Atmospheric
Chemistry
and Physics

Application of linear minimum variance estimation to the multi-model ensemble of atmospheric radioactive Cs-137 with observations

Daisuke Goto[1], **Yu Morino**[1], **Toshimasa Ohara**[2], **Tsuyoshi Thomas Sekiyama**[3], **Junya Uchida**[4], **and Teruyuki Nakajima**[5]

[1]Center for Regional Environmental Research, National Institute for Environmental Studies, Tsukuba, 305-8506, Japan
[2]Fukushima Branch, National Institute for Environmental Studies, Miharu, 963-7700, Japan
[3]Meteorological Research Institute, Tsukuba, 305-0052, Japan
[4]Atmosphere and Ocean Research Institute, University of the Tokyo, Kashiwa, 277-8568, Japan
[5]Earth Observation Research Center, Japan Aerospace Exploration Agency,Tsukuba, 305-8505, Japan

Correspondence: Daisuke Goto (goto.daisuke@nies.go.jp)

Abstract. Great efforts have been made to simulate atmospheric pollutants, but their spatial and temporal distributions are still highly uncertain. Observations can measure their concentrations with high accuracy but cannot estimate their spatial distributions due to the sporadic locations of sites. Here, we propose an ensemble method by applying a linear minimum variance estimation (LMVE) between multi-model ensemble (MME) simulations and measurements to derive a more realistic distribution of atmospheric pollutants. The LMVE is a classical and basic version of data assimilation, although the estimation itself is still useful for obtaining the best estimates by combining simulations and observations without a large amount of computer resources, even for high-resolution models. In this study, we adopt the proposed methodology for atmospheric radioactive caesium (Cs-137) in atmospheric particles emitted from the Fukushima Daiichi Nuclear Power Station (FDNPS) accident in March 2011. The uniqueness of this approach includes (1) the availability of observed Cs-137 concentrations near the surface at approximately 100 sites, thus providing dense coverage over eastern Japan; (2) the simplicity of identifying the emission source of Cs-137 due to the point source of FDNPS; (3) the novelty of MME with the high-resolution model (3 km horizontal grid) over complex terrain in eastern Japan; and (4) the strong need to better estimate the Cs-137 distribution due to its inhalation exposure among residents in Japan. The ensemble size is six, including two atmospheric transport models: the Weather Research and Forecasting – Community Multi-scale Air Quality (WRF-CMAQ) model and non-hydrostatic icosahedral atmospheric model (NICAM). The results showed that the MME that estimated Cs-137 concentrations using all available sites had the lowest geometric mean bias (GMB) against the observations (GMB = 1.53), the lowest uncertainties based on the root mean square error (RMSE) against the observations (RMSE = 9.12 Bq m^{-3}), the highest Pearson correlation coefficient (PCC) with the observations (PCC = 0.59) and the highest fraction of data within a factor of 2 (FAC2) with the observations (FAC2 = 54 %) compared to the single-model members, which provided higher biases (GMB = 1.83–4.29, except for 1.20 obtained from one member), higher uncertainties (RMSE = 19.2–51.2 Bq m^{-3}), lower correlation coefficients (PCC = 0.29–0.45) and lower precision (FAC2 = 10 %–29 %). At the model grid, excluding the measurements, the MME-estimated Cs-137 concentration was estimated by a spatial interpolation of the variance used in the LMVE equation using the inverse distance weights between the nearest two sites. To test this assumption, the available measurements were divided into two categories, i.e. learning and validation data; thus, the assumption for the spatial interpolation was found to guarantee a moderate PCC value (> 0.4) within an approximate distance of at least 70 km. Extra sensitivity tests for several parameters, i.e. the site number and the weighting coefficients in the spatial interpolation, the

time window in the LMVE and the ensemble size, were performed. In conclusion, the important assumptions were the time window and the ensemble size; i.e. a shorter time window (the minimum in this study was 1 h, which is the observation interval) and a larger ensemble size (the maximum in this study was six, but five is also acceptable if the members are effectively selected) generated better results.

1 Introduction

Great efforts have been carried out to simulate atmospheric pollutants, but the spatial and temporal distributions of simulated pollutants are still highly uncertain (e.g. Fuzzi et al., 2015). In contrast, observations are the most reliable method of monitoring the concentrations of atmospheric pollutants with high accuracy, but their spatial networks are usually sporadic. Even if these observations densely cover the target area, they cannot reveal the pathway of pollutants from the source to the sink. To analyse the measurements and deeply understand their behaviours in the atmosphere, we need to improve atmospheric transport models as well as optimal interpolations using observations. To understand the model performance, we have executed model intercomparison projects (MIPs) or multi-model ensembles (MMEs), which provide more reliable results than those by a single model for weather forecasting and climate prediction (e.g. Stensrud et al., 2000). To develop the optimal interpolation, we have also analysed the error and the variance between the simulations and observations to estimate more realistic distributions of the target materials (e.g. Rutherford, 1972; Talagrand, 1997; Robinchand and Ménard, 2014).

