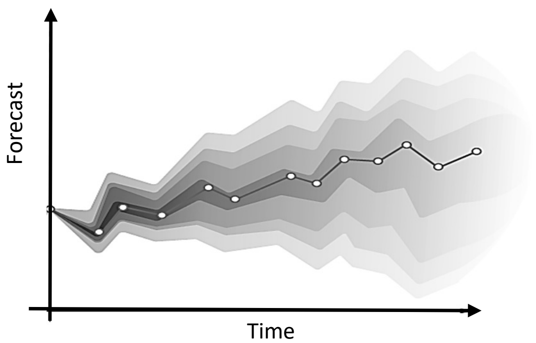
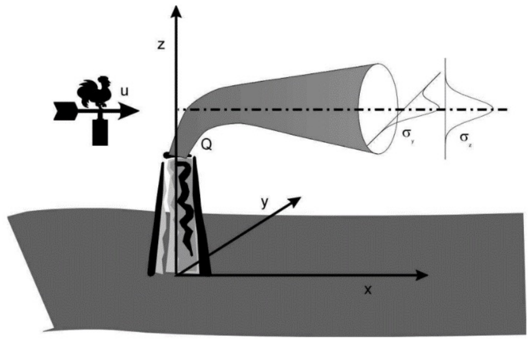
European Radiation Protection Week - 2023

Clustering and selection of relevant meteorological scenarios for short-range atmospheric dispersion during nuclear accident

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





1 : University of Toulouse, CNRS, CNRM-Météo-France, Toulouse, France.
2 : IRSN-BMCA, Fontenay-aux-Roses, France.
3: IRSN-LRC, Cherbourg-en-Cotentin, France.

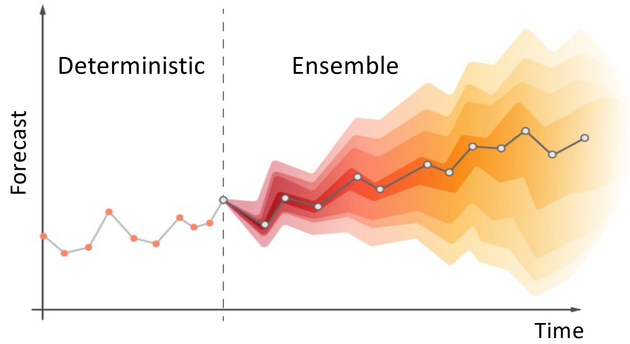
9th-13th October 2023 | UCD, Dublin

Context and motivations

Atmospheric dispersion models and nuclear emergency management

- Atmospheric dispersion models are used to predict the radiological consequences of the nuclear accidents.

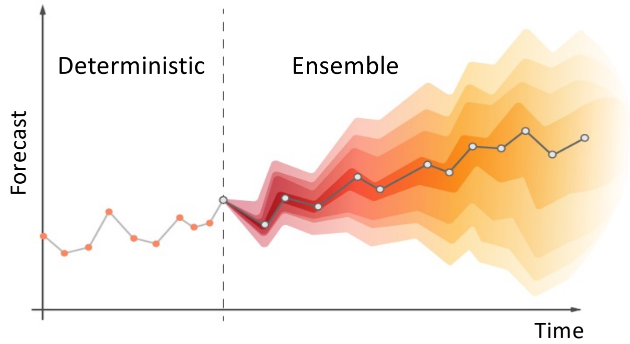
- The use of fine-scale «probabilistic» meteorological forecasts instead of a single deterministic forecast improves atmospheric dispersion forecasting (El-Ouartassy et al., 2022).



Context and motivations

Atmospheric dispersion models and nuclear emergency management

- Atmospheric dispersion models are used to predict the radiological consequences of the nuclear accidents.
- The use of fine-scale ensemble («probabilistic») meteorological forecasts instead of a single deterministic forecast improves atmospheric dispersion forecasting (Leadbetter et al., 2020; El-Ouartassy et al., 2022).



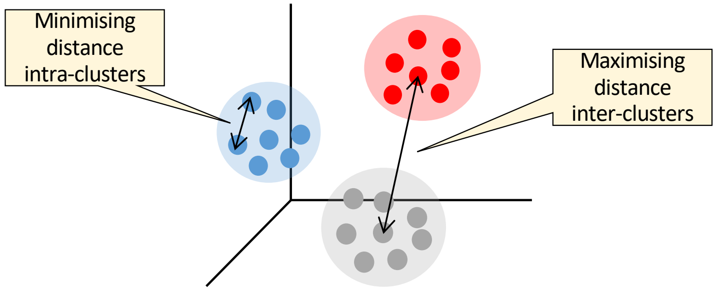
- The use of a high-resolution ensemble approach in an accidental context requires the optimization of the calculation time: data transfer and processing + dispersion calculation.

