Downloaded from orbit.dtu.dk on: Apr 25, 2024

DTU Library

Ensemble calculations for the atmospheric dispersion of radionuclides. The Fukushima case study: ensemble results and indicators to assess the quality of ensembles

Korsakissok, Irène ; Périllat, Raphaël; Andronopoulos, Spyros; Astrup, Poul; Bedwell, Peter; Berge, Erik; Quérel, Arnaud ; Klein, Heiko; Leadbetter, Susan ; Saunier, Olivier

Total number of authors:
13

Published in:
Guidelines for the use of ensemble calculations in an operational context, indicators to assess the quality of uncertainty modelling and ensemble calculations, and tools for ensemble calculation in emergency response

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Korsakissok, I., Périllat, R., Andronopoulos, S., Astrup, P., Bedwell, P., Berge, E., Quérel, A., Klein, H., Leadbetter, S., Saunier, O., Sogachev, A., Tomas, J., & Ulimoen, M. (2019). Ensemble calculations for the atmospheric dispersion of radionuclides. The Fukushima case study: ensemble results and indicators to assess the quality of ensembles. In *Guidelines for the use of ensemble calculations in an operational context, indicators to assess the quality of uncertainty modelling and ensemble calculations, and tools for ensemble calculation in emergency response* (pp. 67-103). Article D9.5.3 European joint programme for the integration of radiation protection research.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

D 9.5.3 – Ensemble calculations for the atmospheric dispersion of radionuclides. The Fukushima case study: ensemble results and indicators to assess the quality of ensembles

Lead Authors: Irène Korsakissok, Raphaël Périllat

With contributions from: Spyros Andronopoulos, Poul Astrup, Peter Bedwell, Erik Berge, Arnaud Quérel, Heiko Klein, Susan Leadbetter, Olivier Saunier, Andrey Sogachev, Jasper Tomas, Magnus Ulimoen

Table of Contents

68

Introduction 69

Simulations setup 69

- Meteorological data 69
- Source terms 70
- Outputs and observation data 73

Ensemble results 73

- Results for ^{137}Cs air concentration 74
- Results for gamma dose rates 86
- Statistical indicators 91

Conclusions 100

Acknowledgement 101

References 101

Introduction

The hypothetical case studies have provided a large database to compare uncertainty assessment provided by different participants (Berge et al. 2019; Korsakissok et al. 2019). The objectives of this comparison were to provide a benchmark in the way uncertainties are taken into account, and to compare the variability between the ensembles (i.e. due to different models and different ways of taking into account the input uncertainties) with the input uncertainties, represented by the ensembles' spreads. In other words, one of the aims of the projects was to determine whether the variability between participants' uncertainty assessments was negligible or not, compared with the global uncertainty. The rationale behind that is that, while inter-model variability might be significant when looking at deterministic outputs, it may be of less importance as far as uncertainties estimations are concerned. Results have shown an important variability between the participants, even when the input uncertainties were the same, although this variability is not of the same order of magnitude as, for instance, source term uncertainties. The questions of how to present the outputs, how to choose the thresholds, percentiles, were also widely investigated. However, to properly evaluate the different ensemble results and possibly the models / approaches used to take into account the input uncertainties, there is the need for comparison with environmental observations. In this regard, the Fukushima case study provides a unique opportunity to complete and shed a new light on the case studies conducted within task 1.2 of WP1. More precisely, it helps infer whether the so-called "inter-model variability" (which, more precisely, should be named "inter-ensemble" to account for the differences in the ensembles' construction as well as between the models) is significant with regard to model-to-measurements comparisons, as well as measurement uncertainty. For instance, if all ensembles are very bad at predicting the observations variability with time at all stations, then the most important issue is to improve the knowledge on input uncertainties. If the performance of all ensembles is similar, then the inter-model variability is probably negligible in a real case. A related question is whether taking into account the combined meteorological and source term uncertainties is sufficient to account for uncertainties in atmospheric dispersion simulations, when comparing them with observations. If the ensembles are globally capable of encompassing the observations, then no significant source of uncertainties was left out. Finally, one of the aims of this report is to present several statistical and graphical indicators of an ensemble's performance by comparison to radiological observations. Probabilistic indicators represent the ability of an ensemble to properly encompass the observations. As a complement, the last part of this report presents deterministic indicators applied to the ensembles' median, which is a possible way of using ensemble results: in complement to the uncertainty estimation, the ensembles' median may perform better than deterministic "best estimates".

Simulations setup

Meteorological data

The ECMWF Integrated Forecast System (cycle 45r1) (Molteni et al. 1996; Leutbecher and Palmer 2008) was used to create an ensemble forecast for the period 00:00UTC 14 to 00:00UTC 17 March 2011. Initial perturbations were constructed from an ensemble data-assimilation with singular vector perturbations and model uncertainty is taken into account using a stochastic physics scheme. The ensemble consists of one run starting from unperturbed initial conditions (the control forecast) and 50 from perturbed initial conditions. Data were extracted over a horizontal domain of 32°N to 43°N and 134°E to 148°E, at a resolution of 0.2° (red region in Figure 51). 36 vertical levels were selected from the

surface to 148hPa with more levels close to the surface and the temporal resolution of the data was hourly.

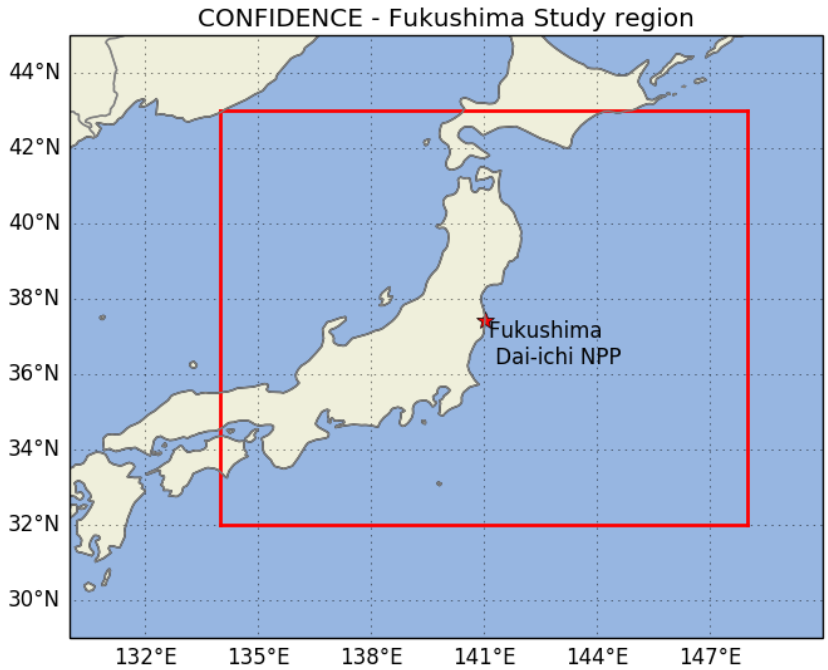


Figure 51 : Area covered by ECMWF data.

Source terms

There are two main families of methods for estimating releases (Mathieu et al. 2018).

1. An approach based on reactor physics and the knowledge of the state of the power plant. To date, there is no complete source term in the literature obtained from this approach which includes a kinetic release model for different radionuclides.

2. Methods coupling the radiological measurements in the environment and the atmospheric dispersion simulations to infer release rates. Inevitably, the quality of the source term correlates to the accuracy of the meteorological fields used, to the kind of atmospheric transport model used and to the relevance of the measurements, since an event can only be reconstructed if it is observed. (Mathieu et al. 2018) distinguish between methods called "simplified" and "inverse" methods. The former are manual or semi-automatic and are based on a limited set of measurements. The inverse modelling techniques are more operational automatic methods based on mathematically rigorous approaches.

Saunier et al. (2016) assessed the realism of several previously published source terms by comparing dispersion model output using these source terms to a dataset of air concentration measurements (the suspended particulate matter – SPM – dataset described below and used in this study) which were not used in these release estimations, having been made public only in 2015. Overall the agreement between the models and measurement is poor and no source term appears better than the others. Saunier et al. (2016) also presented initial attempts to use SPM data to estimate releases and show that these data better reproduce the temporal variability of the air concentration.

Table 13 synthesises the source terms reconstructed using either “simplified” or “inverse” methods. The nine source terms retained for this study are highlighted in green. Some source terms were not selected, either because they did not significantly differ from another source term (to avoid introducing bias in our ensemble, it is better to have source term that are significantly different), or

because they were not deemed relevant due to the spatial and temporal scale of interest of our study, or, because insufficient information was available. Figure 52 shows the release rate of ^{137}Cs for the nine source terms during the 3-day period of interest. Although they are all a posteriori source terms, based on observations, there is still a high variability within the source terms, especially when looking at the kinetics.

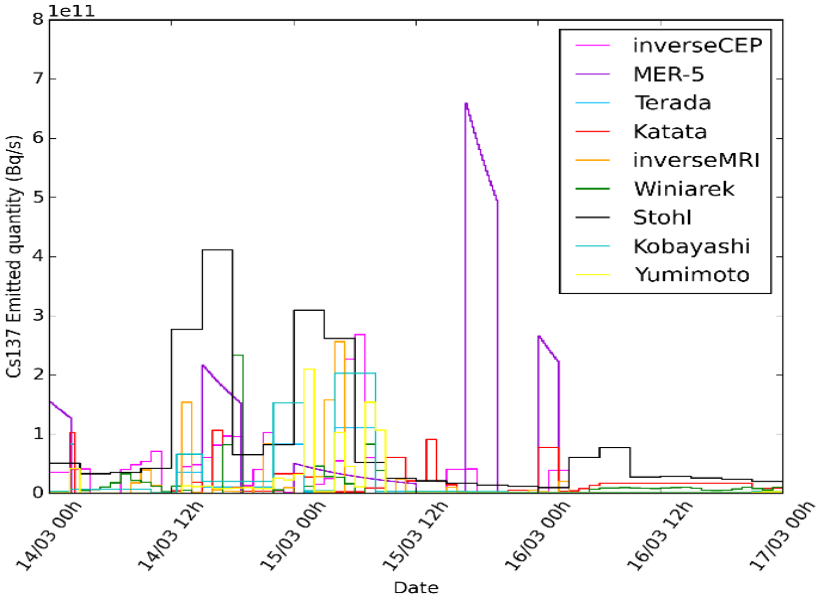


Figure 52: Release rates during the simulated period for the 9 source terms

Table 13 Source terms estimated from measurements in the environment, from (Mathieu et al. 2018). The source terms used in this study are highlighted in green. FDNPP stands for Fukushima Daiichi Nuclear Power Plant.

No.	Source term	Total ^{137}Cs (PBq)	Method	Observations used	Model / meteorology used
1	Chino et al. (2011) JAEA	13	Simplified	Concentrations in the air over Japan. Dose rates close to FDNPP. Facility events.	GEARN (Lagrangian) / GSM (Global Spectral Model)
2	Stohl et al. (2011)	35.7	Inverse	Concentrations in the air over the entire Northern Hemisphere. Facility events.	Flexpart (Lagrangian) / ECMWF 0.18°, 1° and GFS 0.5°
3	Winiarek et al. (2012)	10-19	Inverse	Concentrations in the air over Japan and North America.	Polair3D (Eulerian) / ECMWF 0.25°
4	Terada et al. (2012) JAEA	8.7	Simplified	Concentrations in the air over Japan. Dose rates close to FDNPP. Facility events.	GEARN (Lagrangian) / GSM + MM5
5	Mathieu et al. (2012)	20.6	Simplified	Concentrations in the air over Japan. Dose rates close to FDNPP. Facility events.	Gaussian puff (pX) and Eulerian (ldX) / ECMWF 0.125°
6	Saunier et al. (2013a)	15.5	Inverse	Dose rates in Japan.	ldX (Eulerian) / ECMWF 0.125°
7	Kobayashi et al. (2013) JAEA	13	Simplified	Depositions over the sea. Concentrations in air over Japan. Dose rates close to the facility. Facility events.	SEA-GEARN (ocean) + GEARN (Lagrangian) / GSM + MM5
8	Hirao et al. (2013)	9.6	Simplified	Concentrations in the air. Daily depositions by Prefecture.	LPRM (Lagrangian) / MM5
9	Winiarek et al. (2014)	11.6-19.3	Inverse	Total deposition over Japan. Concentrations in the air over Japan.	Polair3D (Eulerian) / WRF
10	Achim et al. (2014)	10.8	Simplified	Concentrations in the air of CTBTO stations. Dose rate at FDNPP. Facility event.	Flexpart (Lagrangian) / WRF + NCEP (GFS)
11	Katata et al. (2015) JAEA	14.1	Simplified	Concentrations in the air over Japan. Dose rates close to FDNPP. Depositions over the sea.	GEARN Lagrangian / GSM + MM5
12	Yumimoto et al. (2016)	8.12	Inverse	Total deposition over Japan.	CMAQ (Eulerian) / WRF
13	Saunier et al. (2016)	8.1	Inverse	Concentrations in the air over Japan including SPM data.	ldX (Eulerian) / MRI 0.03°
14	Liu et al. (2017)	7-35	Inverse	Concentrations in the air over Japan including SPM data.	Polair3D / WRF

Outputs and observation data

In this work several endpoints are considered. We make use of the hourly concentrations of ^{137}Cs retrieved by Tsuruta et al. (2014) and described in Oura et al. (2015). The air activity concentration is given in Bq/m^3. The data was obtained from the air quality automated monitoring network measuring SPM on filter tapes. They were retrieved too late to detect short-lived radionuclides, but they give crucial information on the temporal variation of ^{137}Cs concentration close to ground level, therefore indicating the passage of different plumes. Figure 53 shows the spatial distribution of the 108 stations; unfortunately, the distribution is not uniform, especially over the mostly contaminated area (in grey). The dataset came with an estimation of the associated measurement uncertainty, mostly of the order of 0.1-0.5 Bq/m^3. In this study, it will be neglected compared to the uncertainty associated with atmospheric dispersion simulations.

The gamma dose rate measurements were provided by automated stations as early as March 11th, with a 10-minutes time step. The dose rate measurements are described in (Saunier et al. 2013b) and most of them are available in the IAEA database (IAEA 2012). There are 88 stations spread over Japan, although the spatial coverage is heterogeneous (Figure 53). The dose rate readings have the advantage of measuring the contribution from all radionuclides, including the short-lived species that could not be detected by the other monitoring systems. They are composed of two parts: the direct plume contribution ("cloud-shine") and the gamma-ray emitted by radionuclides deposited on the ground ("ground-shine"). Hence, the cloud-shine is usually responsible for peak values observed during the plume passage, whereas ground-shine corresponds to a lasting contribution that lingers after the plume has left the area and decreases because of the radioactive decay.