The MME technique is applied for weather forecasting (Stensrud et al., 2000; Gneiting et al., 2005), climate projections (Knutti et al., 2010; Taylor et al., 2012), short-lived climate forcer assessments (Lamarque et al., 2013; Myhre et al., 2013), air quality forecasting (Solazzo et al., 2012; Sessions et al., 2015) and atmospheric dispersion predictions (Draxler et al., 2015; Sato et al., 2018). The members of the ensemble are widely spread for the use of various numerical models with perturbed initial conditions and various physical and chemical modules. The ensemble method is generally divided into three types: pure average scheme (equal weighting), weighting scheme with all members and selected scheme with reduced members. The pure average method is a popular method for MME-based climate studies according to the concept of "one model, one vote" (Knutti et al., 2010; Weigel et al., 2010) or evidence of improvements in the concentrations of pollutants in air quality and atmospheric dispersion simulations (McKeen et al., 2005; van Loon et al., 2007; Sessions et al., 2015; Kitayama et al., 2018). The weighting scheme, especially with a relatively smaller ensemble size, can be adopted to eliminate the common biases and improve the ensemble results in the weather forecast

(Krishnamurti et al., 1999), climate studies (Haughton et al., 2015), air quality forecasts (Casanova and Ahrens, 2009) and atmospheric dispersion predictions (Nakajima et al., 2017; Sato et al., 2018). The selected scheme is used in the air quality simulations (Solazzo et al., 2012, 2013; Solazzo and Galmarini, 2015a) and the atmospheric dispersion simulations (Riccio et al., 2012; Solazzo and Galmarini, 2015b). The use of the non-pure average scheme, i.e. weighting and selected schemes, is increasing, and the technique is a useful tool for estimating the reliable results among MMEs (Kioutsioukis et al., 2016).

In this study, the MME with the weighting scheme that minimizes the variance between the simulations and observations is carried out to derive a more realistic distribution of radioactive caesium (Cs-137) at the surface. The Cs-137 in atmospheric particles was emitted from the Fukushima Daiichi Nuclear Power Station (FDNPS) accident in March 2011. Thus far, many atmospheric dispersion models have simulated Cs-137 aerosols (e.g. Chino et al., 2011; Morino et al., 2011; Stohl et al., 2012), and several MIPs were conducted (Science Council of Japan, 2014; Draxler et al., 2015; Sato et al., 2018). Under the MIPs, the MMEs provided reliable results in the assessed models (10 or more). However, ordinary modellers cannot easily carry out such MMEs using only their own models. For such situations, we propose a useful method for limited ensemble size in MMEs by applying an analytical optimization to determine the weights for the ensemble. The optimization is based on a linear minimum variance estimation (LMVE), which is a classical and basic type of data assimilation similar to the Kalman filter (e.g. Talagrand, 1997; Kalnay, 2003).

This approach is unique from other MMEs of other species based on the following four factors. (1) The observed Cs-137 concentration near the surface is available at approximately 100 sites, providing dense coverage of eastern Japan (Tsuruta et al., 2014; Oura et al., 2015). Since plumes including Cs-137 particles are transported and diffused very heterogeneously, dense measurements are essential to capture such plumes. (2) The spatial distribution of Cs-137 distribution is captured relatively easily since Cs-137 is emitted from the point source. For example, the $PM_{2.5}$ distribution is rather difficult to capture in the atmosphere since it is emitted from complex sources and formed from various chemical reactions. (3) MME studies to identify a simple tracer with a high-resolution (3 km horizontal grid) model over complex terrain, such as Fukushima, are still limited. (4) It is very important for people to properly estimate the spatial and temporal distributions of Cs-137 emitted from the FDNPS. The better estimation of the Cs-137 distribution greatly helps us to understand the impacts of inhalation exposure on residents in Japan.

In a previous study, Nakajima et al. (2017), which comprises the basis of our proposed method, was applied using multi-models, including the Weather Research and Forecasting – Community Multi-scale Air Quality (WRF-CMAQ)

model (Morino et al., 2013) and non-hydrostatic icosahedral atmospheric model (NICAM) (Goto et al., 2018), to derive a better Cs-137 distribution. However, the estimation is still uncertain, and its ensemble results were not greatly improved, which was mainly because the results of the original model were still highly uncertain. In addition, Nakajima et al. (2017) did not discuss the availability of the use of LMVE for more than three members and the uncertainty of the relevant parameters in LMVE. After the study of Nakajima et al. (2017), both models were further developed by using a finer horizontal resolution (3 km) grid and by nudging a new meteorological field provided by Sekiyama et al. (2017) with higher accuracy. In this study, the available results were increased to six members, including two atmospheric transport models and two or four sensitivity experiments. This proposed method using more than two members promises to be applicable for MIPs as a new ensemble method. Furthermore, the estimated Cs-137 concentrations are used for the estimation of inhalation exposure of Cs-137 emitted from FDNPS in March 2011 (Takagi et al., 2020).

