $[hs^{-1}mm^{-1}]$	Uniform/replace	$[10^{-7}, 10^{-4}]$
Scavenging exponent b[a]	Uniform/replace	[0.6,1]
Horizontal diffusion $[m^2s^{-1}]$	Uniform/replace	$[0, 1.5] \times 10^4$
Factor of vertical diffusion	Uniform/ multiplication	[1/3, 3]

[a] Scavenging coefficient: $\Lambda_r = ap_0^b$, where p_0 is the rain intensity.$[mm.h^{-1}]$

(called *emission factor* in Table 2) and on the emission release time (called *emission delay* in Table 2). The values used for these perturbations come from Girard et al. (2014). They are crude perturbations, in the sense that the uncertainty on the released time and quantity is dependent on the release event, and therefore should be time-dependent. Besides, to be representative of *a priori* uncertainties, it would be better to use source terms derived only from installation events, without incorporating information from the observations, if available in the future.

Similarly to Girard et al. (2014), we made use of ℓdX which is a Eulerian model for radionuclide dispersion described in Section 3.1. The simulation domain and grid resolution used by the dispersion model coincide on the horizontal with the discretization of the weather forecasts, that is, a grid cell size of 0.25° horizontal resolution and a domain that spans 40 × 40 grid boxes to cover a great part of Honshu island. As the vertical resolution of the NWP ensemble is very coarse, the meteorological fields are interpolated on a vertical grid used for dispersion by the ℓdX model. The center altitudes are 20 m, 100 m, 220 m, 340 m, 500 m, 700 m, 1000 m, 1500 m, 2200 m, 3000 m, 3850 m and 4650 m. The model computes every hour and in each grid cell the air activity concentration and deposition values for all radionuclides released, taking into account dry deposition, scavenging processes due to precipitation, radioactive decay and progeny. This yields three-dimensional fields (resp. two-dimensional for deposition) varying in time. From these variables, the ambient gamma dose rate can be derived. Simulation are carried out from March 12 to March 30, 2011, with three-weeks atmospheric releases.

As the grid resolution of the model is coarse, the subgrid effect detailed in Section 2.2.1 for meteorological fields also applies to dispersion results. In particular, close to the source, the grid cell size is

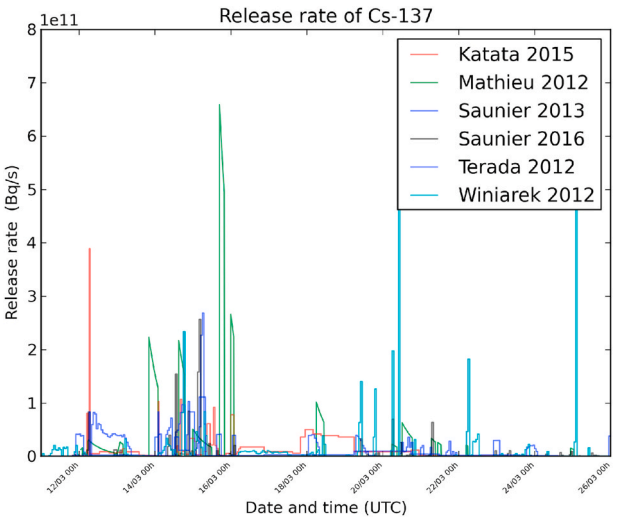


Fig. 6. Time series of release rate (in Bq/s) of the six source terms considered in this study from March 12 to March 26, 2011.

significantly larger than the size of the plume. Therefore, the error made by the Eulerian model in predicting air concentrations close to the source is very large, due to an unrealistic initial dilution of the plume. In general, it is recommended not to use the results of a Eulerian model at distances closer than about 5 cells from the source (Korsakissok and Mallet, 2010).

3.3. Description of the observation datasets

A review of the radiological observations available for the Fukushima case, with their advantages and drawbacks, was made by Mathieu et al. (2018). The following sections briefly describe the datasets used for our comparison purposes.

3.3.1. Air activity concentration

In this paper, we make use of the hourly concentrations of ^{137}Cs retrieved by Oura et al. (2015). The air activity concentration is given in Bq.m^{-3}. The data was obtained from the air quality automated monitoring network measuring suspended particulate matter (SPM) on filter tapes. They were measured too late to detect short-lived radionuclides, but they give crucial information on the temporal variation of ^{137}Cs concentration close to ground level, therefore indicating the passage of different plumes. Fig. 7 shows the spatial distribution of the 108 stations; unfortunately the repartition is not uniform, especially over the mostly contaminated area (in grey).

3.3.2. Deposition

Among the long-lived species, ^{137}Cs abundantly contaminated the Japanese territory and was measured during several airborne and ground measurement campaigns. The airborne measurements obtained from the various surveys conducted jointly by Japanese authorities (Ministry of Education, Culture, Sports, Science and Technologie; MEXT) and the U.S. Department of Energy (DOE) were calibrated against the ground readings to map the total ^{137}Cs deposit over Honshu island Sanada et al. (2014). The values were corrected from the radioactive decay to represent the deposition on April 1st 2011, after the end of all significant releases. While this kind of measurements gives a time-integrated information, there is a very fine spatial coverage (a few hundreds of meters). Here, the observations were averaged over a mesh of resolution 1 km, which amounts to more than 30,000 points. Fig. 8 shows the corresponding deposition map (values below 10 kBq.m^{-2} were not measured).

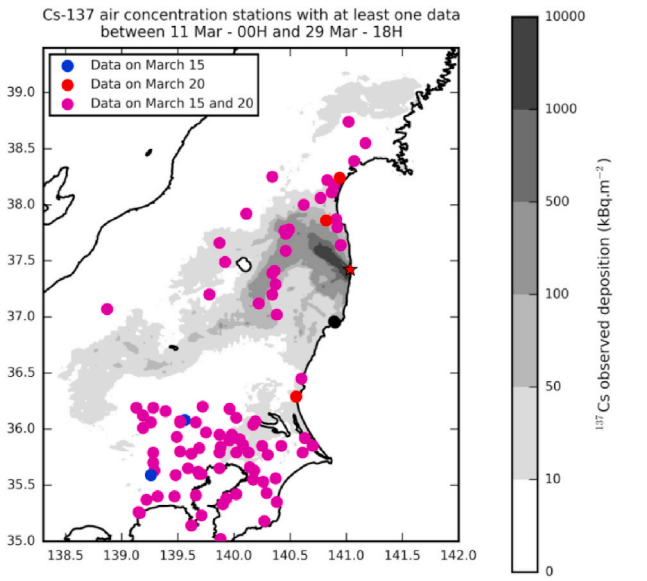


Fig. 7. Position of activity monitoring stations for ^{137}Cs and their availability periods. From Quérel et al. (2016). In grey, the area where ^{137}Cs deposition is higher than 10kBq.m^{-2}.

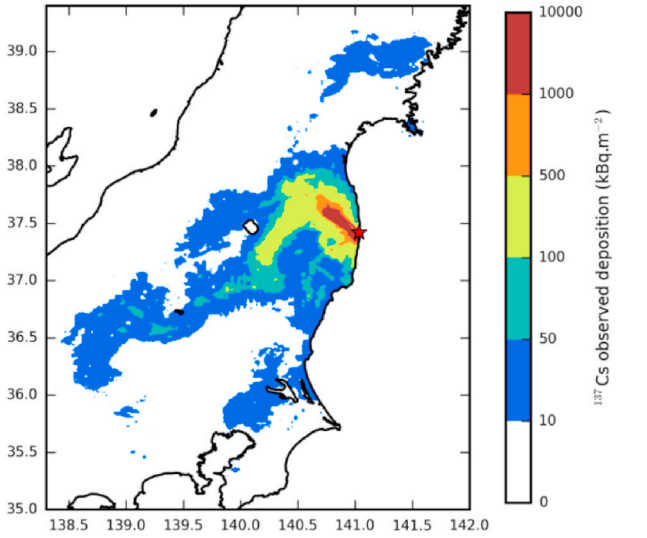


Fig. 8. Observed deposition of ^{137}Cs established for April 1st 2011, Quérel et al. (2016).