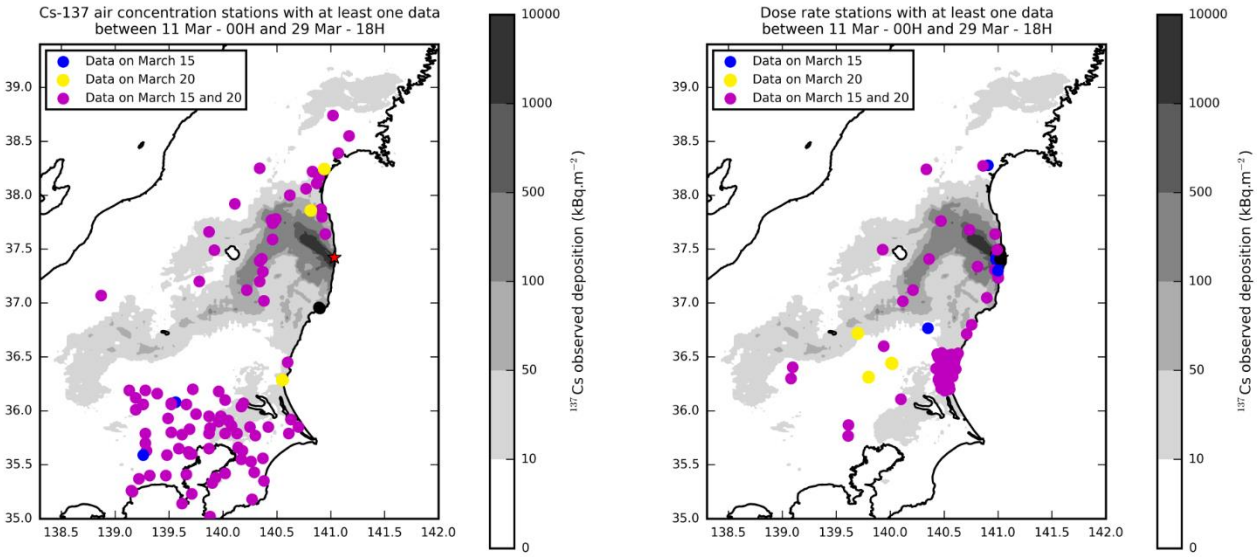


Figure 53 : Localization of ^{137}Cs activity concentration stations (left) and dose rate stations (right) over Japan. The grey shades show the deposition observations at the end of the Fukushima accident.

Ensemble results

The objective of the study is to use the meteorological ensemble (51 members) and the nine source terms and propagate them through the atmospheric dispersion models, in order to compare the output results with the observations. This amounted to many simulations, ranging from 102 to 459, depending on the way the input uncertainties were taken into account (Table 14). Among the participants, 3 of them used all combinations of meteorological members and source terms (459

simulations). The Met Office used randomly 2 different source terms for each meteorological member, which resulted in 100 simulations. IRSN used Monte Carlo sampling and randomly selected meteorological members, source terms, with additional physical perturbations on scavenging coefficients, deposition velocities and vertical diffusion. 200 samples were used. The ranges of variations are given in (Korsakissok et al. 2019).

Participant		Type of model	Number of simulations (perturbations)
France	**IRSN**	ldX – Eulerian	200 simulations (MC)
The Netherlands	**RIVM**	NPK-puff – Gaussian puff	459 simulations (9 ST x 51 MET)
UK	**Met Office**	NAME – Lagrangian particle	102 simulations (2 ST drawn randomly among 9 x 51 MET)
Greece	**EEAE**	DIPCOT – Lagrangian puff	459 simulations (9 ST x 51 MET)
Denmark	**DTU**	RIMPUFF – Gaussian puff	459 simulations (9 ST x 51 MET)
Norway	**NMI MET**	SNAP – Lagrangian particle	459 simulations (9 ST x 51 MET)

Table 14 : Summary of participants, name, type of atmospheric dispersion model used and number of simulations for the Fukushima case.

The observations used are time series of ^{137}Cs air concentrations and gamma dose rates on stations. Time series of the variables of interest at stations where observations are available provide a useful view of the ensembles' performance. While it is difficult to give an overview of the performance on more than a hundred stations, we selected stations located in different areas that were deemed representative of the results. Results are shown first for ^{137}Cs air concentration, then for gamma dose rate stations. In practice, some stations may be selected or be given a larger weight, depending on the contamination episode / area that is deemed more important for the case. For instance, one may decide that representing uncertainties close to the source (i.e. for highest activity or dose rate values) is more important than having a good representation of all uncertainties, including stations with very small values. This will be further discussed in the section devoted to statistical indicators.

Results for ^{137}Cs air concentration

Figures show the observations of ^{137}Cs concentration in Bq/m^3 (in red) and the ensemble results (in blue). The dark blue line is the median of the ensemble, and the dark shade of blue represents the 25-75th percentiles. The lighter shade of blue is the outer range of the ensemble (percentiles 0-25 and 75-100). Simulations were made from March 14th at 00:00 UTC to March 17 at 00:00 UTC, but results are only compared from March 14th at 12:00 UTC, so that any plume released earlier than March 14th at 00:00 UTC would likely to be gone from the stations locations.

North-western area and Abukuma valley

The north-western area is the zone featuring the highest contamination, as shown by the airborne deposition measurements (shades of grey in Figure 54(a)). This contamination was due to wet deposition that occurred on March 14-15, when a significant release, combined with turning winds and a rain episode, induced a significant wash-out. This contamination occurred only within a few hours.

Fukushima city

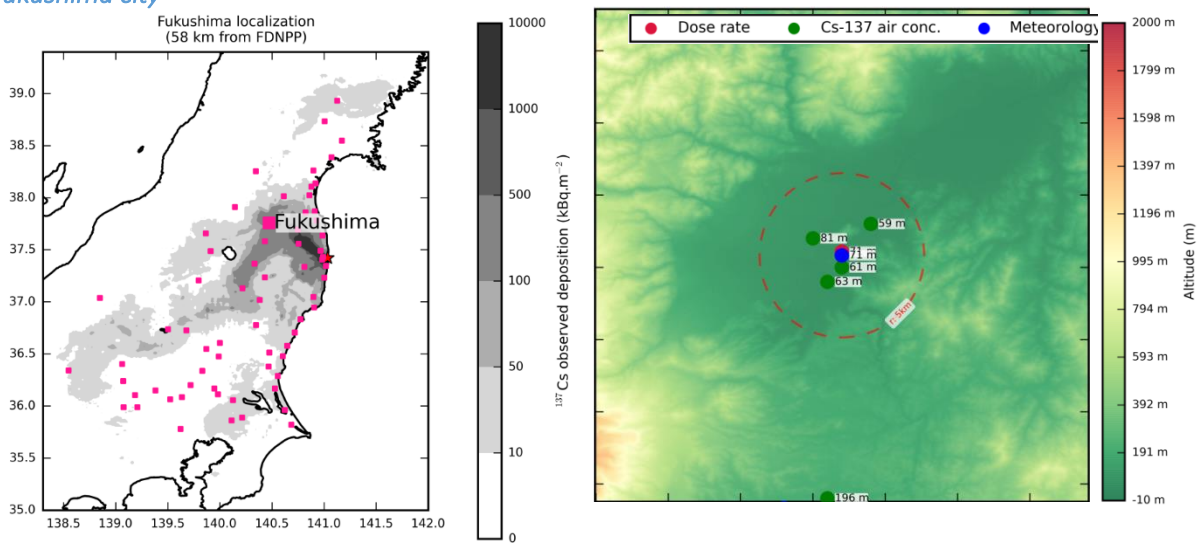


Figure 54 : Location of Fukushima city (a); topography of the area around Fukushima and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)

Fukushima city is located 58 kilometres from the FDNPP (Fukushima Daiichi Nuclear Power Plant), at the furthest point of the high deposition north-western area (Figure 54(a)). Along with gamma dose rate and meteorological measurements, there are four stations of ^{137}Cs measurements within 5 kilometres from each other (Figure 54(b)): Furukawa, Minamimachi, Moriai and Sugitsuma-cho. The four stations show very similar observations, with a plume passage at the same time and a maximum observed concentration of 30 Bq/m^3. However, there are some differences between them, with the plume passage lasting longer in some cases. This is illustrated below with two stations: Minamimachi and Sugitsuma-cho: at Minamimachi, there are activity measurements from March 15th at 06:00 UTC to March 15th at 06:00 UTC, whereas at Sugitsuma-cho, there is no detection after March 15th at 18:00 UTC (except a small peak later, on March 16th at 12:00). Some participants' ensemble predicted a plume duration too short to encompass all observations at Minamimachi (Met Office and, to a lesser extent, DTU and NMI MET) but have a good agreement with observations for Sugitsuma-cho; on the other hand, RIVM and IRSN correctly encompass the observations at Minamimachi but predict activity concentrations where there are no observations in the case of Sugitsuma-cho. For all participants, the modelled concentrations are almost or exactly the same at the two stations. The observed discrepancies between the two stations are probably due to local topography effects that are not represented by the dispersion models, especially with the meteorological resolution used.

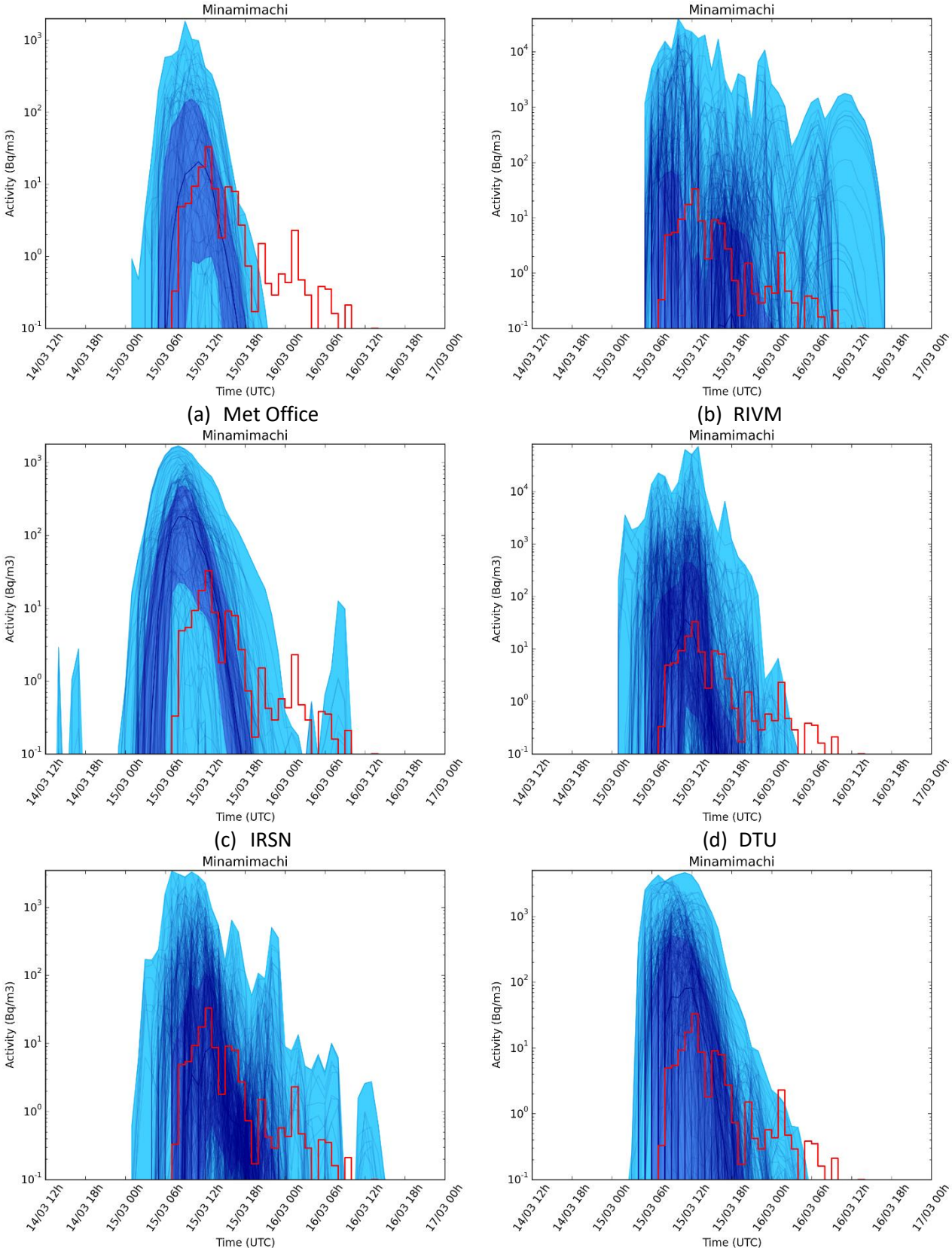
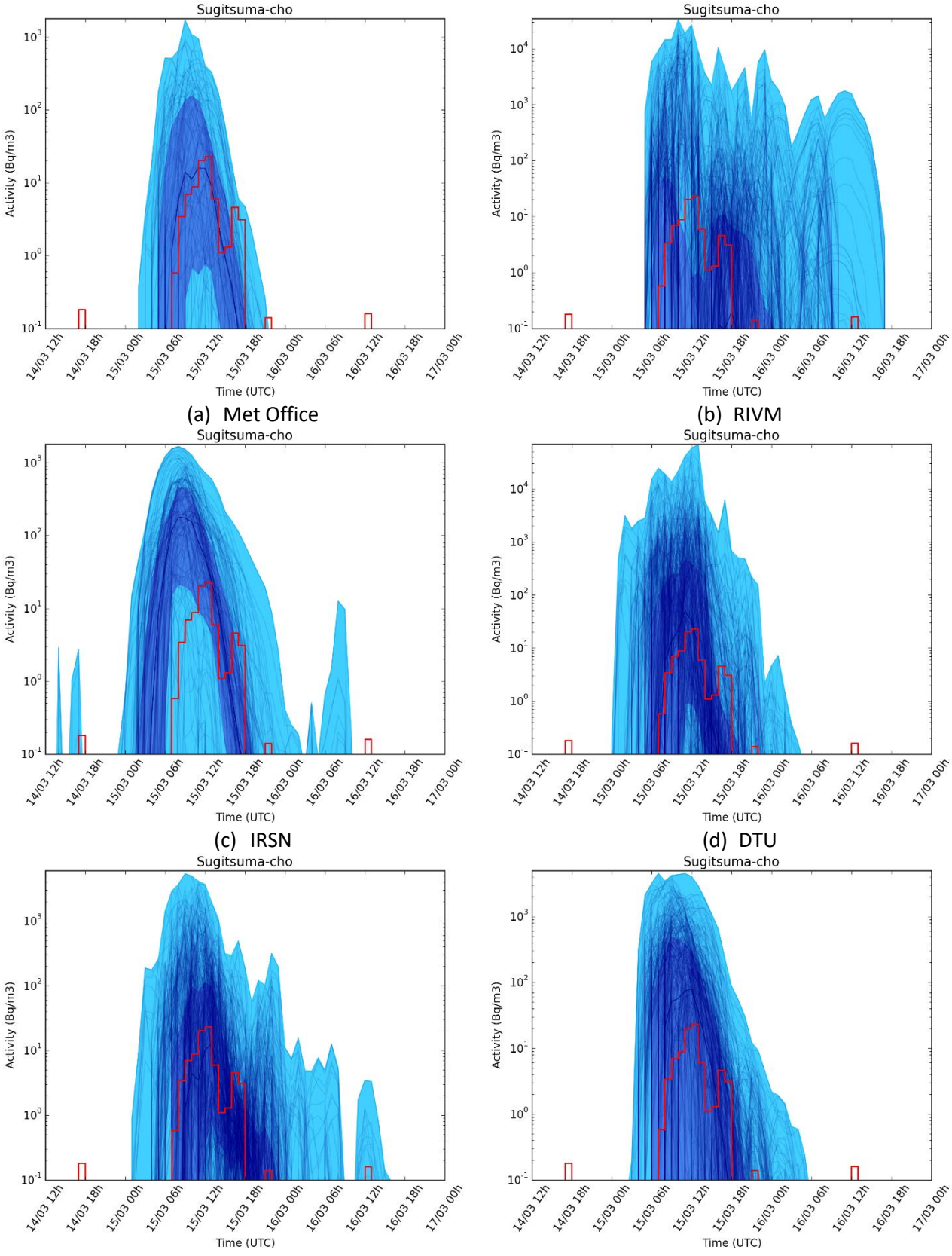


Figure 55 : Ensemble results of ^{137}Cs activity concentration at station Minamimachi, in Fukushima city (58 km NW from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

Figure 56 : Ensemble results of ^{137}Cs activity concentration at station Sugitsuma-cho, in Fukushima city (58 km NW from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.

