Acknowledgements

We thank the reviewers very much for their constructive comments which helped to improve the quality of the paper. In this letter below, we answer to all comments and explain how they have been addressed in the revised manuscript. We hope that this new version may be accepted for publication in Atmospheric Chemistry and Physics.

General Comments reviewer 1

The authors demonstrate the value of ensemble meteorology by showing how it can be used to model the uncertainty in the dispersion of material released from a known source. Where previous studies have focussed on long-range dispersion and used meteorology from global NWP models this study examines the use of a high-resolution (2.5km) ensemble NWP model covering a limited area to provide meteorological input to a dispersion model. In addition, the study focuses on modelling the dispersion of material that is regularly discharged from a reprocessing plant and compares the model results to observations over a period of two months. This field campaign along with the meteorological model and the dispersion model are clearly described within the paper.

My main complaint about this paper is that, for me, it covers too many topics. This has two impacts, first I am distracted from the main results and second the secondary topics are not covered in great detail so I am left with too many questions, as can be seen by the length of the specific comments section. There are two main sections which take my attention away from the main results. The first is the work looking at different methods to compute stability. The second is the consideration of how to model dispersion over time period which are longer than a single meteorological forecast. Both of these are interesting topics but, I feel they would be better placed in separate papers where they can be discussed more fully. I have included my specific comments on both of these sections in the specific comments.

<u>General Comments – answer from authors</u>

In the revised version of the paper, these comments have been addressed in two ways:

1/ less emphasis has been made on the comparison between the Gaussian standard deviation formulas (Pasquill vs. Doury). In order to make the reader less distracted, and given that the main objective of the paper is the evaluation of the ensemble predictions, the section on statistical results (section 4.3.2) has been lightened by focusing only on the Pasquill method.

2/ the explanation of how and why we model dispersion over a time period which are longer than a single meteorological forecast has been improved. We believe it is an important aspect of the simulation set-up, that is addressed in a short section of the manuscript.

Both aspects are covered in detail in the specific comments, and we hoped that this addresses the reviewers' concerns.

Specific Comments

• In the Introduction many papers are mentioned and in some cases the work carried out is described. It would also be helpful to understand the results or outcomes of the work in those papers. For example, the authors note that evaluations of dispersion ensembles were performed by Le et al, (2021) and De Meutter and Delcloo (2022) but they don't say whether the ensembles were found to perform well or whether the use

of ensembles provided more information. Similarly the authors mention the works of Galmarini et al,. (2004a and b) in performing multi-model ensembles but do not say anything about the findings of those works.

Changes made in the text, Line 59 :

In Le et al. (2021) an 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*.

Changes made in the text, Line 74 :

This approach was extensively investigated in Galmarini et al. (2004a, b) 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 showed that the ensemble simulations allows to reduce the uncertainty related to the deterministic simulation.

• Line 50: I'm not sure "coarse" is the appropriate word to use here as one of the studies referenced in the previous paragraph used meteorological data at a resolution of 2.5km which is not generally considered to be a coarse resolution.

Changes made in the text, Line 51 :

All these studies were carried out at long distance and the ensembles used to represent weather uncertainties had coarse spatial and temporal resolution, <u>except Leadbetter et al. (2022) who used</u> <u>also fine-scale weather ensembles with a horizontal resolution of about 2.5 x 2.5 km and 70 vertical levels.</u>

• Line 82: Is it possible to define "reasonable" in reference to the ⁸⁵Kr release? Is the error on the release rate known?

The quantity of 85Kr released to the atmosphere is measured by the operator with a temporal accuracy of 10 minutes (confidential data) and an uncertainty of about 10% on the activity measured (calculated by the difference between the data obtained on 2 measurement channel during the period of release).

Changes made in the text, Line 86 :

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 accuracy (described in section* 2.2).

Paragraph added, Lines from 148 to 151 :

The activity in 85Kr 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 channels 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 ⁸⁵Kr varied from 294 to 379 PBq/year (Orano, 2021).

• Line 115: Is it possible to provide an approximate activity concentration for the amount of ⁸⁵Kr naturally present in the environment or a ratio of the ⁸⁵Kr present in the environment to the amount of ⁸⁵Kr released by the reprocessing?