Section 2 gives a description of two models, WRF-CMAQ and NICAM, including a design of the sensitivity experiments, an explanation of the ensemble method using LMVE, the used measurement datasets with designs to test several assumptions in the proposed ensemble method and statistical metrics for model evaluation. Section 3.1 shows the estimated Cs-137 concentrations and their comparison with the single-model results. Section 3.2 shows the tests used for the assumption of spatial interpolation in the LMVE equation using the distance between the nearest two sites, as shown in Sect. 2.2, by separating the measurement into learning and validation sites. In the proposed method, several parameters are assumed: the size, number and weighting values in the interpolation, the time window in the LMVE and the ensemble size. These parameters are investigated and discussed in Sect. 4. Section 5 shows a conclusion and the implication of this study.

2 Methods

2.1 Description of the two atmospheric transport models

The ensemble size is six, and it includes two different atmospheric transport models, WRF version 3.1 (Skamarock et al., 2008) coupled with CMAQ version 4.6 (Byun and Schere, 2006) and the NICAM (Tomita and Satoh, 2004; Satoh et al., 2008, 2014) coupled with the spectral radiation-transport model for aerosol species (SPRINTARS; Takemura et al., 2005; Goto et al., 2011). According to the rule of Nakajima et al. (2017), these models are hereafter referred to as the W-model and N-model, respectively. The W-model analyses atmospheric processes, such as transport, diffusion and deposition of particles, but it was modified for radioactive particle use, such as Cs-137 emitted from the FDNPS accident in the target area in Japan (Morino et al., 2013). The basic experimental design in this study is widely used in this field, such as in Morino et al. (2013). The N-model is a seamlessly multi-scaled model for air pollutants (Goto et al., 2018) on a global scale with a quasi-uniform grid (Suzuki et al., 2008; Dai et al., 2014), a semi-regional scale with a stretched grid (Goto et al., 2015, 2019) and a perfect regional scale with a diamond grid system (Uchida et al., 2017; Nakajima et al., 2017). The N-model also considers the atmospheric processes of particles and focuses on the target area in this study. The basic experimental design in this study is generally common to Nakajima et al. (2017). Both the W-model and N-model participate in the international MIP for Cs-137 emitted from the FDNPS accident (Sato et al., 2018). The experimental design among the models was harmonized as best as possible; that is, all experiments were carried out by using the same emission inventory in Katata et al. (2015) and nudging the meteorological fields using the operational model for regional weather forecasting around Japan (the non-hydrostatic model, named NHM; Saito et al., 2006) coupled with the local ensemble transform Kalman filter (NHM-LETKF) from Sekiyama et al. (2017), with almost a 3 km grid resolution. Using the W-model and N-model, six experiments were conducted, as shown in Table 1. The W-model was executed in four experiments by considering differences in the meteorological fields used as nudging data, the wet deposition process for Cs-137 and emission scenarios of Cs-137 from Terada et al. (2012). The N-model was performed in two experiments by considering only the difference in the meteorological fields. The hourly Cs-137 concentrations simulated in all the experiments in the lowest layer are linearly interpolated to a 1 km × 1 km grid cell (third mesh grid filled for people's living areas to indicate inhalation exposure in Japan in Takagi et al., 2020) for the ensemble process. The target region is eastern Japan as shown in Fig. 1. The target period is from 11 to 24 March 2011 (Japan Standard Time; JST).

2.2 Ensemble method

One of the optimization methods for the simulated Cs-137 concentration is a multi-model ensemble. When the simulated concentration in model i represents C_i and the number of models is two, the ensemble concentration C_{ens} can be expressed as

$$C_{ens} = a_1 C_1 + a_2 C_2, \qquad (1)$$

where a_i $(i = 1, 2)$ is a weighted coefficient for C_i $(i = 1, 2)$. According to the idea of LMVE used in Nakajima et al. (2017), this study also defines the weighted coefficient as

$$a_i = \left(1/\sigma_i^2\right) / \left(1/\sigma_1^2 + 1/\sigma_2^2\right), \qquad (2)$$

where σ^2 represents a variance between the simulated Cs-137 (C_{sim}) and observed Cs-137 (C_{obs}).

Table 1. Brief model description and the design of the experiments.

Name	W-model (W1)	W-model (W2)	W-model (W3)	W-model (W4)	N-model (N1)	N-model (N2)
Dynamic core	WRF	WRF	WRF	WRF	NICAM	NICAM
Module	CMAQ4.6	CMAQ4.6	CMAQ4.6	CMAQ4.6	Modified SPRINTARS[1]	Modified SPRINTARS[1]
Horizontal grid size (km)	3	3	3	3	3	3
Number of layers (Bottom layer thickness)	34 (19 m)	34 (19 m)	34 (19 m)	34 (19 m)	40 (20 m TS1)	40 (20 m TS2)
Meteorological fields (nudged)[2]	SE17	MSM	SE17	SE17	SE17	MSM
Wet deposition for Cs-137[3]	WSPEEDI	WSPEEDI	CMAQ	WSPEEDI	SPRINTARS	SPRINTARS
Emission scenario of Cs-137[4]	KA15	KA15	KA15	TE12	KA15	KA15

[1] Modified SPRINTARS optimizes SPRINTARS (Takemura et al., 2005) for simulating Cs-137 particles by assuming a one-modal size distribution with a radius centre of 0.24 μm and high hygroscopicity, similar to sulfate (Nakajima et al., 2017). [2] SE17 represents the meteorological fields calculated by NHM-LETKF in Sekiyama et al. (2017). MSM represents mesoscale objective analysis data (MANAL) from the Japan Meteorological Agency (JMA). [3] WSPEEDI is a model for simulating the radioactive materials developed by JAEA (Terada et al., 2008). [4] KA15 and TE12 represent Katata et al. (2015) and Terada et al. (2012), respectively.