Shirakawa

Shirakawa city is located 77 kilometres south-west of the FDNPP, in a region where ^{137}Cs deposit ranges between 50 and 150 Bq/m2 (Figure 57(a)), in the Abukuma valley, surrounded by mountains north-west and south-west from the city (Figure 57(b)).

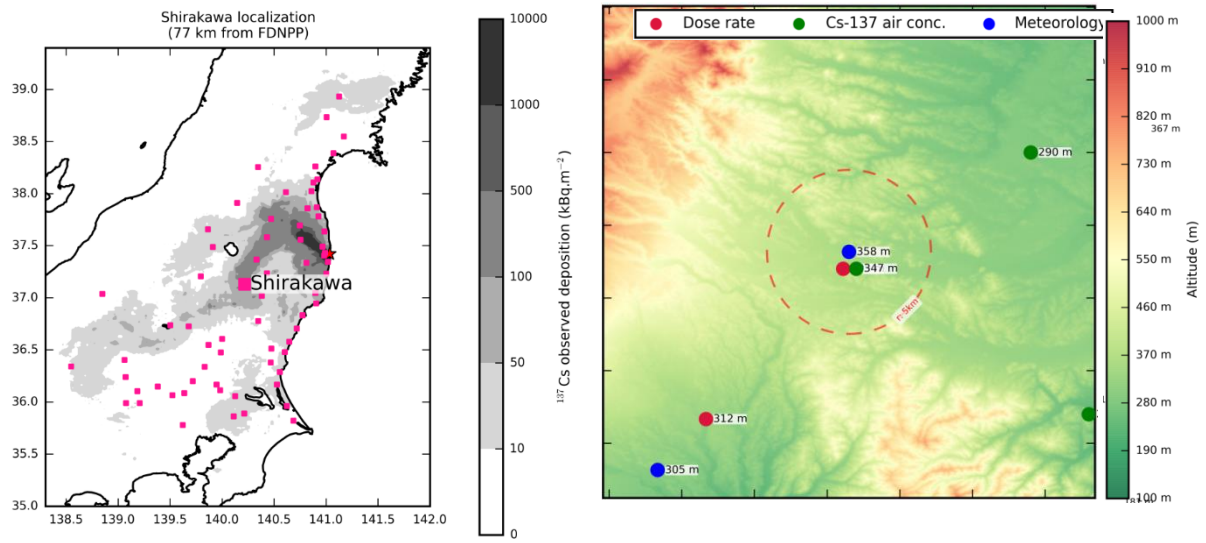


Figure 57 : Location of Shirakawa (a); topography of the area around Shirakawa and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)

On March 15th, the maximum observed activity concentration of ^{137}Cs is over 100 Bq/m^3. The plume is observed from March 15th at 03:00 to March 15th at 16:00, with light rain between 06:00 and 12:00 and more rain between 12:00 and 16:00. Therefore, the deposition corresponds to the scavenging of the plume, mostly due to light rain (Quérel et al, 2016). The ability of a model (or an ensemble of simulations) to correctly predict the deposition in this area is therefore correlated to (a) correctly predicting the plume passage, not only with the maximum value but also with the "tail" of the plume, and (b) accurately predicting the rain, including light rain which is usually difficult to forecast with meteorological models only. Figure 58 shows that the ECMWF ensemble is able to encompass most of the rain observations on the 15th March, including the possibility to have rain earlier than the observation time. Indeed, as Japanese rain gauges have a step of 0.5 mm/h, it is probable that rain with intensity lower than 0.5 mm/h participated in the plume wash-out without being detected by the gauge.

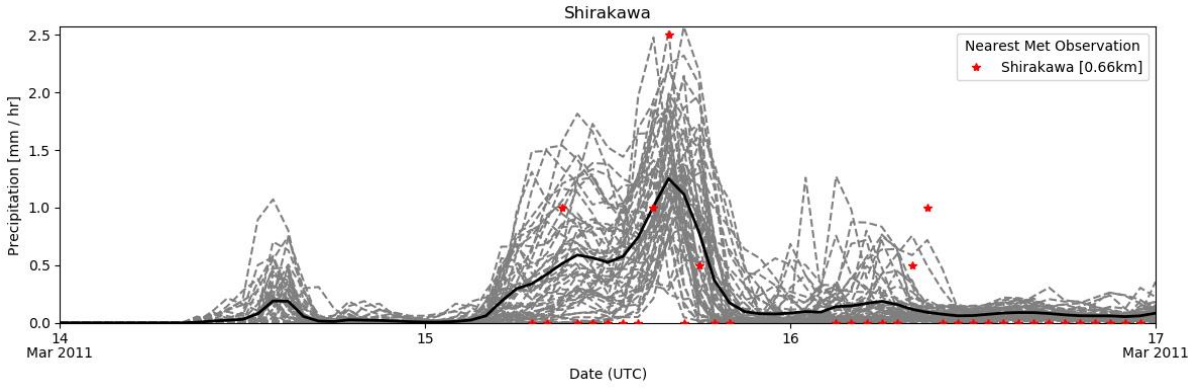
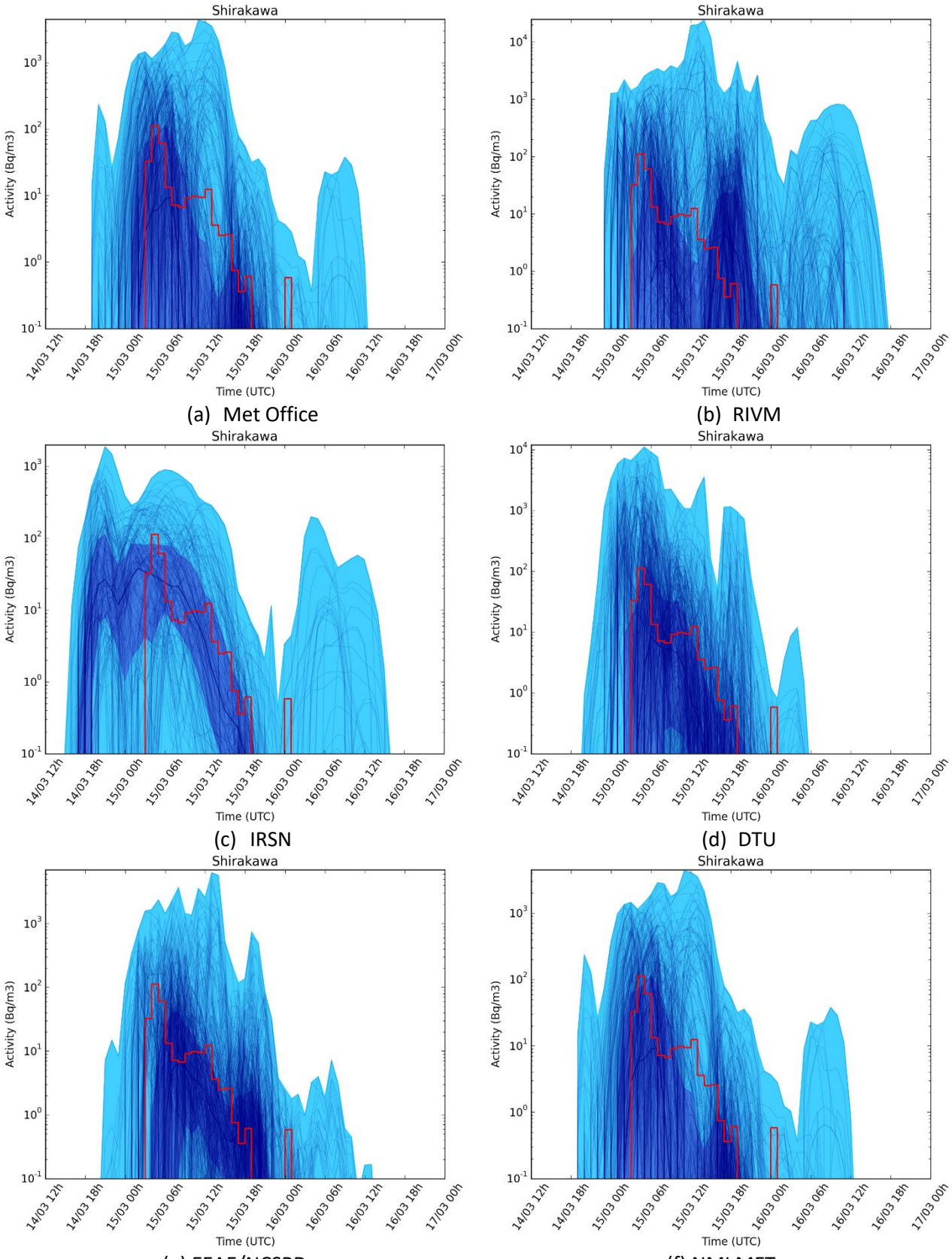


Figure 58 : Ensemble precipitation forecast by the meteorological ensemble at Shirakawa, and comparison with nearest meteorological observations available (in red). The distance to the meteorological observations is given in brackets.



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

Figure 59 : Ensemble results of ^{137}Cs activity concentration at station Shirakawa (77 km SW from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.

Aizuwakamatsu

This station is located within 100 kilometres of the source, close to the highly contaminated north-western area (Figure 60). The contamination in this area occurred mostly within a few hours during March 15th, due to scavenging by rain. The first peak corresponding to the plume passage on March 15th is correctly reproduced by all participants, with a very large spread corresponding to a high overestimation of some members. However, the second peak on March 16th at 05:00, corresponding to a latter release, is only reproduced by one participant, RIVM, and by very few IRSN members. Therefore, the presence or not of this peak is not due to meteorological or source term data. Rather, the station may be located at the edge of the plume and, depending on diffusion schemes, some simulated plumes may "touch" the station while most do not.

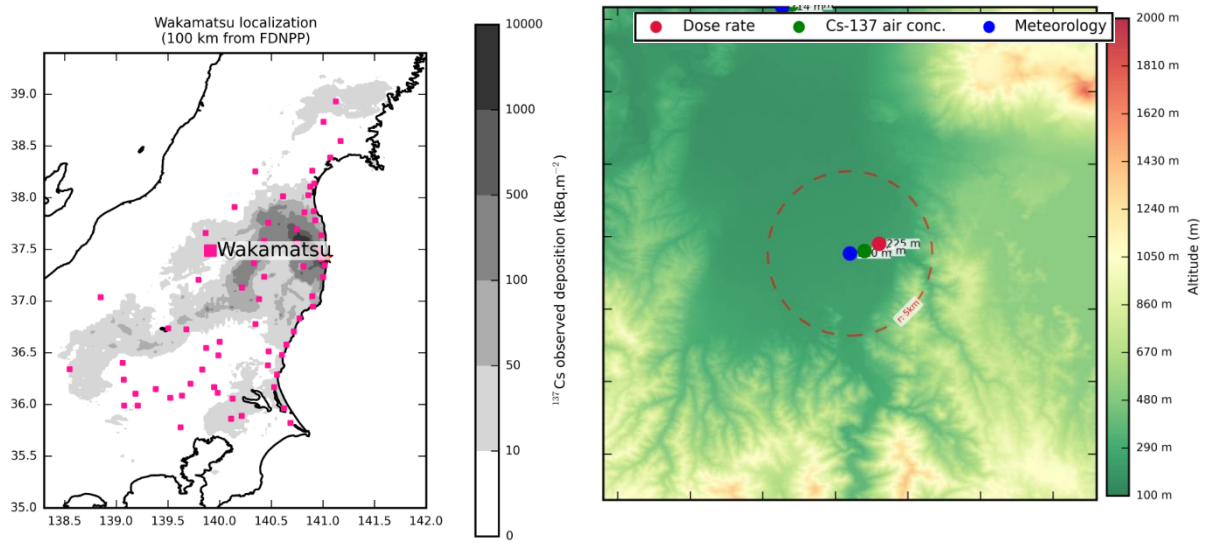
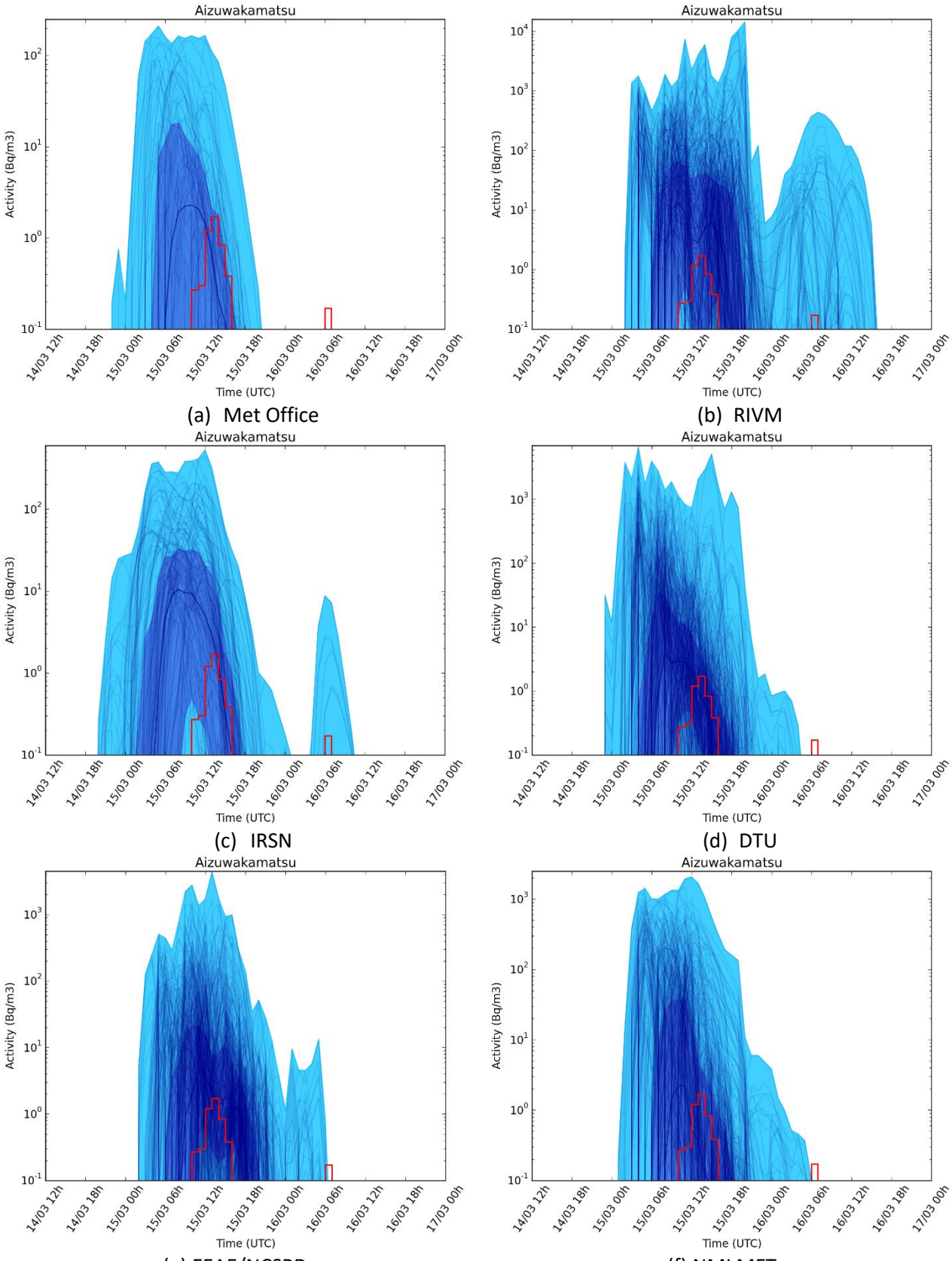


Figure 60 : Location of Aizuwakamatsu (a); topography of the area around Aizuwakamatsu and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

Figure 61 : Ensemble results of ^{137}Cs activity concentration at station Aizuwakamatsu (98 km west from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.