3.3.3. Gamma dose rate

The gamma dose rate measurements were provided by automated stations as early as March 11st, with a 10-min time step. The dose rate measurements are available from the IAEA database IAEA (2012). There are 88 stations spread over Japan, although the spatial coverage is heterogeneous (Fig. 9). The dose rate readings have the advantage of measuring the contribution from all radionuclides, including the short-lived species that could not be detected by the other monitoring systems. It is composed of two parts: the direct plume contribution ("cloud-shine") and the gamma-ray emitted by radionuclides deposited on the ground ("ground-shine"). Hence, the cloud-shine is usually responsible for peak values observed during the plume passage, whereas ground-shine corresponds to a lasting contribution that lingers after the plume has left the area, and decreases because of the radioactive decay.

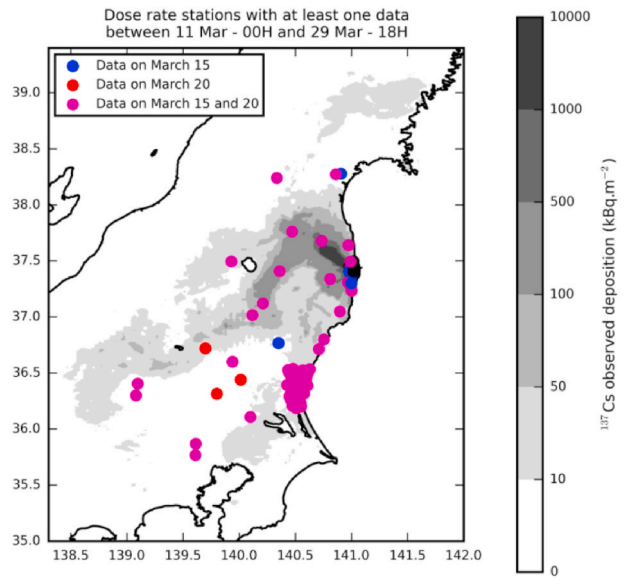


Fig. 9. Position of dose rate monitoring stations for ^{137}Cs and their availability periods. See Quérel et al. (2016). In grey, the area where ^{137}Cs deposition is higher than 10 kBq.m^{-2}.

3.4. Model-to-data comparison for Monte Carlo simulations

The 200 simulations from the MC sampling described in Section 3.2 are now compared to the observation data detailed in Section 3.3. For this purpose, the 3-D ^{137}Cs air concentration (resp. gamma dose rate) fields given by each member of the ensemble were interpolated on the corresponding station locations. As far as deposition was concerned, the 2-D ^{137}Cs deposition maps at the final time step of the simulation were compared to the observation map. The comparison of ensemble results to observation data was carried out using both deterministic indicators and probabilistic scores, as presented in section 2.2. The deterministic scores, such as the Root Mean Square Error (RMSE) and Figure of Merit in Time or Space (FMT, FMS), are used to evaluate the ability for one single simulation result to accurately reproduce the characteristics of the observations (magnitude, spatial and temporal evolution …). For each of the 200 ensemble simulations, these indicators can be computed. Then, the maximum, minimum and median values of this score over the 200 simulations are presented. This gives an idea of the performance of the different members of the ensemble by comparison to observations. It may indicate whether the contamination predicted by one particular selected member, or the ensemble median, would be appropriate for use as a tool for decision making. However, the main goal of the ensemble is to give an estimation of the results uncertainty, which is usually given by the ensemble's spread. Probabilistic scores, such as rank diagrams, BS and DRPS, along with time evolutions (similar to the meteograms shown in section 2.2) are therefore computed to evaluate the ensemble's quality. All these indicators are computed for the three different observation datasets described in section 3.3.

3.4.1. Deposition

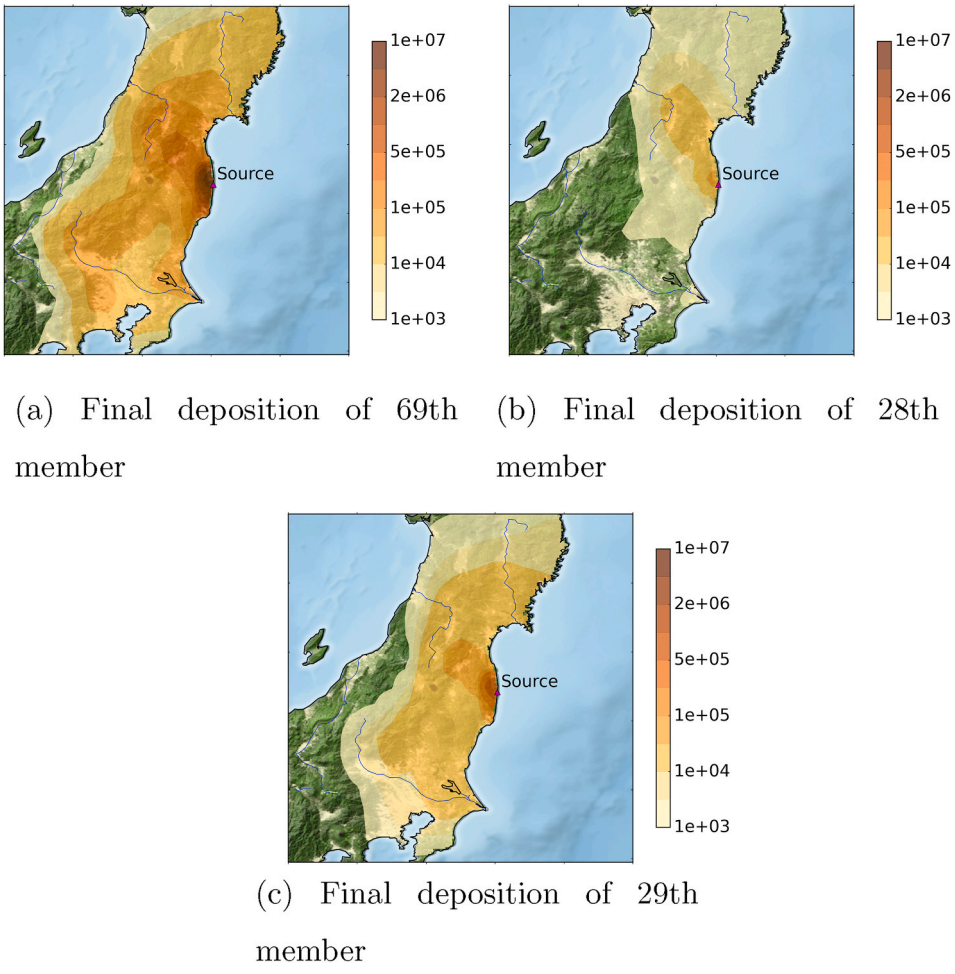
Pollutants can be removed from the air during the transport by various processes. Deposition concerns the transfer from airborne materials, gases or particles, to the soil, water and vegetation. When precipitation is absent, this process is called dry deposition. Deposition by precipitation such as rain or snow is called wet deposition or scavenging. During the Fukushima accident, significant deposition of ^{137}Cs was measured in Honshu island. A substantial proportion comes from wet deposition, especially during the two contamination episodes of March 14–16 and March 20–22 Korsakissok et al. (2013). In the MC ensemble, some simulations provide a deposition map comparable to observations, as illustrated by Fig. 10a. Other simulations show a lower deposition, either because the simulated plume was carried away over the ocean or since precipitation did not occur during the passage of the plume, as in Fig. 10b and c.

