Combining short-range dispersion simulations with fine-scale meteorological ensembles: probabilistic indicators and evaluation during a ^{85}Kr field campaign

Youness El-Ouartassy[1,2], Irène Korsakissok[2], Matthieu Plu[1], Olivier Connan[3], Laurent Descamps[1], and Laure Raynaud[1]

[1]CNRM, University of Toulouse, Météo-France, CNRS, 31057, Toulouse, France.
[2]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-SANTE/SESUC/BMCA, F-92260, Fontenay-aux-Roses, France.
[3]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SRTE/LRC, F-50130, Cherbourg-En-Cotentin, France.

Correspondence: Youness El-Ouartassy (Youness.el-ouartassy@meteo.fr)

Abstract. Numerical models of atmospheric dispersion are used for predicting the health and environmental consequences of nuclear accidents, in order to anticipate countermeasures necessary to protect the populations. However, these simulations suffer from significant uncertainties, arising in particular from input data: weather conditions and source term. Meteorological ensembles are already used operationally to characterize uncertainties in weather predictions. Combined with dispersion models, these ensembles produce different scenarios of radionuclides dispersion, called "members", representative of the variety of possible forecasts. In this study, the fine-scale operational weather ensemble AROME-EPS from Météo-France is coupled to the Gaussian puff model pX developed by IRSN. The source term data is provided at 10-minutes resolution by Orano La Hague reprocessing plant (RP) that regularly discharges ^{85}Kr during the spent nuclear fuel reprocessing process. In addition, a continuous measurement campaign of ^{85}Kr air concentration was recently conducted by the Laboratory of Radioecology in Cherbourg (LRC) of IRSN, within 20 km of the RP in the North-Cotentin peninsula, and is used for model evaluation.

This paper presents a probabilistic approach to study the meteorological uncertainties in dispersion simulations at local and medium distances (2-20 km). First, the quality of AROME-EPS forecasts is confirmed by comparison with observations from both Météo-France and IRSN. Then, the probabilistic performance of the atmospheric dispersion simulations was evaluated by comparison to the ^{85}Kr measurements carried out during a 2-months period, using two probabilistic scores: Relative Operating Characteristic (ROC) curves and Peirce Skill Score (PSS). The sensitivity of dispersion results to the method used to calculate atmospheric stability and associated Gaussian dispersion standard deviations is also discussed.

A desirable feature for a model used in emergency response is the ability to correctly predict exceedance of a given value (for instance, a dose guide level). When using an ensemble of simulations, the "decision threshold" is the number of members predicting an event above which this event should be considered probable. In the case of the 16-members dispersion ensemble used here, the optimal decision threshold was found to be 3 members, above which the ensemble better predicts the observed peaks than the deterministic simulation. These results highlight the added value of ensemble forecasts compared to a single deterministic one, and their potential interest in the decision process during crisis situations.

1 Introduction

Accidental releases of radioactive pollutants into the atmosphere can have serious impact on human health and environment
(Aliyu et al., 2015; Nie et al., 2021). The dispersion of radionuclides released into the atmosphere depends on the physico-
chemical properties of the released substances, the emission parameters (e.g. source elevation, timing and duration of the
release) and meteorological conditions at the accident site (e.g. wind speed and direction) (Girard et al., 2014). In order to fore-
cast the dispersion of radionuclides in the early phase of nuclear accidents and to support decisions and warnings, atmospheric
dispersion models (ADM) are commonly used to predict the transport of radioactive pollutants through the atmosphere as well
as the quantities of radioactive material deposited on the ground (Korsakissok et al., 2013). This information is essential for
decision makers in order to anticipate the countermeasures necessary to protect the population against contamination.

1.1 Uncertainties and ensemble simulations

The outputs from ADM simulations suffer from significant uncertainties that limit the confidence in them when they are used
in an operational context (Korsakissok et al., 2020; Leadbetter et al., 2020). The three main sources of these uncertainties
have been discussed by Rao (2005) and by Mallet and Sportisse (2008). The first one is related to the source term, which is
an essential input data. For prognosis of potential releases, it may be defined from a priori assumptions (pre-defined source
term), modelling of physical processes at stake within the reactor, along with knowledge of the damaged installation status. In
case of an ongoing or past release, when observations are available in the environment, the source term can be reconstructed
by inverse methods. For this purpose, IRSN (French Institute for Radiation Protection and Nuclear Safety) has developed
inverse modelling methods, which are mathematics-based methods aiming to minimize the difference between ADM outputs
and in-situ measurements (Saunier et al., 2013, 2020).

The second main source of uncertainty is related to the meteorological forecasts that are given as input to ADM. Weather
information used for dispersion prediction is, frequently, provided by numerical weather predictions (NWP) as 3-D or 4-D
physical fields. To take into account the meteorological uncertainties on dispersion simulations, two methods have commonly
been used. The first one is by adding random perturbations to weather inputs (Girard et al., 2014, 2020). The second one is by
using meteorological ensembles (Straume et al., 1998).

Some studies have used operational ensemble prediction systems (EPS) as input for dispersion models in the case of the
Fukushima accident (Sørensen et al., 2016; Kajino et al., 2019; Le et al., 2021), and others for hypothetical nuclear accident
scenarios (Sørensen et al., 2019, 2020; Korsakissok et al., 2020; Leadbetter et al., 2022). These studies include either meteoro-
logical uncertainties only, or sometimes both meteorological and source term uncertainties. All these studies were carried out
at long distance and the ensembles used to represent weather uncertainties had coarse spatial and temporal resolution, except
Leadbetter et al. (2022) who used also fine-scale weather ensembles with a horizontal resolution of about 2.5×2.5 km and
70 vertical levels. For example, De Meutter et al. (2016) studied the use of meteorological ensembles at hemispheric scale
to predict radioxenon peaks coming from radiopharmaceutical facilities, and De Meutter and Delcloo (2022) at continental
scale for the same application, using ECMWF ERA5 ensemble at a horizontal resolution of 63 km. Le et al. (2021) investi-

gated the dispersion of radionuclides at Japan scale in the case of the Fukushima disaster, using the operational ECMWF-ENS (Leutbecher and Lang, 2014) with a spatial resolution of approximately 25×25 km and 3 h time steps, using several source terms from literature. In Le et al. (2021) and De Meutter and Delcloo (2022), an evaluation of the dispersion ensembles was performed by comparison to radiological observations in the environment, and the results illustrate the added value of the use of weather ensembles for dispersion simulations. In a hypothetical case study, Leadbetter et al. (2022) explored the uncertainties coming only from weather conditions at synoptic scale, by using the operational Met Office's EPS named MOGREPS-G (Tennant and Beare, 2014) with a spatial resolution of approximately 20×20 km and 3-h time steps. Although this approach allowed to demonstrate the ability of meteorological ensembles to perform atmospheric dispersion results more skillfully than results produced with deterministic meteorology, it did not evaluate the performance of the ensemble dispersion simulations in the case of a realistic release. While most applications of meteorological ensembles cited above were focused on hemispheric or continental scale, the impact of meteorological uncertainty on dispersion forecasts at local scale (in the range of 2 to 20 km) has received less attention. With the development of kilometer-scale EPS (Bouttier et al., 2012), the feasibility and interest for such studies are rising. High-resolution meteorological ensembles were used in the case of a fictious nuclear release (Sørensen et al., 2017, 2020; Korsakissok et al., 2020), but no comparison to observations was made. The realistic performance of ADM outputs can be assessed only by using well-known real source term combined with reliable tracer measurements appropriate for the studied scale, as discussed in Section 1.2.

The third source of uncertainty arises from approximations for resolving atmospheric processes in the ADM, such as, for instance, turbulent diffusion and deposition (Leadbetter et al., 2015; Girard et al., 2016). A possible approach to include these model-related uncertainties is to use a multi-model ensemble. This approach was extensively investigated in Galmarini et al. (2004b, a) by using a set of different ADM to construct an ensemble of simulations, either with identical or different input data, to represent the modelling uncertainties and the results shown that the ensemble simulations allows to reduce the uncertainty related to the deterministic simulation. This approach has been used for various applications, including the Fukushima accident (Draxler et al., 2015; Sato et al., 2018). However, this multi-model approach differs from the more systematic method based on meteorological ensembles, in the sense that the latter are built for each member to have the same probability. In this paper, we use a single ADM, but the influence of model-related variables such as atmospheric turbulent parameters is discussed.

1.2 Dispersion datasets at local scale

There is a large panel of atmospheric dispersion tracer experiments for model validation at local scale (Olesen, 1998), both for rural and urban areas. However, most of the experiments were conducted within a few kilometers of the source. There is a lack of tracer measurement experiments studies ranging from the short to medium distances (2 km-20 km). At such scale, the ^{85}Kr can be a good tracer since it is an inert gas with a long half-life ($\tau_{1/2}$=10.7 years) and its radioactive decay negligible at these distances. The main sources of the ^{85}Kr in the atmosphere are reprocessing plants of spent nuclear fuel, from which the ^{85}Kr release can be known with a good accuracy (described in section 2.2).

As an example, the work conducted in the Laboratory of Radioecology in Cherbourg (LRC) of IRSN, presented by Connan et al. (2013), is one of the rare studies that explored the dispersion of radionuclides at distances between 5 km and 50 km. In this

previous paper, continuous ^{85}Kr measurements at 1 minute time period were carried out at three stations and combined with well-known discharge data which were provided by the nuclear fuel reprocessing plant of Orano La Hague (later called RP), located in the North-Cotentin peninsula (north-western France). These data were then used to perform dispersion simulations, but, with only three stations, the dataset was not large enough to capture the spatial spread of released radioactive material. From these previous studies arose the need for a campaign using more observation stations, spatially representative of the area, along with a longer time-period to make the conclusions more statistically robust. These previous studies also showed that the assumption of homogeneous meteorological data, using a single meteorological observation as input, was responsible for a large part of simulation errors (Korsakissok et al., 2016), thus highlighting the need to account for meteorological uncertainties.

1.3 Objectives of the paper

The main purpose of the present article is to investigate the impact of the meteorological uncertainties on local-scale dispersion. The operational high-resolution meteorological ensembles AROME-EPS (Bouttier et al., 2012) and AROME deterministic NWP (Seity et al., 2011) of Météo-France are used as input of the IRSN short-range Gaussian puff model pX (Soulhac and Didier, 2008; Mathieu et al., 2012; Korsakissok et al., 2013) around RP facility at local scales (less than 20 km). In this area, there is a dense weather observation network (from both IRSN and Météo-France) that has been used to validate AROME-EPS ensembles before combining them with the dispersion model. Measurements of ^{85}Kr air concentration at eight fixed points located at various distances, from 2 km to 20 km, and at various orientations from RP facility were carried out by IRSN in the framework of the DISKRYNOC project (DISpersion of KRYpton in the NOrth-Cotentin). This dataset is presented for the first time in this paper, and is used to evaluate the probabilistic performance of ensemble dispersion simulations. Thus, the originality of this work can be summarized in three points: (i) the use of a unique and original dataset of continuous data of ^{85}Kr air concentration measurements (every 1 minute or 10 minutes) over a relatively long period, (ii) the evaluation of an ensemble of dispersion simulations using a fine-scale meteorological ensemble with in-situ observations, (iii) an innovative method developed to assess the probabilistic performance of the dispersion ensembles.

The outline of the article is as follows: in Section 2, the source term, the observations and the models used in the study are described. Section 3 presents the verification of AROME-EPS against wind measurements, and then in Section 4 the ensemble dispersion simulations are presented and discussed. Conclusion and perspectives are provided in Section 5.

2 Case study, data and models

2.1 Case study

The present study focuses on the dispersion of ^{85}Kr at short and medium distances (less than 20 km), in the North-Cotentin peninsula located in the North-West of France territory (Fig. 1). This geographical area is where the nuclear fuel reprocessing plant of Orano La Hague (later called RP) is located (Fig. 1). ^{85}Kr is a β^- and γ emitting radioactive noble gas that is naturally present in the environment, but mainly released into the atmosphere during the reprocessing of spent nuclear fuel. Background
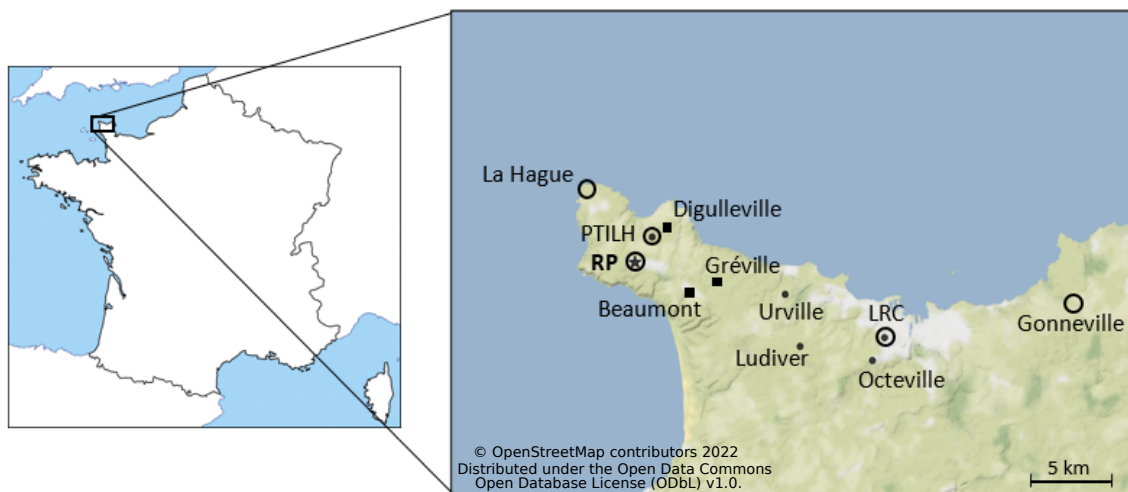
Figure 1. Location of North-Cotentin peninsula (left panel) and map of the monitoring sites (right panel). The dots and squares indicate the locations of the ^{85}Kr measurement stations carried out by IRSN and RP, respectively, as part of the DISKRYNOC campaign. The RP facility location is marked with a star. The circles indicate the locations of the 3D-wind measurement sites (from IRSN or Météo-France).

levels of ^{85}Kr in the atmosphere, excluding an industrial plume, are currently below 2 Bq.m^{-3} (Bollhöfer et al., 2019). Within 2 kilometers of the La Hague RP, air activity concentrations of ^{85}Kr can reach 100 000 Bq.m^{-3} (Connan et al., 2014). At distances of the order of 20 km, the maximum measurable activity concentrations are generally less than 10 000 Bq.m^{-3} and beyond a few tens of km of RP, the activities in ^{85}Kr are too low to be measurable in real time (Connan et al., 2013).