Changes made in the text Lines from 120 to 124:

Background levels of 85Kr in the atmosphere, excluding an industrial plume, are currently below 2 Bq.m-3 (Bollhofer et al, 2019). In nearby fields in the plume around the RP of La Hague (about 0-2 km), activities can reach 100,000 Bq/m3 (Connan et al 2014). At distances of the order of 20 km, the maximum measurable activities are generally less than 10000 Bq/m3 and beyond a few tens of km of RP, the activities in 85Kr are too low to be measurable in real time (Connan et al 2013).

• Line 117: Similar to line 82; is it possible to define "reasonable" in reference to the ⁸⁵Kr release? Is the error on the release rate known?

See comments above

• Line 125: In describing the terrain around La Hague as complex is it possible to provide values for the maximum and minimum elevations to provide meteorological readers with a reference point for how the terrain might affect the wind speed and direction?

Changes made in the text, Line 133 :

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).



• Line 136 and 137: For me the availability of data at a 10-minute resolution doesn't, on its own, constitute an accurate and reliable source term. I would be interested to know the uncertainty on the measurements relative to the amount of material released.

The following sentence is deleted: Line 136 and 137 [*The sum of the amounts of 85Kr released from UP2 and UP3 units, over regular 10 minutes.... and reliable source term.*] Measurement uncertainty relative to source term was added in Line 149 (see comment above).

• Table 2: Would it possible to add the temporal resolution of the met data to the table? I think this is mentioned later on in the text but it would be helpful to include it in this table too.

Changes made in the table 2.

• Section 3.1: There are a large number of different skill scores which could be used for the verification of both deterministic and ensemble predictions. Would it be possible for the authors to include an explanation of why bias and spread-skill were chosen?

The choice is based on subsequent work by the meteorological community. The most commonly used scores for evaluating the reliability of ensembles (the ability of the meteorological ensembles to represent realistic uncertainties) are the spread-skill ratio and rank diagrams (not shown in the paper), which are complementary scores. On the other hand, the bias allows to identify the systematic errors of the weather predictions, as explained in the text.

Changes made in the text, Line 267:

For this purpose, two common scores, among others, used by the meteorological community for the evaluation of ensemble reliability, have been calculated based on the observations of 3D-wind speed and direction...

• Figure 3: In the text the authors mention that there is a diurnal cycle in the bias, but I find this difficult to see because the bias shares the same axis as the mean values. Would it be possible to place the bias on a separate axis to the mean values?

Changes made in the Figure 3.

• Section 4.1.1: I am very surprised that it is necessary to use more than one 24-hour forecast for this study. The furthest observation point is situated <20km from the release location and assuming direct transport it would only take more than 24 hours to travel this distance if the mean winds for the whole 24 hours were less than 0.8m/s. In addition, 9 further hours of the first forecast were also still available so it would be possible to carry out a 36-hour forecast without needing to combine meteorological data from different days.

The release of krypton-85 from the source is almost permanent. Thus, the dispersion calculation is done over long continuous periods where the plume goes over the measurement sites (c.f. Table 5). In order to have continuous weather forecasts that cover the whole calculation period, several forecasts starting from different initial instants must be combined successively. Different choices can be done to combine weather forecasts. The method used in this paper is to cover every day by a single forecast by taking into account real-time running constrains, since operational warning is the final purpose of such study. Some details are provided in the two comments below. The section (4.1.1) that described this aspect has been modified to better explain the choice done in the study.

• Line 319: I am curious to understand why the first 8 forecast hours were skipped? Is this a recommendation of the developers of AROME-EPS or is it due to the location of the release relative to the edge of the meteorological model domain?

In addition to the explanations given in the previous comment, the first 8 forecast hours are skipped to take into account the availability and transfer time of AROME-EPS data, which take on average

about 6 hours to be available from the beginning of the run. Thus, to get closer to an operational situation, it is interesting to choose the most "recent and available" forecast to the beginning of a day D (00h), which is the forecast of 1500 UTC of the day D-1.