❖ How can we reduce the number of meteorological members used at the input of dispersion models while retaining some statistical properties of the complete ensemble?

Clustering

Definition and principle

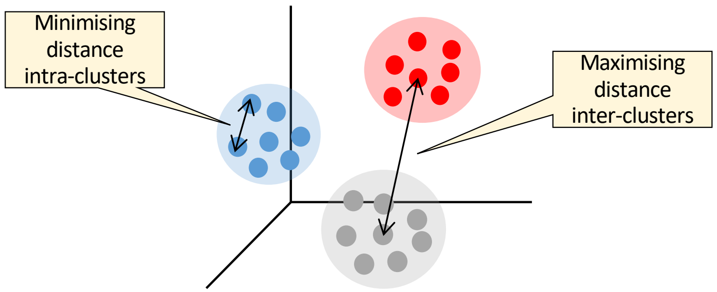
- A set of machine learning algorithms designed to identify homogeneous groups within a population.
- Given a set of points (the meteorological members, in our case) and a similarity metric defined between them, find a number of clusters (classes, groups, segments) such that :
  - The points within the same cluster are very similar to each other,
  - Points belonging to different groups are very dissimilar.



Clustering

Definition and principle

- ❑ A set of machine learning algorithms designed to identify homogeneous groups within a population.
- ❑ Given a set of points (the meteorological members, in our case) and a similarity metric defined between them, find a number of clusters (classes, groups, segments) such that :
    - The points within the same cluster are very similar to each other,
    - Points belonging to different groups are very dissimilar.



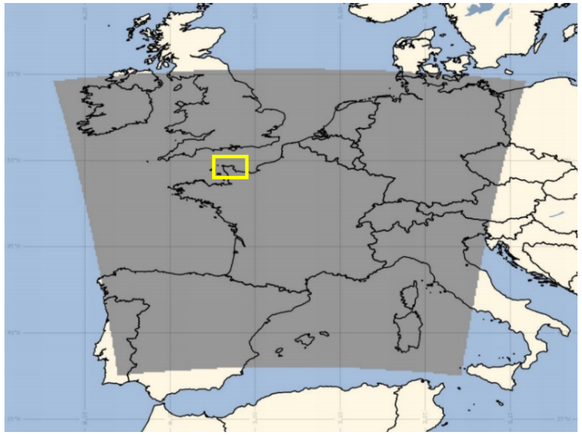
Issues

- ❑ What is the nature of the data (binary, qualitative, numerical, etc.)?
- ❑ What is the appropriate metric for measuring similarity (Euclidean, DTW, Wasserstein, etc.)?
- ❑ How many clusters can be identified in the data set (K=1, 2, 3,...)?
- ❑ Which learning algorithm?
- ❑ Comparison of different clustering results ...

Case study

AROME Esemble Prediction System (AROME-EPS, Météo-France)

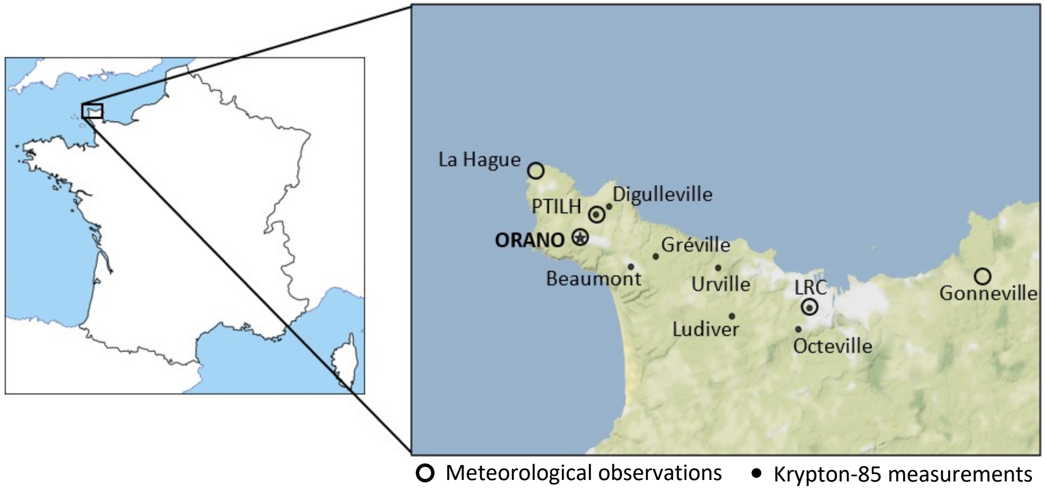
- 16 members at a horizontal resolution of 2.5 km.
- 25 vertical levels [10 – 3000m].
- Hourly forecasts.