125   The potential interest of the La Hague area is that the release rate of ^{85}Kr emitted by the RP into the atmosphere is known with a good accuracy. In addition, there is a sufficient density of meteorological measurements combined with ^{85}Kr radiological air concentration measurements. Meteorological measurements are carried out by Météo-France on a regular basis. IRSN's LRC laboratory regularly performs meteorological and radiological measurements in the framework of measurement campaigns. Additional meteorological and air concentration measurements are carried out by Orano for the environmental monitoring of

130   the RP. For these reasons, the La Hague experimental site is an ideal environment for the study and validation of atmospheric dispersion simulations.

    Past validation studies conducted in this framework have shown that dispersion simulation results are quite sensitive to the meteorological data used as input (Connan et al., 2013). The North-Cotentin peninsula of La Hague is a rocky area of approximately 15 km located at 190 m a.s.l above cliffs, surrounded by the sea less than 5 km in most directions (Fig. 1).

135   Such a complex terrain leads to spatially heterogeneous wind fields that may be difficult to accurately forecast. Therefore, this case study should provide good insights to examine the influence of meteorological uncertainties on atmospheric dispersion simulations.
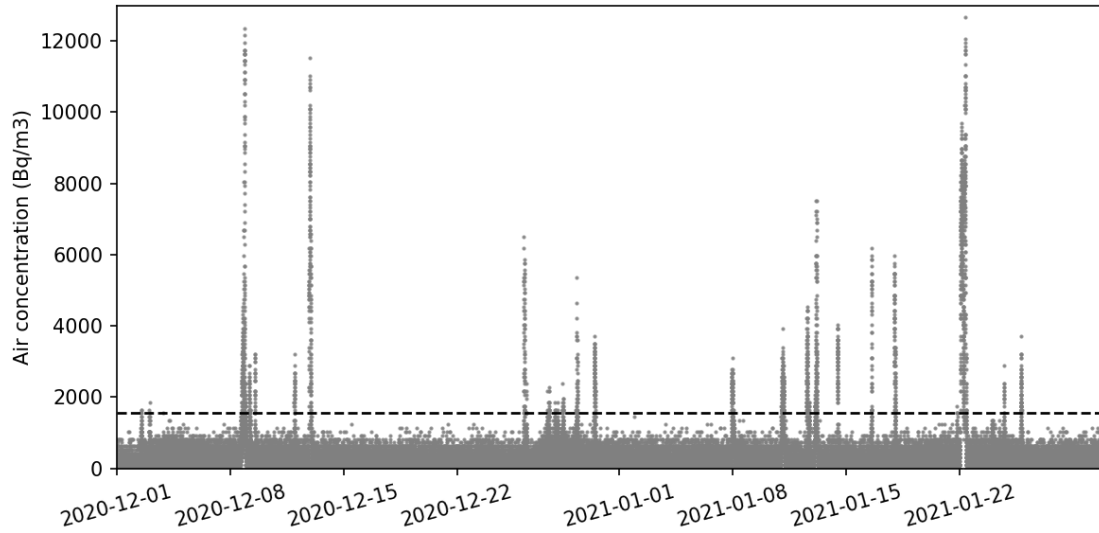
Figure 2. ^{85}Kr air concentration measurements carried out at LRC (Figure 1) from 01 Dec. 2020 to 31 Jan. 2021. The horizontal dashed line shows the air concentration threshold (1545 $Bq.m^{-3}$) above which peaks are considered.

2.2 Source term of ^{85}Kr

The La Hague RP has two production units called UP2-800 (1.87941° W, 49.67705° N) and UP3 (1.87606° W, 49.67705° N). Each of the two units has a stack for the discharge of ^{85}Kr with a height of 100 m and the two stacks are 200 m apart (Leroy et al., 2010).

During the reprocessing process of spent nuclear fuel, ^{85}Kr is intermittently released into the atmosphere from stacks of the plant during 30 to 45 min, separated by approximately 10-minutes intervals without releases. Depending on the industrial activity, long periods (a few hours to a few days or weeks) without releases are frequent. Both plants can operate separately and the release can come from one or both stacks. ^{85}Kr release fluxes (measured at a frequency of 10 min) were provided by Orano RP for the study period. The radioactive concentration of the released ^{85}Kr depends on the burn-up of the reprocessed spent fuel and the processing rate of the plant (Connan et al., 2013).

The activity in ^{85}Kr released from the factory by the stacks (confidential data) is known with a time step of 10 minutes and an uncertainty of measurement of the order of 10% in period of release (two channel of measurements for each stack). The discharge being intermittent, this 10-min time step ensures a precision indispensable for atmospheric dispersion studies. From 2019 to 2021, annual releases of the ^{85}Kr varied from 294 to 379 PBq/year (Orano, 2021).

In this paper, the atmospheric dispersion of ^{85}Kr is studied along the continuous two-months period ranging from 01 December 2020 to 31 January 2021. This period comprises the detection of an important number of ^{85}Kr events at all measurement sites (Fig. 2) due to favourable wind direction.

Stations	Lon/Lat	Height above ground	Distance from RP facility	Measurements time step
PTILH	1.8733° W, 49.6949° N	1.5 m	2 km	1 min
Urville	1.7431° W, 49.6607° N	1.5 m	10.4 km	1 min
Ludiver	1.7283° W, 49.6297° N	1.5 m	12.7 km	1 min
Octeville	1.6579° W, 49.6211° N	1.5 m	17.7 km	1 min
LRC	1.6458° W, 49.6347° N	2 m	18 km	1 min
Digulleville	1.8595° W, 49.7001° N	2 m	2.6 km	10 min
Beaumont	1.8358° W, 49.6613° N	2 m	4.2 km	10 min
Gréville	1.8097° , 49.6682° N	2 m	5.2 km	10 min

Table 1. Description of the 8 localisations of ^{85}Kr air concentration measurement stations used in this article.

2.3 Measurements campaign of ^{85}Kr in the North-Cotentin

The IRSN routinely monitors ^{85}Kr air concentrations close to the RP to study the transfer of radionuclides in the environment, but also to validate the atmospheric dispersion models and improve the understanding of radionuclides dispersion in various atmospheric conditions (Maro et al., 2002, 2007; Leroy et al., 2010; Connan et al., 2014). ^{85}Kr is a very good tracer of atmospheric dispersion in short and medium distances since it is an inert gas (noble gas), it is not chemically or physically reactive, so it does not get depleted by rain (wet scavenging) or by dry deposition processes. In addition, ^{85}Kr has a sufficiently long half-life ($\tau_{1/2}$=10.7 years) for its radioactive decay to be negligible at short and medium distances.

Since November 2020, IRSN has been carrying out a continuous ^{85}Kr air measurement campaign in several locations chosen at different distances and directions from RP, as part of the DISKRYNOC project. This project aims to provide a comprehensive new observational dataset for model validation purposes. In this study, this dataset is used to acquire feedback on the use of meteorological ensembles to quantify the associated uncertainties. The 8 closest air sampling locations from RP used in this article, described in Table 1 and shown in Fig. 1, are : PTILH, Urville, Ludiver, Octeville, LRC, Digulleville, Beaumont and Gréville. The IRSN is the owner of the first five stations, while the last three are Orano stations. The measurements have been carried out since mid-November 2020 and are expected to extend over approximately 18 months. This extended time period should provide a significant number of observations at all measurement sites. It should compensate for periods without reprocessing activity or with a wind direction towards the sea. Typical values of ^{85}Kr air concentrations in these stations range from tens to thousands $Bq.m^{-3}$ (depending on the distance from RP, wind direction, plant reprocessing activities and atmospheric conditions). Continuous measurements are being performed every minute in the IRSN stations and every 10 minutes in the RP stations.

The activity concentration in the air is determined by β counting in a Berthold LB123 or LB134 gas proportional counter calibrated with a common ^{85}Kr source specially fabricated (Gurriaran et al., 2001). This method is only useful in near fields

(less than 30 km) where the ^{85}Kr air concentration is sufficiently high. The same method has been used by Connan et al. (2014) and it has been documented by Gurriaran et al. (2004).

2.4 3-D wind observations

Evidence from past studies has shown that (in the absence of deposition) 3-D wind field is one of the most sensitive meteorological parameter for atmospheric dispersion models (Girard et al., 2014). For this reason, the performance of the AROME-EPS forecasts in terms of wind speed and direction should be assessed before they may be used for atmospheric dispersion. For this purpose, four kinds of observation data of wind have been used:

- The real-time ground observation acquisition network of Météo-France called RADOME, which has about 550 automatic ground observation stations spread over the whole French territory, among which two stations are located in the study area and are shown in Fig. 1: La Hague (1.9398° W, 49.7251° N) located ∼2.5 km from the RP plant, and Gonneville (1.4635° W, 49.6526° N) located ∼31 km from the RP plant. These stations provide continuous hourly measurement data of 10 m wind, temperature, humidity, rainfall and surface solar radiation fields.

- Vertical wind profile measured by Doppler Lidars (LIght Detection And Ranging). Atmospheric Lidars are currently used for atmospheric measurements of aerosols and wind, and thus allow for climate monitoring, air quality or cloud monitoring (Werner, 2005; Wu et al., 2022). A Doppler Lidar (version Leosphere Windcube 2) was recently installed by the IRSN's LRC laboratory, on the PTILH (Instrumented Technical Platform of La Hague) measurement site, located ∼2 km from the RP plant (Table 1). This Lidar provides wind data (speed and direction) at 10 minutes intervals on 13 vertical levels: 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260 and 280 m.

- Ultrasonic measurements acquired by Sodar (Sonic Detection And Ranging), which is a remote sensing instrument often used in meteorology for the 3-D acquisition of wind fields (speed and direction) on several vertical levels, using Doppler effect on sound waves levels (Tamura et al., 2001).
The Sodar measurements used in this work come from the instrument located ∼200 m West (1.8901° W, 49.6800° N) of the RP facility, which provides measurements on six vertical levels: 0, 10, 50, 100, 150 and 200 m.

- Ultrasonic measurements by LRC's anemometer (1.6458° W, 49.6347° N) installed at a height of 13 m above the ground. This instrument provide 10 minutes wind measurements.

Thus, five wind measurement points are available and used to evaluate the AROME-EPS meteorological ensemble over the two-month period of this work. This validation process has been done near the surface and on several vertical levels of the atmospheric boundary layer (ABL), as the knowledge of the evolution of the meteorological fields through out the lower atmosphere with a good accuracy is beneficial for a short distance atmospheric dispersion model to describe the physical processes (e.g. turbulence) occurring in the ABL.

	AROME	AROME-EPS
Domain	Western Europe, centred on France (∼1000 x 1000 km)	
Size	Deterministic (1 forecast)	16 members
Vertical levels	90 [5 m-10 hPa]	90 [5 m-10 hPa]
Horizontal resolution	1.3 km	2.5 km
Temporal resolution	1 h	1 h
Forecast initial time	0000, 0600, 1200, 2100 UTC	0300, 0900, 1500, 2100 UTC
Forecast range	48 h, 42 h, 48 h, 42 h	45 h, 51 h, 45 h, 51 h

Table 2. Description of AROME and AROME-EPS.

2.5 Description of AROME and AROME-EPS

The Météo-France NWP model AROME used in this study is summarized in Table 2 and extensively documented in Seity et al. (2011). AROME is a non-hydrostatic kilometer scale NWP limited area model. This model covers a geographical domain of about 1000 × 1000 km centred over the French territory with 90 vertical levels and horizontal resolution of 1.3 km. The

210 lateral and upper boundary conditions are provided by the operational global NWP model ARPEGE (Courtier et al., 1991) of Météo-France. AROME runs four times per day up to at least 42 h range, starting from the initial times 0000, 0600, 1200 and 1800 UTC. Its 3D-VAR data assimilation scheme (Brousseau et al., 2011) is a state-of-the art assimilation algorithm, that produce analyses at 1.3 km resolution by correcting the model state at hourly time steps using different kinds of meteorological observations (in-situ ground-based measurements, radar reflectivities and winds, satellite radiances, among others.).

215 The AROME-EPS (Bouttier et al., 2016) used in this article is a 16 members ensemble based on the AROME model at 2.5 km (Table 2). The ensemble runs four time per day, up to at least 45 h range, at 0300, 0900, 1500 and 2100 UTC. The 16 AROME-EPS perturbed initial conditions are built from the AROME 3D-var analyses on which perturbations from the Ensemble Data Assimilation (EDA) at 3.25 km resolution are added (Raynaud and Bouttier, 2016). The AROME EDA comprises 25 members that are obtained by perturbing the observations and the model state during the assimilation process. The outputs from both

220 are combined and interpolated to 2.5 km to produce the initial conditions of the AROME-EPS. The lateral and upper boundary conditions (Bouttier and Raynaud, 2018) are provided by the Météo-France ARPEGE-EPS operational global EPS (Descamps et al., 2015).

Besides the initial errors, forecast uncertainty also arises from the dynamic part of the model (e.g. spatial and temporal discretization of the equations that represent phenomena whose characteristic scale is larger than the mesh size), or from the

225 physical part of the model (e.g. corrective terms added to the dynamic equations to take into account the effect of phenomena whose scale is smaller than the mesh size). To account for model uncertainties in the AROME-EPS forecasts, the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme is used (Palmer et al., 2009; Bouttier et al., 2012). This method consists in adding random perturbations to the model physics tendencies.

2.6 Description of the pX model

230 The IRSN's Gaussian puff model pX used in this work is part of the operational platform C3X (Tombette et al., 2014), which is used by IRSN Emergency Response Center in case of an accidental radioactive release. pX is used to simulate the atmospheric dispersion of radionuclides on short and medium distances [500 m-50 km] (Korsakissok et al., 2013; Mathieu et al., 2012). The principle of such a dispersion model is based on the following assumptions:

- The release comes from a point source,

235 - A continuous release can be discretized into a series of puffs transporting a given amount of pollutants,

- Within each puff, the meteorological variables can be considered homogeneous,

- The concentration of pollutant in the puff can be represented by a Gaussian law in each of the three directions (Appendix B).