South coast and Kanto area

Kamisu-yokose (Tonosho)

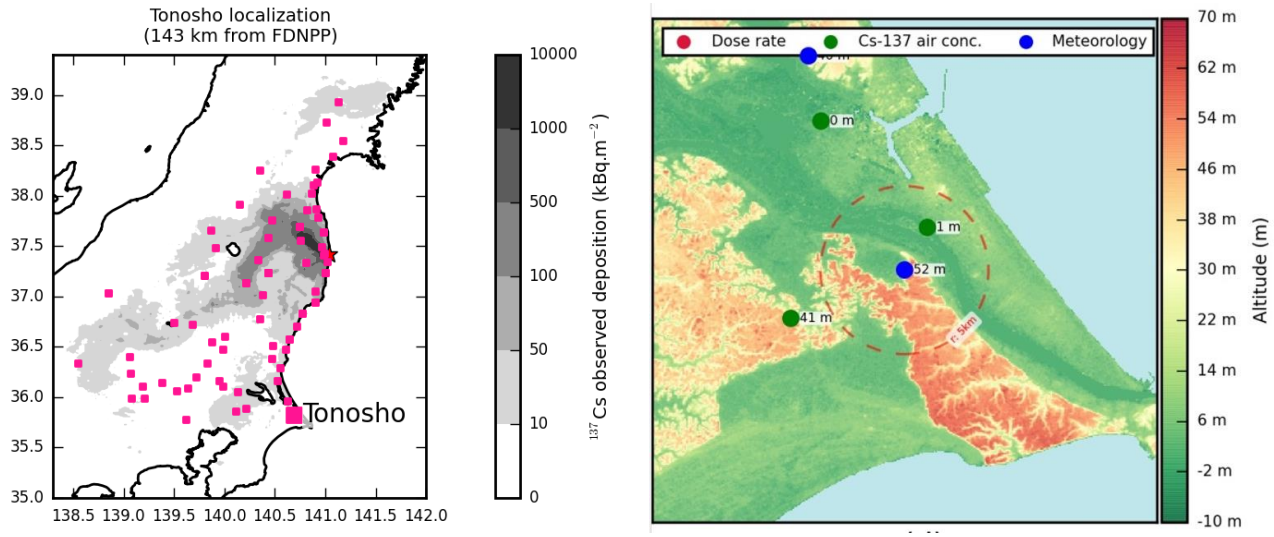


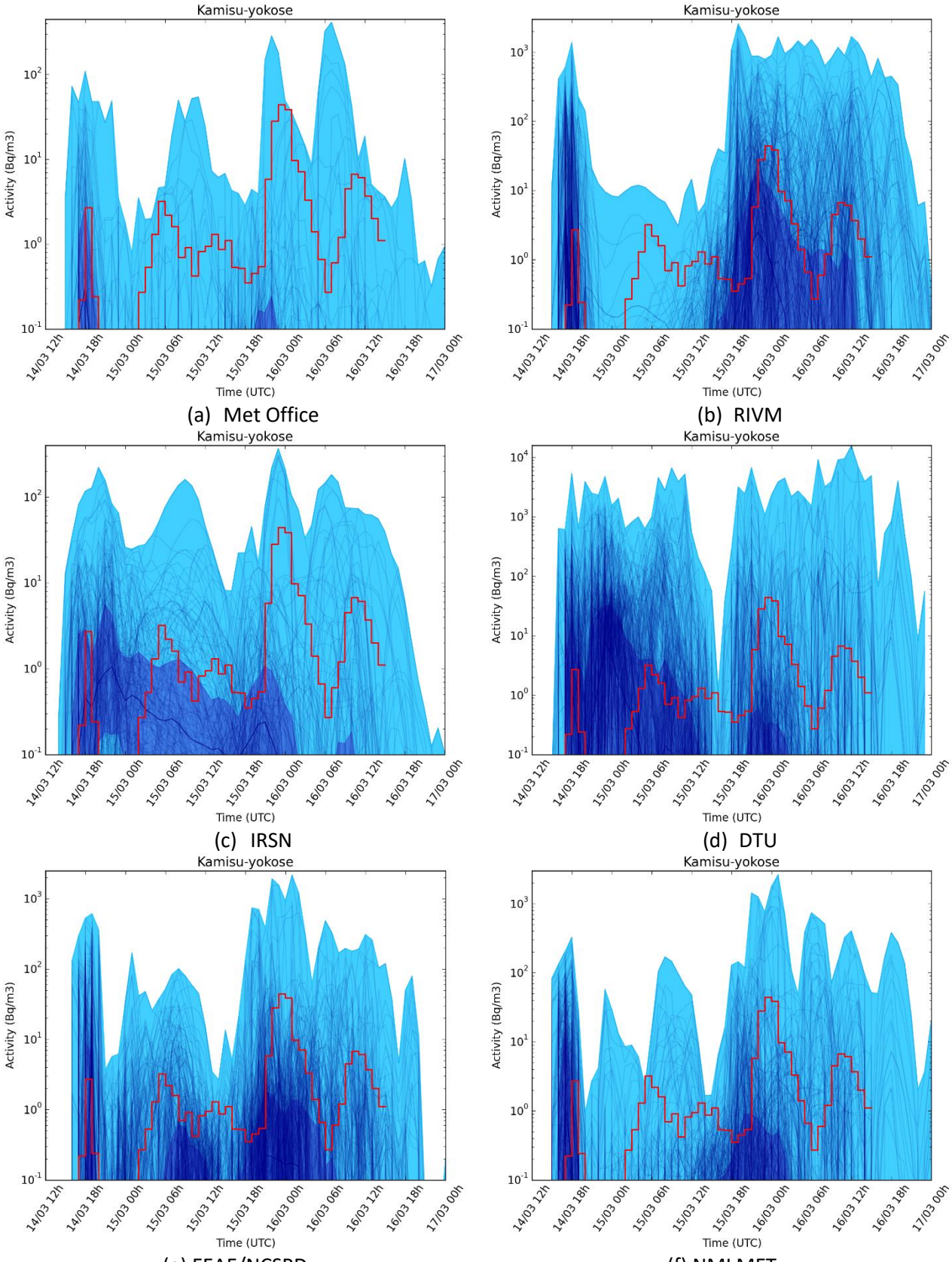
Figure 62 : Location of Tonosho and Kamisu-yokose (a); topography of the area around Tonosho and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)

The meteorological station of Tonosho and the ^{137}Cs concentration station Kamisu-yokose are located about 140 kilometres south from FDNPP, along the coast (Figure 62). The main plume is detected by this station. There are four different plumes: (1) on March 14th at 18:00 UTC there is a small peak, (2) on March 15th between 03:00 and 21:00 UTC (both with values well below 10 Bq/m^3), (3) on March 15th between 21:00 UTC and March 16th at 06:00 UTC, with concentrations over 40 Bq/m^3, and (4) on March 16th between 09:00 and 15:00 UTC, with concentrations lower than 10 Bq/m^3. Of all four plumes, only the 3rd concurs with rain observations and is therefore eligible for undergoing wet deposition.

Here, all ensembles globally encompass the observations, with the exception of IRSN's ensemble which is a bit late for the 3rd episode. However, when looking at the darker blue, which represents the 25-75th percentile envelope, tendencies are different:

- RIVM and NMI MET represent reasonably well the 1st and 3rd plumes, with some underestimation of plume 2;
- DTU and IRSN represent plumes 1 and 2 (between March 14th at 12:00 to March 15th at 12:00) well except that plume 2 is observed longer than what the ensembles feature;
- Met Office only represents plume 1, and show a very small peak for plume 3;
- EEAE represents well plumes 1 and 3, and underestimates plume 2;
- Only RIVM models a peak for plume 4.

This highlights the importance of taking into account all members of the ensemble, and not focusing only on the 25-75th percentile. It also illustrates the variability between the different ensembles, due to the ensembles constructions and the different models and physical parameterizations.



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

Figure 63 : Ensemble results of ^{137}Cs activity concentration at station Kamisu-yokose (140 km South from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.

Abiko

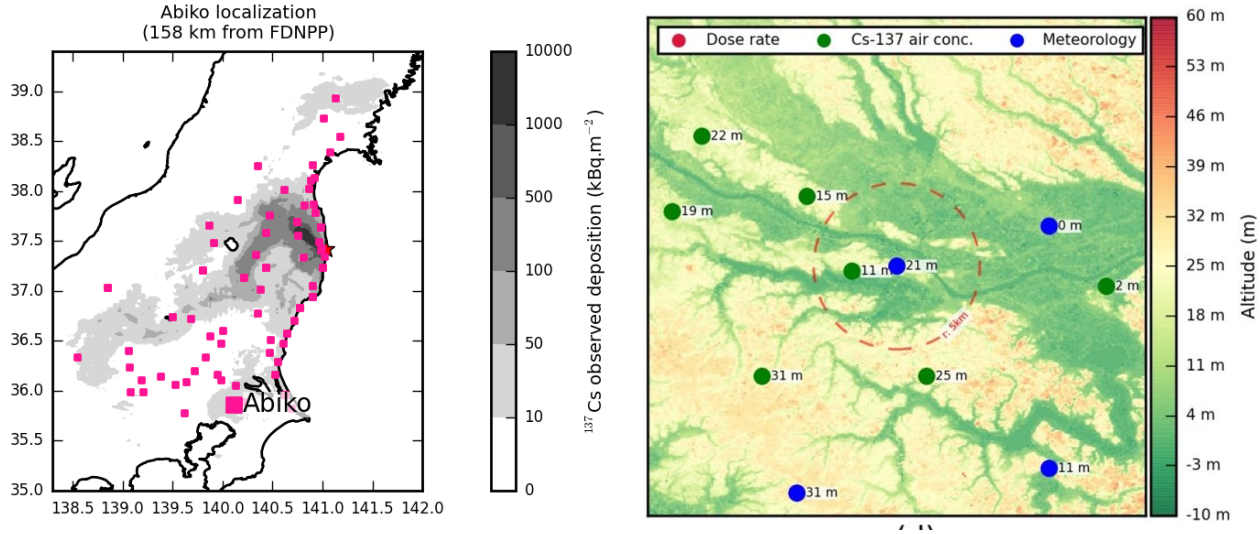
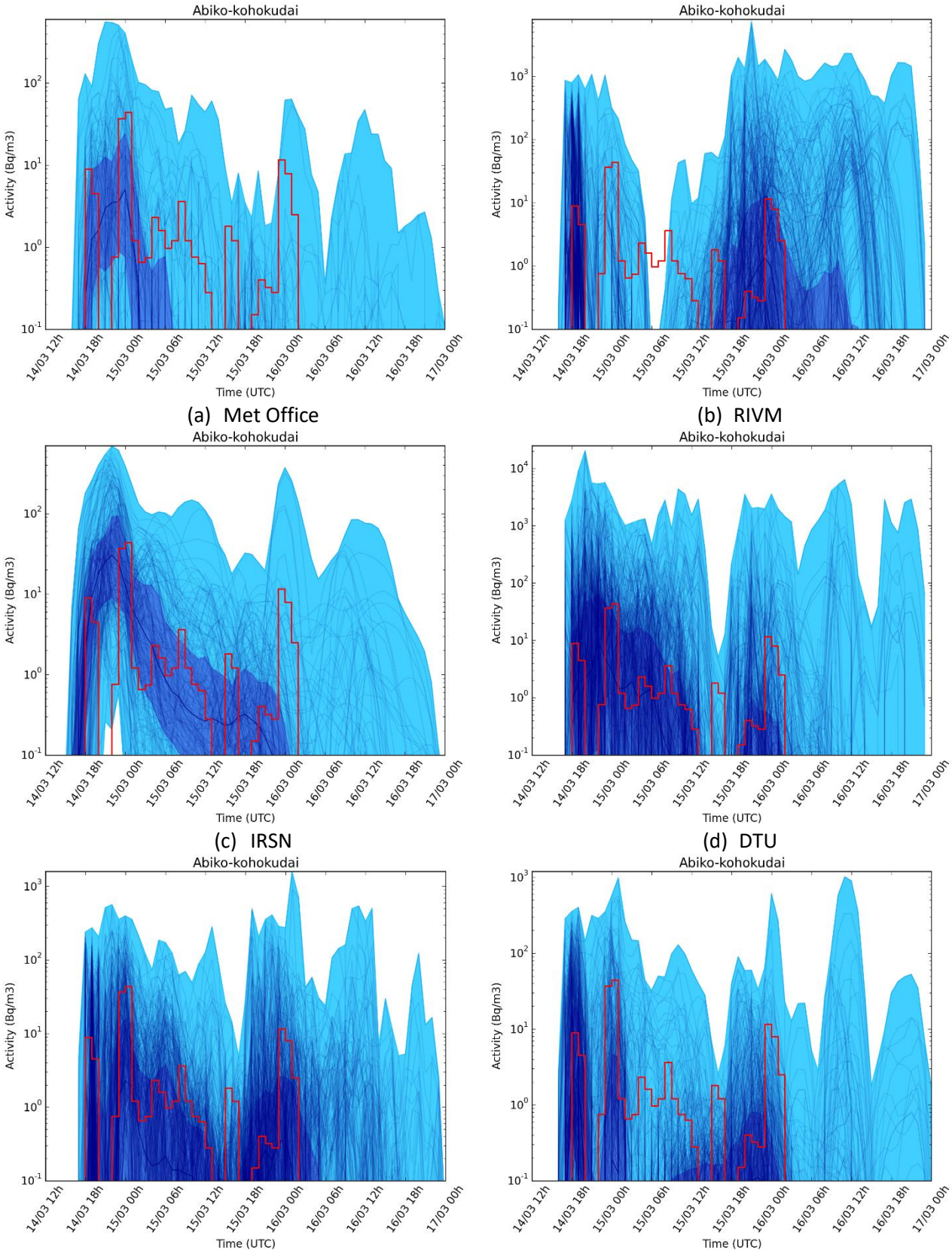


Figure 64 : Location of Abiko (a); topography of the area around Abiko and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)

Abiko is located in the Kanto plain, 158 kilometres south-west of FDNPP, in an urban area surrounded by croplands (Figure 64). It is located in a zone with ^{137}Cs deposition between 10 and 50 Bq/m^2. On March 14-16th, there are four different plumes, the main being between March 14th at 22:00 UTC and March 15th at 03:00 UTC (over 40 Bq/m^3). No rain observations coincide with the passage of these plumes, therefore only dry deposition occurs on these dates; the deposition featured on the map is likely to come from wet deposition on March 20th (Quérel et al, 2016).