AROME domain and the study area

Study area: La Hague experimental site

- Regular release of ^{85}Kr, which is a good tracer of atmospheric dispersion (no deposition, $\tau_{1/2}$=10.7 years).
- Well known Source Term (Orano La Hague).
- The AROME-EPS-pX dispersion ensembles have already been validated in this area during the period [Dec. 2020 – Jan. 2021] : (DISKRYNOC project, El-Ouartassy et al.,2022).



Methodology

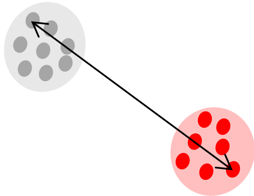
Number of clusters

- The number of clusters chosen in this work is $K = 4$.

Clustering algorithms used :

- ***Agglomerative clustering:***
  - Complete Linkage: The similarity between two clusters is the distance between their most distant individuals.
  - Ward : The similarity between two clusters is the variance of their union.
- **Partitional clustering:**
  - K-means : Based on the minimization of a cost function..

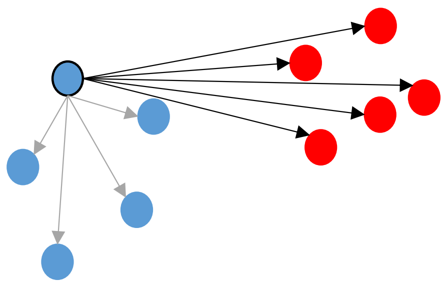
Complete Linkage



Representative member of each cluster:

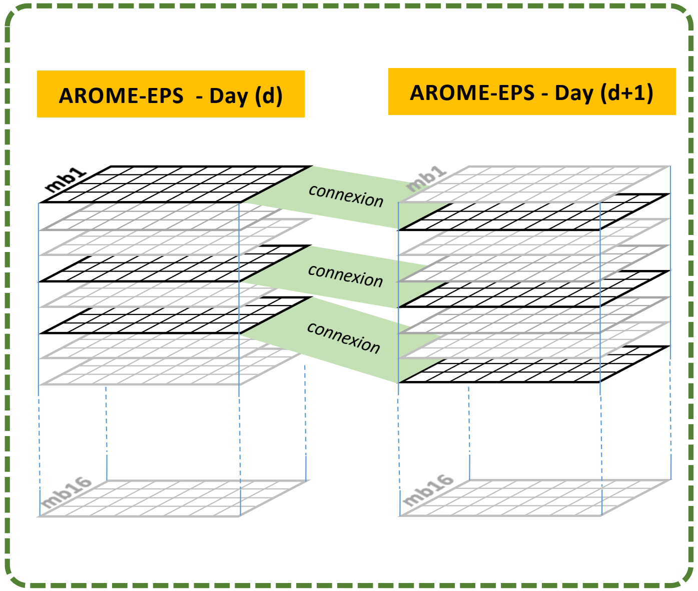
- the member who minimizes the representativeness index : $\mathrm{I} = \frac{a}{b}$

  where : a is the average distance of the member from members of the same cluster,
  b is the average distance of the member from members of other clusters.