RMSE. The RMSE is a statistical indicator widely used for model-to-data comparison (see the definition in Section Appendix A.5). The computation of the RMSE on deposition was carried out with all MC members (Table 3). The minimum value over all ensemble members is given as an illustration of the best score that could be obtained if a proper member selection was applied, for instance by minimization of such a score. Of course, this step could only be achieved *a posteriori*, when a sufficient number of observations is available. The RMSE has the same unit as the variable of interest, and therefore it cannot be compared between different variables. Thus, a ratio between the "best estimate" (given by the minimum RMSE of the ensemble) and the observation mean is also computed in order to have a normalized indicator. In addition, the mean and median RMSE are also given as indicative of the average performance of the ensemble. Firstly, these indicators were calculated over all measurement points. The value obtained is much higher than the observational average, as shown in the first line of Table 3. Secondly, it was calculated over the cell average, i.e. model-to-data error is then the difference between the prediction in a cell and the observational mean in this cell. The ratio "Min/Obs mean" decreases from 3.24 to 1.42 because there are not subgrid effects in the second case. The RMSE values of the third line use only the measurement points whose distances from the source are higher than 80 km. The ratio is even better because the short-range error is eliminated. In the last line, both subgrid effects and short-range discrepancies are filtered out and the RMSE falls at 67% of the observational mean.

FMS. Figures of merit, like RMSE, are analytic parameters used for evaluating the prediction accuracy in relation to observations. FMS is calculated at a fixed threshold, and it is defined as the amount of overlap between the observational and predicted areas where values are greater than the threshold. It is given by the ratio between the intersection of the areas and their union (see Appendix Appendix A.3). The value 100% in FMS represents a perfect cover of prediction over observation. In Table 4, two thresholds are used. The threshold of 10 kBq/m^2 corresponds to the detection limit, and therefore the whole contaminated area displayed Fig. 8 is involved; the threshold of 37 kBq/m^2 is a Chernobyl reference level. When the threshold increases, the performance of the model decreases because there are less measurement points and a small shift in deposited direction may drastically change the FMS values. Following Chang and Hanna (2004), a forecast whose FMS value reaches about 50% is satisfactory. Therefore, the results shown in tab: fms deposition for the ensemble mean and median are acceptable according to that statistical indicator for the proposed thresholds.

Rank histogram. According to Fig. 11a, the ensemble shows a bias as it tends to underestimate the observations. The rank map rank map Fig. 11b indicates the overestimated (blue points) and underestimated (red points) areas. According to Mathieu et al. (2018), the long red branch, first north-west and then south-west, corresponding to the Abukuma Valley, was mainly contaminated during March 14–16. The presence of precipitation and/or fog, together with a concentrated plume, caused a strong deposition that most simulations failed to reproduce. In particular, fog and light rains are not well predicted by weather forecasts and these uncertainties may not be well represented by the meteorological ensemble. Besides, overestimation in the north-west (the blue area) and underestimation in the south-west (the little red area) can be due to the subgrid effect. The plume came into a valley and was wedged there but our ensemble can not reproduce this because of the coarse resolution.

BS and DRPS. As the model-to-data comparison on weather forecast, we plotted the BS following several thresholds and computed the DRPS of three radiological outputs. The BS tends towards zero when the threshold increases: with a sufficiently high threshold, all observations



(a) Final deposition of 69th member

(b) Final deposition of 28th member

(c) Final deposition of 29th member

Fig. 10. Illustration of the final deposition of ^{137}Cs, computed by three ensemble members, in Bq.m^{-2}: (a) Final deposition of 69th member, (b) Final deposition of 28th member and (c) Final deposition of 29th member.

Table 3
RMSE on the final deposition [kBq/m^2] .

	Min	Mean	Median	Obs mean	Min/Obs mean
RMSE all points	263.27	269.05	282.46	81.06	3.24
RMSE all cells (cell average)	67.18	92.36	91.5	47.44	1.42
RMSE points $>$ 80 *km*	16.69	22.11	21.63	22.15	0.75
RMSE cells $>$ 80 *km*	10.56	17.49	14.72	15.72	0.67

Table 4
FMS [%] for final deposition at all stations.

Threshold [kBq/m^2]	Max	Mean	Median
10	99.16	93.08	82.07
37	57.64	48.20	42.39

and simulations are below the threshold. This situation arises from the lack of events above the threshold and not from the effectiveness of the prediction ensemble. In some cases, such as deposition and gamma dose rate, Figs. 12 and 15, the BS curve takes on a hat shape which reaches a maximum. For deposition, the maximum BS value is a little above 0.3, which indicates a good ensemble performance for predicting threshold exceedance. The DRPS score of deposition computed from Fig. 12 is 6 Bq/m^2.

3.4.2. Gamma dose rate

Fig. 13 presents the time evolution of the gamma dose rate measured at two stations, Shirakawa and Kita Ibaraki, as well as the ℓdX predictions by MC method in these stations. The sharp increase generally corresponds to the plume arrival at the station location, or to the beginning of rain (in the case of an elevated plume not detected by the station but washed out by precipitation). After the peak, the plume departure generally drives to a decrease of measured dose rate. In the case of Shirakawa, the remaining gamma dose rate stays high because of wet deposition and slowly decreases due to radioactive decay as shown in Fig. 13a. When there is no precipitation and only dry deposition is involved, there are less radionuclides deposited, so gamma dose rate will decrease brutally after the plume departure, as in Fig. 13b. Shirakawa is a station located in the western area where the contamination by deposition is high (Fig. 3). The rain meteogram in Shirakawa cell (Fig. 3) shows that most meteorological members have a tendency to forecast late the beginning of the rain episode on March 15th. This episode is responsible for the wet deposition in this area and corresponds in timing to the abrupt increase of gamma dose rate observed on this station (Fig. 13a). Therefore, the delay in the simulation of the gamma dose rate peak observed Fig. 13a can be explained by the difficulty to accurately forecast rain. It should be noted that several members (dark blue lines) seem to predict peaks with a correct arrival time, but lower amplitude than the observed gamma dose rate peak. Besides, it is probable that both rain observations and simulations are not accurate enough for light rain events, and that light rain might actually have occurred before the detection by rain gauges. It is therefore possible that wash-out of the

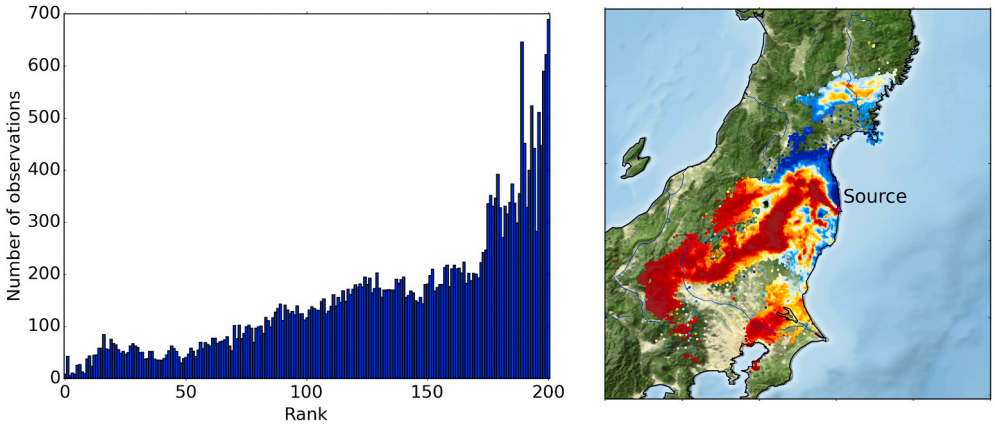


Fig. 11. Rank study of final deposition of ^{137}Cs over all points with a threshold at 5 kBq.m^{-3}. (a) Rank histogram. (b) rank map: the color of each point corresponds to the rank of the observation. The dark blue points are overestimated, the dark red are underestimated. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

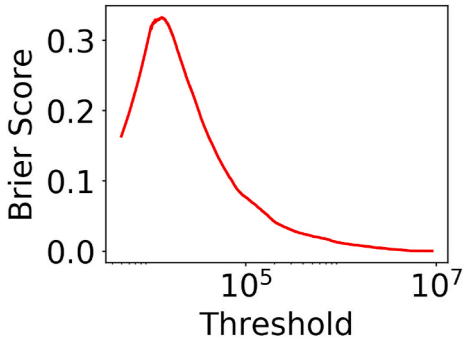


Fig. 12. BS of deposition [kBq/m^2].

elevated plume by light rain may have been sufficient to trigger an increase on gamma dose rate station earlier than expected, or a higher peak than modelled by our wash-out coefficients (Quérel et al., 2016).