For an instantaneous release of a mass Q of a given radionuclide, the concentration c at a given point (x, y, z) and a time t is 240 given by:

$$c(x,y,z,t) = \frac{Q}{(2\pi)^{2/3}\sigma_x\sigma_y\sigma_z} \exp\left[-\frac{1}{2}\left(\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2} + \frac{(z-z_0)^2}{\sigma_z^2}\right)\right] \tag{1}$$

Where σ_x, σ_y and σ_z are the Gaussian standard deviations of the diffusion of the puffs over time in the three directions of space and (x_0, y_0, z_0) is the position of the mass center of the puff, which is transported by the mean wind flow. If x is the mean wind direction, and U is the mean wind speed in the x direction, then the position of the centre of the puff at each time 245 $t + dt$ from its position at time t is :

$$\boldsymbol{x_0}(t + \Delta t) = \boldsymbol{x_0}(t) + \boldsymbol{U}(t)\Delta t \tag{2}$$

This advection scheme allows to transport the puffs' mass centers through a non-stationary and heterogeneous wind field. In addition, the puffs are growing over time, to represent the plume's mixing by atmospheric turbulence. This is represented in Eq. (1) by the standard deviation σ_x, σ_y and σ_z that increase over time. This increase of plume spread depends on the 250 atmospheric stability, and is described by empirical standard deviation laws. In the pX model, the laws of Doury (Doury, 1976) or Pasquill (Pasquill, 1961) can be used. In this work, Pasquill stability was determined using two methods : (i) Pasquill-Turner (Turner, 1969) and (ii) the temperature gradient between 10 m and 100 m in the meteorological forecasts (Seinfeld and Pandis, 1998). That is, three stability diagnoses which are compared in this work (Appendix A).

For a continuous emission of release rate q_s (in $Bq.s^{-1}$) that is discretized into a series of N puffs, each puff i containing a 255 mass $Q_i = q_s\Delta t$, the concentration c at a given point is computed as the sum of the contribution of all puffs:

$$c(x,y,z,t) = \sum_{i=1}^{N} c_i(x,y,z,t) \tag{3}$$

where c_i is given by Eq. (1).

Finally, the mass of the material transported by the puff is depleted by wet and dry deposition as well as by radioactive decay. In our case, the transported mass will be assumed to remain constant over time, since ^{85}Kr is an inert gas (no deposition) who has a long half-life (no radioactive decay at short-distance) as shown previously. Equation (1) is also modified to take into account reflections on the ground and ABL height under certain conditions. Specifically, reflections on the ABL height are considered in unstable situations (when a capping inversion is assumed).

3 AROME-EPS verification

3.1 Scores for AROME-EPS verification

Before coupling the numerical weather predictions from AROME-EPS to the pX model, it is necessary to evaluate them in order to have an exhaustive overview of their quality and to take it into account in the interpretation of atmospheric dispersion simulations. For this purpose, two common scores, among others, used by the meteorological community for the evaluation of ensemble skill, have been calculated based on the observations of 3D-wind speed and direction described in Section 2.4:

Bias: In order to identify the systematic deviations of AROME-EPS meteorological ensemble forecasts from the observations, the bias over all days of the period of interest for a variable X is calculated at each forecast range t, by the following equation:

$$Bias(t) = \frac{1}{N_{day}} \sum_{d=1}^{N_{day}} \left(\langle X_{mod}^{(t,d)} \rangle - X_{obs}^{(t,d)} \right) \tag{4}$$

Where N_{day} is the number of days of the interest period. $\langle X_{mod}^{(t,d)} \rangle$ is the AROME-EPS ensemble mean at forecast range t on day d and $X_{obs}^{(t,d)}$ is the observed value at the same instant.

Spread-skill: As shown by Fortin et al. (2014), the ability of an ensemble to represent simulation errors can be evaluated by comparing, at each forecast range, the skill (or Root Mean Square Error, RMSE) of the ensemble mean and its spread (Spd), the latter calculated relative to the ensemble mean (Raynaud et al., 2012; Charrois et al., 2016). For a variable X, the ensemble Spd and RMSE terms are defined, at each forecast range t, as follows:

$$Spd(t) = \sqrt{\frac{1}{N_{day}} \sum_{d=1}^{N_{day}} \frac{1}{N_{ens} - 1} \sum_{n=1}^{N_{ens}} \left(X_{mod,n}^{(t,d)} - \langle X_{mod}^{(t,d)} \rangle \right)^2} \tag{5}$$

$$RMSE(t) = \sqrt{\frac{1}{N_{day}} \sum_{d=1}^{N_{day}} \left(\langle X_{mod}^{(t,d)} \rangle - X_{obs}^{(t,d)} \right)^2} \tag{6}$$

Where N_{ens} represents the ensemble size (N_{ens}=16 in the case of AROME-EPS). The value of variable X given by the ensemble member n at the forecast range t is $X_{mod,n}^{(t,d)}$. This diagnostic can be summarized by calculating the spread–skill

ratio, which should be as close to 1 as possible. Values less than 1 (respectively greater than 1) indicate that the ensemble is underdispersive (respectively overdispersive).

3.2 Model-to-data comparison of AROME-EPS

The evaluation of the quality of AROME-EPS predictions was carried out over the two-months period considered in this study (Dec. 2020-Jan. 2021). This evaluation process is done independently for each of the 0300, 0900, 1500 and 2100 UTC forecasts for all stations described in Section 2.4. The results for all forecasts and stations are similar. Therefore, only the results of the 1500 UTC forecast are shown here for two configurations: (i) at 10 m height (two RADOME stations: La Hague and Gonneville) and (ii) at several levels of ABL for stations where wind vertical profile measurements are available (Lidar at PTILH and Sodar at the RP site).

Figure 3 shows the ensemble biases in terms of 10 m wind speed and direction aggregated from La Hague and Gonneville stations. In the case of wind speed, the ensemble mean is above the observation for most of the forecast ranges, resulting in a slight systematic bias which varies between 0.71 and 1.45 m.s^{-1}. Both forecasts and observations (and eventually the resulting bias) evolution shows a marked diurnal cycle. The maximum bias is around 1500 UTC, corresponding to forecast range 25 h, with approximately no bias in the first forecast range. For the wind direction there is a good average performance of the model as shown by the good agreement between the averages of the forecasts and observations. The mean bias of the model oscillates around zero, with maximum and minimum values of $+7.8°$ and $-13.3°$, respectively.

Figure 4 shows the ensemble spread and skill evolution over forecast range. For wind speed, the spread of the ensemble is consistent with the RMSE with respect to the observations, with a slight underdispersion, while for wind direction the ensemble spread is above the RMSE at all forecast ranges. This overdispersion in terms of wind direction should be kept in mind when interpreting the ensemble atmospheric dispersion simulations.

To complement this evaluation at 10 m height, it is worth examining the quality of the AROME-EPS meteorological ensembles at different vertical levels in the lower atmosphere. To do so, the bias and the spread-skill ratio have been calculated at several vertical levels above ground. The results at the PTILH station are presented in Fig. 5 and summarized in Table 3. The wind speed forecasts are slightly less biased and overdispersive at higher altitudes than at lower ones, with an overdispersion more pronounced in the earlier forecast ranges. The latter is probably due to an imperfect accounting of modeling and/or initial conditions uncertainties in the perturbation process. However, the bias at 40 m (Lidar measurements) and 10 m (in-situ measurements) are consistent, which means that probably the high bias in the lower layers is not due to Lidar measurement errors. It may be due to the representation of surface processes in AROME in this area which is characterised by a complexe orography and heterogeneous surfaces (sea and land). For the wind direction there is no significant dependency of biases and spread of the ensembles with respect to the altitudes.

To summarise, the assessment of the consistency of AROME-EPS forecasts showed that they perform well by comparison to wind speed and direction measurements in the North-Cotentin area for the selected period, despite slight errors in the wind speed forecast. The same conclusion was reached for deterministic AROME forecasts (not shown here), by calculating the bias from wind observations.
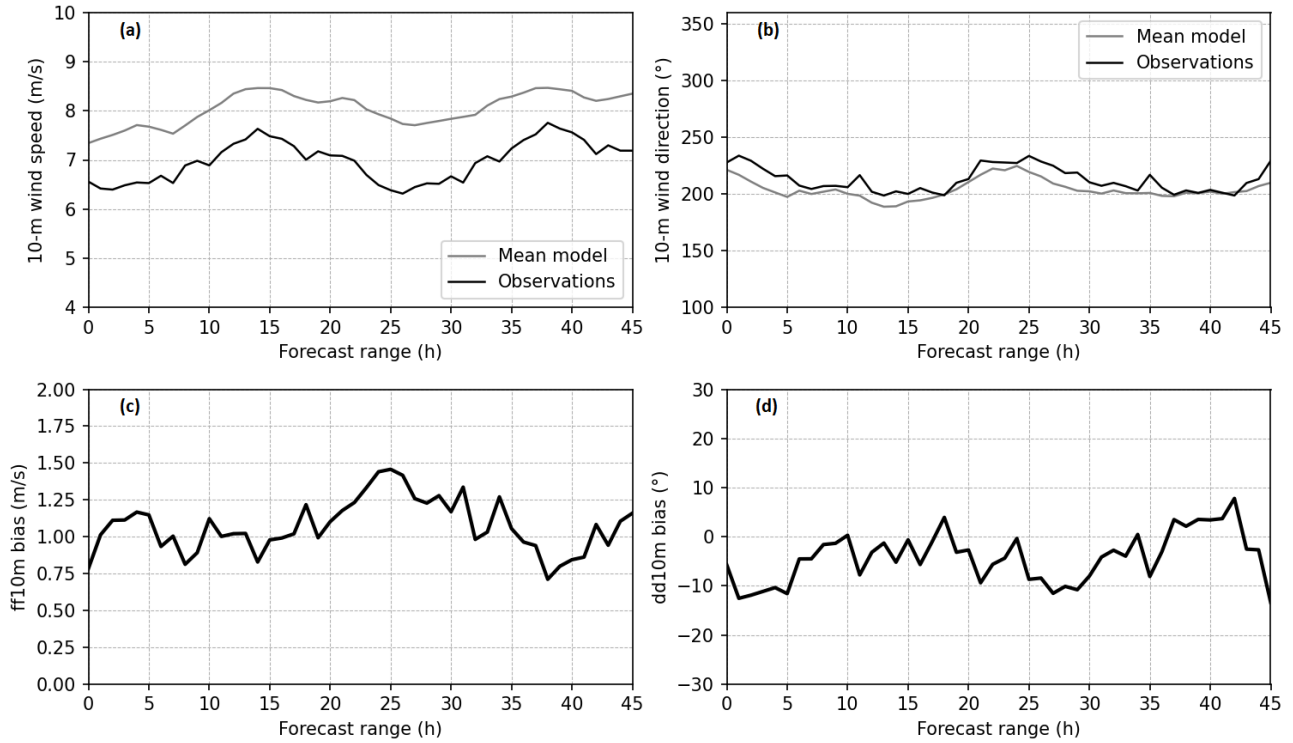
Figure 3. Ensemble mean and observations (a, b) as a function of forecast range, and the resulting bias (c, d) for both 10 m wind speed (referred to as ff10m) and direction (referred to as dd10m), aggregated from the two ground measurement stations La Hague and Gonneville.

4 Analysis of the ensemble dispersion simulations

4.1 Coupling AROME-EPS and pX model

Once meteorological forecasts from AROME-EPS have been qualified as shown in section 3, they are coupled to the Gaussian dispersion model pX (section 2.6). This process consists in running in parallel several simulations with the pX model, each using a different member of the AROME-EPS ensemble as input, along with the source term data provided by RP La Hague. This allows to generate an ensemble of dispersion simulations composed of 16 members (hereafter called the pX ensemble). Furthermore, in order to quantify the benefit of using ensembles instead of deterministic simulations, an additional pX simulation was performed using the deterministic weather forecast from AROME as input of the model. This simulation is called deterministic pX in the following.
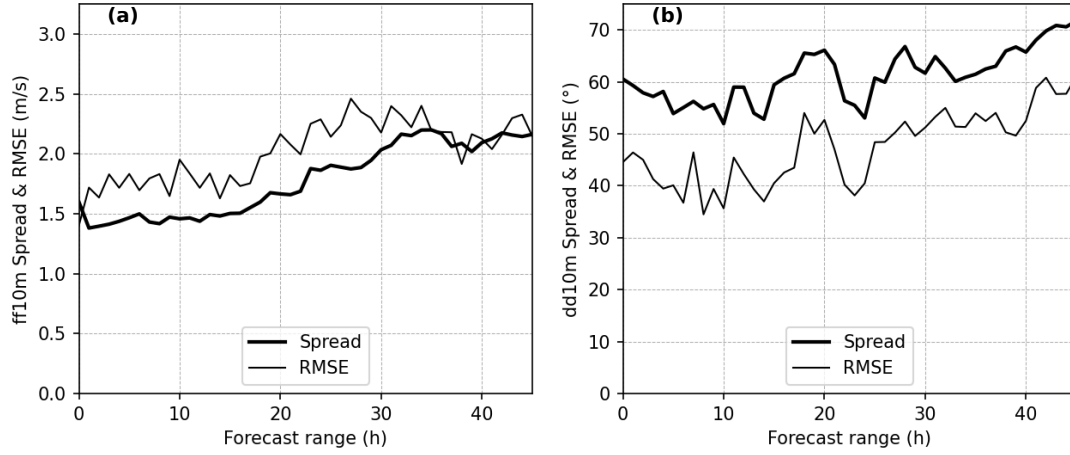
Figure 4. Ensemble spread and RMSE of the ensemble mean forecast for both 10 m wind speed (referred to as ff10m) (a) and direction (referred to as dd10m) (b) aggregated from the two ground measurement stations La Hague and Gonneville, as a function of forecast range.