σ^2 is defined as

$$\sigma^2 = \sum_{\mathrm{d}S, \mathrm{d}t} (C_{\mathrm{sim}} - C_{\mathrm{obs}})^2, \qquad (3)$$

where $\mathrm{d}S$ is a spatial window, i.e. a specific domain, and $\mathrm{d}t$ is a time window, i.e. a specific period. This formulation is classical and widely used for various subjects, such as data assimilation (e.g. Rutherford, 1972; Talagrand, 1997; Kalnay, 2003), and can be applicable for MMEs as shown in theoretical works (Potempski and Galmarini, 2009; Kioutsioukis and Galmarini, 2014). This method leads to the minimization of the difference between the simulation and observation results without a large amount of computer resources.

Different from the previous study of Nakajima et al. (2017), the number of ensemble members is not only two but also more than two. In this case, Eqs. (1) and (2) are generalized as follows:

$$C_{\mathrm{ens}} = \sum_{i=1}^{N} a_i C_i \qquad (4)$$

$$a_i = \left(1/\sigma_i^2\right) / \sum_{j=1}^{N} \left(1/\sigma_j^2\right), \qquad (5)$$

where N is the number of ensemble members. Equation (5) represents the non-biased conditions and the weight is satisfied as follows:

$$\sum_{j=1}^{N} a_j = 1. \qquad (6)$$

In this study, the term C_i in Eqs. (3) and (4) is the logarithmic scale of Cs-137 concentrations because the Cs-137 concentration is very heterologous and ranges from 0.1 to 100 000 Bq m^{-3} (Tsuruta et al., 2014). When the observed Cs-137 concentration is less than the detection limit, it is assumed to be 0.01 Bq m^{-3} in Eq. (3).

In grids where an observation site or observed data at an observation site are missing, σ^2 is calculated by interpolating σ^2 values at the available observation sites. The interpolation of σ^2 is applied for calculating C_{ens} at all grids in the whole domain. In this study, C_{ens} is not directly interpolated because it varies abruptly in space and time. The interpolation of σ^2 at grid k (σ_k^2) depends on the distance according to the inverse distance weighting (IDW) used in previous studies (e.g. Rutherford, 1972; Hollingsworth and Lönnberg, 1986) and can be expressed as follows:

$$\sigma_k^2 = \sum_{j=1}^{M} \left[\left\{ \left(\frac{1}{r_{j,k}} \right)^m \Big/ \sum_{i=1}^{M} \left(\frac{1}{r_{i,k}} \right)^m \right\} \sigma_j^2 \right], \qquad (7)$$

where $r_{j,k}$ represents the distance between grid j and the grid k, M ($= 1, 2, 3$) is the number of grids used in the interpolation, and m ($= 1, 2$) is the weighting power. The values of M and m are determined by the interpolation methods shown in Table 2. The standard experiment adopts $M = 2$ and $m = 1$; i.e. the weighted coefficient at grid k depends on the inverse distance between the target grid (k) and the two nearest observation grids shown in Table 2 as LIP1. To confirm this assumption, an additional four IDW methods were carried out by changing the values of M and m as shown in Table 2 (discussed in Sect. 4.1). One set of values is the nearest neighbours ($M = 1$ and $m = 1$), which is a popular method for interpolation; however, the spatial pattern of the σ^2 value is not smoothly distributed. In two of the sensitivity tests, the dependence of the σ^2 value on distance is stronger than that of the other method based on an inverse square distance ($m = 2$) to the interpolation.

In Eq. (3), $\mathrm{d}t$ is set to 1 h, which is an assumption and is tested by setting $\mathrm{d}t$ to 3–48 h (discussed in Sect. 4.2); $\mathrm{d}S$

Table 2. Test experiments in the ensemble method for linear interpolation (LIP).

Name	Nearest	LIP1	LIP2	LIP3	LIP4
Number of grids (M)	1	2	2	3	3
Method	Nearest neighbours	LIP	LIP	LIP	LIP
Weighting	No ($m = 1$)	Inverse distance ($m = 1$)	Inverse square distance ($m = 2$)	Inverse distance ($m = 1$)	Inverse square distance ($m = 2$)
Spatial seamlessness	No	Yes	Yes	Yes	Yes

The parameters M and m are defined in Eq. (7).

is set to one grid, i.e. 1 km by 1 km, where the observation site is located. The ensemble size, N, is set to six, which is also tested by changing the number from six to two, three, four and five (discussed in Sect. 4.3). Hereafter, the proposed method is referred to as the "LMVE ensemble method".