• Line 320: In table 2, the authors state that AROME-EPS is run four times a day, so I was wondering why model runs which are 24 hours apart are stitched together to build a continuous time series rather than model runs which are only 6 hours apart. My feeling is that using model runs which are 6 hours apart would reduce jumps at the forecast joins.

In general, forecasts are not available immediately, they take on average about 6 hours to be available since the beginning of the run. Thus, to cover a day D, taking into account the accidental context, it is interesting to choose the most "recent and available" forecast the most which is used to anticipate the next 24 hours, which is the period when decision-making is required. Moreover, recently we tested the method using the 4 forecasts with a change of the ensembles every 6 hours, and the results were not better than the method used in the paper.

• Line 333: Were the comparisons made in section 3 carried out using the unprocessed or processed meteorological fields?

The meteorological evaluation was done on unprocessed data. Here, the word "processing" refers to the process of projecting and interpolating the meteorological data (which are defined on a lon/lat regular grid) onto a Cartesian grid so that they fit the form readable by the dispersion model used.

• Line 337 and 338: Are the authors able to comment on the impact of setting the minimum ABL height to 200m and/or provide evidence that this is a reasonable minimum ABL for the study area?

In addition to the explanations given in the text about the ABL from AROME, the time series of the ABL from AROME-EPS (c.f. figure below) confirms that there are times when it reaches unrealistic levels down to below 10 m. However, values below 200 meters are reached only a few times within the two-months period of interest, which means that the threshold value of 200 m should not significantly alter the simulations. This parameter is often not very influential on the pX simulations at short distance, because it is only used in cases where there are reflections on the inversion layer (not in stable situations), and only if the plume is sufficiently developed on the vertical. Therefore, this threshold is only set to ensure that there are no cases where the release is above the ABL, because it would then be considered in the pX code that the ground concentration is zero.



• In Figure 7 and 8 I find it difficult to determine where the peaks in the ensemble are as the grey lines overlap a lot. Would it be possible to plot the ensemble as a shaded area rather than individual grey lines?

Changes made in the Figure 7 and 8.

• Line 362. The authors state that the use of a stack height of 100m does not allow them to accurately predict concentrations at 2km from the source in stable situations. Please could they expand on why stable conditions are problematic for the dispersion model they are using.

Stable situations were found to be more tricky to reproduce in previous dispersion studies in this area (Connan et al, 2013, Korsakissok et al, 2016), due to the difficulty of the model to accurately simulate vertical plume spread and the fact that building downwash effects are not taken into account.

Changes made in the text, Line 376:

In our simulations, the use of the stack height (100m) as release height does not allow to accurately predict significant ground concentrations at this distance, <u>due to approximations made in the</u> <u>Gaussian model that does not include building downwash effects.</u> This is especially <u>the case when</u> <u>using</u> Doury standard deviations, ...

• Line 365: What does the word "this" refer to in the sentence which begins in this line?

Changes made in the text Line 378:

This phenomenon, that characterizes pX-Doury simulations in stable situations, was specifically shown in the case of La Hague RP.

• Line 375, figure 9 and table 4: Reading table 4 I think that peaks 2, 3 and 4 are much smaller than peaks 1 and 5. For me it would be helpful for this to mentioned in the text.

Changes made in the text Line 390:

Table 4 summarizes the five 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.

• Line 381: I think, in this sentence, the authors are arguing that the peaks are small because they are located close to the edge of the plume where the concentration gradients are high. It would be helpful to see a figure showing this. In addition, the authors appear to be suggesting that the solution to the underprediction is simply to increase the width of the plume which could be done by changing the stability category. Firstly, I would be interested to see why the authors believe that the inability of the model to predict the peaks is due to the stability and not to the wind speeds and directions along the path the puff has taken from the source location. Secondly, increasing the spread of the plume may help the model to capture the peaks where they are located at the edge of the plume, but this will be at the expense of the magnitude of the peaks where they are located at or close to the centre of the plume. Finally, given the emphasis placed on the stability within the second half of the paper

I would be interested to see comment in the first half of the paper on the meteorological variables which impact on the calculation of the stability.