Methodology

Overview

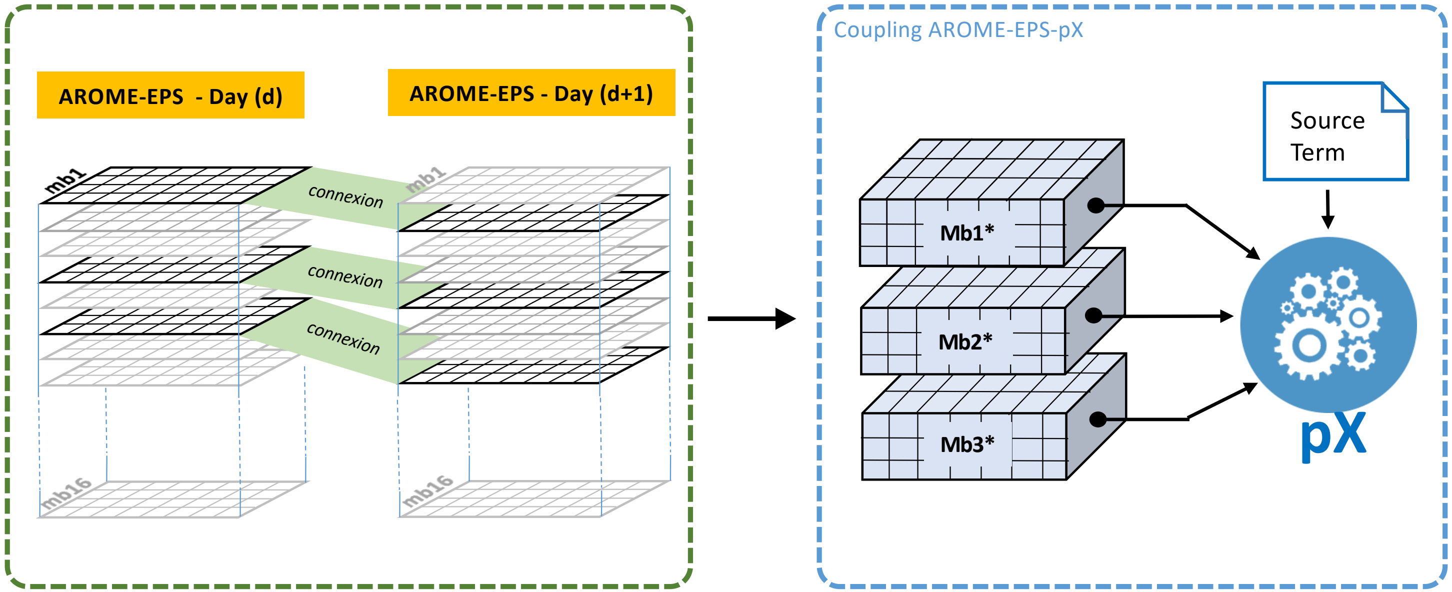


Clustering and calculation of the representative member

1

Methodology

Overview



Clustering and calculation of the representative member

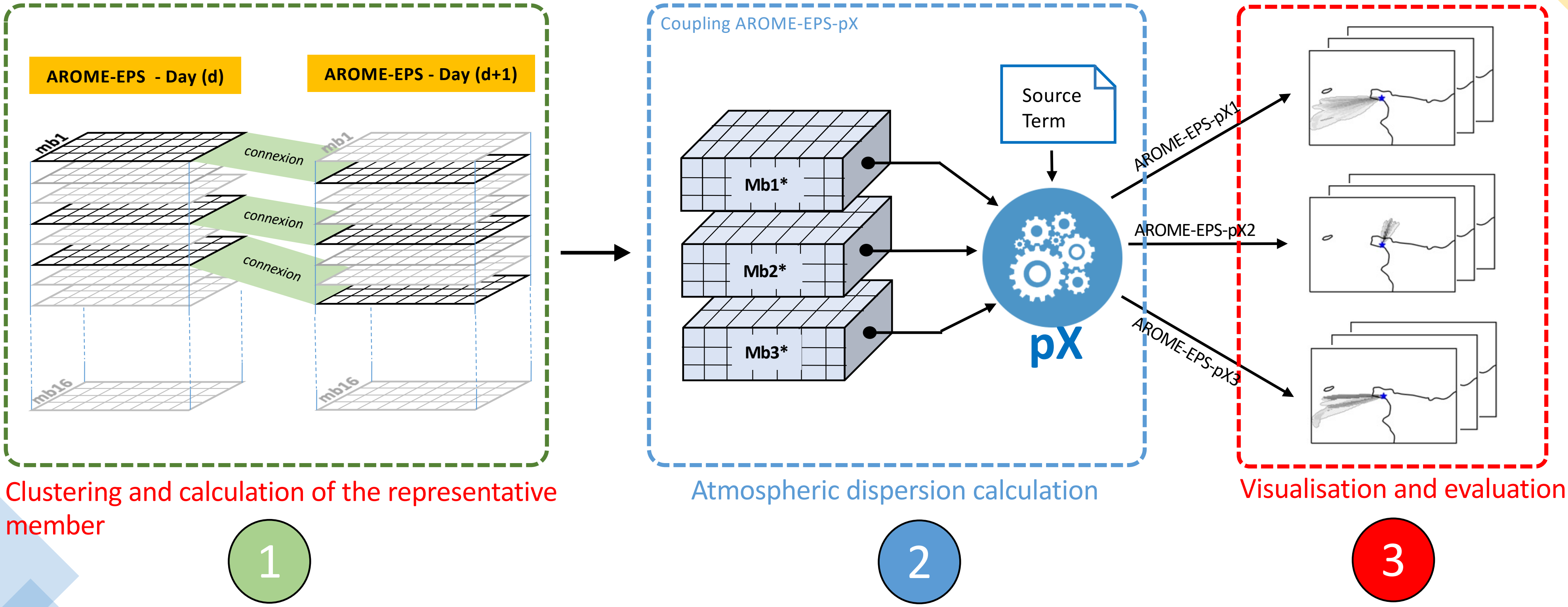
1

Atmospheric dispersion calculation

2

Methodology

Overview

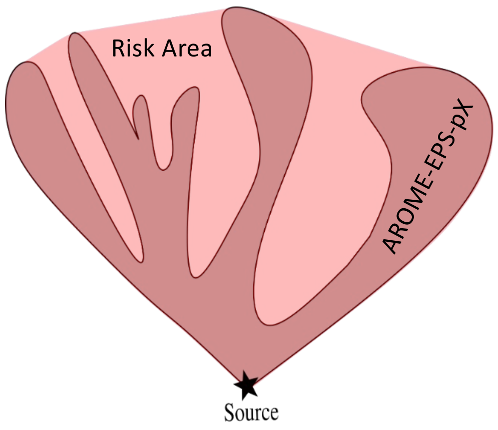


Methodology

Evaluation strategy for dispersion maps

What areas would be potentially contaminated in the next few hours following an atmospheric radioactive release ?

- ❑ Each ensemble/subensemble is assigned a risk area defined as the smallest convex surface surrounding the ensemble.