Levels (m)	Bias		Spread-skill ratio	
	Wind speed (m.s^{-1})	Wind direction (°)	Wind speed	Wind direction
40	0.948	1.696	0.959	1.416
60	0.606	0.852	1.034	1.410
80	0.448	0.913	1.062	1.412
100	0.364	1.002	1.089	1.416
120	0.217	0.581	1.116	1.418
140	0.141	0.490	1.134	1.412
160	0.095	-0.421	1.145	1.407
180	0.053	-0.468	1.151	1.402
200	0.062	-0.305	1.171	1.386
220	0.042	-0.915	1.182	1.356
240	0.023	0.247	1.200	1.393
260	-0.021	-0.790	1.216	1.395
280	-0.078	-1.246	1.226	1.444

Table 3. The AROME-EPS ensemble bias and spread-skill ratio averaged over forecast range in the vertical levels observed by Lidar located in the PTILH site.
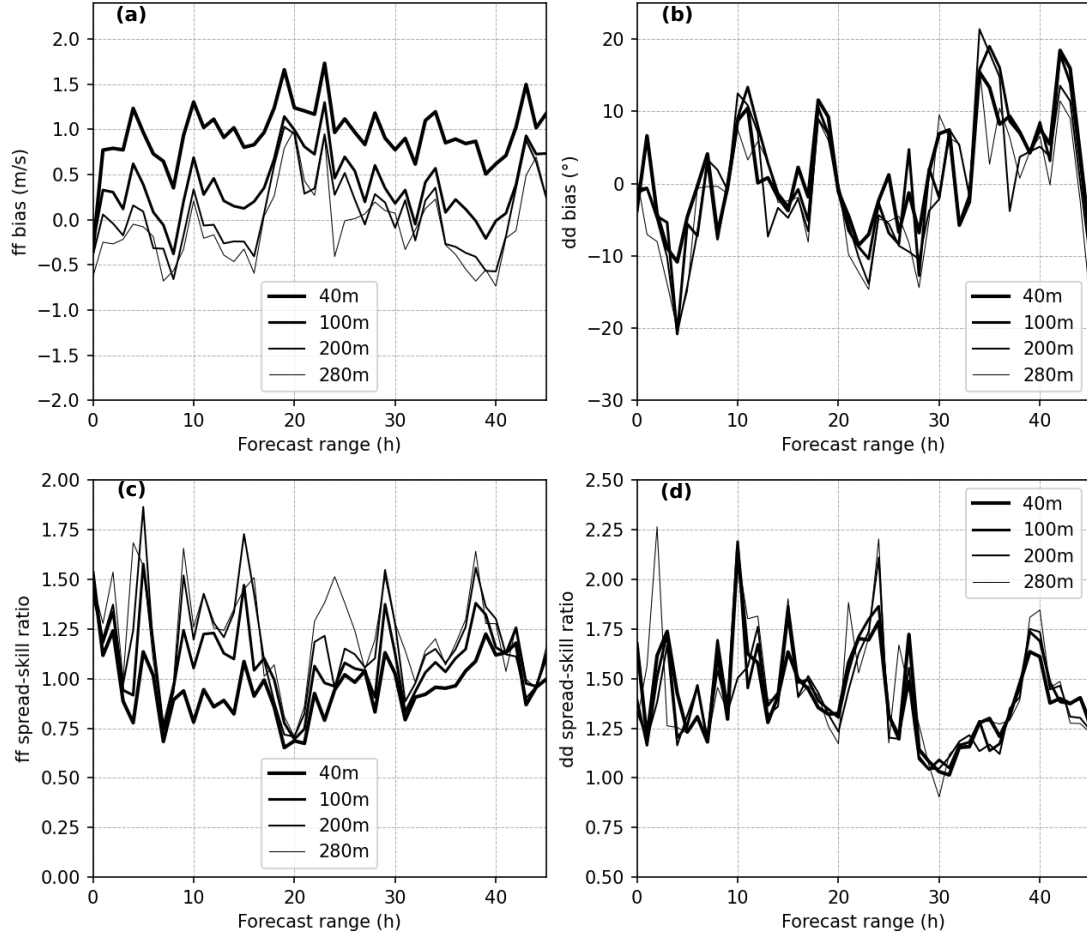
Figure 5. Bias (a, b) and spread-skill ratio (c, d) of the ensemble between 40 m and 280 m above ground, for both wind speed (referred to as ff) (a, c) and direction (referred to as dd) (b, d) measured by Lidar at the PTILH station, as a function of forecast range.

4.1.1 Temporal continuity of AROME-EPS members

In the case of accidental releases that span a long time period, it is important to properly build continuous time series from several forecasts made at different initial times, without causing brutal jumps in spread between forecasts from two consecutive days. Besides, fine-scale weather forecasts are, usually, not available until few hours after their start. To deal with this issue,
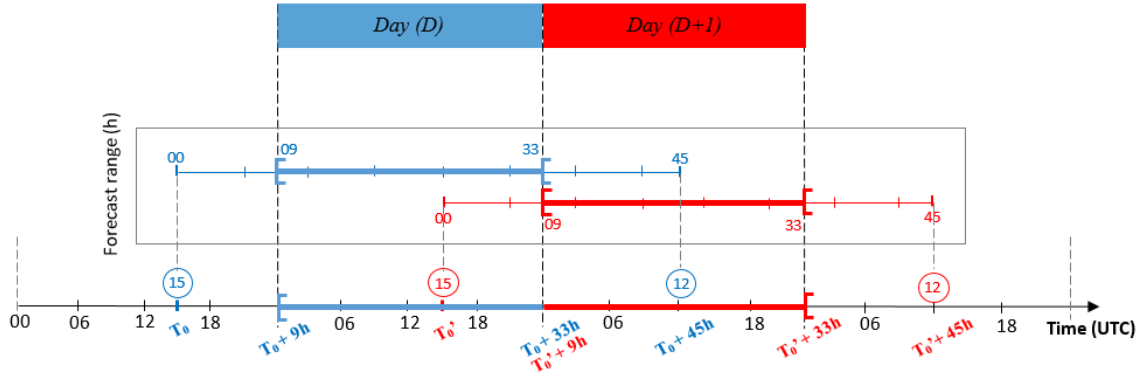
Figure 6. Illustration of meteorological forecasts used from AROME-EPS (in bold) as input to pX: the forecast starting from 1500 UTC on a day D is used to cover the next day $D+1$.

330 this study proposes to use the closest available forecast at the beginning of a day. In other words, to simulate a release occurring from 00 h to 23 h of a day D, the AROME-EPS forecasts starting from 1500 UTC of the day before $(D-1)$ are used. Thus, the first 8 forecast hours are skipped and the next 24 ones [09-32h] are used to cover the entire day. In the same way, the forecast starting from 1500 UTC of day D is used to cover the day $D+1$. Figure 6 illustrates this cycle. Then, these intervals are combined to cover the three simulation periods detailed in Table 5, by connecting each member i of day D with member i
335 of day $D+1$.

Note that to perform pX deterministic simulations, the deterministic weather forecasts from AROME are built in the same way, by using the forecast starting from 1200 UTC. In this case the first 11 forecast hours were skipped and the next 24 h were used [12-35h].

4.1.2 pX simulation set-up

340 The calculation domain of pX is defined by the grid of meteorological forecasts, and the puffs that leave this domain no longer participate to the concentration calculations. Moreover, the concentration calculated on a point located on the border of the meteorological domain will only account for the contribution of the puffs inside. It is, therefore, necessary to define a simulation domain whose borders are sufficiently far from the calculation points (i.e. ^{85}Kr measurements sites in Fig. 1). Thus, a 60×60 km domain centred on the source (i.e. RP) was defined, where the meteorological forecasts were interpolated
345 on a Cartesian grid with 2.5 km of horizontal resolution, leading to a horizontal mesh of 24×24 cells. This process was accomplished by a weather pre-processor that was developed as part of this study.

Considering the objectives of this work, only the first 25 vertical levels outputs from AROME [10 m-3000 m] are used here to cover the entire ABL. The ABL height is diagnosed from AROME forecasts as the lowest altitude where the turbulent kinetic energy is below $0.01 \, \mathrm{m^2.s^{-2}}$. As a result, it may happen that this diagnostic reaches unrealistically low values, as low as 10 m.

350 In order to avoid such low values and to ensure that the source emission does not occur above the ABL, a minimum value of 200 m is imposed to the ABL height before being applied to the pX simulations.

Even though the NWP forecasts are given with an hourly frequency, the pX simulations were performed in this study with a time step of 10 minutes in order to better capture the temporal variations of the plume. Sensitivity tests showed that the pX simulations with the two Pasquill-based stability diagnoses (Pasquill-Turner and temperature gradient) gave very similar
355 results with a very slight better performance of the diagnosis of temperature gradient. In the following, simulations with Pasquill (called pX-Pasquill) standard deviations will thus be computed with the latter stability diagnosis and will be compared with simulations using Doury standard deviations (called pX-Doury).

Finally, the effects of the complex topography (coastline, rocky terrain) and buildings on the plume dispersion may lead to downwash effects that are not explicitly taken into account by the Gaussian puff model. To compensate for this limitation, an
360 effective height that differs from the physical stack height may be used as input. In this case, five values of effective height have been tested: 20, 50, 100, 150 and 200 m, and the most optimum simulations were obtained by using the physical stack height of 100 m.

4.2 Qualitative results and discussion

In this section, we illustrate the behavior of the dispersion simulations at two stations located at short and medium distances
365 from the source. For this purpose, a short time period where the occurrence of few marked events was selected.

4.2.1 Comparison of ensemble and deterministic dispersion results

Figure 7 and Figure 8 show an example of dispersion results at PTILH (2 km from the source) and LRC stations (18 km from the source), respectively. There are observed peaks that the deterministic simulation does not reproduce while some ensemble members simulate them with acceptable accuracy. This highlights the potential interest of the ensemble approach compared
370 to the deterministic one. In addition, visual analysis of the results shows that the pX-Pasquill simulations correctly predict most peaks at PTILH station (Fig. 7), although with a tendency to underestimate their maximum value. At the same station, pX-Doury simulations show a strong underestimation of concentrations, resulting in a failure to forecast most observed peaks. As the PTILH station is located only 2 km from the source, the release conditions (initial buoyancy and building downwash effects) largely influence the concentrations at this short distance.
375 In our simulations, the use of the stack height (100 m) as release height does not allow to accurately predict significant ground concentrations at this distance, due to approximations made in the Gaussian model that does not include building downwash effects. This is especially the case when using Doury standard deviations, which simulate a very narrow plume on the vertical, resulting in an underestimation of ground concentrations in the case of elevated release. This phenomenon, that characterizes pX-Doury simulations in stable situations, was specifically shown in the case of La Hague RP (Connan
380 et al., 2014; Korsakissok et al., 2016). At LRC station (Fig. 8), located farther from the source, this underestimation is much less visible and peaks are better reproduced. There are still, however, peaks that are missed by the deterministic simulations
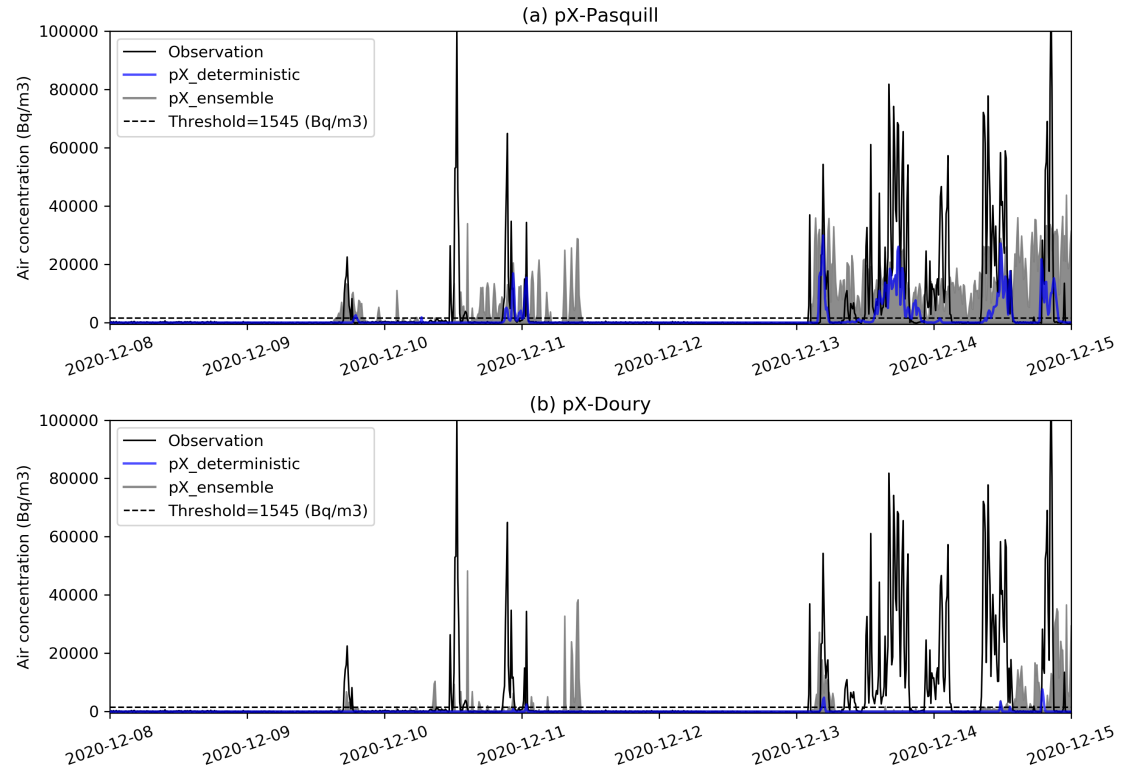
Figure 7. pX-Pasquill (a) and pX-Doury (b) ensemble and deterministic simulations of ^{85}Kr air concentration compared to the observation in the first aggregated period (Table 5), at the PTILH station. The horizontal dashed line shows the air concentration threshold (1545 $Bq.m^{-3}$) above which peaks are considered.

and forecast by some members of the ensemble. This can be explained by the meteorological uncertainties, as detailed in the following section.