The requested figure is added (Fig. 11), and the following sentence is added in Line 397:

Figure 11 illustrates this issue in the case of the 3rd 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.

Firstly, it is assumed that the model failures are probably more related to the stability than to the wind, because we find that the wind forecasts are sufficiently accurate, and that the wind direction values given by the different members are very similar while the stability is more variable (Fig.9 and Fig.10).

We agree with the reviewer's statement that, if the stability diagnosis leads to an increase the spread of the plume, the intensity of the peaks detected near the plume's centreline will decrease. Although we have a fair spatial coverage of the area, our network density does not allow us to determine whether a sensor located within the plume would be simulated with less accuracy, should the plume spread increase in this particular case.

Finally, in the evaluation of the meteorological ensembles, particular attention was paid to wind data, since an extensive database of wind measurements is available. On the other hand, for the temperature (which is a key parameter in the stability calculation), we have no *in situ* measurements to justify the temperature gradient on the vertical in the study area. Thus, we think it is all the more interesting to highlight the importance of the stability diagnosis since it is difficult to evaluate *a priori* the ensemble's quality on this respect – this leaves room for future work on this subject.

• Line 403: Within the literature there are a number of different techniques proposed for the assessment of the performance of ensembles. Would it be possible for the authors to briefly explain why they selected the method of Querel, 2022 which is designed for the assessment of deterministic simulations?

In our case, we are particularly interested in the evaluation of the capacity of the ensembles to anticipate the exceedance of a given threshold. In this case, there are no other scores that can be used other than the scores based on the contingency table, which are extensively described in Wilks, 2006. The choice of some of the many scores based on these tables depends on the purpose of the study, and the particular features of the ensemble that we wish to evaluate. In the case of deterministic simulations, the method presented in Quérel, 2022 is the most suitable for our application, except that it is not possible to implement it in the case of ensemble simulations for the reasons described in the paper (section 4.3.1). Thus, we chose to use similar scores, but within a framework adapted to probabilistic forecast. To our knowledge, there is no other work in the literature that proposes other methods to evaluate the atmospheric dispersion ensembles in terms of occurrence/non-occurrence of threshold exceedance.

• Line 462, 463: Can I just check that the statement made on these two lines refers to the assessment carried out with parameters $\Delta T=3h$, $\tau=2h$?

Yes.

• Lines 523-525: Suggest removing this paragraph or expanding it substantially. Clustering has been tried with dispersion ensembles (Klonner, 2013) and was not found to be useful with the boundary layer.

The paragraph was removed.

Technical Corrections

Line 13: Replace "As first step" with "As a first step"

Changes made in the text

Line19: Replace "than deterministic one" with "than the deterministic one"

Changes made in the text

Line 60: "demonstrate" rather than "examine"?

Changes made in the text

Line 60: "skillfully" rather than "skillful"

Changes made in the text

Line 146: Suggest replacing "which means it does not generate chemical or physical reactions" with "it is not chemically or physically reactive".

Changes made in the text

Line 166: Suggest adding "(in the absence of deposition)" between "shown that" and "3-D wind field".

Changes made in the text

Line 281: The range -0.2 to 1.75 m/s doesn't appear to match the range in Figure 3.

Changes made in the text : <u>0.71 to 1.45 m/s.</u>

Line 285: +10 and -15 don't appear to match the minimum and maximum values in Figure 3.

Changes made in the text : <u>-13.3 and 7.8 °</u>.

Figure 4: In the y-axis labels what do the "dd" and "ff" mean?

Explanation added in the caption of Figure 4: "ff" and "dd" means the wind speed and direction, respectively.

Table 3: Please could the authors separate the bias and spread-skill columns and place separate wind direction and wind speed labels above them as I'm not sure what each column represents.

Changes made in Table 3.

Line 399: I think this reference should be Wilks, 2006 not Daniel and Wilks, 2006.

Changes made in the reference.

Line 406: Replace "because the" with "because there"

Changes made in the text.

Figure 10: Would it be possible to explain the meaning of "dd" and "ff" in the figure caption?

Explanation added in the caption of Figure 10.

Line 419: For clarity I suggest using the same language here as in the definition of "TN" in line 421

Changes made in the text.