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

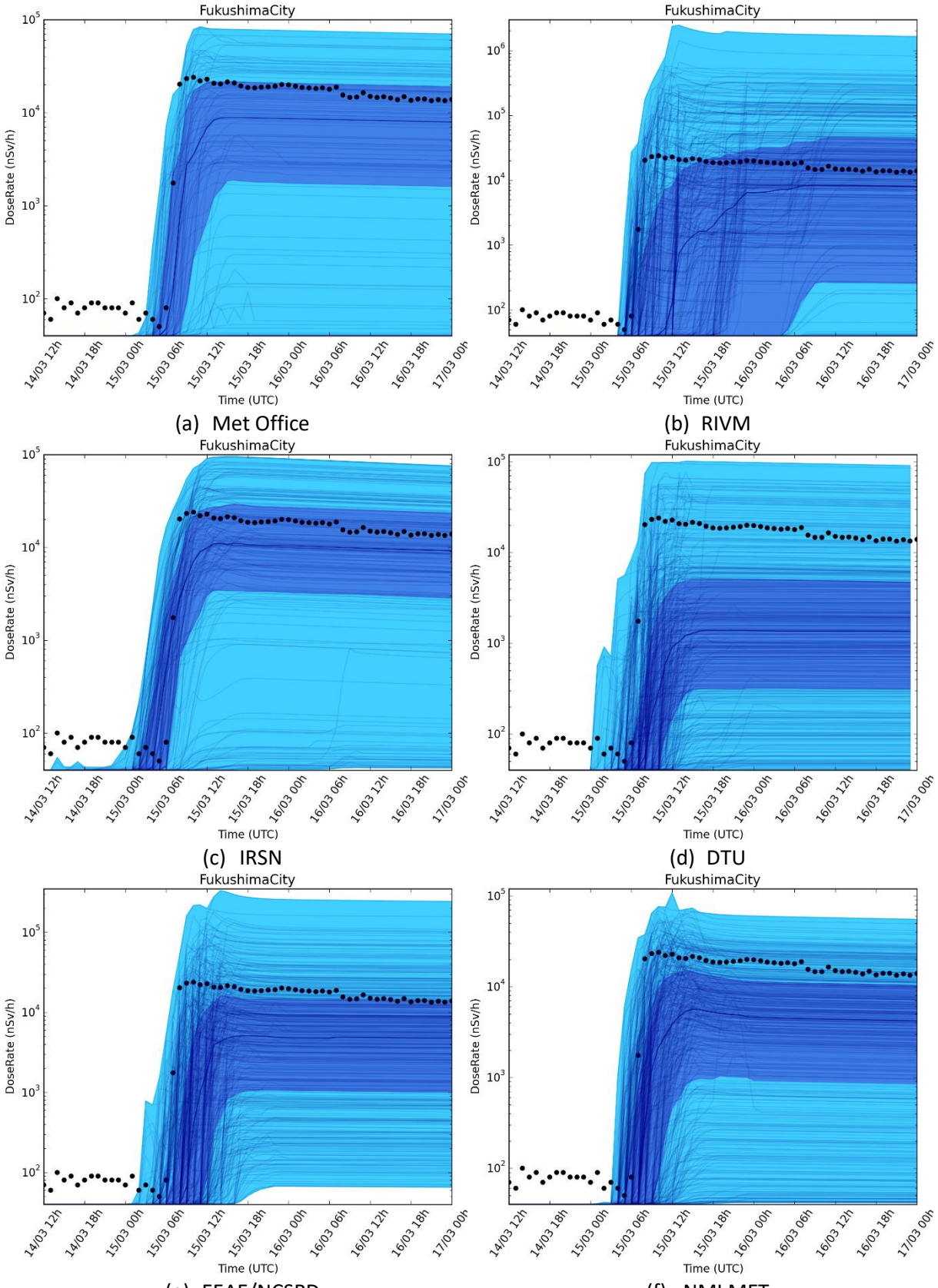
Figure 65 : Ensemble results of ^{137}Cs activity concentration at station Abiko (158 km South from FDNPP) for the Fukushima case study for several project participants, compared with observations (red line). The y-axis are not the same on all figures.

Results for gamma dose rates

Gamma dose rate monitoring stations measure the gamma radiation coming from the air and the ground nearby the station. Therefore, they include the deposition from plumes released before March 14th, which was not modelled by our studies. Although, deposition that occurred earlier than this date is negligible compared to the contamination that occurred during March 14-16th, it appears that, in many cases, an offset corresponding to the background radiation should be added to the simulations in order to compare the ensembles with the observations. In the following results, an offset equal to the minimum value for each station on the period March 14th at 12:00 UTC and March 15th at 15:00 UTC was added to the modelled values. For some stations (e.g. Iwaki, Figure 70) there was no contamination due to the accident before March 14th, and this value represents the background noise coming from cosmic and terrestrial radiations. For others (e.g. Fukushima, Figure 66), this value is higher due to previous contamination of this area.

Fukushima city

Fukushima city is located in the highly contaminated north-western area (Figure 54). The deposition in this area occurred within a few hours on March 15th, due to the scavenging by rain of the plume that reached the station between 06:00 and 15:00 UTC, as demonstrated by air concentration measurements (Figure 55 and Figure 56). This episode was particularly difficult to simulate by models, since it is a conjunction between turning winds, rainfall and plume passage within a few hours ((Korsakissok et al. 2013); (Mathieu et al. 2018)). The gamma dose rate observations shown in Figure 66 are typical of a wet deposition pattern: there is a first increase that corresponds to the beginning of the plume passage and/or rain, but no significant decrease after the plume has left the area, due to the important contribution of radionuclides deposed on the ground to the total measured gamma dose rate. It should also be noted that, in the case of Fukushima, the initial increase of the gamma dose rate corresponds to the beginning of the rain (between 03:00 and 07:00 UTC on March 15th) and is likely due to the scavenging of a plume in altitude, since concentrations at ground level are not high enough to explain the deposition values and subsequent gamma dose rate measured (Quérel et al. 2016). This is consistent with the fact that Fukushima city is surrounded with high mountains, and the plume trajectory is likely to have crossed some of them. This is still another pitfall for modelling this episode, especially with a low resolution meteorological dataset. In this respect, results presented in Figure 66 are correct, since all ensembles encompass the observations. This means that most ensemble members manage to more or less reproduce this wet deposition episode, although there is an expected high variability between them. When looking at the dark blue lines, representing the 25-75th percentile, globally the observations are underestimated. The timing of the gamma dose rate increase is well encompassed by the models, which is usually difficult for deterministic simulations, as light rains are not well forecasted (Mathieu et al. 2018). This is allowed by the meteorological ensemble which has a sufficient variability in the rain timing (Figure 67).



(a) Met Office (b) RIVM

(c) IRSN (d) DTU

(e) EEAE/NCSRD (f) NMI MET

Figure 66 : Ensemble results of dose rate (nSv/h) at Fukushima city (60 kilometres north-westof FDNPP) for the Fukushima case study for several project participants, compared with observations (black dots). The y-axis are not the same on all figures.

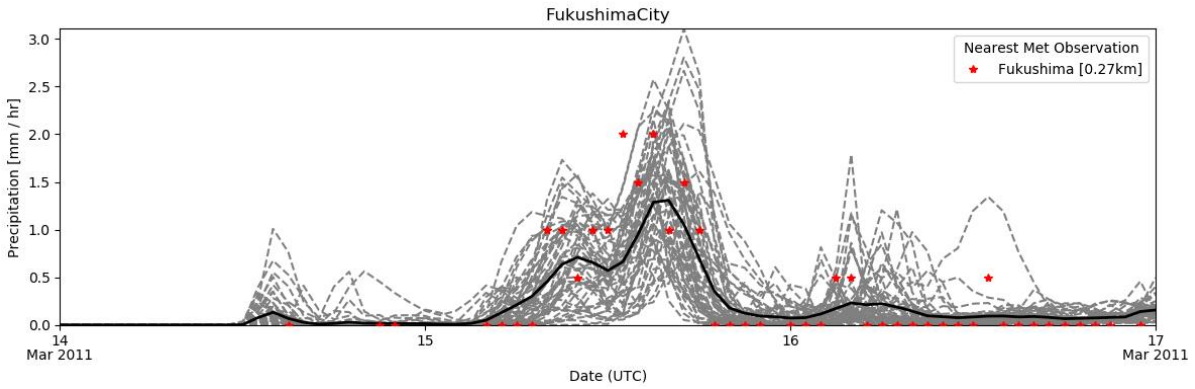


Figure 67: Ensemble precipitation forecast by the meteorological ensemble at Fukushima city, including meteorological observations at the nearest station (in red).

Iwaki

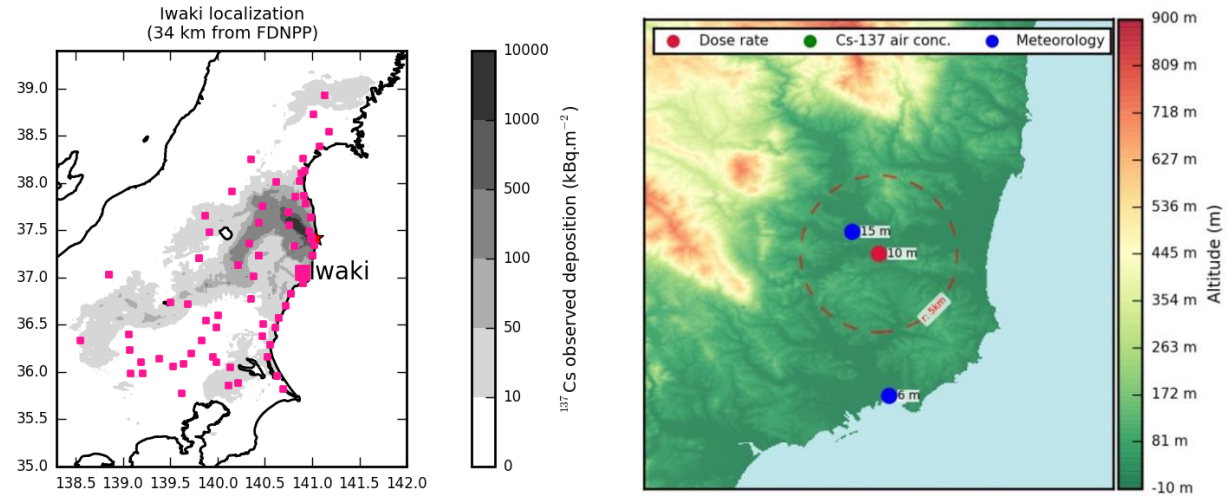


Figure 68 : Location of Iwaki (a); topography of the area around Iwaki and measuring stations (b). Courtesy of A. Quérel (Quérel et al. 2016)

Iwaki station is located 34 kilometres south of FDNPP, along the coast (Figure 68). Here, gamma dose rate observations on March 14-16th are typical of a plume passage without wet deposition, with a large peak during the plume passage and then a much lower gamma dose rate, due to dry deposition. There is observed precipitation from 16:00 to 22:00 UTC on March 15th, but at this moment, the plume that was scavenged contained a significantly lower concentration of radioactivity than during the initial peak that occurred between 16:00 and 23:00 UTC on March 14th. For this initial peak, the gamma dose rate is as high as 22 mSv/h, whereas it is about 5 mSv/h during the rain episode on March 15th. However, it appears that the meteorological ensemble forecasts rain in the night between March 14th and March 15th (Figure 69).

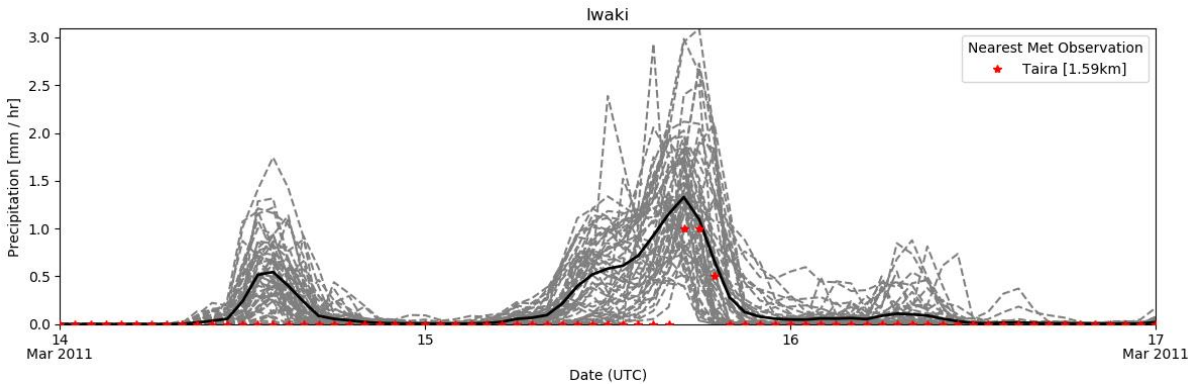


Figure 69: Ensemble precipitation forecast by the meteorological ensemble at Iwaki, and comparison with nearest meteorological observations available (in red).

This explains why the ensemble results for gamma dose rates feature a behaviour similar to the one showed at Fukushima city, with a large ground shine due to wet deposition (Figure 70). This also explains why the 25-75th percentiles have a tendency to overestimate the gamma dose rates after the first peak. Globally the ensembles encompass the observations, but it is difficult to see whether the peaks occurring on March 15th are only forecasted by some ensemble members. In summary, most members may be correct, not because they are accurately simulating the peaks, but because there is already a large variability in the gamma dose rate due to deposition after the initial increase of gamma dose rate.

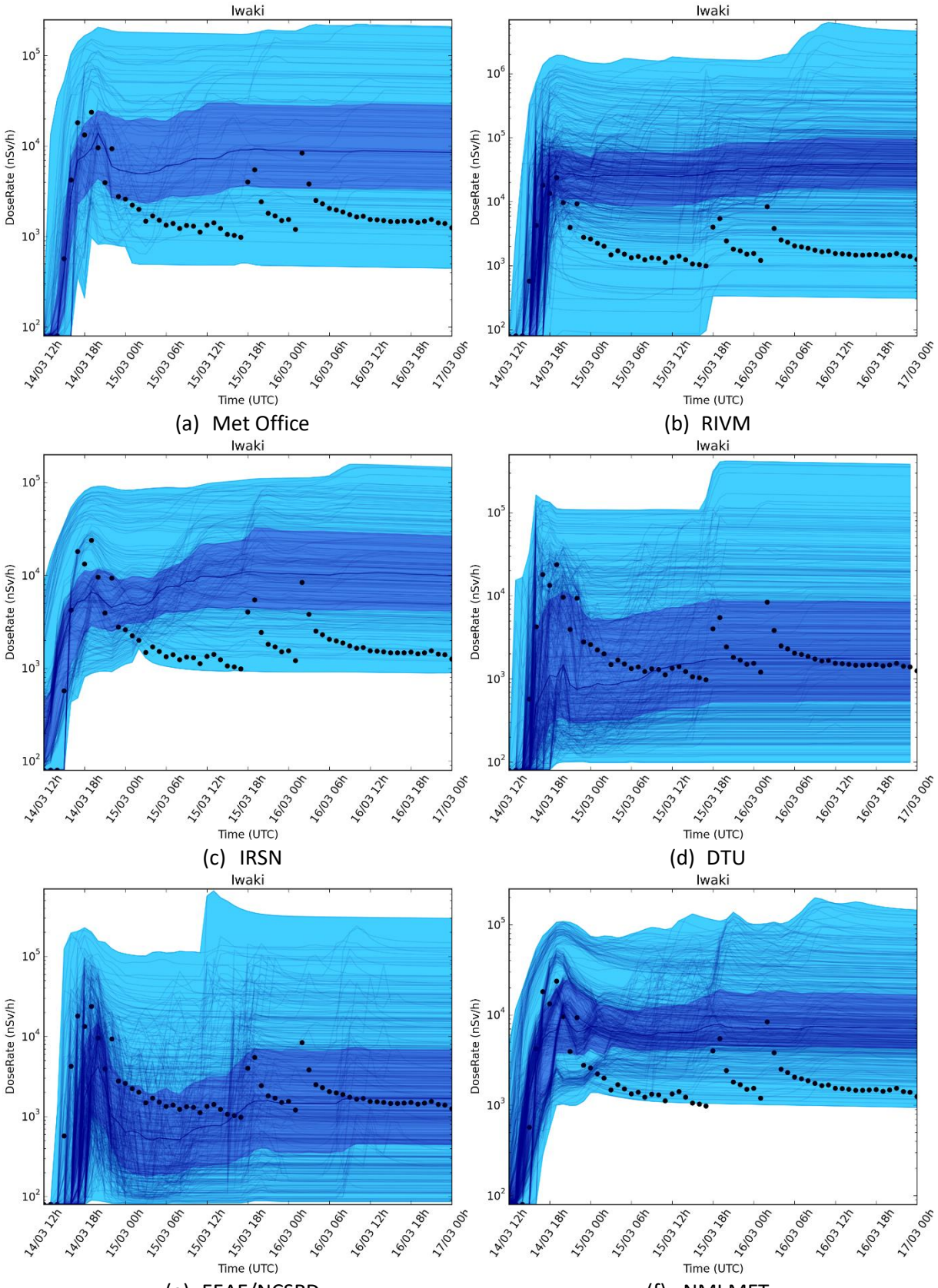


Figure 70 : Ensemble results of dose rate (nsv/h) at Iwaki (42 kilometres south-west of FDNPP) for the Fukushima case study for several project participants, compared with observations (black dots). The y-axis are not the same on all figures.