4.2.2 Sensitivity to wind and atmospheric stability

385   In this section, the sensitivity to meteorological variables such as wind and stability is detailed. The aim is to illustrate how small variations in these parameters affect the outputs of atmospheric dispersion simulations. For this purpose, three members of the pX-Pasquill ensemble which have different behaviors are shown. The study is carried out at LRC station, where both air concentration and wind measurements are available. This station is representative of the model behaviour at medium distance, where release conditions are of relatively less importance than meteorological uncertainties. Table 4 summarizes the five

390   observed peaks (with peaks 2, 3 and 4 are much smaller than peaks 1 and 5) from 08 Dec. 2020 to 12 Dec. 2020, when the ensemble behaviour is studied.
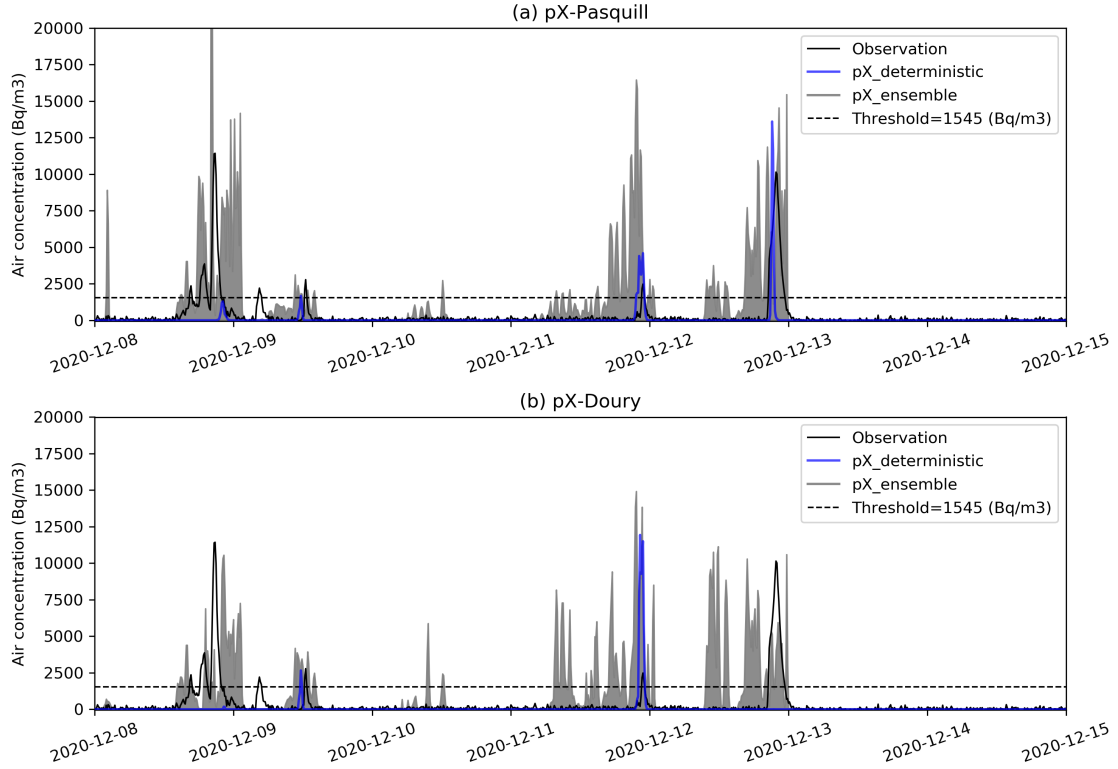
Figure 8. The same as in Figure 7, but in the LRC station.

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
Date (UTC)	08 Dec. 2020, 20:45	09 Dec. 2020, 04:30	09 Dec. 2020, 12:30	11 Dec. 2020, 22:50	12 Dec. 2020, 21:50
Observed activities	11432	2204	2780	2482	10145
Member 1 [timing error]	8000 [+1h15min]	_	_	11335 [-2h00min]	6000 [+1h30min]
Member 2 [timing error]	13795 [+3h20min]	_	_	1714 [+0h50min]	10945 [-3h10min]
Member 3 [timing error]	9400 [-2h05min]	_	2900 [-1h50min]	4690 [-1h20min]	10468 [+0h10min]

Table 4. The five observed peaks of ^{85}Kr at LRC station (Figure 9) from 08 Dec. 2020 to 12 Dec. 2020, and the simulated peaks in the three selected members (in $Bq.m^{-3}$), for pX-Pasquill configuration.
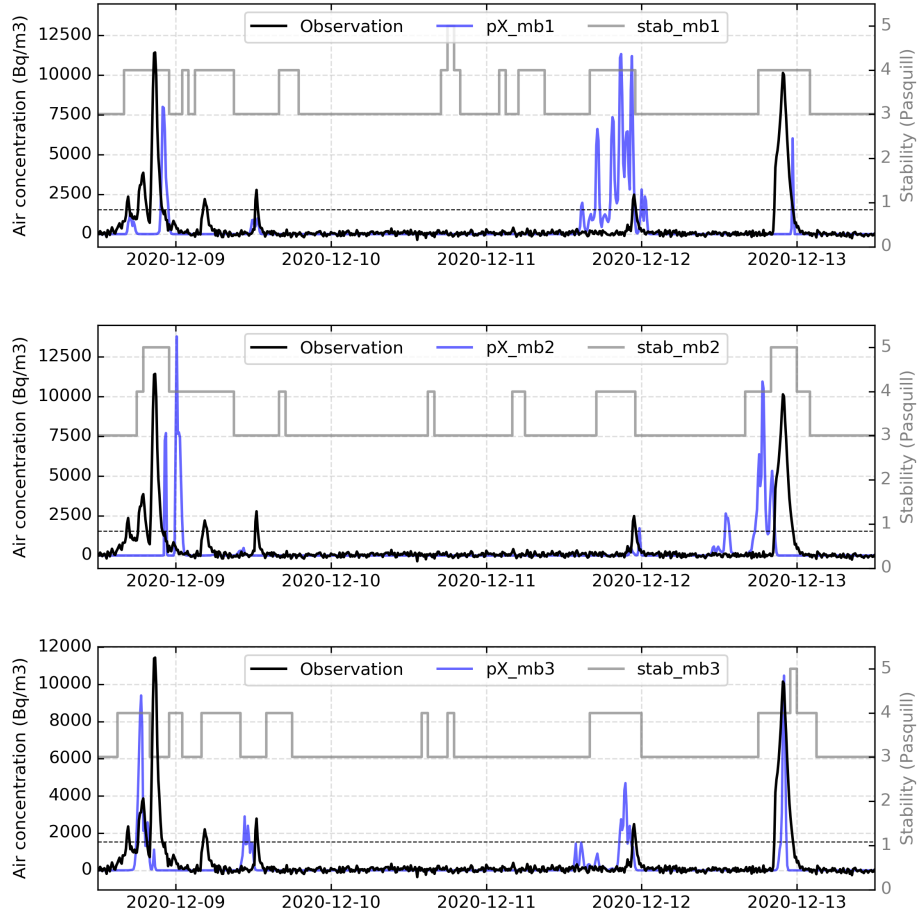
Figure 9. The three selected pX ensemble members (referred to as pX_mb1, pX_mb2 and pX_mb3) of ^{85}Kr activities, at LRC site, from 08 Dec. 2020 at 12:00 to 13 Dec. 2020 at 12:00. The gray curves represents the hourly Pasquill stability classes (referred to as stab_mb1, stab_mb2 and stab_mb3) ued to generate each of the three pX members. The horizontal dashed line shows the air concentration threshold (1545 $Bq.m^{-3}$) above which peaks are considered.

Although the wind forecasts used to generate the three pX simulations in Fig. 9 are sufficiently close to the observation at LRC station around the time of peaks occurrence (as shown in Fig. 10), some events are reproduced either with small errors in timing (i.e. delay/advance of two to three hours) or with errors in intensity (i.e. underestimation/overestimation). This can be a result of local effects on the dispersion simulations. In other words, when one is interested in calculating the activity concentration at a point in space, a small error in the wind speed and/or direction can have a significant impact on the estimation of the peaks timing and intensity, due to the sharp concentration gradient. Figure 11 illustrates this issue in the case of the 3rd

395
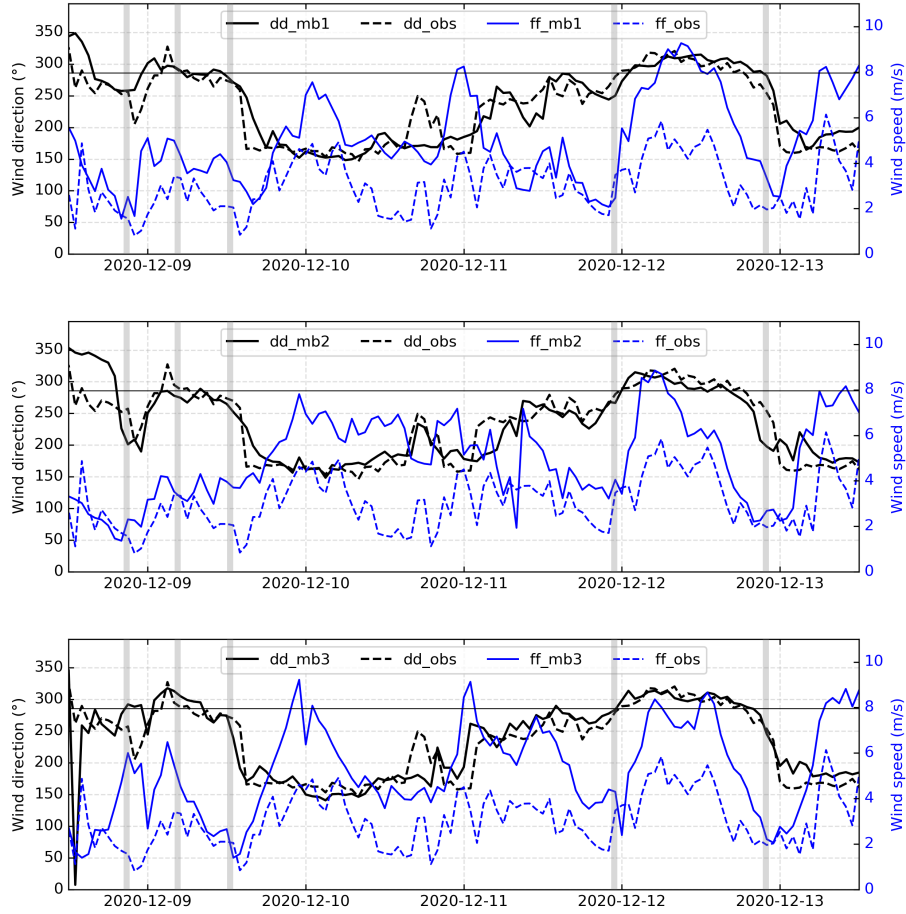
Figure 10. The three wind forecasts from AROME-EPS (solid curves) used to generate the pX simulations presented in Figure 9, compared to wind observations stored at LRC (dashed curves). ff and dd denotes, respectively, wind speed and direction, while mb1, mb2 and mb3 denotes each of the three AROME-EPS members. The vertical gray lines shows the occurrence time of the five peaks presented in Table 4. The horizontal line shows the angle of the LRC station with respect to the source (286°), which corresponds to the wind direction that transports the plume from the source to LRC.

peak of member 1 in Fig 9 and Table 4. This peak underestimates the air concentration because it is located close to the edge of the plume where the concentration gradients are expected to be high.

400   Figure 9 also shows the effect of the diagnosed stability classes on the dispersion simulations. Almost all the simulated peaks are associated with stable conditions of the atmosphere (4th or 5th Pasquill stability class, corresponding to classes E and F respectively in Appendix A). This may explain the failure of members 1 and 2 to reproduce the third observed peak which is associated with neutral stability conditions (3rd stability class, corresponding to class D of Pasquill).
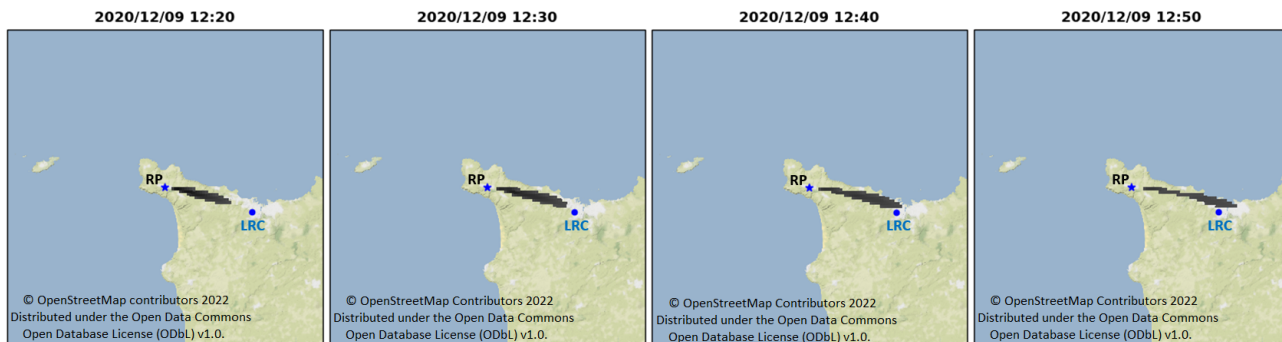
Figure 11. Simulation of the dispersion of the ^{85}Kr exceeding 1545 $Bq.m^{-3}$ in the case of the member 1 presented in Figure 9 and Table 4, surrounding the moment of the 3rd observed peak at LRC station.

In order to better understand the effect of the stability conditions on the pX simulations, a test was carried out by using stationary stability classes in the simulations, using each time one of the three classes obtained by the temperature gradient diagnosis in the period shown in Fig. 9. The result, shown in Fig. 12, confirms that the diagnosed stability classes has a significant effect, mainly on the simulated intensity. It allows to explain some model failures, such as the peak 3 for the member 1, but not others. In most cases the 4th stability class gives the highest ^{85}Kr concentration.

In summary, while some detections / non-detections can be easily explained by examining wind speed and direction time series, other features are less predictable on this sole basis, due to the interaction between the variables and the possible accumulation of small direction errors over the plume trajectory.

4.3   Statistical evaluation of the dispersion ensemble

4.3.1   Evaluation procedures

It is often a desirable feature for a dispersion model to be able to correctly predict a threshold exceedance. It is particularly useful for decision-making purposes, when protective actions for the population are based on the prediction of zones where a given dose threshold could be exceeded. Evaluating the model performance for this kind of purpose is often based on contingency tables (Wilks, 2019) allowing to compare the series of observations and simulations by counting four features: (i) true positive (TP) when a peak is observed and well simulated, (ii) false negative (FN) when a peak is observed but not simulated, (iii) false positive (FP) when there is no observed but simulated peak and (iv) true negative (TN) when there is no observed and no simulated peak.