Kita Ibaraki is a station on the south-west coast as shown in Fig. 13c and d. The time evolution of the gamma dose rate observations shows three consecutive peaks between March 15th and 17th, corresponding to different plume passages. The sharp decrease of gamma dose rate after each peak shows that, contrary to Shirakawa, there was no wet deposition during this episode. However, some simulation members feature a very large gamma dose rate due to deposition, leading to a significant overestimation of the observed values. Those may be members where either the rain episode, or the wind direction change (first blowing toward North-West, then South-West) on March 15th were not accurately forecast in time. Additionally, in the MC simulations, a perturbation was also applied to the release time on top of choosing various source terms. This illustrates how the uncertainties linked to release time and meteorology (especially rain) may have a very large impact on some stations.

RMSE. The RMSE of the gamma dose rate is shown in Table 5. The subgrid effect is of relatively lesser importance for gamma dose rate than for deposition. Indeed, the measurement stations of gamma dose rate can detect gamma radiation within a few kilometers, therefore smoothing over the spatial variability observed with deposition. For stations at more than 80 km from the source, the ensemble minimum, average, median and the observational mean are similar. The ratio "Min/Obs mean" decreases significantly when we remove the stations at less than 80 km from the source, where the numerical diffusion due to the grid resolution is too high.

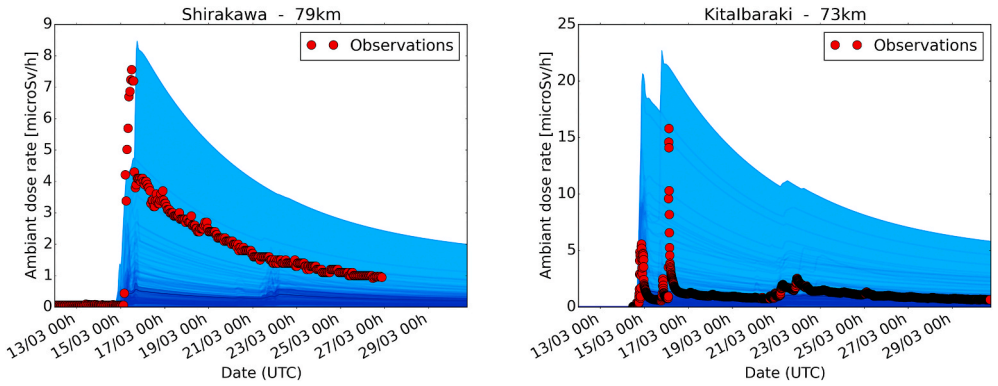
FMT. In the same way as FMS, the computation of FMT is based on a threshold. It represents the ratio of overlap between prediction and observation over time (see the definition in appendix Appendix A.4). Table 6 confirms that the ℓdX results over the area beyond 80 km are better than those over entire Japanese territory or closer to the source. The FMT value of the ensemble average is similar to the maximum FMT, and 58% is the result of most ensemble members. Following Korsakissok et al. (2013), this is a satisfactory value.

Rank histogram. The triangular shape of the rank histogram of gamma dose rates shows that the ensemble spreads out largely in comparison to the observations. A threshold of 0.1μSv h^{-1} was applied, to account for background noise. Contrary to the rank histogram of deposition, the rank histogram of gamma dose rate has low bars on the left, as shown Fig. 14. This shape suggests that the ensemble is over-dispersed by comparison to this variable. This was illustrated by the stations shown Fig. 13 and some reasons for this overestimation were already discussed. Each gamma dose rate station provides many measurements, but they mostly bring similar, correlated information. After the plume has left the area, the only difference between two consecutive measurement points comes from the radioactive decay of the deposited radionuclides. Therefore, in that case, most observations get similar ranks, as it is the case at Shirakawa station (see Fig. 13a). This issue creates peaks in the rank histogram.

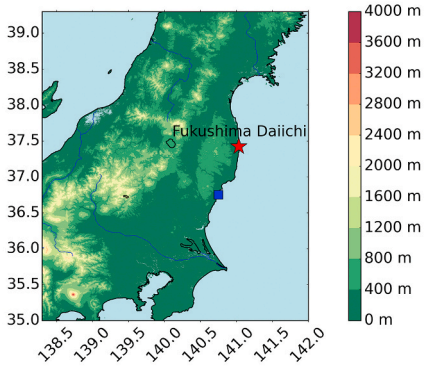
BS and DRPS. The maximum BS shown Fig. 15 is a little over 0.3, which again indicates a good ensemble performance for predicting threshold exceedance. The DRPS score of gamma dose rate is 0.012 μSv/h with a threshold of 0.1 μSv/h to take into account natural background radiation. By comparison, Korsakissok et al. (2018) compared several ensembles (constructed with the same meteorological ensemble and set of source terms, but different dispersion models) with Fukushima observations on March 15th–17th. The DRPS for this episode were globally lower, ranging from 0.003 to 0.011. The discretization of the range of variation, the way background noise was taken into account and the time period which differ from the present study prevent from drawing general conclusions from this comparison. However, the better ensembles' performance given in Korsakissok et al. (2018) is consistent with the fact that the meteorological ensemble used had a better resolution, spatially (0.2° with 36 vertical levels) and temporally (hourly values).

3.4.3. Air activity concentration

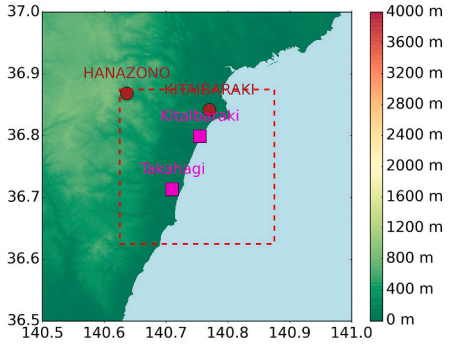
During the Fukushima accident, two main episodes of contamination over the Honshu island were observed and will be studied here: March 14–16 and March 20–22 Mathieu et al. (2018). During the first period, the observational amplitude was correctly represented by the ensemble but at some stations, a delay in time and a discrepancy in the duration was observed. The highest values of most stations are included in the



(a) Shirakawa (b) Kita Ibaraki



(c) Location of the cell containing Kita Ibaraki



(d) Zoom in on this cell

Fig. 13. Illustration for gamma dose rate temporal variations in Shirakawa (left), Kita Ibaraki (right), location of the cell containing Kita Ibaraki and zoom in on this cell.

Table 5
RMSE for gamma dose rates μSv/h.

	Min	Mean	Median	Obs mean	Min/Obs mean
All stations	1.85	2.86	2.19	0.68	2.72
Stations > 80 km	0.33	0.34	0.39	0.34	0.97
Stations < 80 km	4.43	6.97	5.3	2.4	1.84

Table 6
FMT [%] for gamma dose rate.

	Max	Mean	Median
All stations	41.1	33.09	28.22
Stations > 80 km	58.83	58.61	42.30
Stations < 80 km	35.74	23.40	21.17

ensemble envelop. Some members largely overestimate the observations. However, at 57 out of 108 stations, the plumes simulated by all members do not stay long enough to cover all observations, as illustrated by Fig. 16 for Yabuki station. All plumes modelled by the ensemble arrive too late at 22 stations. For the second episode, simulations tend to underestimate observations and a delay in time is observed for several stations. At 39 stations out of 108, the ensemble envelop does not reach the highest observed values. The plumes arrive late at 32 stations.