Statistical indicators

The Fukushima case allows for the presentation and discussion of additional outputs and indicators, not designed for decision making, but oriented toward ensemble verification. These indicators are briefly presented in this section.

Rank histograms

In a synthetic way, the rank histogram is a good method to estimate if an ensemble is representative of the observed uncertainty. The rank of an observation is determined by counting how many members of the ensemble are below this observation (Figure 71). Therefore, observations of rank zero are below all members, etc. The rank histogram shows the number of observations of each rank (ranks range between zero and the number of members in the ensemble), counted on all stations and time steps. Theoretically, a "perfect" rank histogram would be flat, indicating no bias in the model results.

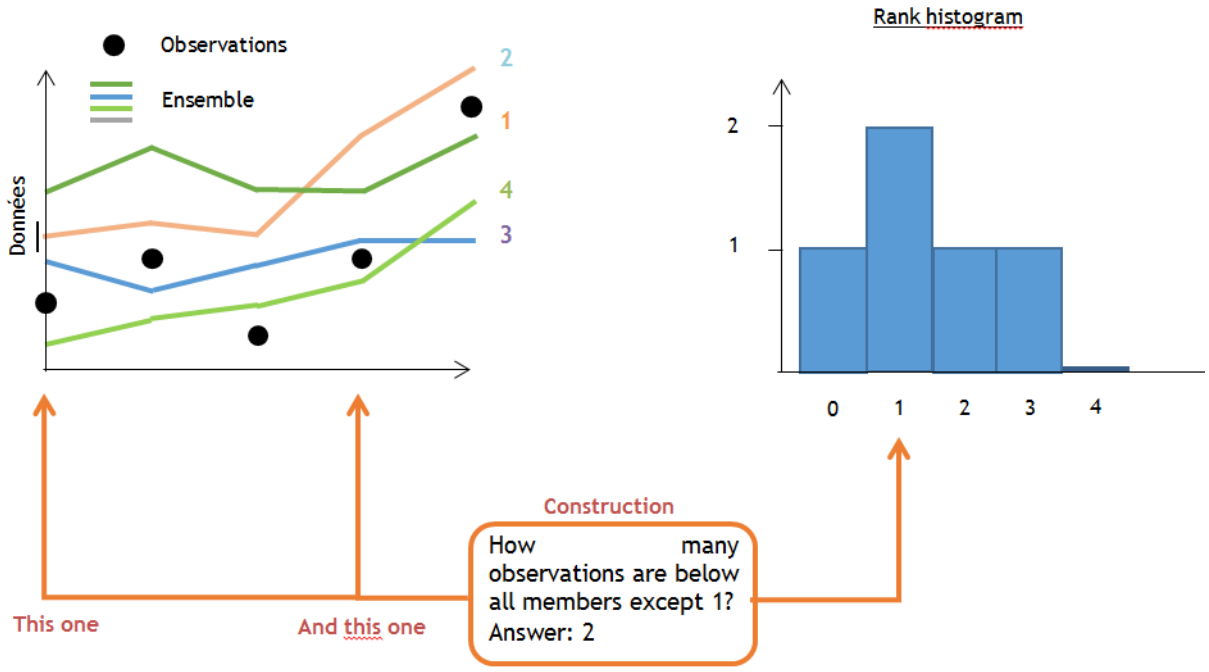


Figure 71 : Construction of a rank histogram.

The rank histograms of ^{137}Cs air concentrations (Figure 72) show a bias towards underestimation (the right-side bar is higher) for all participants. This was not obvious when looking at time series, although at some stations, there are peaks not encompassed by the ensemble (mostly on March 16th). It should also be noted that the look of the diagram is slightly different when there are more members in the ensemble; for instance, for DTU and RIVM the x-axis ranges from 0 to 460 while for Met Office there are only 102 simulations; therefore, the "bins" of the histogram are wider for the latter, since the number of observations is the same for all participants. It should also be noted that these rank diagrams are much better than those presented in a previous study with IRSN's model IdX, on the same dataset, with an ECMWF at a cruder spatial resolution (Le et al. 2017). However, those rank diagrams were made on the full duration of the accident (3 weeks).

An important point related to the rank diagram is whether the observation dataset is truly representative of what we are trying to simulate with the ensemble. For instance, Figure 53 shows that there is a lot of air concentration stations located in the southern area, while few of them are in the highly contaminated north-western area. The rank diagram gives the same weight to all observations, regardless of their actual value. Therefore, the rank diagram is not necessarily representative of the ensembles' performance, for instance, for the high contamination area. The use of a threshold on the

observations and/or the selection of a suitable subset of stations might be a useful complement for further analysis, with the drawback that enough observations have to be left to provide a useful rank diagram.

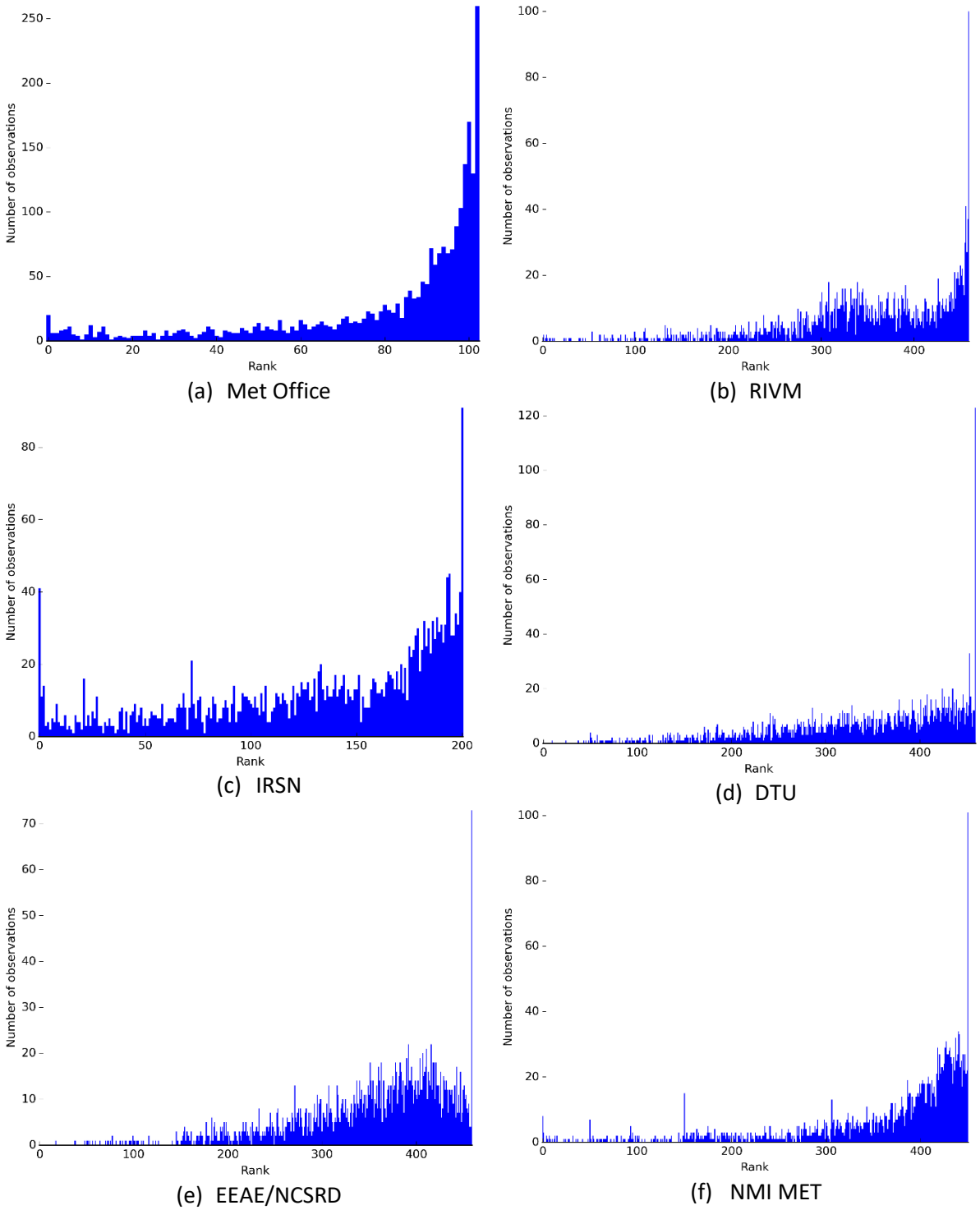


Figure 72 : Rank histogram for observations of ^{137}Cs air concentrations on all stations, for 6 participants.

The rank diagrams for gamma dose rates are shown Figure 73. Here, the rank diagrams are good although not totally flat. However, there is a problem of redundancy in the observations than for air concentrations. Indeed, for each station, the time series of gamma dose rate due to deposition gives several times the same information, the only difference between one time step and the next being the

radioactive decay. This explains the presence of some peaks corresponding to a large number of observations that have the same rank.

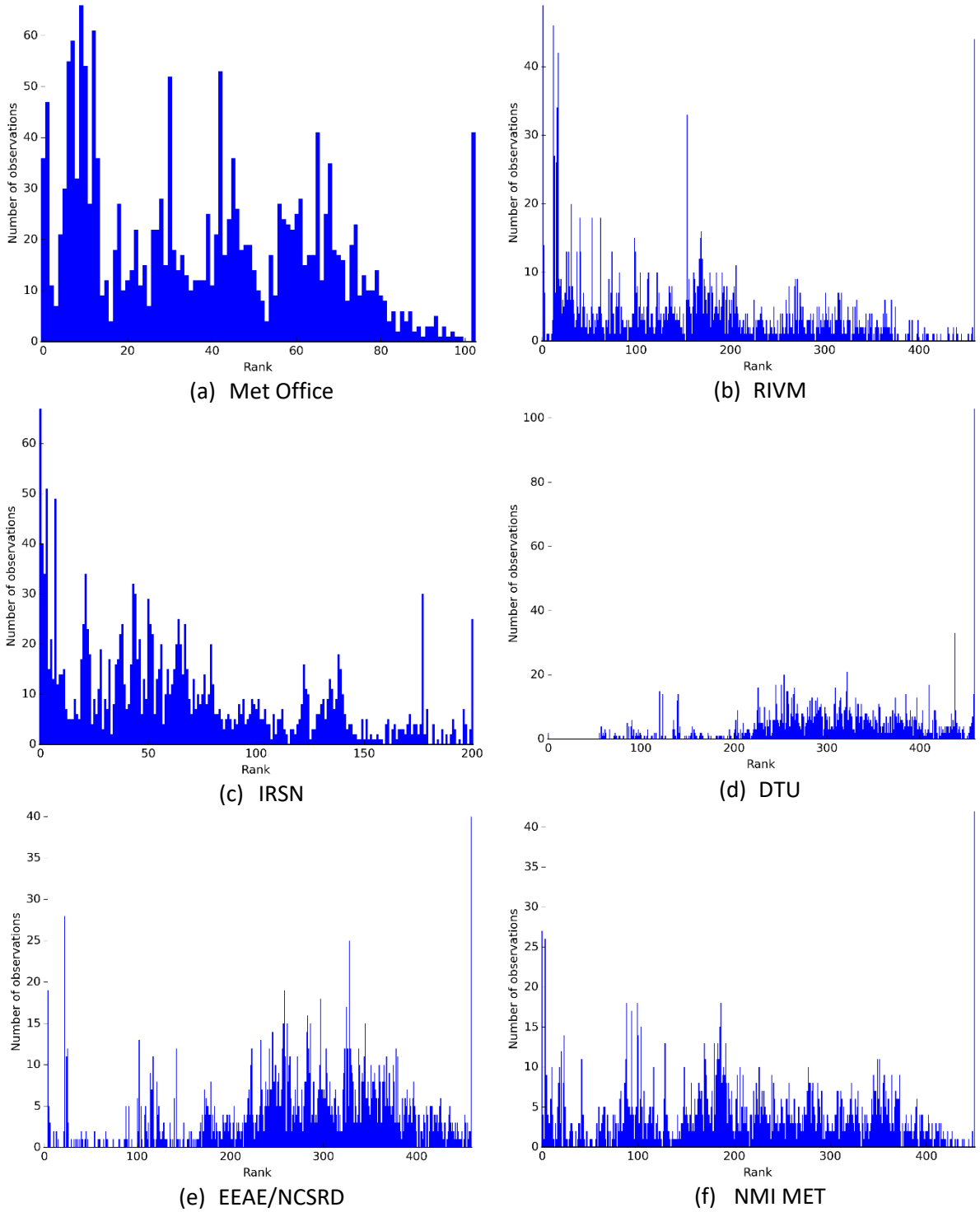


Figure 73 : Rank histogram for observations of gamma dose rates on all stations, for 6 participants.

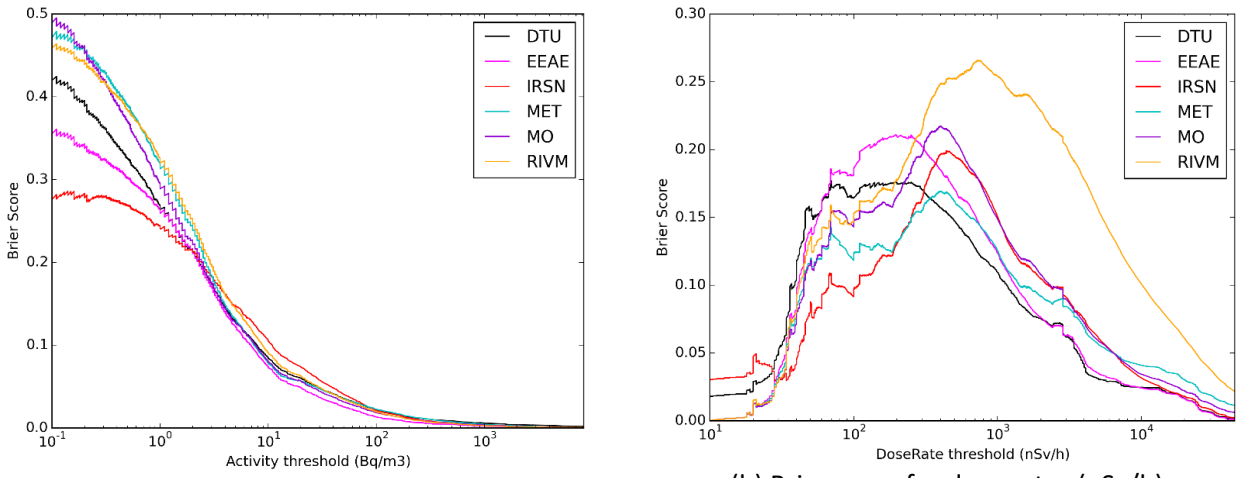
Brier score and CRPS / DRPS

The Brier score is defined as follows:

$$\mathrm{BS} = \frac{1}{N}\sum_{t=1}^{N}(p_i - o_i)^2$$

Where p_i is the probability of exceeding a given threshold according to the ensemble (between 0 and 1) and o_i is the outcome, given by the observation (0 if the threshold is not exceeded, 1 otherwise). Therefore, the closer to zero the score is, the better the ensemble is at accurately forecasting the probability of threshold exceedance.