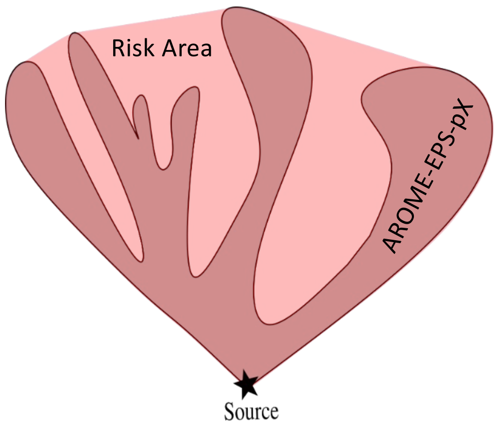


Methodology

Evaluation strategy for dispersion maps

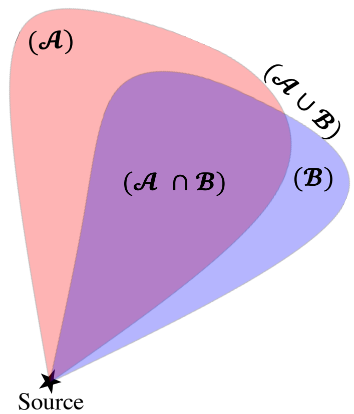
What areas would be potentially contaminated in the next few hours following an atmospheric radioactive release ?

- Each ensemble/sub-ensemble is assigned a risk area defined as the smallest convex surface surrounding the ensemble.



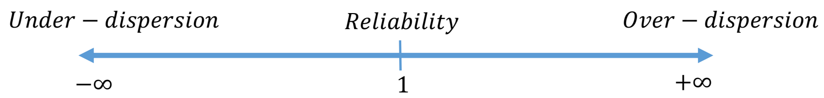
- We calculate the FMS (Figure of Merit in Space) temporal evolution of the risk areas of the clustering sub-ensemble ($\mathcal{B}$) in relation to the risk zone of the PEARO-pX set ($\mathcal{A}$):