In Fig. 16, two measurement stations of air concentration are shown: Sukagawa and Yabuki. These stations are located in the same cell, about

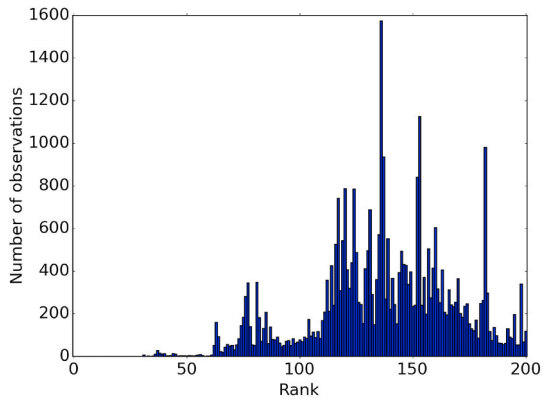


Fig. 14. Rank histogram of gamma dose rate over all measurement stations with a threshold of 0.1 μSv/h.

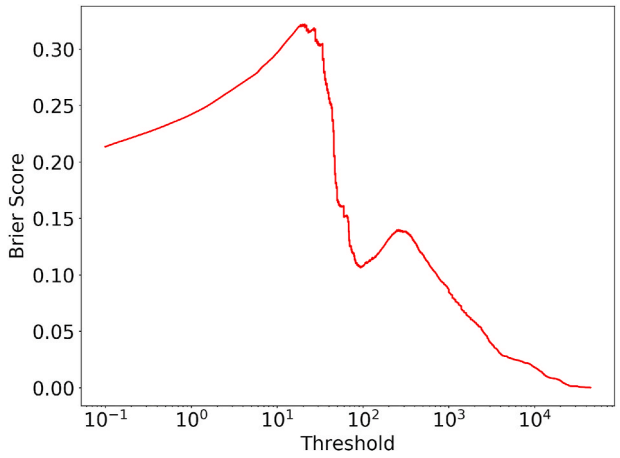


Fig. 15. BS of gamma dose rate [μSv/h].

70 km in the south-west of the Fukushima plant (Fig. 3). This figure is a good illustration of the subgrid effect. ℓdX computes the cell average of air concentrations. Hence, the simulations give the same result at these two stations. However, the mesh used is coarse, and the plume does not fill entirely the cell, as illustrated by Fig. A19 in Appendix Appendix A.2. During the first episode, Yabuki station detected several plumes whereas Sukagawa station observed only one plume from 03:00 to 21:00 on March 15th. The same phenomenon took place during the second episode. Therefore, stations inside the plume give measurements that differ from those given by stations outside the plume, whereas the simulations are identical. This can produce a significant model-to-data discrepancy.

3.4.3.1. RMSE. Table 7 presents for air concentrations: the best RMSE among all ensemble members, the RMSE of the ensemble average and that of the median. Overall, the "best" RMSE is about 3–4 times the observational average, which confirms that the ensemble members do not show a very good performance for air concentrations, compared to the other observation datasets.

The RMSE can vary a lot from one member to another (from 13.43 Bq m^{-3} and 116.06 Bq m^{-3}), which is a sign of high variability in the ensemble. There are very few members whose RMSE values are higher than 40 Bq m^{-3}. Note that a high RMSE does not necessary mean that an ensemble member is useless. It can still help in capturing some part of the uncertainties, and produce the forecasts at some locations and times.

3.4.3.2. Rank histogram. Fig. 17 presents a rank histogram for air concentrations. This histogram does not use all measurements. We selected stations whose distances from the source are greater than 80 km because ℓdX is a long-range model. The resolution used is about 25 km × 25 km and the model results are reliable only beyond about three to five cells from the source, to account for unrealistic numerical diffusion close to the source Korsakissok and Mallet (2010). We also put a threshold at 1 Bq m^3 in order to discriminate stations with a significant signal. This rank histogram is U-shaped with two external bars that are very high. The left bar indicates some overestimation of the ensemble, partly because of the coarse resolution that causes plume dilution, inducing unrealistic near-source concentrations at some stations. The right bar corresponds to some underestimation. Activity concentration is very sensitive to the arriving time of the plume. If there is a time delay in the simulations, the measured concentration at time t can be higher than all members values, leading to a high right bar. The concentration measurements are very difficult to reproduce because of the coarse (spatial and temporal) resolution, which leads to a rather bad rank histogram.

3.4.3.3. BS and DRPS. The DRPS of activity concentration is 2×10^{-3}Bq m^{-3} for a threshold of 1 Bq m^{-3}. The maximum BS is 0.6, which, again, represents a lesser ability to forecast threshold exceedance than for deposition or gamma dose rates (Fig. 18).

4. Synthesis and perspectives

Uncertainties carried by the weather forecasts influence very substantially the results of atmospheric dispersion models: air concentration activity, deposition and gamma dose rate. Among all meteorological variables, wind direction and speed, and precipitations are obviously two major factors that influence the plume trajectory and subsequent contamination patterns. The wind direction determines direction and spread of plumes, and precipitation has a strong influence on the deposited quantities. The model-to-data comparison to meteorological observations provided a first insight into the suitability of weather ensemble forecasts as a descriptor of uncertainties related to these boundary layer variables. In this study, the operational ECMWF-ENS forecast was retrieved and used for the three weeks of the Fukushima disaster (12–30 March 2011). This raised several issues. First, the horizontal and vertical resolution of this ensemble, as stored by ECMWF, is quite coarse, with a 0.25° horizontal grid resolution, only five pressure levels within the boundary layer, and a 3-h time step. This in turn raises the concern of subgrid variability which will create discrepancies with local variables at meteorological stations. Another issue was the fact that the ECMWF ensemble prediction system is tuned to be representative of large-scale errors. This spread is not sufficient to represent the uncertainty within the boundary layer height, as illustrated in this paper by the rank diagrams for meteorological variables. Finally, another challenge was to combine different ensemble forecasts from various initial times to span the three-week period of the accident. This was done by simply juxtaposing successive 24-h forecasts. This may be viewed as representative of what could be used operationally in case of an emergency, as successive forecasts would be used to infer of complement recommendations for the protection of the population, as the accidental sequence develops. However, this leaves clear room for future research in order to ensure a smoother transition between different ensembles. The ensemble performance shown in this study is therefore representative of the quality of an ensemble that could realistically be used for operational purposes.

The other question addressed in this study was then to what extend this ensemble may be used, despite its limitations, to evaluate the uncertainties in the dose evaluations made by atmospheric dispersion models. The study of Girard et al. (2014) ascertained that three uncertainty sources are meteorological forecasts, source term and some parameters used inside the model. The perturbation by MC was carried out over ten input variables of ℓdX. Among those, the weather forecasts and