The Brier score can be plotted for different thresholds. The 'mean' of all Brier scores for a threshold's range of variation is called the "continuous rank probability score" (CRPS). It can be obtained by calculating the area below the brier curve, divided by the length of the threshold variation range. In our case, since the variation over the threshold is discrete, the area below the curve is not calculated by an integral, but as a sum over several "bins" that discretize the x-abscissa. Then, the associated score is called the "discrete rank probability score" (DRPS). The Brier score tends toward zero when the threshold increases: for a high enough threshold, all observations as well all simulated values will be below the chosen value, reflecting the lack of events above the threshold rather an actual skill of the ensemble. In the case of a variable such as the ambient gamma dose rate, the Brier score for a very low threshold also tends toward zero, since the background radiation ensures that all observed and simulated values will be above this threshold. This is not necessarily the case for air activity concentrations: the values cannot go below zero, but there can be many occurrences where observations are equal to zero but some members give a positive value, which may result in a high Brier score.



(a) Brier score for air concentration of ^{137}Cs (Bq/m3)

(b) Brier score for dose rates (nSv/h)

Figure 74 : Brier score for 6 participants, for ^{137}Cs concentrations (left) and dose rates (right).

The Brier scores for air concentrations (Figure 74(a)) are consistent between the 6 participants. For air concentrations, there is some variability between the participants for lower thresholds (between 0.1 and 1 Bqm^3), where the score ranges from 0.28 to 0.5. Above 1 Bq/m^3, which corresponds to a significant observation value, all results are very close to each other. As far as dose rates are concerned, there is a peak in Brier score, which is reached around 200 nSv/h (EEAE) to 1000 nSv/h (RIVM). Some participants are better when low dose rates are considered (mostly due to dry deposition, or small wet

deposition), others are better for high dose rates, i.e., forecasting peaks of dose rates. For instance, EEAE and DTU have the highest Brier scores (meaning that they have more difficulty in predicting threshold exceedance) for low thresholds, but are among the best scores for thresholds above 1000 nSv/h. Conversely, DTU and, to a lesser extent, IRSN and Met Office, have very good scores for low thresholds and perform less well for high thresholds. Looking at the DRPS, which represents the value averaged over all thresholds, gives an overview of the overall performance of the ensembles. Table 15 shows that the best DRPS (i.e. closest to zero) are obtained by EEAE for air concentration and dose rates, although that was not necessarily visible on the figures, which are drawn in logarithmic scale.

Table 15: DRPS for air concentration and dose rates, for all participants. The best scores are highlighted in green.

DRPS	DTU	EEAE/NCSRD	IRSN	NMI MET	Met Office	RIVM
Air Concentration (× 10^{-3})	4.01	1.54	2.36	2.56	2.22	2.01
Dose Rate (× 10^{-2})	2.98	2.80	3.66	3.04	3.71	11.46

Skill-spread diagram

The skill-spread diagram consists in comparing the "skill" of a reference simulation by comparison to observations, to the "spread" of the ensemble. The rationale behind this is that the e-del-to-observations discrepancy is linked to the model uncertainties, which is precisely what is supposed to be represented by the spread of the ensemble (Jones et al. 2019; Kaufmann and Rüdisühli 2019). Therefore, comparing the Root Mean Square Error (RMSE) of the "deterministic" simulation (the "skill") with the ensemble's spread given by its standard deviation is a way to evaluate whether the uncertainties are properly taken into account in the ensemble construction. The question arises how to properly determine the reference simulation to which the ensemble's spread will be compared. One possibility is to use the ensemble median (or mean) as a reference, which was done here. Indeed, it may be possible to use the control member of the meteorological ensemble (if available) for a reference dispersion simulation, but there is no "reference" source term, and source terms are responsible for a huge part of the overall uncertainty. A good skill-spread diagram would be close to the y = x axis.

Figure 75 shows the skill-spread diagrams of all participants for ^{137}Cs air activity concentrations on all stations. It globally shows that the ensembles' spreads have a tendency to underestimate the error of the ensemble median, especially for high concentrations. This may be related to the fact that the rank diagrams for this variable show a bias (Figure 72). The skill-spread diagrams for dose rates (not shown) feature the same tendency to underestimate the error of the median. However, this indicator is to be taken with care, as the use of a "reference" simulation (as done in meteorology with the control member) would have much more sense. One may, for instance, prefer to draw skill-spread diagrams for each individual source term separately (not shown here).

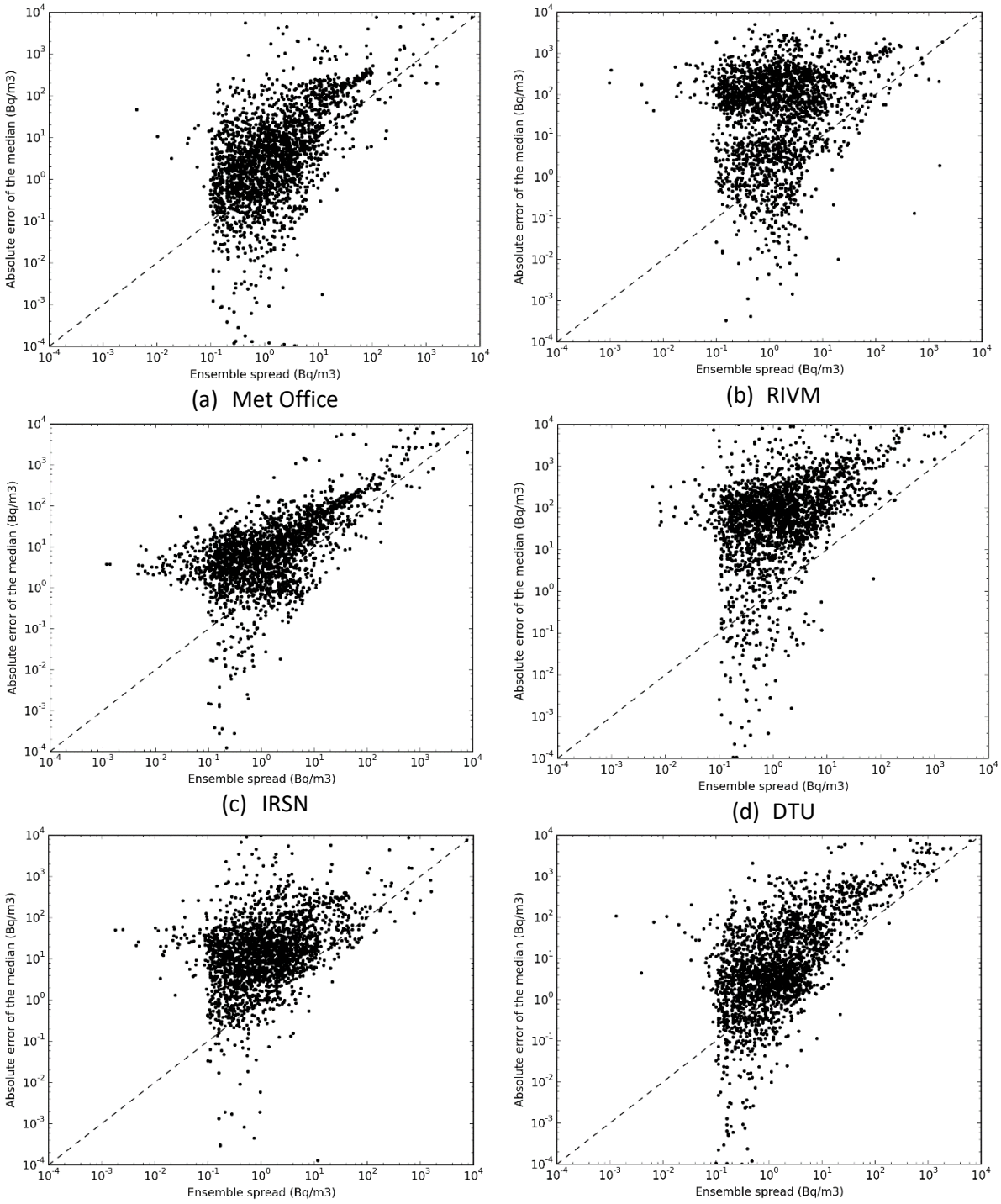
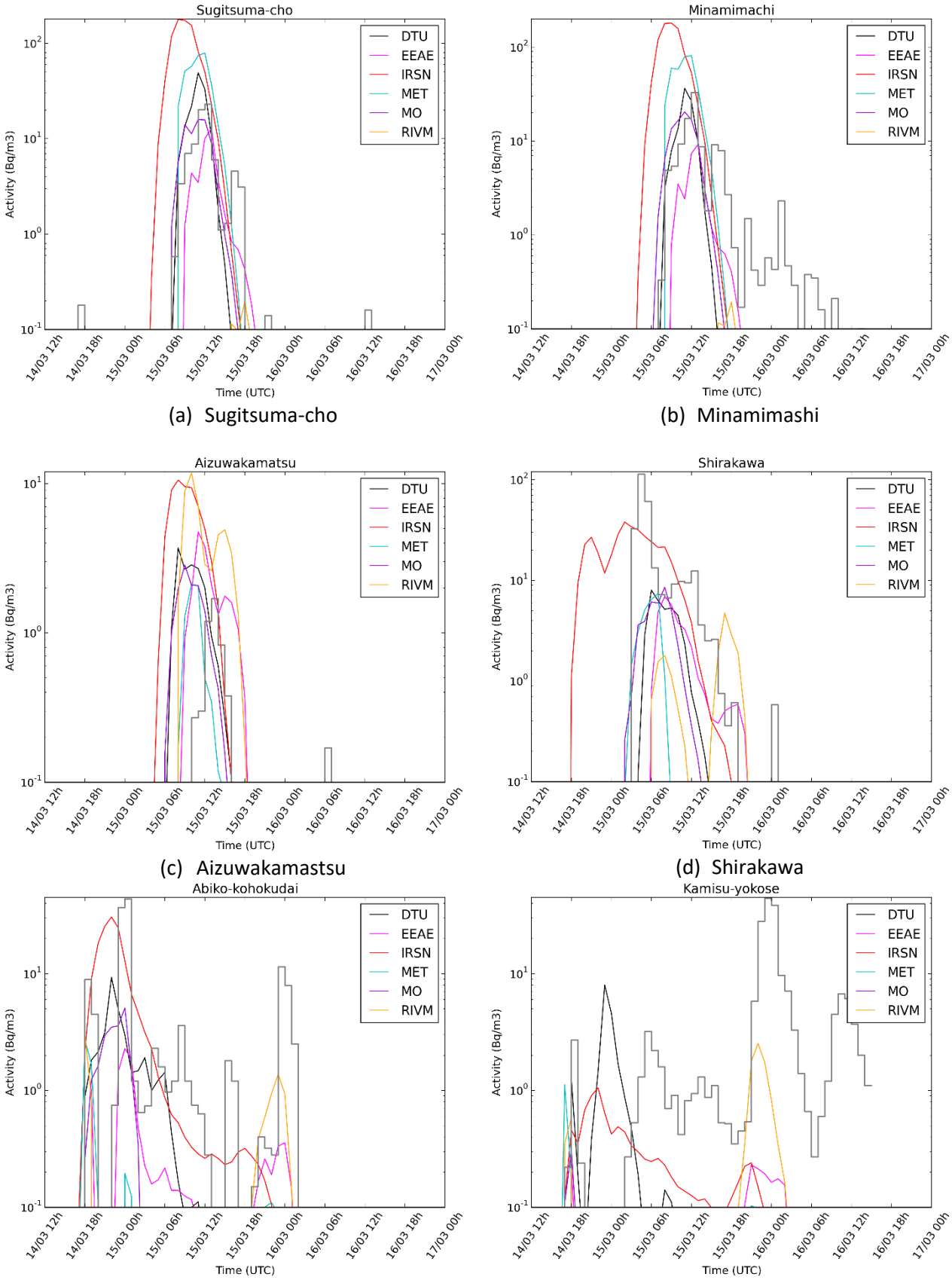


Figure 75 : Skill-spread diagram for observations of ^{137}Cs air concentrations on all stations, for 6 participants.

Performance evaluation of the ensembles median

In addition to the ensemble scores described in the previous sections, it is interesting to compute some deterministic model-to-observations statistical scores. For this, the ensemble median may be used, as a “deterministic” simulation resulting from taking into account all uncertainties. It is interesting to determine whether these scores may be higher than those of a “best estimate”.



(a) Sugitsuma-cho (b) Minamimashi

(c) Aizuwakamastsu (d) Shirakawa

(e) Abiko-kohokudai (f) Kamisu-yokose

Figure 76 : Median of ^{137}Cs activity at 6 stations for the Fukushima case study for several project participant, compared with observations (in grey).

Figure 76 shows the different stations for air activity concentrations that have already been presented in the previous sections. Here, the observations are in grey, and the median of the six participants' ensembles are given in various colours. As noted before, the observations are well represented at Sugitsuma-cho by most participants, but not at Minamimashi, while the two stations are very close to each other (in the same meteorological cell), denoting the influence of subgrid-scale processes not taken into account by the simulations. The ensembles' medians are better at representing the high values of stations located in the north-western area (Figures (a) to (d)) than low values farther south (Figures (e) and (f)). Globally the RIVM ensemble tends to have a very low median compared to the other ensembles on these stations, although it is one of the most widespread ensembles with a good ensemble score on air concentrations.

The statistical scores of the ensemble medians obtained by the six participants are computed in this section. The chosen scores are the Root Mean Square Error (RMSE), Pearson Correlation Coefficient (Corr), Figure of Merit in Time (FMT), and FAC2 (respectively FAC5) which represent the proportion of simulated values that are within a factor 2 (respectively 5) of the observations. These scores are defined below.

If $(o_i)_{i\in[1,N]}$ is the series of all scalar observations and $(s_i)_{i\in[1,N]}$ the values of the simulation corresponding to those observations, then we can define the RMSE and the Pearson correlation by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(s_i - o_i)^2}$$

$$\mathrm{Corr} = \frac{\sum_{i=1}^{N}(s_i - \bar{s_i})(o_i - \bar{o_i})}{\sqrt{\sum_{i=1}^{N}(s_i - \bar{s_i})^2}\sqrt{\sum_{i=1}^{N}(o_i - \bar{o_i})^2}}$$

Also, FAC *N* is defined as the percentage of s_i which are located between $N \times o_i$ and o_i/N.

The Figure of Merit in Time (FMT) is a score based on a threshold. If $\boldsymbol{A}_m$ and $\boldsymbol{A}_p$ are the ensembles of time steps where respectively the measurement and the prediction are above the fixed threshold, and if $|\,.\,|$ is the cardinality of an ensemble (i.e. the number of elements in the ensemble), then:

$$\mathrm{FMT} = 100 \times \frac{|\boldsymbol{A}_m \cap \boldsymbol{A}_p|}{|\boldsymbol{A}_m \cup \boldsymbol{A}_p|}$$

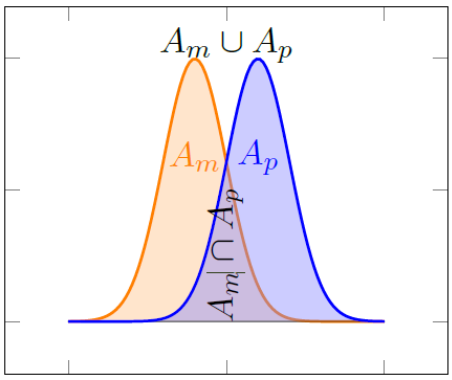


Figure 77 : Illustration of the Figure of Merit in Time between two time series. The abscissa usually represents the time, while the ordinate represents the variable of interest (here, concentration or dose rate). From Le et al. (2019).