The method used by Quérel et al. (2022) is based on this principle and used to evaluate a series of peaks from deterministic simulation against observations. This method consists in evaluating the success/failure of the model for each observed or simulated peak, including a defined temporal tolerance. However, in the case of an ensemble, the same procedure cannot apply because there are multiple simulations and so unobserved events cannot be well-defined. In addition, it often occurs that the FPs from different members constitute a series that exceeds the correlation timescale between peaks (i.e. the temporal offset
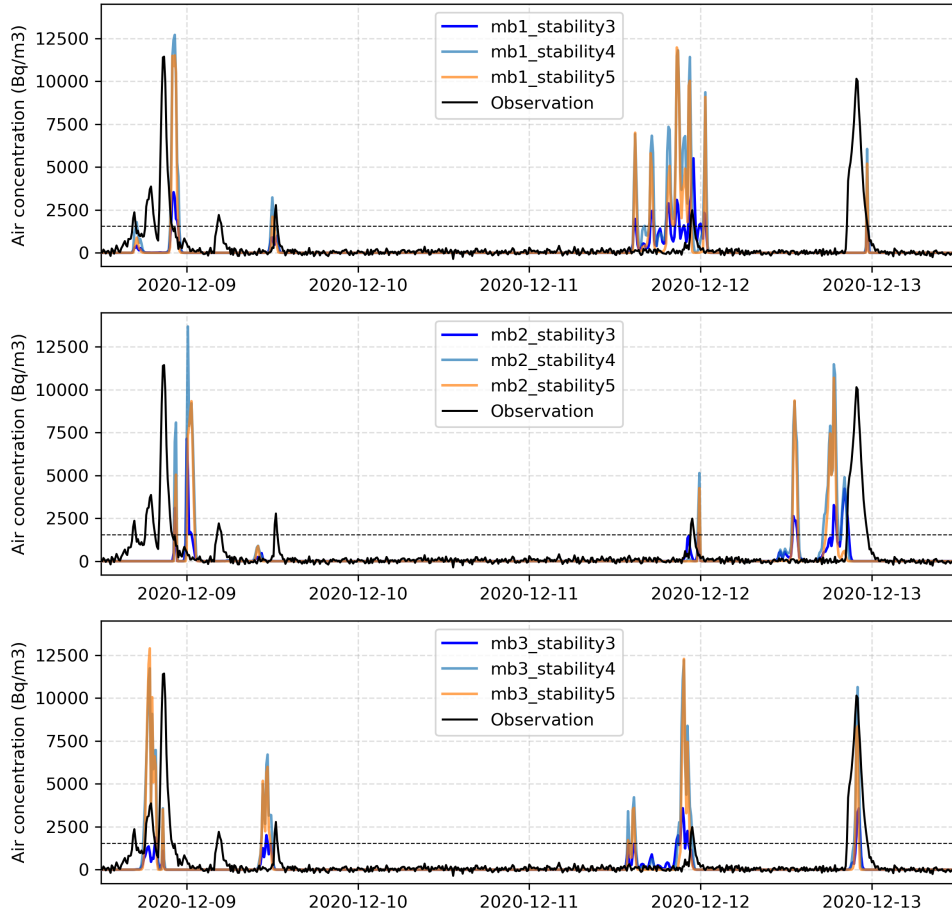
Figure 12. The same experiments as in Figure 9, but with three stationary stability classes of Pasquill: 3rd, 4th and 5th class. mb1, mb2 and mb3 denotes, respectively, each of the three members of pX ensemble simulations.

between two peaks at which they no longer correspond to the same event). Hence, one cannot decide whether all these TPs correspond to the same event or several. To deal with this problem, the method used in this work consists in discretizing the time series by sliding intervals of length Δt and moving time τ (Fig. 13). For the kth discretization step, the evaluation interval is $[t', t' + \Delta t]$, such that:

$$t' = t_0 + (k - 1)\tau \tag{7}$$

Where t_0 is the initial time of the time series.

Then, the maximum values from each ensemble simulation are compared to the maximum observed value in each discretization interval. Thus, considering a threshold of 1545 $Bq.m^{-3}$ (corresponding to the detection threshold for air concentration of
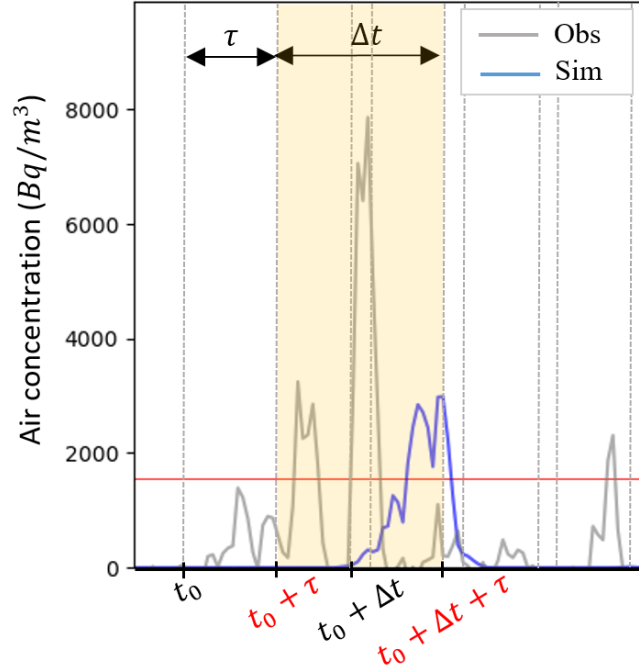
Figure 13. Illustration of the temporal discretization method by sliding intervals used in this study. This figure shows the case of the second time interval of length Δt: $[t_0 + \tau, t_0 + \Delta t + \tau]$.

^{85}Kr) and a given decision threshold x (i.e. the number of members at which the success/failure of the ensemble is considered), the four features of the contingency tables are defined in this case as follows:

- TP: when a peak is observed and well simulated by at least x members,

- FN: when a peak is observed but simulated by a number of members less than x,

- FP: when there is no observed peak but simulated by at least x members,

- TN: when there is no observed peak but simulated by a number of members less than x.

This method also allows to decrease the number of TN without having a statistically significant impact on the scores thanks to the normalization in the contingency tables. This advantage gives the possibility of using scores that integrate the number of TN, contrary to the classical method which is not adapted to the case of rare events (i.e. the case of non continuous events in time).

Then, the performance of the ensemble is measured using hit rate (H) and false alarm rate (F) metrics (Quérel et al., 2022; Wilks, 2019). The hit rate (also called recall) is the fraction of the observed events that are successfully reproduced (Eq. (8)). The false alarm rate is the fraction of the simulated peaks that are not observed (Eq. (9)).

$$H = \frac{TP}{TP + FN} \tag{8}$$

Aggregated periods	08 Dec. 2020 to 15 Dec. 2020	26 Dec. 2020 to 30 Dec. 2020	07 Jan. 2021 to 26 Jan. 2021	Total
Observed peaks	116	92	200	408 peaks/30days

Table 5. Aggregated time periods for calculating the probabilistic scores of evaluation of pX ensemble and deterministic simulations.

$$F = \frac{FP}{FP + TN} \tag{9}$$

To choose the most representative combination $(\Delta t, \tau)$, we consider that two events are independent if they are more than 3 hours apart. Thus, six combinations $(\Delta t, \tau)$ are tested by the statistical scores below: $(\Delta t = 1\ h, \tau = 1\ h)$, $(\Delta t = 2\ h, \tau = 1\ h)$, $(\Delta t = 2\ h, \tau = 2\ h)$, $(\Delta t = 3\ h, \tau = 1\ h)$, $(\Delta t = 3\ h, \tau = 2\ h)$ and $(\Delta t = 3\ h, \tau = 3\ h)$.

In the case of the AROME-EPS-pX ensemble, there are 16 possible decision thresholds ($x = 1, 2, ..., 16$). In order to identify the most optimal ones, the ROC (Relative Operating Characteristic) curves are commonly used as a graphical summary of the decision-making skill of an ensemble, by connecting all points $[F(x), H(x)]$ for each decision threshold x (Swets, 1973; Wilks, 2019; Raynaud and Bouttier, 2016). In addition, to better capture the internal variation of the performance of the model according to the decision thresholds, the Peirce Skill Score (PSS) (Peirce, 1884; Wilks, 2019) was calculated for each x, as follows:

$$PSS(x) = H(x) - F(x) = \frac{TP \times TN - FP \times FN}{(TP + FN) \times (FP + TN)} \tag{10}$$

Note that the $PSS(x)$ corresponds to the vertical distance between the point $[F(x), H(x)]$ of the ROC curve and the no-skill line (i.e. the bisector line, $H = F$). That means that the threshold that presents a better compromise between the probability of detection and the probability of false detection of events corresponds to the one that maximizes the PSS (the closest point to $[F = 0, H = 1]$ in the ROC) (Manzato, 2005, 2007).

Finally, the verification process was performed by aggregating the measurements and simulations of all stations in three periods where a high density of events was recorded (Table 5). This gives a total number of observed threshold exceedance events of 408 over 30 days which is sufficient for the metrics to be statistically robust.

4.3.2 Statistical results

Simulations and observations at all stations were aggregated in order to investigate the probabilistic performance of the ensembles, using ROC curves and PSS. Figure 14 shows the results for three combinations $(\Delta t, \tau)$ (the three other cases that were not shown are similar to $(\Delta t = 1\ h, \tau = 1\ h)$ and $(\Delta t = 2\ h, \tau = 1\ h)$). For the deterministic simulations, all the discretization configurations give almost the same false alarm rate (around 6 %) but with large difference for hit rates, with a difference of about 20 % between the best and the worst configuration. The best scores were obtained with the discretization parameters $(\Delta t = 3\ h, \tau = 2\ h)$ for both pX-Doury and pX-Pasquill simulations. This configuration also gave the closest results to the
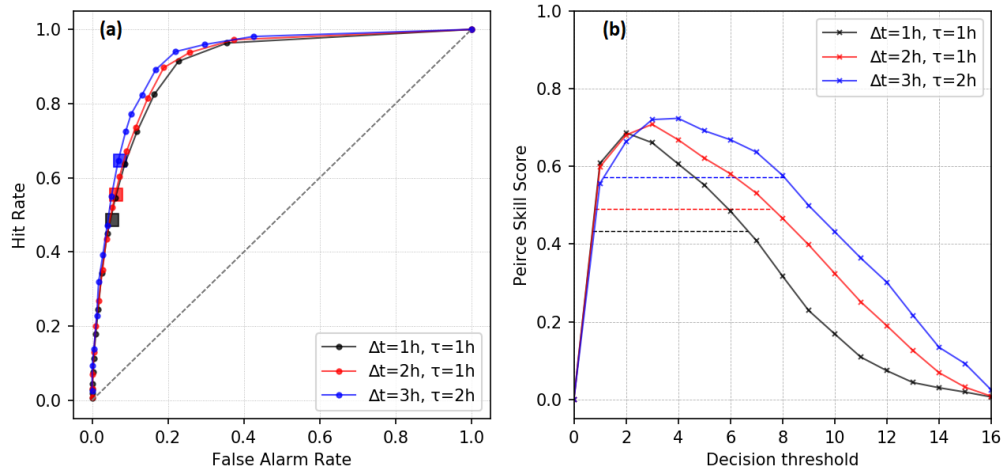
Figure 14. ROC curves (a) and the PSS as a function of decision thresholds (b) of the pX ensemble simulations performed with Pasquill stability classes by aggregating simulations and observations at all stations. There is one curve for each $(\Delta t, \tau)$: $(\Delta t = 1\ h, \tau = 1\ h)$, $(\Delta t = 2\ h, \tau = 1\ h)$, $(\Delta t = 3\ h, \tau = 2\ h)$. The values of the scores for the deterministic pX simulation are indicated by squares in the ROC curves and by horizontal dashed lines in the PSS curves. The diagonal dashed lines are the no-skill lines $(H = F)$.

scores obtained with the method of Quérel et al. (2022) for the deterministic simulation (not shown here). In the following, only the results with the Pasquill standard deviations will be shown.

475   The pX-Pasquill ensembles have a maximum value of PSS $PSS_{max} = 0.72$ corresponding to an optimal decision thresholds of 3 and 4 members. The ensemble performs better than the deterministic simulation in a range of seven decision thresholds, which represents almost 50 % of the possible values of the decision thresholds. In addition, the ensemble simulations allows to optimize the decision threshold (Richardson, 2001). These results highlight the robustness of the probabilistic simulations compared to the deterministic simulation in the process of the prediction of threshold exceedances.

480   To go further into the analysis of the probabilistic performance of the ensembles, the effect of the distance from the source is investigated in Fig. 15. The most representative discretization parameters $(\Delta t = 3\ h, \tau = 2\ h)$ were used to generate dispersion simulations by aggregating data for two groups of stations. The first is $-10km$ group which contains stations at distances less than 10 km: PTILH (2 km), Digulleville (2.6 km), Beaumont (4.2 km) and Gréville (5.2 km). The second is $+10km$ group which contains stations beyond 10 km: Urville (10.4 km), Ludiver (12.7 km), Octeville (17.7 km) and LRC (18 km).

485

     The model performs better in the near-field stations. In this case, the ensemble is more efficient than the deterministic simulation in 50 % (8 members) of the decision thresholds, against 37.5 % (6 members) for stations located beyond 10 km. In both cases the optimal threshold is 3 members. This better performance in the near-field area may be related to the accumulation
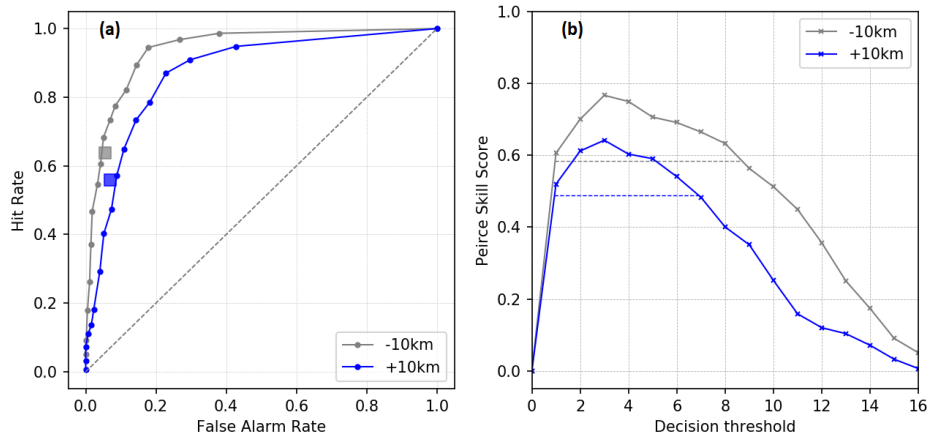
Figure 15. ROC curves (a) and the PSS as a function of decision thresholds (b) of the pX ensemble simulations performed with Pasquill stability classes by aggregating data in the two groups of stations: $-10km$ (Beaumont, Digulleville, Gréville, and PTILH) and $+10km$ (LRC, Ludiver, Octeville and Urville). There is one curve for each group of stations. The scores are calculated using the most optimal discretization configuration ($\Delta t = 3\ h, \tau = 2\ h$).

of small errors along the plume trajectory, that could make the far-range forecast of the peaks more accurate. In addition, peaks
490 at distances farther than 10 km are closest to the background noise, and some errors in the peaks detection may be linked to the difficulty to discriminate a peak from the background noise.