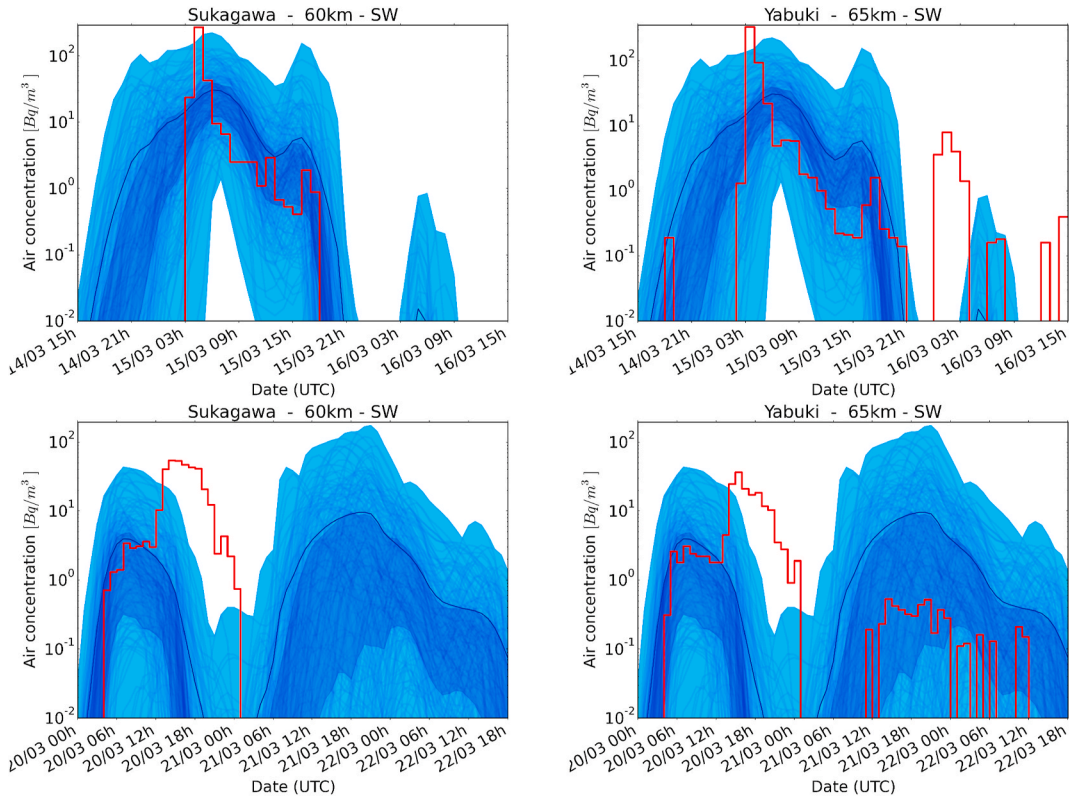


Fig. 16. Model-to-data comparison of air concentrations at two stations located in the same cell: Sukagawa (left) and Yabuki (right), during two episodes: 14–16 March 2011 (top) and 20–22 March 2011 (bottom). The red line is the observation. The simulations are represented by the light-blue lines. The dark blue line is the average of the simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 7
RMSEs on air activity concentrations of ^{137}Cs during entire March 2011, for the best RMSE among all ensemble members (Min), the RMSE of the ensemble average (Mean) and that of the median.

Thres.	Min	Mean	Median	Obs. mean	Min/Obs mean
0.1 Bq/ m^3	13.43	18.76	14.03	3.49	3.84
1 Bq/ m^3	15.18	21.21	15.86	4.38	3.47
10 Bq/ m^3	20.87	29.24	21.83	7.14	2.92

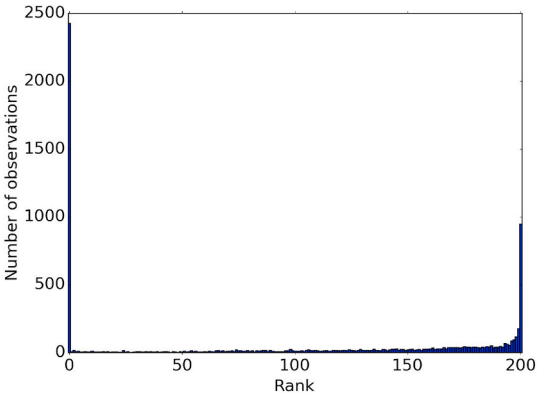


Fig. 17. Rank histogram of air concentrations in March 2011.

the source term are uniformly selected in discrete sets. Other variables follow different probability distributions. Two hundred perturbed data samples were generated, and as many dispersion simulations were

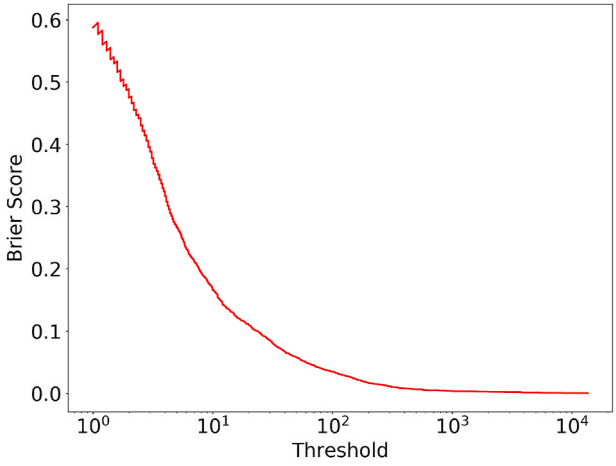


Fig. 18. BS of activity concentration [Bq/m^3].

computed. Observations of air concentrations, deposited quantities and gamma dose rates allowed us to evaluate this simulation ensemble with several statistical indicators, both deterministic (RMSE, FMS and FMT) and probabilistic (rank diagram, BS, DRPS).

The aim of this study was to implement a method of perturbation on inputs and analyze the uncertainty propagation in ℓdX. In the weather ensemble assessment, we understood that the coarse resolution of the forecasts was responsible for a significant part of the model-to-data discrepancy. With the MC perturbations, the simulations' output spread was much larger than what was observed in the meteorological ensemble. Although the ensemble did not prove sufficient to represent uncertainties for air activity concentrations, the results encompassed

reasonably well the deposition and gamma dose rate observations despite the resolution. In addition, the removal of the area within 80 km of the source confirmed the ℓdX effectiveness for long distances. One of the questions when dealing with ensemble simulations, especially in an emergency context, is the computational time. Therefore, the use of a coarse-resolution meteorological ensemble may be an acceptable choice, compared to a high-resolution ensemble which may improve the accuracy but would significantly increase the computational burden. To take into account the subgrid effect, suitable perturbations may be studied in the future to compensate for this uncertainty not taken into account in our ensemble.

A next step is to improve the perturbations on the key inputs so that the ensemble better fits the observations. The comparison should rely on probabilistic scores, like the BS or the rank histogram. Several statistical methods are available to carry out an ensemble calibration. Among them are Bayesian inference schemes, for instance, based on Monte Carlo Markov Chain (MCMC) algorithms. This approach makes use of observations to calibrate the model inputs and samples from the *a posteriori* probability distribution of each input variable. However, this approach requires a huge number of simulations, of the order of several millions. It is not reachable with current models. A future study should therefore focus on meta-models' construction. A meta-model is a mathematical approximation of some model but with a much lower computational time. The results of the meta-model must be similar to those of the model, so that they can replace the model in the MCMC algorithms. A more complex issue, raised in this study, is the redundancy in the observational data. When two stations are close, they may essentially bring the same information. This may also happen at one given station because of high correlation in time, such as gamma dose rate stations when they are measuring the radioactive decay of the surrounding deposition.

Credit author statement

Ngoc Bao Tran Le: simulation, writing, reviewing, editing.
Irène Korsakissok: methodology, writing, reviewing.
Vivien Mallet: writing, reviewing.
Raphaël Périllat: simulation, methodology.
Anne Mathieu: reviewing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A.1. Interpreting by using rank histogram

The shape analysis of the histograms allows to conclude on potential biases in the ensemble as well as over- or under-dispersions. In the rank histogram, the height of the bar i is the number of observations that are higher than exactly i ensemble members. If an ensemble is reliable, the observations are indistinguishable from any other ensemble member, and the rank histogram is flat. If an ensemble does not spread out enough, the external bars will be higher than the internal bars. If an ensemble spreads out too much, the observations' dispersion is smaller than the ensemble's spread, and the rank histogram is shaped like a triangle. In case the ensemble has a bias, e.g., overestimates (resp. underestimates) the variable, the histogram will show higher bars on the left (resp. on the right) than on the other side. In practice, the rank histograms are rarely flat because all uncertainties are not taken into account.