In practice, the FMT represents the ability of a simulated time series to overlap the time series of the observations (Figure 77). If, for instance, a peak in the observations is reproduced by the model with an incorrect timing, and there is no overlap of the peaks, then FMT will be zero. In general, a "perfect" model-to-observations comparison (assuming there are no errors) would be given by RMSE=0, Corr=1, FMT=1, FAC2=FAC5=1. The RMSE represents the ability of the model to be unbiased, but gives a large weight to high values, as it is not normalized. The correlation and FMT both represent the ability of the model to correctly represent the temporal variation of the data, regardless of the magnitude of the observations. FAC2 and FAC5 represent the ability of the model to give close values to the observations, but grant the same weight to all observations. The selection of a particular subset of stations could be discussed with regard to whether it is more important to accurately represent the observations close to the source, farther from the source, or in particular regions.

It should be noted that, here, the observation errors are not taken into account, since they are considered negligible compared to the simulations' uncertainties. Table 16 shows the statistical indicators of all participants' median for ^{137}Cs air concentrations, and Table 17 for ambient gamma dose rates. Here, the threshold for FMT is chosen so that all values are above the threshold (i.e. zero). It is interesting to note that the "best score" (highlighted in green) is not obtained by the same participant, depending on the statistical score considered. Overall, four out of six participants obtain the "best score" at least for one score and one variable, denoting that no particular model and configuration is significantly better than the others. Some may be better at forecasting the plume passage, others at simulating deposition; some may be better close to the source, others at farther locations.

Table 16: Ensemble median scores for ^{137}Cs air concentration results. The best score is highlighted in green.

Air concentration scores	DTU	EEAE/NCSRD	IRSN	NMI MET	Met Office	RIVM
RMSE (Bq/m^3) (× 10^2)	5.57	1.83	2.07	1.90	2.54	1.93
Correlation	0.23	0.36	0.22	0.45	0.23	0.26
FMT (%)	41.7	46.5	63.8	26.3	32.6	28.9
FAC 2 (%)	10.3	11.8	18.7	4.9	8.5	6.6
FAC 5 (%)	24.5	26.4	41.7	13.2	18.8	16.2

Table 17: Ensemble median scores for ambient gamma dose rate results. The best score is highlighted in green.

Dose rate scores	DTU	EEAE/NCSRD	IRSN	NMI MET	Met Office	RIVM
RMSE (nSv/h) (× 10^3)	4.95	4.85	4.92	6.70	4.81	9.23
Correlation	0.33	0.34	0.37	0.15	0.38	0.04
FMT (%)	81.3	79.3	97.8	93.5	97.2	89.7
FAC 2 (%)	37.7	38.6	41.4	41.4	43.4	20.1
FAC 5 (%)	74.1	73.3	77.5	83.8	77.1	47.7

It is interesting also to compare these scores with those obtained by "best estimate" configurations. For instance, in Saunier et al. (2016), statistical scores are computed using IRSN's model ldX, various source terms and meteorological data. These scores are computed on the three week period of the Fukushima accident; nonetheless, it may be noted that for ECMWF deterministic data at 0.125° resolution, the FAC5 obtained on air concentration ranges from 25 to 32 depending on the source term used (all of those being included in the simulations made in the present study). Therefore, it is probable that all source terms bring some information and contribute to improving the ensemble median, leading to a better score (for the same dispersion model *ldX*) for the ensemble than for the "best estimate" with any separate source term. It is also interesting to note that the "best" ensemble, in

terms of encompassing all observations, which corresponds to EEAE if CRPS/DRPS is to be chosen as an indicator, is not necessarily the "best" median in terms of scores.

Also, it is worth mentioning that among the source terms presented in Table 13, two of them were constructed using IRSN's model *IdX*. This may bias the results in favour of this model, although the meteorological data used were not the same: ECMWF deterministic meteorology was used to inverse gamma dose rates in Saunier et al. (2013a), while a high-resolution meteorological data provided by the Japan Meteorological Research Institute was used to inverse air concentrations in (Saunier et al. 2016). Therefore, there is no clear evidence that IRSN's ensemble should perform better than others with these source terms. Besides, another study using a similar ensemble of source term and a meteorological ensemble from ECMWF with a more crude resolution, and carried out with the IdX model on the 3 weeks of the accident, showed far lower performance on the air concentration stations (Le et al. 2017; Le et al. 2019).

Conclusions

The Fukushima case study shows a good consistency between the participants' ensembles performance. When using the ECMWF meteorological ensemble and the nine source terms, without additional perturbations, the ensembles show a very large spread and encompass the observations, although there is a bias toward underestimation. Four out of six participants carried out cross-simulations with all possible combinations of source terms and meteorological members (DTU, NMI, RIVM and EEAE), and therefore, the differences between their results only stem from the use of different dispersion models and configurations. Additional perturbations on physical parameters introduced by IRSN do not seem to significantly change the ensembles' spread.

This report presented several indicators to evaluate the performance of an ensemble, that is, its ability to represent the overall uncertainty. The rank diagram is a synthetic way to evaluate how well an ensemble can encompass the observations variability. A good rank diagram is a necessary, but not sufficient condition for an ensemble to be deemed correct. In practice, in meteorology for instance, rank diagrams are never flat, since the sources of uncertainties are never totally and properly taken into account. In that respect, the rank diagrams for air concentrations and gamma dose rates presented here seem to present a good performance, despite a bias on air concentrations. Other indicators are the Brier score and the DRPS; the Brier score represents the ability of the ensemble to forecast the probability of exceeding a given threshold, and the DRPS is its average over all thresholds. The drawback of this score is that it is difficult to interpret in itself, but it is useful to compare several ensembles. Here, the ensembles performance seem good but show some significant variability, denoting that the inter-model variability described in Korsakissok et al. (2019) still has its importance here. Finally, the skill-score diagrams represent the ability of the ensemble to represent the model-to-observation error, which itself is a measure of the uncertainties associated to the simulation. The drawback, here, is that a "reference" simulation has to be defined. Although the meteorological ensemble has a control member, there is no "reference" source term. The skill-score diagrams shown in this report were drawn using the ensembles' medians as reference. They tend to show an underestimation of the modelling errors, meaning that not all sources of uncertainties are taken into account. However, further investigation should be made by taking into account the different source terms separately.

Finally, the performance of the ensembles' medians was assessed by comparison to the observations, using traditional deterministic statistical scores. It is interesting to note that, depending on the score and the variable of interest (gamma dose rate or air concentrations), the "best" ensemble was not the

same. Four out of six ensembles obtained the best score at least for one indicator and one variable. It is also worth mentioning that these scores seem to be better than those obtained with the same model and deterministic configurations with different meteorological data and source terms. Although this finding has to be further analysed, it seems to infer that there is an additional value provided by the ensemble median by comparison to the "best estimates" given by deterministic simulations. In other words, even configurations that are not optimum in themselves provide useful information and can improve the overall results.

Future work includes refining the meteorological ensembles. In particular, here, a single 3-day forecast was used, but, to simulate the three-week period of the accident, it would be necessary to combine several ensemble forecasts together, which raises the issue of temporal continuity between the ensemble members. Other aspects of future research would include working on reducing the number of ensemble members (a topic addressed in Bedwell et al. (2019)) and better taking into account all sources of uncertainties, including those highlighted by the variability between the participants.

Acknowledgement

We acknowledge use of ECMWF's computing and archive facilities for the re-running of the ensemble prediction system.

References

Achim, P., M. Monfort, G. Le Petit, P. Gross, G. Douysset, T. Taffary, X. Blanchard and C. Moulin (2014). "Analysis of Radionuclide Releases from the Fukushima Dai-ichi Nuclear Power Plant Accident Part II." Pure and Applied Geophysics **171**(3): 645-667.

Bedwell, P., I. Korsakissok, S. Leadbetter, R. Périllat, C. Rudas, J. Tomas and J. Wellings (2019). D 9.5.5 Guidelines for the use of ensembles in the description of uncertainty in atmospheric dispersion modelling: operational applications in the context of an emergency response. D9.5 Guidelines for the use of ensemble calculations in an operational context, indicators to assess the quality of uncertainty modeling and ensemble calculations, and tools for ensemble calculation for use in emergency response, European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE project.

Berge, E., H. Klein, M. Ulimoen, S. Andronopoulos, O. C. Lind, B. Salbu and N. Syed (2019). D 9.5.2 Ensemble calculations for the atmospheric dispersion of radionuclides. Hypothetical accident scenarios in Europe: the Western Norway case study. D9.5 Guidelines for the use of ensemble calculations in an operational context, indicators to assess the quality of uncertainty modeling and ensemble calculations, and tools for ensemble calculation for use in emergency response, European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE project.

Chino, M., H. Nakayama, H. Nagai, H. Terada, G. Katata and H. Yamazawa (2011). "Preliminary estimation of release amounts of ^{131}I and ^{137}Cs accidentally discharged from the Fukushima Daiichi Nuclear Power Plant into the atmosphere." Journal of nuclear science and technology **48**(7): 1129-1134.

Hirao, S., H. Yamazawa and T. Nagae (2013). "Estimation of release rate of iodine-131 and cesium-137 from the Fukushima Daiichi nuclear power plant." Journal of nuclear science and technology **50**(2): 139-147.

IAEA. (2012). "Fukushima monitoring database." from https://iec.iaea.org/fmd/search_by_dataset.aspx.

Jones, A., S. Leadbetter and B. Evans (2019). Assessing the performance of a global met ensemble for radiological dispersion. 19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Bruges, Belgium.

Katata, G., M. Chino, T. Kobayashi, H. Terada, M. Ota, H. Nagai, M. Kajino, R. Draxler, M. C. Hort, A. Malo, T. Torii and Y. Sanada (2015). "Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and oceanic dispersion model." Atmos. Chem. Phys. **15**(2): 1029-1070.

Kaufmann, P. and S. Rüdisühli (2019). Dispersion model uncertainty simulation using a limited area ensemble model. 19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Bruges, Belgium.

Kobayashi, T., H. Nagai, M. Chino and H. Kawamura (2013). "Source term estimation of atmospheric release due to the Fukushima Dai-ichi Nuclear Power Plant accident by atmospheric and oceanic dispersion simulations." Journal of nuclear science and technology **50**(3): 255-264.

Korsakissok, I., A. Mathieu and D. Didier (2013). "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.

Korsakissok, I., R. Périllat, G. Geertsema, H. De Vries, T. Hamburger, J. Tomas, S. Andronopoulos, P. Bedwell, T. Charnock, I. Ievdin, S. Leadbetter, T. Pazmandi, C. Rudas, R. Scheele, A. Sogachev, P. Szanto and J. Wellings (2019). D 9.5.1 Ensemble calculations for the atmospheric dispersion of radionuclides. Hypothetical accident scenarios in Europe: the REM case studies. D9.5 Guidelines for the use of ensemble calculations in an operational context, indicators to assess the quality of uncertainty modeling and ensemble calculations, and tools for ensemble calculation for use in emergency response, European Joint Program for the Integration of Radiation Protection Research CONCERT - CONFIDENCE project.

Le, N. B. T., I. Korsakissok, V. Mallet, A. Mathieu, R. Périllat and D. Didier (2019). "Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a Monte Carlo approach, applied to the Fukushima nuclear accident." Atmospheric Environment(To be submitted).

Le, N. B. T., R. Périllat, I. Korsakissok, V. Mallet, A. Mathieu and D. Didier (2017). Uncertainty studies on an atmospheric dispersion model with Monte Carlo method. 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Bologna, Italy.

Leutbecher, M. and T. N. Palmer (2008). "Ensemble forecasting." J. Comp. Phys **227**: 3515-3539.

Liu, Y., J.-M. Haussaire, M. Bocquet, Y. Roustan, O. Saunier and A. Mathieu (2017). "Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima Daiichi accidental releases of radionuclides." Quarterly Journal of the Royal Meteorological Society **143**(708): 2886-2901.

Mathieu, A., M. Kajino, I. Korsakissok, R. Périllat, D. Quélo, A. Quérel, O. Saunier, T. T. Sekiyama, Y. Igarashi and D. Didier (2018). "Fukushima Daiichi–derived radionuclides in the atmosphere, transport and deposition in Japan: A review." Applied Geochemistry **91**: 122-139.

Mathieu, A., I. Korsakissok, D. Quélo, J. Groëll, M. Tombette, D. Didier, E. Quentric, O. Saunier, J.-P. Benoit and O. Isnard (2012). "Atmospheric dispersion and deposition of radionuclides from the Fukushima Daiichi nuclear power plant accident." Elements **8**: 195-200.

Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis (1996). "The ECMWF Ensemble Prediction System: methodology and validation." Q. J. R. Meteorol. Soc. **122**: 73-119.

Oura, Y., M. Ebihara, H. Tsuruta, T. Nakajima, T. Ohara, M. Ishimoto, H. Sawahata and W. Nitta (2015). "A Database of Hourly Atmospheric Concentrations of Radiocesium (134Cs and 137Cs) in Suspended Particulate Matter Collected in March 2011 at 99 Air Pollution Monitoring Stations in Eastern Japan." Journal of Nuclear and radiochemical sciences **15**(2): 1-12.

Quérel, A., D. Quélo, Y. Roustan, A. Mathieu, M. Kajino, T. Sekiyama, K. Adachi, D. Didier, Y. Igarashi and T. Maki (2016). Fukushima: Lessons Learned on Wet Deposition from a Combined Analysis of Radiation Dose Rate and Volume Activity Measurements of 137Cesium. Goldschmidt, Yokohama.

Saunier, O., A. Mathieu, D. Didier, M. Tombette, D. Quélo, V. Winiarek and M. Bocquet (2013a). "An inverse modeling method to assess the source term of the Fukushima nuclear power plant accident using gamma dose rate observations." Atmos. Chem. Phys. Discuss. **13**(6): 15567-15614.

Saunier, O., A. Mathieu, D. Didier, M. Tombette, D. Quélo, V. Winiarek and M. Bocquet (2013b). "An inverse modeling method to assess the source term of the Fukushima Nuclear Power Plant accident using gamma dose rate observations." Atmos. Chem. Phys. **13**(22): 11403-11421.

Saunier, O., A. Mathieu, T. Sekiyama, M. Kajino, K. Adachi, M. Bocquet, T. Maki, T. Higarashi and D. Didier (2016). A new perspective on the Fukushima releases brought by newly available 137Cs air concentration observations and reliable meteorological fields. 17th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. Budapest (Hungary).

Stohl, A., P. Seibert, G. Wotawa, D. Arnold, J. F. Burkhart, S. Eckhardt, C. Tapia, A. Vargas and T. J. Yasunari (2011). "Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition." Atmospheric Chemistry and Physics Discussions **11**(10): 28319-28394.

Terada, H., G. Katata, M. Chino and H. Nagai (2012). "Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part II: verification of the source term and analysis of regional-scale atmospheric dispersion." Journal of Environmental Radioactivity **112**: 141-154.

Tsuruta, H., Y. Oura, M. Ebihara, T. Ohara and T. Nakajima (2014). "First retrieval of hourly atmospheric radionuclides just after the Fukushima accident by analyzing filter-tapes of operational air pollution monitoring stations." Sci. Rep. **4**.

Winiarek, V., M. Bocquet, N. Duhanyan, Y. Roustan, O. Saunier and A. Mathieu (2014). "Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations." Atmospheric Environment **82**(0): 268-279.

Winiarek, V., M. Bocquet, O. Saunier and A. Mathieu (2012). "Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant." J. Geophys. Res. **117**(D5): D05122.

Yumimoto, K., Y. Morino, T. Ohara, Y. Oura, M. Ebihara, H. Tsuruta and T. Nakajima (2016). "Inverse modeling of the 137Cs source term of the Fukushima Dai-ichi Nuclear Power Plant accident constrained by a deposition map monitored by aircraft." Journal of Environmental Radioactivity **164**: 1-12.