2.3 Observation data

The hourly measured Cs-137 concentrations at the surface are directly estimated by using the aerosol sampling tapes of the national suspended particulate matter (SPM) network (Tsuruta et al., 2014, 2018). There are almost 400 SPM sites in eastern Japan, but now 101 sites have available data for Cs-137 (Oura et al., 2015). In this study, the measured Cs-137 data at 100 sites (one site, Futaba, was eliminated because the site is located in the same model grid as FDNPS, which is known as the change-of-support problem; Gotway and Young, 2002) are used for the ensemble. In addition, at an extra site, Tokai (36.45° N, 140.59° E), which covers the missing area of Oura et al. (2015), daily measured Cs-137 concentrations (Furuta et al., 2011) are used. All 101 sites used in this study are plotted using four different colours in Fig. 1. All measurements shown in Fig. 1 (all colour sites) are used as learning data in the LMVE ensemble method as a control experiment (CTL). In the sensitivity tests discussed in Sect. 4, we used all observation sites.

The colours represent the sites used to learn or validate the data in the extra ensemble estimation to examine the effect of the spatial interpolation for the σ^2 value on the MME concentrations. In the test experiments (SEN1, SEN2 and SEN3), some of the datasets are used as learning data and others are used as validation data, as shown in Table 3. The validation data are selected by choosing the sites with the largest number of observed samples as the representative site in the specific domain. When the specific domain is defined by a lattice of 0.125° by 0.125°, the sites in yellow are used as validation data for the experiments (SEN3). When the specific domain is defined as a lattice of 0.25° by 0.25°, the sites

in yellow and green are used as validation data in the experiments (SEN2 and SEN3, respectively). When the specific domain is defined as a lattice of 0.5° by 0.5°, the sites in yellow, green and blue are used as validation data for the experiments (SEN1, SEN2 and SEN3). This means that in the SEN2 experiment, the 56 sites in blue and green can be used as validation data, whereas the 45 sites in red and yellow are used as learning data. Since the area where Cs-137 is emitted from is the point source of FDNPS, a method used to randomly choose the learning/validation sites is not applied for accurately evaluating the spatial interpolation of the ensemble method.

2.4 Statistic metrics for the MME evaluation

The model evaluation should be carried out using multiple statistical metrics (e.g. Chang and Hanna, 2004). In this study, we introduce the geometric mean bias (GMB), root mean square error (RMSE) using the geometric variance (GV), Pearson correlation coefficient (PCC), and the fraction of data within a factor of 2 of observations (FAC2):

$$\text{GMB} = \exp\left(\overline{\log C_{\text{obs}}} - \overline{\log C_{\text{sim}}}\right) \tag{8}$$

$$\text{RMSE} = \ln\left(\text{GV}^2\right) \tag{9}$$

$$\text{GV} = \exp\left\{\overline{\left(\log C_{\text{obs}} - \log C_{\text{sim}}\right)^2}\right\} \tag{10}$$

$$\text{PCC} = \frac{\sum\left(\log C_{\text{obs}} - \log \overline{C_{\text{obs}}}\right)\left(\log C_{\text{sim}} - \log \overline{C_{\text{sim}}}\right)}{\sqrt{\sum\left(\log C_{\text{obs}} - \log \overline{C_{\text{obs}}}\right)^2 \sum\left(\log C_{\text{sim}} - \log \overline{C_{\text{sim}}}\right)^2}} \tag{11}$$

FAC2 = fraction of data that satisfy;