Line 592: As mentioned previously I thin this reference should be Wilks, 2006 not Daniel and Wilks, 2006.

Changes made in the reference.

Line 653: Leadbetter, Jones and Hort has now been published and can be found here: https://acp.copernicus.org/articles/22/577/2022/

Changes made in the reference.

General Comments reviewer 2

The paper describes a probabilistic approach to study effects of meteorological uncertainties on atmospheric dispersion prediction at a scale of 2–20 km from source. A case study is performed using data of a two-month measurement campaign of the noble gas Kr-85 released from a reprocessing plant. These data are employed to evaluate a dispersion model driven by results of a high-resolution numerical weather prediction (NWP) model run in ensemble mode. The results of the study emphasize the value of introducing a probabilistic approach in dispersion modelling as compared to deterministic modelling. For the evaluation, two probabilistic scores are used, and for the dispersion modelling, two stability classifications are employed, and results are compared. It could be argued that the study on the two stability formulations is outside the main focus of the paper; however, I feel that it is still interesting to compare the results based on them.

The paper is well written, relevant and interesting both from a scientific and an application point of view

General Comments – answer from authors

The authors demonstrate the value of ensemble meteorology by showing how it can be used to model the uncertainty in the dispersion of material released from a known source. Where previous studies have focussed on long-range dispersion and used meteorology from global NWP models this study examines the use of a high-resolution (2.5km) ensemble NWP model covering a limited area to provide meteorological input to a dispersion model. In addition, the study focuses on modelling the dispersion of material that is regularly discharged from a reprocessing plant and compares the model results to observations over a period of two months. This field campaign along with the meteorological model and the dispersion model are clearly described within the paper.

My main complaint about this paper is that, for me, it covers too many topics. This has two impacts, first I am distracted from the main results and second the secondary topics are not covered in great detail so I am left with too many questions, as can be seen by the length of the specific comments section. There are two main sections which take my attention away from the main results. The first is the work looking at different methods to compute stability. The second is the consideration of how to model dispersion over time period which are longer than a single meteorological forecast. Both of these are interesting topics but, I feel they would be better placed in separate papers where they can be discussed more fully. I have included my specific comments on both of these sections in the specific comments.

Specific comments

I think the abstract needs to be rephrased. In general, the standard of English language in the paper is good; however, this does not apply fully to the abstract. In addition, certain parts of the abstract are incomprehensible unless one has in fact read the paper, and thus the abstract does not comply with the intention that an abstract should be self-explanatory. As an example, the abstract contains the following sentence: "The results show that the stability diagnostics of Pasquill provides better dispersion simulations." Better than what? Furthermore: "In addition, the ensemble dispersion performs better than deterministic one, and the optimum decision threshold (PSS maximum) is 3 members." Members of what? Please rewrite the abstract to ensure that it is self-consistent.

The abstract was rewritten according to the reviewers' comments. It is now self-consistent.

• At a few places, reference is made to the work by Galmarini et al. using a multi-model approach. A brief discussion would be in place on the difference between using such approach and the probably more systematic approach constructing a dispersion model ensemble by using an NWP model ensemble.

The following sentence is added in the text, Line 78:

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 section 1.1 Uncertainties and ensemble simulations, reference is given to earlier work on the use of ensemble techniques for atmospheric dispersion modelling including the work by Sørensen et al. (2016, 2017 and 2019). It would be appropriate, e.g. in lines 33 and 48, to include also, or as appropriate to replace by, the paper:

Sørensen, J.H., Bartnicki, J., Blixt Buhr, A.M., Feddersen, H., Hoe, S.C., Israelson, C., Klein, H., Lauritzen, B., Lindgren, J., Schönfeldt, F., Sigg, R. Uncertainties in atmospheric dispersion modelling during nuclear accidents. J. Environ. Radioact. 222 (2020) 1-10. https://doi.org/10.1016/j.jenvrad.2020.106356

Changes made in the reference.

• In section 2.1 Case study, lines 116-117, it is mentioned that the release rate of Kr-85 is known with good accuracy. Please elaborate on this. What was the actual release rate, how was it measured, and how was the associated uncertainty estimated?