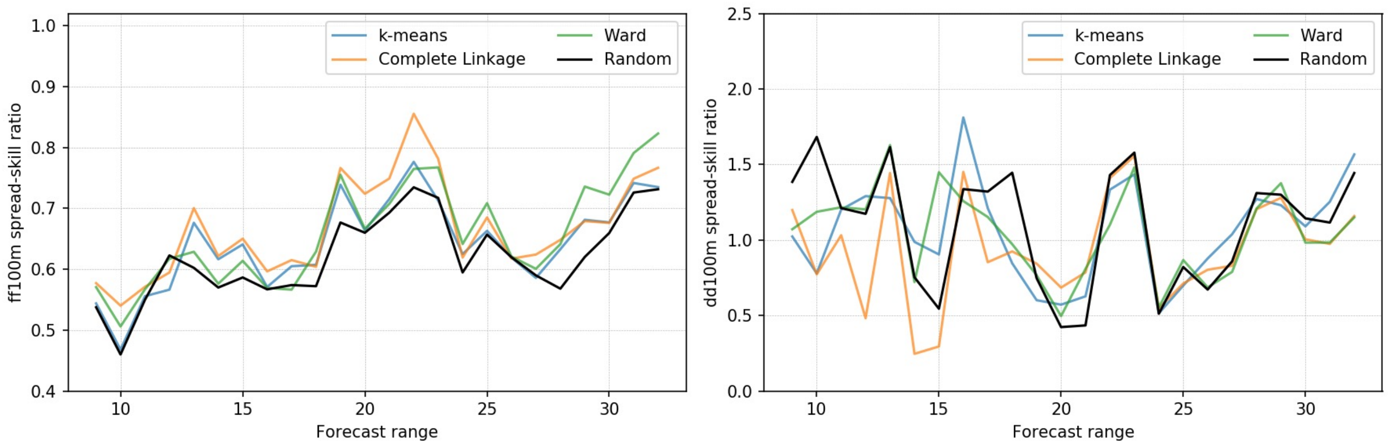
$$FMS = \frac{\mathcal{A} \cap \mathcal{B}}{\mathcal{A} \cup \mathcal{B}} \times 100$$



Impact of clustering on AROME-EPS sub-ensembles

- ❑ ***Spread-skill ratio:***





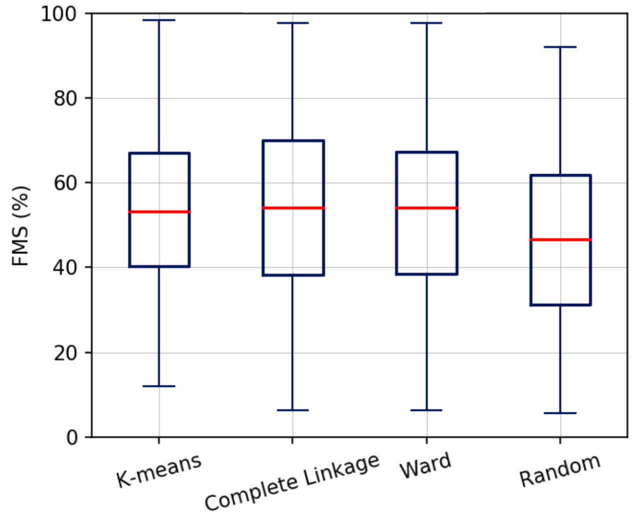
✓ In terms of wind speed, the clustering algorithms improve the dispersion of the subsets constructed to calculate dispersion, with a slight preference for the « complete-Linkage ».

✓ In terms of wind direction, the impact of the clustering algorithms is not obvious.

Results

Impact of clustering on the AROME-EPS-pX sub-ensembles

- ❑ ***Evaluation of ^{85}Kr dispersion maps***



✓ Clustering algorithms improve atmospheric dispersion forecasts on average.

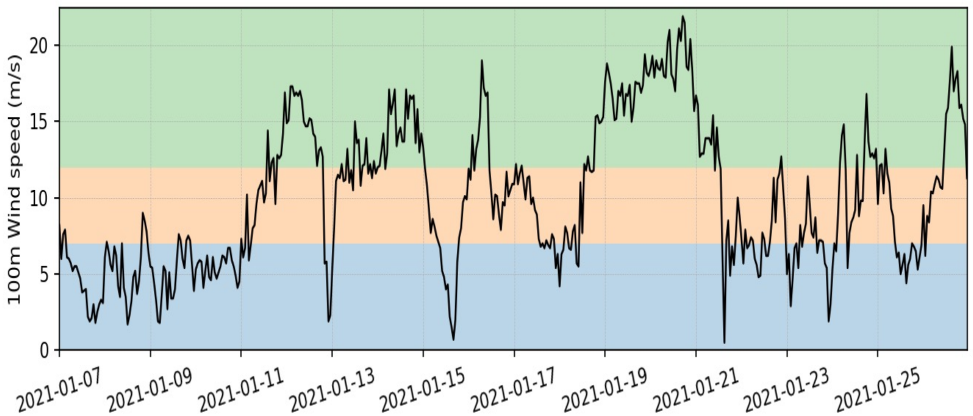
Results

Impact of clustering on the AROME-EPS-pX sub-ensembles

❑ *Sensitivity to wind conditions*

- A comparison of wind speed evolution and the impact of clustering on dispersion forecasts (FMS) shows that there is a correlation between these two variables.