$$0.5 \leq \frac{\log C_{\text{sim}}}{\log C_{\text{obs}}} \leq 2.0. \tag{12}$$

Because Cs-137 concentrations vary by several orders (e.g. Tsuruta et al., 2014), GMB and GV are preferred over the fractional bias method because they are suitable for extreme high and low values (Chang and Hanna, 2004). The RMSE is an indicator of uncertainty. In this study, the PCC is estimated using logarithmically scaled Cs-137 concentrations.
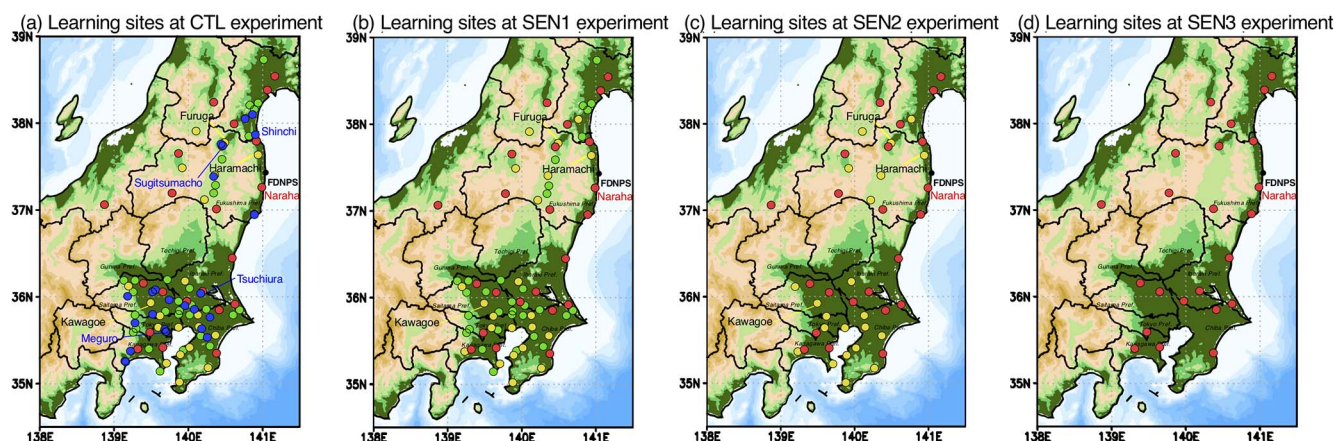
Figure 1. Cs-137 observation sites used as a learning site in **(a)** the standard (CTL), **(b)** SEN1, **(c)** SEN2 and **(d)** SEN3 experiments in this study. The closed black circle is the location of the FDNPS. The closed circles in red, yellow, green and blue are the locations of the learning data sites used in the CTL experiment. The closed circle in yellow is the learning data location used for the ensemble in CTL, SEN1 and SEN2. The closed circle in green is the learning data location used for the ensemble in CTL and SEN1. The closed circle in blue is the learning data location used for the ensemble in the CTL only. The number of sites is 23 (red), 22 (yellow), 32 (green) and 24 (blue). The details are also explained in Table 3. The words in italics are the names of prefectures. The Kantō region includes seven prefectures: Tokyo, Kanagawa, Chiba, Saitama, Ibaraki, Tochigi and Gunma. The background map for the elevation is obtained from the ETOPO (Amante and Eakins, 2009).

Table 3. Test experiments in the ensemble method using the selected sites.

Description		CTL	SEN1	SEN2	SEN3
Number of learning sites		101	77	45	23
Number of validation sites		0	24	56	78
Maximum distance between the learning and validation sites		–	0.125°	0.25°	0.5°
Colours in the used learning sites in Fig. 1	Red	Yes	Yes	Yes	Yes
	Yellow	Yes	Yes	Yes	No
	Green	Yes	Yes	No	No
	Blue	Yes	No	No	No

Because the logarithmic Cs-137 is undefined for zero values, a minimum threshold of logarithmic Cs-137 is set to the lower threshold in the measurement, i.e. $0.01\,\mathrm{Bq\,m^{-3}}$. The PCC using logarithmically scaled values is actually not a robust measure because the lower limitation is sensitive to the results. However, the PCC using logarithmically scaled values is a useful indicator because it becomes applicable even for highly skewed distributions; therefore, we also use the PCC for the model evaluation. FAC2 becomes the most flexible metric for evaluating the model results even if its probability distribution frequency is highly skewed, and it can be an indicator of precision. These statistical metrics are calculated for the case in which the observed Cs-137 exceeds the detection limit. The total sampling number is 7056 for all available sites and times in the CTL experiment, whereas it is 1865 (minimum value) for the SEN1 experiment. Because the sampling number is adequately large, a direct comparison of these statistical metrics among the different experiments can be performed.

3 Results

3.1 LMVE ensemble method using all available observations

Cs-137 simulated by parts of the ensemble six members, i.e. W1 and N1, is evaluated under an MIP (Sato et al., 2018). The performances of these two members are moderate among the MIP-participating models. Here, all of the ensemble members and the ensemble results are compared with the measurements of the surface Cs-137 concentrations. Figure 2 shows the temporal variation in both simulated and observed Cs-137 at the sites near FDNPS and in the Kantō region. Generally, the ensemble results at these sites are the closest to the observations compared to the results of the single-member model. For example, at Naraha (Fig. 2a), which is the closest site to FDNPS, the first and second largest peaks in the observed Cs-137 are noteworthy during 15–17 March. The results of the ensemble members are largely dispersed, so the ensemble results become very close to the obser-
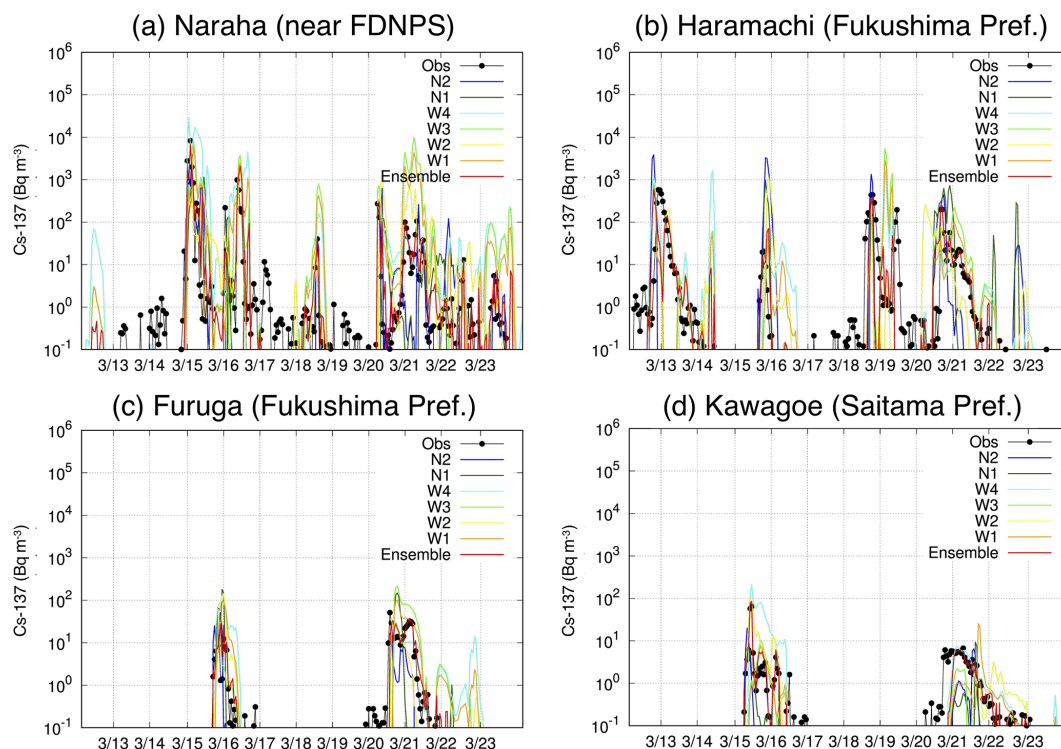
Figure 2. Temporal variations in Cs-137 at the relevant sites (Naraha, Haramachi, Furuga and Kawagoe). The locations in brackets represent the names of prefectures. The results are shown for the observations ("obs" in black), ensemble members (W1, W2, W3, W4, N1 and N2 in colours) and the ensemble model (red). The time is JST.