Clarification was provided in the text in 1.2 and 2.1 paragraph.

The quantity of 85Kr released from the stacks was provided by RP at a measurement time of 10min. These data are confidential and cannot be described in detail in the paper. The monitoring and control of the procedures leading to the provision of release data for the 2 stacks of the 2 plants in operation are subject to checks by the French authorities (ASN; Nuclear Safety Authority), and only annual aggregate data are public in the plant's annual environmental reports.

What it is possible to specify to you:

The release varies according to the industrial activity according to the years related to the tonnage of reprocessed fuels: so from 2019 to 2021 (Orano, 2021), the annual quantity of 85Kr released was between 294 and 379 PBq/year, which leads on average to activities in average of the order of 9.3 10^9 to 1.2 10^{10} Bq/s. However, it should be noted that the release is intermittent even during periods of activity and can be completely stopped for several weeks. Having the data "in real time" at a 10-minute time step therefore justifies the term in the text "release with good accuracy".

In terms of uncertainty, the difference between a period without release (with the plant in operation) and a period with release is extremely clear, with a ratio of about 2 orders of magnitude on average (factor 100 to 150). It will be possible to go from value of the order of $3 \ 10^8$ to $3 \ 10^{10}$ Bq/s in a few minutes for a stack for example between 2 periods of fuel shearing. For each stack in each unit, 2 measurement channels are in place. The mean error between the two measurement channels varies from 7 to 10% for both units when 85 Kr is detected.

If the plant is completely shut down for several weeks (maintenance, failure), the release will be measured at even lower activity of the order of 10^6 Bq/s.

• In the first paragraph of section 4.1.1, a way to build a continuous time series of NWP model data from consecutive forecast series is described involving skipping the first eight hours of a forecast series. However, I fail to see the point in the proposed method. In my understanding, modern data assimilation techniques ensure that NWP models are initialized very well and thus consistent also at short forecast lengths. I encourage the authors to argue for their method.

The first 8 forecast hours were skipped to take into account the availability and transfer time of AROME-EPS data, wich take on average about 6 hours to be available since the run. Thus, to get closer to an operational situation, it is interesting to choose the most "recent and available" forecast, which is the one of 1500 UTC of the day D-1.

Changes made in the text, Line 330:

In other words, to simulate a release occurring <u>from 00h to 23h of</u> a day D, the AROME-EPS forecasts starting from 1500 UTC of the day before (D-1) are used.

• In lines 335 and 336, the method used to diagnose the ABL height is mentioned supplemented by imposing a minimum of 200 m. However, no reference is given. Please, add a reference or elaborate on the method.

In addition to the explanations given in the text about the ABL from AROME, the time series of the ABL from AROME-EPS (c.f. figure below) confirms that there are times when it reaches unrealistic levels down to below 10 m. However, values below 200 meters are reached only a few times within the two-months period of interest, which means that the threshold value of 200 m should not significantly alter the simulations. This parameter is often not very influential on the pX simulations at short distance, because it is only used in cases where there are reflections on the inversion layer (not in stable situations), and only if the plume is sufficiently developed on the vertical. Therefore, this threshold is only set to ensure that there are no cases where the release is above the ABL, because it would then be considered in the pX code that the ground concentration is zero. There is no reference in the literature to justify the chosen value (200m). It is a usual threshold used operationally at IRSN for the above considerations. This value was found to be consistent with AROME data (see figure below).



• In Figs. 9, 10 and 11 appear a number of abbreviations, e.g. pc_mb1, stab_mb1, ..., dd_mb3, ff_obs, ..., mb3_stability5. Please explain these in figure captions.

Clarifications was provided in the captions of Figures 9, 10 and 12:

• In section 5. Conclusions and perspectives, line 498, is mentioned: "(...) allow them to correctly represent the uncertainties within ABL". Please elaborate on this. What is meant by "correctly represent"?

Changes made in the text, Lines from 510 to 513:

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.

• In line 549, a mathematical equivalence is presented introducing a new mathematical function φ. This seems unnecessary to me. Please rephrase.

Changes made in the text, Line 560 and Equation B2.