- We define 3 wind intervals:

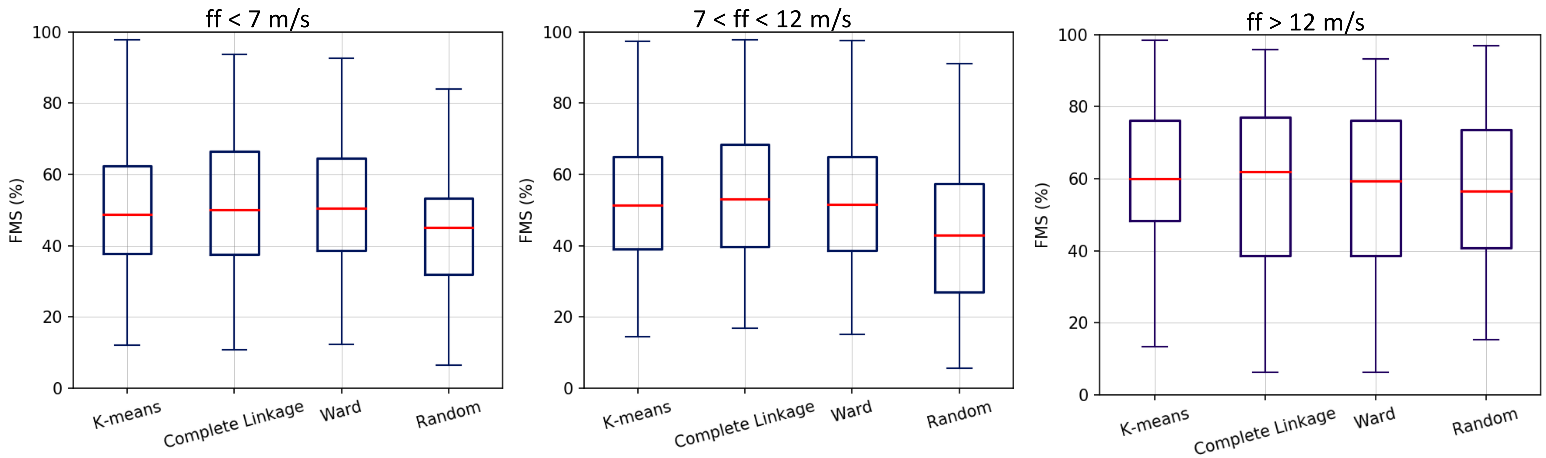


ff100 < 7 m/s : 136 samples
7 < ff100 < 12 m/s : 153 samples
ff100 > 12 m/s : 143 samples