vations. In the other peaks, such as those on 16 and 20–21 March at Furukawa (Fig. 2c), the ensemble results are almost completely matched with the observations. However, the ensemble results are not close to the observations when all the results of the ensemble members are underestimated compared to the observation, as is the case on 13 March at Haramachi (Fig. 2b). In this case, the ensemble result has a peak on 12 March, which is earlier than the timing captured by the observation. In contrast, on 12 March at Naraha, some of the members have a peak in the simulated Cs-137; thus, the ensemble result also has a small peak, whereas the observation does not have such a peak. These cases are also shown on 20 March at Kawagoe (Fig. 2d), which can be explained as follows: when the Cs-137 simulated by all members is underestimated compared to the observations, the variance in Cs-137 between the observations and the simulations, as defined in Eq. (3), must be too large, and thus the weighted coefficient of the members, as defined in Eq. (5), becomes very small. Because the cross terms of the Cs-137 concentration and the weighted coefficient are small, the Cs-137 concentrations estimated by the ensemble must be underestimated. In contrast, when Cs-137 simulated by some members is overestimated compared to the observations, the weighted coefficient becomes very small. However, because the cross terms of the Cs-137 concentration and the weighted coefficient are not small, the Cs-137 concentrations estimated by

the ensemble are overestimated, which represents one of the disadvantages of the LMVE ensemble method and prevents it from obtaining more accurate ensemble results relative to the observations.