Appendix A.2. Illustration of subgrid effects

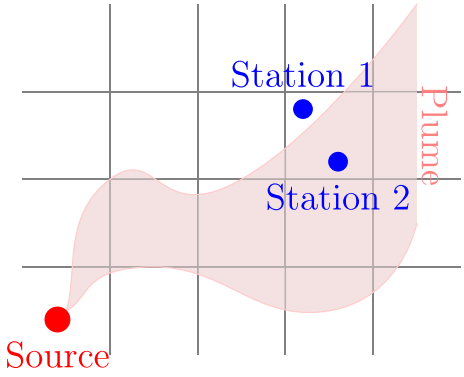


Fig. A.19. subgrid effect description.

Appendix A.3. FMS

The FMS formula is

$$FMS = \frac{A_m \cap A_p}{A_m \cup A_p} 100$$

where A_m is the observed area above the threshold and A_p is the predicted area above the same threshold.

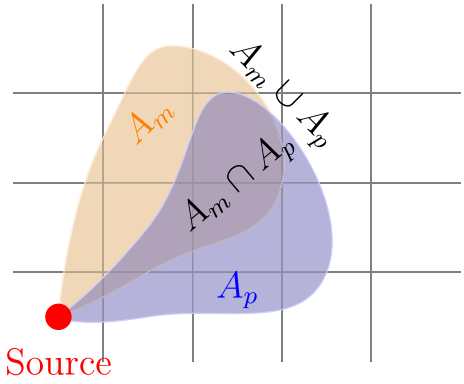


Fig. A.20. Illustration of FMS.

Appendix A.4. FMT

The FMT formula is the same as the FMS formula. Fig. A21 shows how to find the formula components. The intersection of A_m and A_p can be written as the minimum between them and the union is the maximum.

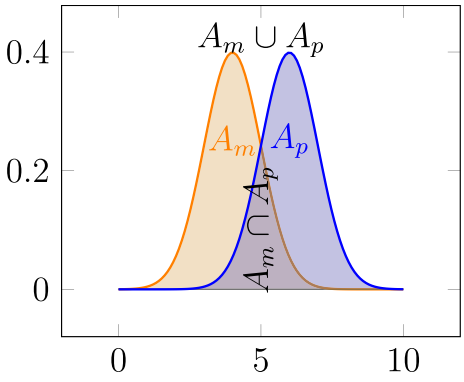


Fig. A.21. Illustration of FMT.

Appendix A.5. RMSE

Indicator RMSE is the standard deviation of model-to-data error, it aggregates the magnitudes of the difference between simulation and observation in space for various time into a single value. It allows to compare accuracy of different simulations. RMSE is always positive and the value 0 represents a perfect prediction. In practice, this value is never reached because of various reasons: model error, observational error, etc. Let $Y = (y_1, y_2, \ldots, y_n)$ the observed vector and $\widehat{Y} = (\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_n})$ the predicted vector of variable y. The RMSE formula reads

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n}}$$

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aeaoa.2021.100112.

References

Armand, P., Brocheton, F., Poulet, D., Vendel, F., Dubourg, V., Yalamas, T., 2014. Probabilistic safety analysis for urgent situations following the accidental release of a pollutant in the atmosphere. Atmos. Environ. 96, 1–10.

Arnold, D., Maurer, C., Wotawa, G., Draxler, R., Saito, K., Seibert, P., 2015. Influence of the meteorological input on the atmospheric transport modelling with flexpart of radionuclides from the fukushima daiichi nuclear accident. J. Environ. Radioact. 139, 212–225.

Bedwell, P., Wellings, J., Leadbetter, S., Tomas, J., Andronopoulos, S., Korsakissok, I., Périllat, R., Mathieu, A., Geertsema, G., Klein, H., Vries, H.D., Hamburger, T., Pázmándi, T., Rudas, C., Sogachev, A., Szántó, P., 2018. Guidelines Detailing the Range and Distribution of Atmospheric Dispersion Model Input Parameter Uncertainties. European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE Project.

Buizza, R., Leutbecher, M., Isaksen, L., 2008. Potential use of an ensemble of analyses in the ecmwf ensemble prediction system. Q. J. R. Meteorol. Soc. 134 (637), 2051–2066.

Chang, J., Hanna, S., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87, 167–196.

De Meutter, P., Camps, J., Delcloo, A., Deconninck, B., Termonia, P., 2016. On the capability to model the background and its uncertainty of ctbt-relevant radioxenon isotopes in europe by using ensemble dispersion modeling. J. Environ. Radioact. 164, 280–290.

Draxler, R., Arnold, D., Chino, M., Galmarini, S., Hort, M., Jones, A., Leadbetter, S., Malo, A., Maurer, C., Rolph, G., Saito, K., Servranckx, R., Shimbori, T., Solazzo, E., Wotawa, G., 2015. World meteorological organization's model simulations of the radionuclide dispersion and deposition from the fukushima daiichi nuclear power plant accident. J. Environ. Radioact. 139, 172–184.

French, S., 2015. Cynefin: uncertainty, small worlds and scenarios. J. Oper. Res. Soc. 66, 1635–1645.

French, S., Haywood, S., Oughton, D.H., Turcanu, C., 2020. Different types of uncertainty in nuclear emergency management. Radioprotection 55, S175–S180. https://doi.org/10.1051/radiopro/2020029.

Girard, S., Armand, P., Duchenne, C., Yalamas, T., 2020. Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations. Atmos. Environ. 224, 117313.

Girard, S., Korsakissok, I., Mallet, V., 2014. Screening Sensitivity Analysis of a Radionuclides Atmospheric Dispersion Model Applied to the Fukushima Disaster. Atmospheric Environment, vol. 95. Elsevier, pp. 490–500.

Girard, S., Mallet, V., Korsakissok, I., Mathieu, A., 2016. Emulation and sobol' sensitivity analysis of an atmospheric dispersion model applied to the fukushima nuclear accident. J. Geophys. Res.: Atmos. Am. Geophys. Union J. Geophys. Res. Atmos. Am. Geophys. Union 3484–3496.

Haiden, T., Janousek, M., Vitart, F., Ferranti, L., Prates, F., 2019. Evaluation of Ecmwf Forecasts, Including the 2019 Upgrade (853).

Hudson, D., Ebert, B., 2017. Ensemble verification metrics. In: ECMWF Annual Seminar.

IAEA, 2012. In: Fukushima Monitoring Data Base.

Kajino, M., Sekiyama, T., Igarashi, Y., Katata, G., Sawada, M., Adachi, K., Zaizen, Y., Tsuruta, H., Nakajima, T., 2019. Deposition and dispersion of radio-cesium released due to the fukushima nuclear accident: sensitivity to meteorological models and physical modules. J. Geophys. Res.: Atmospheres 124 (3), 1823–1845.

Katata, G., Chino, M., Kobayashi, T., Terada, H., Ota, M., Nagai, H., Kajino, M., Draxler, R., Hort, M.C., Malo, A., Torii, T., Sanada, Y., 2015. Detailed source term estimation of the atmospheric release for the fukushima daiichi nuclear power station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and oceanic dispersion model. Atmos. Chem. Phys. 15, 1029–1070.

Korsakissok, I., Andronopoulos, S., Astrup, P., Bedwell, P., Chevalier-Jabet, K., Vries, H. D., Geertsema, G., Gering, F., Hamburger, T., Klein, H., Leadbetter, S., Mathieu, A., Pazmandi, T., Périllat, R., Rudas, C., Sogachev, A., Szanto, P., Tomas, J., Twenhöfel, C., Wellings, J., 2019. Comparison of ensembles of atmospheric dispersion simulations: lessons learnt from the confidence project about uncertainty quantification. In: 19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Bruges, Belgium.

Korsakissok, I., Mallet, V., 2010. Subgrid-scale treatment for major point sources in an eulerian model: a sensitivity study on the european tracer experiment (etex) and chernobyl cases. J. Geophys. Res.: Atmospheres 115 (D3).