Results

Evaluation of 85Kr dispersion maps

- ***Sensitivity to wind conditions***

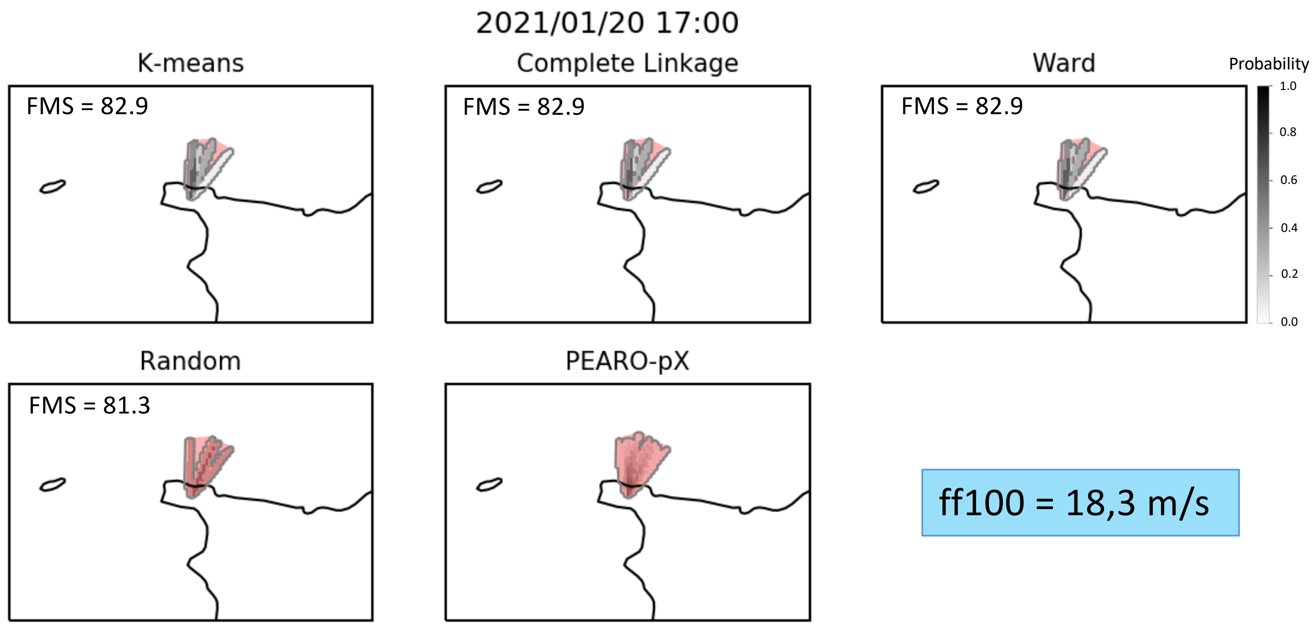


✓ Clustering has a higher impact in low/moderate wind situations.

Results

Evaluation of ^{85}Kr dispersion maps

- ***Sensitivity to wind conditions***

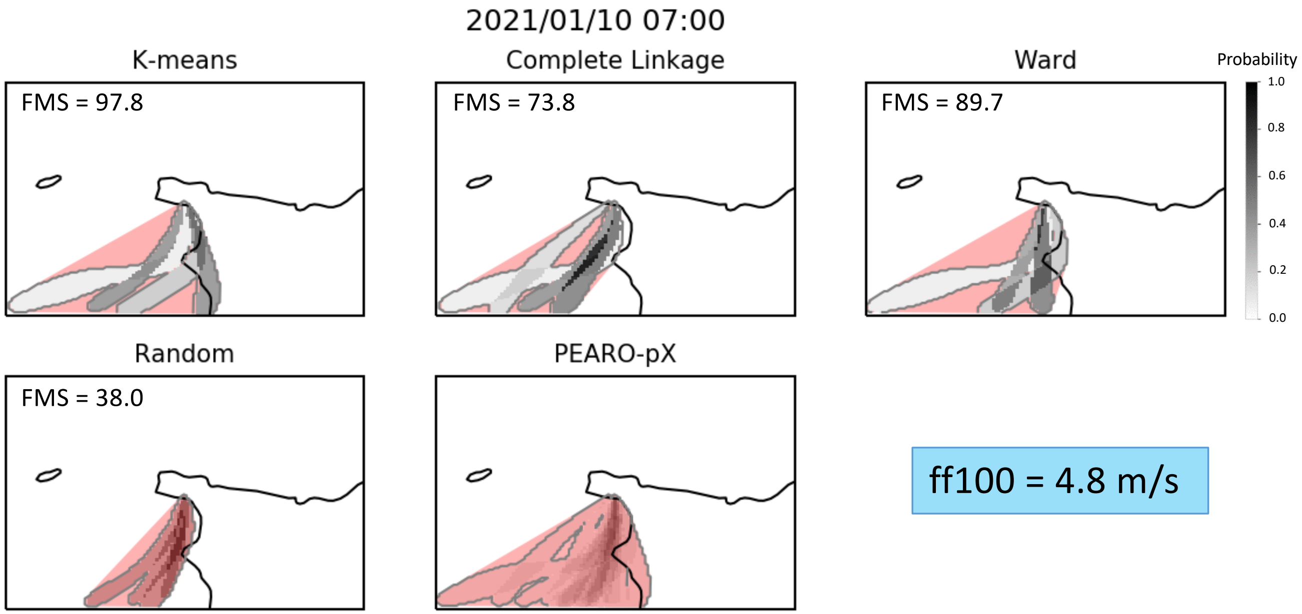


ff100 = 18,3 m/s

Results

Evaluation of 85Kr dispersion maps

- ***Sensitivity to wind conditions***



ff100 = 4.8 m/s

Conclusions and perspectives

Conclusions

- Clustering algorithms have an advantage over random sampling in predicting short-range atmospheric dispersion,
- Wind is the appropriate predictor variable for the calculation of clustering,
- Clustering results are efficient in low and medium wind conditions.

Conclusions and perspectives

Conclusions

- ❑ Clustering algorithms have an advantage over random sampling in predicting short-range atmospheric dispersion,
- ❑ Wind is the appropriate predictor variable for the calculation of clustering,
- ❑ Clustering results are efficient in low and medium wind conditions.

Perspectives

- ❑ Application to an ensemble containing a relatively large number of members (PEARP, 35 members, 10km resolution),
- ❑ Implement more complicated algorithms wherever possible (calculation time!), using more efficient distances (Wasserstein distance) for calculating inter- and intra-cluster similarity,
- ❑ Study the sensitivity of the representative member of each cluster to the calculation method.

Thank you !

Youness El-Ouartassy
Youness.el-ouartassy@meteo.fr