With the diffusion laws of Doury (not shown here), the best scores are obtained also for the group of nearest stations to the source in the case of deterministic simulation. However, for the ensembles there is no significant dependency of the probabilistic scores with respect to the distance from the source.

495 Taking into account both meteorological and model uncertainties would imply generating an ensemble by also perturbing model parameters (Pasquill/Doury, source elevation, stability). In this perspective, a 32-member super-ensemble was generated by combining pX-Pasquill and pX-Doury ensembles. The result (not shown here) is very similar to the pX-Pasquill ensemble, without any notable added value.

5   Conclusions and perspectives

500 In this study we explored the potential value of using fine-scale spatial and temporal meteorological ensembles to represent the inherent meteorological uncertainties in ADM outputs. To do so, the high-resolution operational forecasts AROME-EPS of Météo-France have been coupled to the Gaussian puff short-range dispersion model pX of IRSN to generate a 16-member dispersion ensemble, which accounts for meteorological uncertainties. This paper also proposes an original method to evaluate

the ability of a dispersion ensemble to forecast threshold exceedances, using probabilistic scores. For this purpose, we used an original data set of continuous ^{85}Kr air concentration measurements (DISKRYNOC campaign recently conducted by IRSN), along with a well-known source term (every 10 minutes, provided by Orano La Hague RP) and meteorological data (NWP from Météo-France and continuous observations from Météo-France/IRSN).

As a first step, the assessment of the quality of the AROME-EPS forecasts, in the North-Cotentin peninsula (north-western France) was carried out, using meteorological observations, over the two-month period of interest (Dec. 2020-Jan. 2021). Wind speed and direction are the most influential variables on the transport of a plume through the atmosphere. For this reason, the meteorological ensembles were evaluated in terms of these two meteorological variables in 25 vertical levels within the ABL. The results of this evaluation showed that the AROME-EPS ensembles represent the wind in the ABL with a very acceptable accuracy, despite the slight systematic errors present in the lower layers.

Then, an ensemble dispersion modeling chain was implemented using the AROME-EPS forecasts as inputs to the pX model. At this stage, it was necessary to set up a way to combine several 45 h forecasts from different initialization times, and that could be used in the early phase of a nuclear accident. The method proposed in this paper is to use the newest forecast available at the beginning of a day (at 0000 UTC). This approach can be used to span long period, in the case of an emergency, by juxtaposing 24-h successive forecasts. Then, two configurations of dispersion simulations were run, with Pasquill and Doury Gaussian standard deviations. A qualitative assessment of the simulations was first presented, to illustrate the ability of some members of the ensemble to forecast peaks while the deterministic simulations failed. The sensitivity of the results to atmospheric stability diagnosis was also highlighted. The probabilistic consistency of the resulting dispersion ensembles was then compared using an innovative method of temporal discretization by sliding intervals, and by calculating two probabilistic scores: ROC curves and PSS. This evaluation process was performed in two parts. First, by comparing the overall performance of the ensemble by aggregating the data from all the measurement stations. Secondly, by comparing the performance of the two configurations in the near fields (stations located less than 10 km from the source) and far fields (stations beyond 10 km from the source). In all cases studied, the best decision threshold was found to be 3 members, and the ensembles performed better than the deterministic simulations. For operational purposes during emergency situations, this result would imply that when 3 or more members of the ensemble forecast a threshold exceedance, protective actions should be recommended.

One of the limitations of this study is that it evaluates the performance of the dispersion ensembles by only considering its ability to forecast a value above a given threshold, with a temporal tolerance between the simulated and observed peaks. To complete this evaluation, it would be interesting to develop complementary indicators that evaluate the consistency of dispersion ensembles in term of intensity between the simulated and observed peaks. In addition, since ^{85}Kr is a noble gas, this work do not investigate the deposition of radionuclides on the ground, a parameter that would be sensitive to uncertainties in rain forecasts. Thus, it will be interesting to apply the approaches developed in this study to the case of another atmospheric tracer that is not an inert gas, such as radon-222 (Quérel et al., 2022).

10-m wind speed (m.s^{-1})	Day Surface solar radiation downwards w/m2			Night Total cloud cover (%)	
	Strong]700, +∞]	Moderate [350, 700]	Low [0, 350[	[4/8, 7/8]	[0, 3/8]
<2	A	A-B	B	F	F
3	A-B	B	C	E	F
3 - 5	B	B-C	C	D	E
5 - 6	C	C-D	D	D	D
>6	C	D	D	D	D

Table A1. Pasquill classes according to method of Turner.

Appendix A: Atmospheric stability by classes

Stability classes of Doury (Doury, 1976)

This diagnosis consists in the discretization of the atmospheric stability in two classes: Normal Diffusion (ND) and Low Diffusion (LD). ND corresponds to unstable and neutral situations. LD corresponds to stable situations. In addition, this method is based only on the vertical temperature gradient, which means that it does not take into account the turbulence of mechanical origin:

$$\frac{\partial T}{\partial z} \begin{cases} \leq -0.5(°C/100m) & : \text{ND} \\ > -0.5(°C/100m) & : \text{LD} \end{cases} \tag{A1}$$

Stability classes of Pasquill (Pasquill, 1961)

The Pasquill classes allows to discretize the atmospheric stability in six classes from A (very unstable, coded by 0) to F (very stable, coded by 5). In this article we have used two diagnostics of calculating of the Pasquill classes:

- **Method of Turner** (Turner, 1969): Based on the 10-m wind speed and the surface solar radiation downwards (during daytime) or total cloud cover (at night) (Table A1). This diagnostic has the advantage of taking into account the two origins of turbulence: Mechanical (wind) and thermal (solar radiation).

- **Method of temperature gradient** (Seinfeld and Pandis, 1998): Based on the temperature difference over 100m (Table A2). This diagnosis does not take into account the turbulence of mechanical origin, but it captures variations in stability conditions better than Doury diagnosis.

Atmospheric stability	Stability classes of Pasquill	$\partial T/\partial z$	Stability classes of Doury
Unstable	A	]$-\infty$, -1.9[	Normal Diffusion (ND)
Unstable	B	[-1.9, -1.7[	Normal Diffusion (ND)
Unstable	C	[-1.7, -1.5[	Normal Diffusion (ND)
Neutral	D	[-1.5, -0.5[	
Stable	E	[-0.5, 1.5[	Low Diffusion (LD)
Stable	F	[1.5, $+\infty$[	Low Diffusion (LD)

Table A2. Pasquill classes according to method of temperature gradient, and correspondence with the classes of Doury.

Appendix B: Formulas for Gaussian standard deviations

The aim is to calculate the evolution of standard deviations of the concentration distribution, which is given by:

$$\sigma(t+dt) = \sigma(t) + \frac{d\sigma}{dt}dt = \sigma(t) + \frac{d\sigma}{dx}U\,dt \tag{B1}$$

555 Where t and x are time and distance since the emission of the puff, respectively. U is the speed of advection of the puff.

Thus, The problem is deferred to the determination of $\frac{d\sigma}{dx}$ (or $\frac{d\sigma}{dt}$) in each time step t. For a given standard deviation law, we have: $\sigma(x) = f(x, \alpha_1, \alpha_2, ..., \alpha_n)$, where α_i are parameters that depend on the atmospheric stability, and are determined empirically. Then, $\frac{d\sigma}{dx}$ can be expressed as follows:

$$\begin{cases} \frac{d\sigma}{dx} = \frac{\partial f}{\partial x}(x, \alpha_1, \alpha_2, ..., \alpha_n) \\ x = f^{-1}(\sigma) \end{cases} \Leftrightarrow \frac{d\sigma}{dx} = \frac{\partial f}{\partial x}(f^{-1}(\sigma), \alpha_1, \alpha_2, ..., \alpha_n) \tag{B2}$$

560 A similar reasoning is possible for $\frac{d\sigma}{dt}$. Depending on the complexity of the function f, it will be more or less easy to express the function $\frac{d\sigma}{dx}$.

– **Pasquill laws**:

$$\sigma = ax^b + c \Rightarrow \frac{d\sigma}{dx} = ab\left(\frac{\sigma}{a}\right)^{\frac{b-1}{b}} \tag{B3}$$

The parameters a, b and c are determined according to the Pasquill stability classes.

565 – **Doury laws**:

$$\sigma = At^k t^{k-1} \Rightarrow \frac{d\sigma}{dx} = kA\sigma^{\frac{k-1}{k}} \tag{B4}$$

The parameters A and k are determined according to the Doury stability classes, and defined by step of time.

Code and data availability. The Arome and Arome-EPS data can be accessed by https://donneespubliques.meteofrance.fr/. IRSN data (radiological and meteorological) are available on demand. To get statistical calculation code and the plotting code please contact the corresponding author.

Author contributions. The project was conceptualized and supervised by MP and IK. Formal analysis and development of the calculation codes were carried out by YE. The radiological measurement campaign was conducted by OC. All the authors contributed to the discussion of the results and to writing the article.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors thank Orano RP for providing source term and environmental measurement data (^{85}Kr and meteorological measurements). The authors thank Emmanuel Quentric for his review, and Arnaud Quérel and Pierrick Cébron for discussions and suggestions on the use of statistical indicators.

References

Aliyu, A. S., Evangeliou, N., Mousseau, T. A., Wu, J., and Ramli, A. T.: An overview of current knowledge concerning the health and environmental consequences of the Fukushima Daiichi Nuclear Power Plant (FDNPP) accident, Environment international, 85, 213–228, https://doi.org/10.1016/j.envint.2015.09.020, 2015.

Bollhöfer, A., Schlosser, C., Schmid, S., Konrad, M., Purtschert, R., and Krais, R.: Half a century of Krypton-85 activity concentration measured in air over Central Europe: Trends and relevance for dating young groundwater, Journal of environmental radioactivity, 205, 7–16, https://doi.org/10.1016/j.jenvrad.2019.04.014, 2019.

Bouttier, F. and Raynaud, L.: Clustering and selection of boundary conditions for limited-area ensemble prediction, Quarterly Journal of the Royal Meteorological Society, 144, 2381–2391, https://doi.org/10.1002/qj.3304, 2018.

Bouttier, F., Vié, B., Nuissier, O., and Raynaud, L.: Impact of stochastic physics in a convection-permitting ensemble, Monthly Weather Review, 140, 3706–3721, https://doi.org/10.1175/MWR-D-12-00031.1, 2012.

Bouttier, F., Raynaud, L., Nuissier, O., and Ménétrier, B.: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX, Quarterly Journal of the Royal Meteorological Society, 142, 390–403, https://doi.org/10.1002/qj.2622, 2016.

Brousseau, P., Berre, L., Bouttier, F., and Desroziers, G.: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var, Quarterly Journal of the Royal Meteorological Society, 137, 409–422, https://doi.org/10.1002/qj.750, 2011.

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., and Picard, G.: On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model, The Cryosphere, 10, 1021–1038, https://doi.org/10.5194/tc-10-1021-2016, 2016.

Connan, O., Smith, K., Organo, C., Solier, L., Maro, D., and Hébert, D.: Comparison of RIMPUFF, HYSPLIT, ADMS atmospheric dispersion model outputs, using emergency response procedures, with 85Kr measurements made in the vicinity of nuclear reprocessing plant, Journal of Environmental Radioactivity, 124, 266–277, https://doi.org/10.1016/j.jenvrad.2013.06.004, 2013.

Connan, O., Solier, L., Hébert, D., Maro, D., Lamotte, M., Voiseux, C., Laguionie, P., Cazimajou, O., Le Cavelier, S., Godinot, C., et al.: Near-field krypton-85 measurements in stable meteorological conditions around the AREVA NC La Hague reprocessing plant: estimation of atmospheric transfer coefficients, Journal of environmental radioactivity, 137, 142–149, https://doi.org/10.1016/j.jenvrad.2014.07.012, 2014.

Courtier, P., Freydier, C., Geleyn, J.-F., Rabier, F., and Rochas, M.: The Arpege project at Meteo France, in: Seminar on Numerical Methods in Atmospheric Models, 9-13 September 1991, vol. II, pp. 193–232, ECMWF, ECMWF, Shinfield Park, Reading, https://www.ecmwf.int/node/8798, 1991.

De Meutter, P. and Delcloo, A.: Uncertainty quantification of atmospheric transport and dispersion modelling using ensembles for CTBT verification applications, Journal of environmental radioactivity, 250, https://doi.org/10.1016/j.jenvrad.2022.106918, 2022.

De Meutter, P., Camps, J., Delcloo, A., Deconninck, B., and Termonia, P.: On the capability to model the background and its uncertainty of CTBT-relevant radioxenon isotopes in Europe by using ensemble dispersion modeling, Journal of environmental radioactivity, 164, 280–290, https://doi.org/10.1016/j.jenvrad.2016.07.033, 2016.

Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., and Cébron, P.: PEARP, the Météo-France short-range ensemble prediction system, Quarterly Journal of the Royal Meteorological Society, 141, 1671–1685, https://doi.org/10.1002/qj.2469, 2015.

Doury, A.: Une méthode de calcul pratique et générale pour la prévision numérique des pollutions véhiculées par l'atmosphère, Tech. Rep. CEA-R-4270, CEA, 1976.

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., and Wotawa, G.: World Meteorological Organization's model simulations of the radionuclide dispersion and deposition from the Fukushima Daiichi nuclear power plant accident., Journal of Environmental Radioactivity, pp. 172—184, https://doi.org/10.1016/j.jenvrad.2013.09.014, 2015.

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R.: Why should ensemble spread match the RMSE of the ensemble mean?, Journal of Hydrometeorology, 15, 1708–1713, https://doi.org/10.1175/JHM-D-14-0008.1, 2014.

Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J., Bellasio, R., Buckley, R., Champion, H., Chino, M., R., D., Davakis, E., Eleveld, H., Glaab, H., Manning, A., Mikkelsen, T., Pechinger, U., Polreich, E., Prodanova, M., Slaper, H., Syrakov, D., Terada, H., Der Auwera, L., Valkama, I., and Zelazny, R.: Ensemble dispersion forecasting—Part II: application and evaluation., Atmospheric Environment, p. 4619–4632, https://doi.org/10.1016/j.atmosenv.2004.05.031, 2004a.

Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., R., D., Davakis, E., Eleveld, H., Geertsema, G., Glaab, H., Kollax, M., Ilvonen, M., Manning, A., Pechinger, U., Persson, C., Polreich, E., Potemski, S., Prodanova, M., Saltbones, J., Slaper, H., Sofiev, M., Syrakov, D., Sørensen, J., Der Auwera, L., Valkama, I., and Zelazny, R.: Ensemble dispersion forecasting—Part I: concept, approach and indicators., Atmospheric Environment, pp. 4607—-4617, https://doi.org/10.1016/j.atmosenv.2004.05.030, 2004b.

Girard, S., Korsakissok, I., and Mallet, V.: Screening sensitivity analysis of a radionuclides atmospheric dispersion model applied to the Fukushima disaster, Atmospheric Environment, 95, 490–500, https://doi.org/10.1016/j.atmosenv.2014.07.010, 2014.

Girard, S., Mallet, V., Korsakissok, I., and Mathieu, A.: Emulation and Sobol' sensitivity analysis of an atmospheric dispersion model applied to the Fukushima nuclear accident, Journal of Geophysical Research: Atmospheres, 121, 3484–3496, https://doi.org/10.1002/2015JD023993, 2016.

Girard, S., Armand, P., Duchenne, C., and Yalamas, T.: Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations, Atmospheric Environment, 224, 117 313, https://doi.org/10.1016/j.atmosenv.2020.117313, 2020.

Gurriaran, R., Maro, D., and Solier, L.: Etude de la dispersion atmosphérique en champ proche en cas de rejet en hauteur–étalonnage des appareils de mesure nucléaires. IPSN/Département de protection de l'environnement, Tech. rep., Rapport DPRE/SERNAT/2001-08, 2001.

Gurriaran, R., Maro, D., Bouisset, P., Hebert, D., Leclerc, G., Mekhlouche, D., Rozet, M., and Solier, L.: In situ metrology of 85Kr plumes released by the COGEMA La Hague nuclear reprocessing plant, Journal of environmental radioactivity, 72, 137–144, https://doi.org/10.1016/S0265-931X(03)00195-4, 2004.

Kajino, M., Sekiyama, T. T., Igarashi, Y., Katata, G., Sawada, M., Adachi, K., Zaizen, Y., Tsuruta, H., and Nakajima, T.: Deposition and dispersion of radio-cesium released due to the Fukushima nuclear accident: Sensitivity to meteorological models and physical modules, Journal of Geophysical Research: Atmospheres, 124, 1823–1845, https://doi.org/10.1029/2018JD028998, 2019.

Korsakissok, I., Mathieu, A., and Didier, D.: Atmospheric dispersion and ground deposition induced by the Fukushima Nuclear Power Plant accident: A local-scale simulation and sensitivity study, Atmospheric environment, 70, 267–279, https://doi.org/10.1016/j.atmosenv.2013.01.002, 2013.

Korsakissok, I., Contu, M., Connan, O., Mathieu, A., and Didier, D.: Validation of the Gaussian puff model pX using near-field krypton-85 measurements around the AREVA NC La Hague reprocessing plant: comparison of dispersion schemes, in: 17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes.(Budapest), 2016.

650   Korsakissok, I., Périllat, R., Andronopoulos, S., Bedwell, P., Berge, E., Charnock, T., Geertsema, G., Gering, F., Hamburger, T., Klein, H., et al.: Uncertainty propagation in atmospheric dispersion models for radiological emergencies in the pre-and early release phase: summary of case studies, Radioprotection, 55, S57–S68, 2020.

Le, N. B. T., Korsakissok, I., Mallet, V., Périllat, R., and Mathieu, A.: Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a Monte Carlo approach, applied to the Fukushima nuclear accident, Atmospheric Environment: X, 10,

655   100 112, https://doi.org/10.1016/j.aeaoa.2021.100112, 2021.

Leadbetter, S., Andronopoulos, S., Bedwell, P., Chevalier-Jabet, K., Geertsema, G., Gering, F., Hamburger, T., Jones, A., Klein, H., Korsakissok, I., et al.: Ranking uncertainties in atmospheric dispersion modelling following the accidental release of radioactive material, Radioprotection, 55, S51–S55, https://doi.org/10.1051./radiopro/2020012, 2020.

Leadbetter, S. J., Hort, M. C., Jones, A. R., Webster, H. N., and Draxler, R. R.: Sensitivity of the modelled deposition of Caesium-137 from

660   the Fukushima Dai-ichi nuclear power plant to the wet deposition parameterisation in NAME, Journal of environmental radioactivity, 139, 200–211, https://doi.org/10.1016/j.jenvrad.2014.03.018, 2015.

Leadbetter, S. J., Jones, A. R., and Hort, M. C.: Assessing the value meteorological ensembles add to dispersion modelling using hypothetical releases, Atmospheric Chemistry and Physics, 22, 577–596, https://doi.org/10.5194/acp-22-577-2022, 2022.

Leroy, C., Maro, D., Hébert, D., Solier, L., Rozet, M., Le Cavelier, S., and Connan, O.: A study of the atmospheric dispersion of a high

665   release of krypton-85 above a complex coastal terrain, comparison with the predictions of Gaussian models (Briggs, Doury, ADMS4), Journal of environmental radioactivity, 101, 937–944, https://doi.org/10.1016/j.jenvrad.2010.06.011, 2010.

Leutbecher, M. and Lang, S.: On the reliability of ensemble variance in subspaces defined by singular vectors, Quarterly Journal of the Royal Meteorological Society, 140, 1453–1466, 2014.

Mallet, V. and Sportisse, B.: Air quality modeling: From deterministic to stochastic approaches, Computers Mathematics with Applications,

670   55, 2329—2337, https://doi.org/10.1016/j.camwa.2007.11.004, 2008.

Manzato, A.: An odds ratio parameterization for ROC diagram and skill score indices, Weather and Forecasting, 20, 918–930, https://doi.org/10.1175/WAF899.1, 2005.

Manzato, A.: A note on the maximum Peirce skill score, Weather and Forecasting, 22, 1148–1154, https://doi.org/10.1175/WAF1041.1, 2007.

675   Maro, D., Crabol, B., Germain, P., Baron, Y., Hebert, D., and Bouisset, P.: A study of the near field atmospheric dispersion of emissions at height: comparison of Gaussian plume models (Doury, Pasquill-Briggs, Caire) with krypton 85 measurements taken around La Hague nuclear reprocessing plant, Radioprotection, 37, 1277–1282, 2002.

Maro, D., Chechiak, B., Tenailleau, L., Germain, P., Hebert, D., and Solier, L.: Analysis of experimental campaigns on atmospheric transfers around the AREVA NC spent nuclear fuel reprocessing plant at La Hague: comparison between operational models and measurements, in:

680   11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes.(Cambridge), 2007.

Mathieu, A., Korsakissok, I., Quélo, D., Groëll, J., Tombette, M., Didier, D., Quentric, E., Saunier, O., Benoit, J.-P., and Isnard, O.: Atmospheric dispersion and deposition of radionuclides from the Fukushima Daiichi nuclear power plant accident., Elements, pp. 195–200, https://doi.org/10.2113/gselements.8.3.195, 2012.

Nie, B., Fang, S., Jiang, M., Wang, L., Ni, M., Zheng, J., Yang, Z., and Li, F.: Anthropogenic tritium: Inventory, discharge, environmental

685   behavior and health effects, Renewable and Sustainable Energy Reviews, 135, 110 188, https://doi.org/10.1016/j.rser.2020.110188, 2021.

Olesen, H.: Model Validation Kit - Status and Outlook, in: 5th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes.(Rhodes), 1998.

Orano, l. H.: Rapport d'information du site Orano La Hague, Tech. rep., https://www.orano.group/docs/default-source/orano-doc/groupe/publications-reference/tsn-la-hague-2021.pdf?sfvrsn=7865d99c_8, 2021.

690 Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., and Weisheimer, A.: Stochastic parametrization and model uncertainty, Tech. Rep. 598, ECMWF Reading, UK, https://www.ecmwf.int/node/11577, 2009.

Pasquill, F.: The estimation of the dispersion of windborne material, Met. Mag., 90, 33, 1961.

Peirce, C. S.: The numerical measure of the success of predictions, Science, pp. 453–454, https://doi.org/10.1126/science.ns-4.93.453.b, 1884.

695 Quérel, A., Meddouni, K., Quélo, D., Doursout, T., and Chuzel, S.: Statistical approach to assess radon-222 long-range atmospheric transport modelling and its associated gamma dose rate peaks, Advances in Geosciences, 2022.

Rao, K. S.: Uncertainty analysis in atmospheric dispersion modeling, Pure and applied geophysics, 162, 1893–1917, https://doi.org/10.1007/s00024-005-2697-4, 2005.

Raynaud, L. and Bouttier, F.: Comparison of initial perturbation methods for ensemble prediction at convective scale, Quarterly Journal of
700 the Royal Meteorological Society, 142, 854–866, https://doi.org/10.1002/qj.2686, 2016.

Raynaud, L., Berre, L., and Desroziers, G.: Accounting for model error in the Météo-France ensemble data assimilation system, Quarterly Journal of the Royal Meteorological Society, 138, 249–262, https://doi.org/10.1002/qj.906, 2012.

Richardson, D.: Ensembles using multiple models and analyses, Quarterly Journal of the Royal Meteorological Society, 127, 1847–1864, https://doi.org/10.1002/qj.49712757519, 2001.

705 Sato, Y., Takigawa, M., Sekiyama, T. T., Kajino, M., Terada, H., Nagai, H., Kondo, H., Uchida, J., Goto, D., Quélo, D., Mathieu, A., Quérel, A., Fang, S., Morino, Y., von Schoenberg, P., Grahn, H., Brännström, N., Hirao, S., Tsuruta, H., Yamazawa, H., and Nakajima, T.: Model Intercomparison of Atmospheric 137Cs From the Fukushima Daiichi Nuclear Power Plant Accident: Simulations Based on Identical Input Data, Journal of Geophysical Research: Atmospheres, 123, 11,748–11,765, https://doi.org/10.1029/2018JD029144, 2018.

Saunier, O., Mathieu, A., Didier, D., Tombette, M., Quélo, D., Winiarek, V., and Bocquet, M.: An inverse modeling method to assess the
710 source term of the Fukushima Nuclear Power Plant accident using gamma dose rate observations, Atmospheric Chemistry and Physics, 13, 11 403–11 421, https://doi.org/10.5194/acp-13-11403-2013, 2013.

Saunier, O., Korsakissok, I., Didier, D., Doursout, T., and Mathieu, A.: Real-time use of inverse modeling techniques to assess the atmospheric accidental release of a nuclear power plant, Radioprotection, 55, 107–115, https://doi.org/10.1051/radiopro/2020044, 2020.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, Wiley-Interscience, p. 1326,
715 https://doi.org/10.1080/00139157.1999.10544295, 1998.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, Monthly Weather Review, 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Sørensen, J. H., Amstrup, B., Feddersen, H., Bartnicki, J., Klein, H., Simonsen, M., Lauritzen, B., Hoe, S., Israelson, C., and Lindgren, J.: Fukushima accident: uncertainty of atmospheric dispersion modelling (FAUNA), Tech. Rep. 360, Nordic Nuclear Safety Research,
720 http://www.nks.org/en/nks_reports/view_document.htm?id=111010213440189, 2016.

Sørensen, J. H., Amstrup, B., Bøvith, T., Feddersen, H., Gill, R., Sørensen, M., Vejen, F., Astrup, P., Davis, N.and Lauritzen, B., Hoe, S., Dyve, J., and Lindahl, P.: MEteorological uncertainty of ShOrt-range dispersion (MESO), Tech. Rep. 380, Nordic Nuclear Safety Research, https://www.nks.org/en/nks_reports/view_document.htm?id=111010214043891, 2017.

Sørensen, J. H., Schönfeldt, F., Sigg, R., Pehrsson, J., Lauritzen, B., Bartnicki, J., Klein, H., Cordt Hoe, S., and Lindgren, J.: Added Value of uncertainty Estimates of SOurce term and Meteorology (AVESOME), Tech. Rep. 420, Nordic Nuclear Safety Research, https://www.nks.org/en/nks_reports/view_document.htm?id=111010214696230, 2019.

Sørensen, J. H., Bartnicki, J., Buhr, A. M. B., Feddersen, H., Hoe, S. C., Israelson, C., Klein, H., Lauritzen, B., Lindgren, J., Schönfeldt, F., et al.: Uncertainties in atmospheric dispersion modelling during nuclear accidents, Journal of Environmental Radioactivity, 222, 106 356, 2020.

Soulhac, L. and Didier, D.: Projet pX, note de principe pX 1.0. Note technique IRSN, Tech. rep., DEI/SESUC/08-39 (available upon request at http://www.irsn.fr/EN/Contact), 2008.

Straume, A. G., Koffi, E. N., and Nodop, K.: Dispersion modeling using ensemble forecasts compared to ETEX measurements, Journal of Applied Meteorology, 37, 1444–1456, https://doi.org/10.1175/1520-0450(1998)037<1444:DMUEFC>2.0.CO;2, 1998.

Swets, J. A.: The Relative Operating Characteristic in Psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition., Science, 182, 990–1000, https://doi.org/10.1126/science.182.4116.990, 1973.

Tamura, Y., Suda, K., Sasaki, A., Iwatani, Y., Fujii, K., Ishibashi, R., and Hibi, K.: Simultaneous measurements of wind speed profiles at two sites using Doppler sodars, Journal of wind engineering and industrial aerodynamics, 89, 325–335, https://doi.org/10.1016/S0167-6105(00)00085-4, 2001.

Tennant, W. and Beare, S.: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS, Quarterly Journal of the Royal Meteorological Society, 140, 1150–1160, https://doi.org/10.1002/qj.2202, 2014.

Tombette, M., Quentric, E., Quélo, D., Benoit, J., Mathieu, A., Korsakissok, I., and Didier, D.: C3X: A software platform for assessing the consequences of an accidental release of radioactivity into the atmosphere, in: International Radiation Protection Association congress, Geneva., 2014.

Turner, D. B.: Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling, CRC press, 1969.

Werner, C.: Doppler wind lidar, in: Lidar, pp. 325–354, Springer, 2005.

Wilks, D. S.: Statistical methods in the atmospheric sciences (Fourth Edition), Elsevier, 2019.

Wu, S., Sun, K., Dai, G., Wang, X., Liu, X., Liu, B., Song, X., Reitebuch, O., Li, R., Yin, J., et al.: Inter-comparison of wind measurements in the atmospheric boundary layer and the lower troposphere with Aeolus and a ground-based coherent Doppler lidar network over China, Atmospheric Measurement Techniques, 15, 131–148, https://doi.org/10.5194/amt-15-131-2022, 2022.