Korsakissok, I., Mathieu, A., Didier, D., 2013. Atmospheric dispersion and ground deposition induced by the fukushima nuclear power plant accident: a local-scale simulation and sensitivity study. Atmos. Environ. 70, 267–279.

Korsakissok, I., Périllat, R., Andronopoulos, S., Astrup, P., Bedwell, P., Berge, E., Quérel, A., Klein, H., Leadbetter, S., Saunier, O., Sogachev, A., Tomas, J., Ulimoen, M., 2018. Ensemble Calculation for a Past Accident Scenario: the Fukushima Case Study. European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE Project.

Korsakissok, I., Périllat, R., Andronopoulos, S., Bedwell, P., Berge, E., Charnock, T., Geertsema, G., Gering, F., Hamburger, T., Klein, H., Leadbetter, S., Lind, O.C., Pázmándi, T., Rudas, C., Salbu, B., Sogachev, A., Syed, N., Tomas, J.M., Ulimoen, M., de Vries, H., Wellings, J., 2020. Uncertainty propagation in atmospheric dispersion models for radiological emergencies in the pre- and early release phase: summary of case studies. Radioprotection S57–S68.

Leadbetter, S., Andronopoulos, S., Bedwell, P., Chevalier-Jabet, K., Geertsema, G., Gering, F., Hamburger, T., Jones, A., Klein, H., Korsakissok, I., Mathieu, A., Pázmándi, T., Périllat, R., Rudas, C., Sogachev, A., Szántó, P., Tomas, J., Twenhöfel, C., de Vries, H., Wellings, J., 2020. Ranking uncertainties in atmospheric dispersion modelling following the accidental release of radioactive material. Radioprotection 55, S51–S55.

Leadbetter, S., Hort, M., Jones, A., Webster, H., Draxler, R., 2015. Sensitivity of the modelled deposition of caesium-137 from the fukushima dai-ichi nuclear power plant to the wet deposition parameterisation in name. J. Environ. Radioact. 139, 200–211.

Leutbecher, M., Lang, S.T.K., 2014. On the reliability of ensemble variance in subspaces defined by singular vectors. Q. J. R. Meteorol. Soc. 140 (682), 1453–1466.

Louis, J., 1979. A parametric model of vertical eddy fluxes in the atmosphere. Boundary-Layer Meteorol. 17.

Mallet, V., Quélo, D., Sportisse, B., de Biasi, M.A., Debry, E., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., Foudhi, H., 2007. Technical note: the air quality modeling system polyphemus. Atmos. Chem. Phys. 7 (20), 5479–5487.

Mathieu, A., Kajino, M., Korsakissok, I., Périllat, R., Quélo, D., Quérel, A., Saunier, O., Sekiyama, T.T., Igarashi, Y., Didier, D., 2018. Fukushima daiichi–derived radionuclides in the atmosphere, transport and deposition in Japan: a review. Appl. Geochem. 91, 122–139.

Mathieu, A., Korsakissok, I., Leadbetter, S., Andronopoulos, S., Bedwell, P., Wellings, J., Geertsema, G., Jones, A., Tomas, J., Vries, H.D., Klein, H., Périllat, R., Chevalier-Jabet, K., Stéphanie, F., Fougerolle, S., Créach, V., Cogez, E., Hamburger, T., Pazmandi, T., Rudas, C., Sogachev, A., Szanto, P., 2017. D9.1 -guidelines Ranking Uncertainties for Atmospheric Dispersion. Tech. rep., European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE Project, Report D9.1 125.

Mathieu, A., Korsakissok, I., Quélo, D., Groëll, J., Tombette, M., Didier, D., Quentric, E., Saunier, O., Benoit, J.-P., Isnard, O., 2012. Atmospheric dispersion and deposition of ra-dionuclides from the fukushima daiichi nuclear power plant accident. Elements 8, 195–200.

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. Q. J. R. Meteorol. Soc. 143 (702), 408–422.

Oura, Y., Ebihara, M., Tsuruta, H., Nakajima, T., Ohara, T., Ishimoto, M., Sawahata, H., Nitta, W., 2015. A database of hourly atmospheric concentrations of radiocesium (134cs and 137cs) in suspended particulate matter collected in march 2011 at 99 air pollution monitoring stations in eastern Japan. J. Nucl. Radiochem. Sci. 15, 1–12.

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A., 2009. Stochastic Parametrization and Model Uncertainty (598), 42.

Périllat, R., Korsakissok, I., Mallet, V., Mathieu, A., Sekiyama, T., Kajino, M., Adachi, K., Igarashi, Y., Maki, T., Didier, D., 2016. Using Meteorological Ensembles for Atmospheric Dispersion Modelling of the Fukushima Nuclear Accident. 17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Budapest, Hungary.

Quérel, A., Quélo, D., Roustan, Y., Mathieu, A., Kajino, M., Sekiyama, T., Adachi, K., Didier, D., Igarashi, Y., Maki, T., 2016. Fukushima: lessons learned on wet deposition from a combined analysis of radiation dose rate and volume activity measurements of cesium-137. In: Goldschmidt Conferences. Yokohama. Japon.

Quérel, A., Roustan, Y., Quélo, D., Benoit, J., 2015. Hints to discriminate the choice of wet deposition models applied to an accidental radioactive release. Int. J. Environ. Pollut. 58 (4), 268–279.

Raynaud, L., Bouttier, F., 2017. The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. Q. J. R. Meteorol. Soc. 143 (709), 3037–3047.

Sanada, Y., Sugita, T., Nishizawa, Y., Kondo, A., Torii, T., 2014. The aerial radiation monitoring in Japan after the fukushima daiichi nuclear power plant accident. Prog. Nucl. Sci. Technol. 4, 76–80.

Saunier, O., Mathieu, A., Didier, D., Tombette, M., Quélo, D., Winiarek, V., Bocquet, M., 2013. An inverse modeling method to assess the source term of the fukushima nuclear power plant accident using gamma dose rate observations. Atmos. Chem. Phys. 13, 11403–11421.

Saunier, O., Mathieu, A., Sekiyama, T., Kajino, M., Adachi, K., Bocquet, M., Maki, T., Higarashi Didier, D., 2016. A new perspective on the fukushima releases brought by newly available 137cs air concentration observations and reliable meteorological fields. In: 17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. The publisher, pp. 75–174.

Sørensen, J., Amstrup, B., Feddersen, H., Bartnicki, J., Klein, H., Simonsen, M., Lauritzen, B., Hoe, S., Israelson, C., Lindgren, J., 2016. Fukushima Accident: Uncertainty of Atmospheric Dispersion Modelling (Fauna). Tech. rep., Nordic Nuclear Sakety Research (NKS), Rapport NKS-360.

Terada, H., Katata, G., Chino, M., Nagai, H., 2012. Atmospheric discharge and dispersion of radionuclides during the fukushima dai-ichi nuclear power plant accident. part ii: verification of the source term and analysis of regional-scale atmospheric dispersion. J. Environ. Radioact. 112, 141–154.

Troen, I., Mahrt, L., 1986. A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. Boundary-Layer Meteorol. 37.

Wellings, J., Bedwell, P., Leadbetter, S., Tomas, J., Andronopoulos, S., Korsakissok, I., Périllat, R., Mathieu, A., Geertsema, G., Vries, H.D., Klein, H., Hamburger, T., Gering, F., Pázmándi, T., Szántá, P., Rudas, C., Sogachev, A., Davis, N., Twenhöfel, C., 2018. Guidelines for ranking uncertainties in atmospheric dispersion. In: D9.1 Guidelines Ranking Uncertainties for Atmospheric Dispersion. European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE Project.

Winiarek, V., Bocquet, M., Saunier, O., Mathieu, A., 2012. Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: application to the reconstruction of the cesium-137 and iodine-131 source terms from the fukushima daiichi power plant. J. Geophys. Res.: Atmospheres 117 (D5).