Figure 3 shows scatterplots for the observed and simulated Cs-137 at all sites using the ensemble results (a), the results of each member (b–g) and the median results (h) among the ensemble members. The statistical metrics are listed in Fig. 4, including the bias (GMB), uncertainty (RMSE), correlation (PCC) and FAC2. The perfect model presents values of GMB $= 1$ ($C_{\text{sim}} = C_{\text{obs}}$), RMSE $= 0\,\text{Bq m}^{-3}$, PCC $= 1$ and FAC2 $= 1$ (100 %). The observation values at all 101 sites are used in the LMVE ensemble method as learning data. Compared with the results of the other ensemble members, the ensemble result is the closest to the observations, with GMB $= 1.53$, RMSE $= 10^{1.709}$ ($= 9.42$) Bq m^{-3} (i.e. the lowest uncertainty), PCC $= 0.59$ and FAC2 $= 54\,\%$. The ensemble members produce slightly overestimated results ranging from GMB $= 1.21$ to GMB $= 2.54$ (for W1, W2, W3 and W4) and considerably overestimated results ranging from GMB $= 4.20$ to GMB $= 4.30$ (for N1 and N2), and they have low to moderate correlations ranging from PCC $= 0.29$ to PCC $= 0.45$ and uncertainties ranging from $10^{1.283}$ ($= 19.2$) to $10^{1.709}$ ($= 51.2$) Bq m^{-3}. The median result using all ensemble members has a moderate correlation of PCC $= 0.42$ but a high GMB of 4.39 and high RMSE
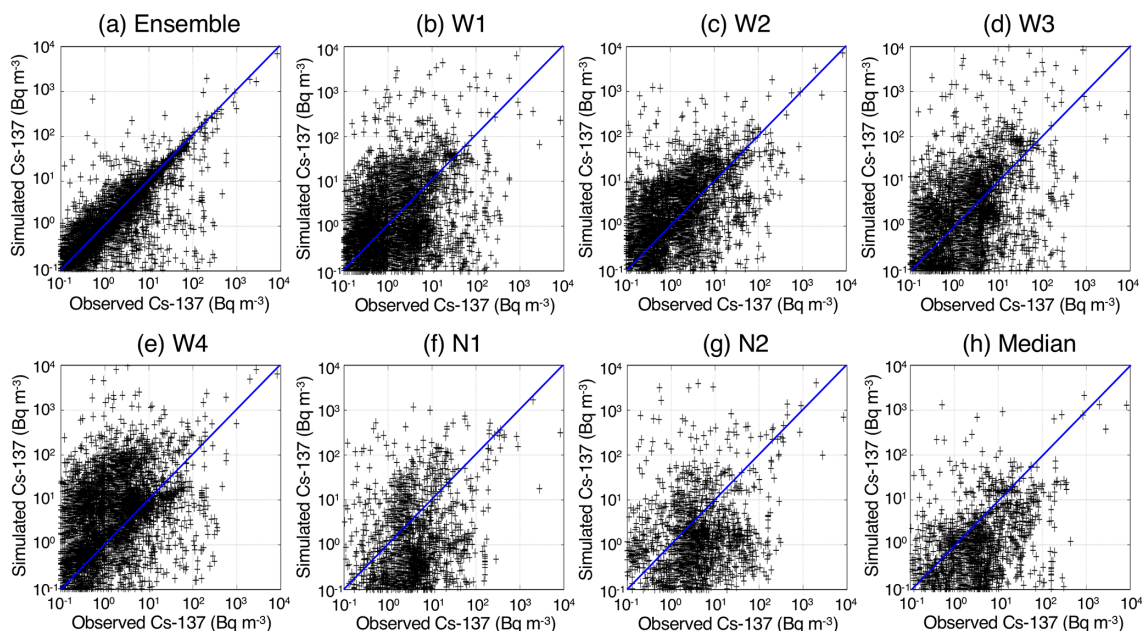
Figure 3. Relationship between the simulated Cs-137 and the observed Cs-137 at all available sites using the **(a)** ensemble model with all models (W1, W2, W3, W4, N1 and N2), **(b–g)** original one-member model (W1, W2, W3, W4, N1 and N2) and **(h)** median model using all models (W1, W2, W3, W4, N1 and N2). The blue line is the 1 : 1 line.
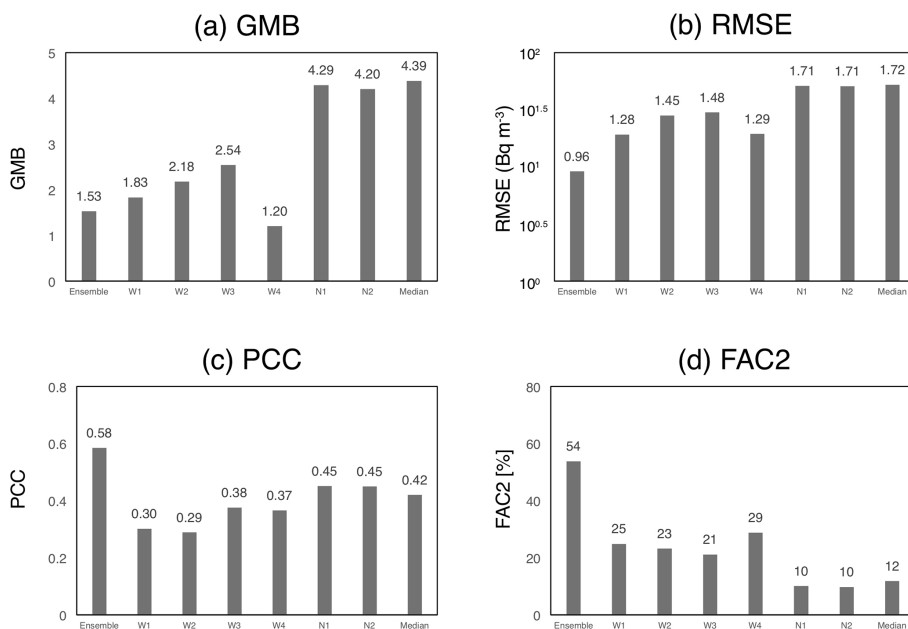
Figure 4. Statistical metrics defined in Sect. 2.4 using the simulated Cs-137 and observed Cs-137 at the available 101 sites. The metrics show the **(a)** GMB, **(b)** RMSE, **(c)** PCC and **(d)** FAC2. The results correspond to Fig. 3.

of $10^{1.717}$ ($= 50.7$) Bq m^{-3}. The median and average values using many members is generally close to the best estimate compared to the original models (e.g. Draxler et al., 2015), although in this study, the median among the six members does not provide the best results.

3.2 LMVE ensemble method using limited observations

In Sect. 3.1, the results of the CTL are shown, and in this section, the sensitivity tests (SEN1, SEN2 and SEN3 as shown in Sect. 2.3) for the separation of learning and validation sites
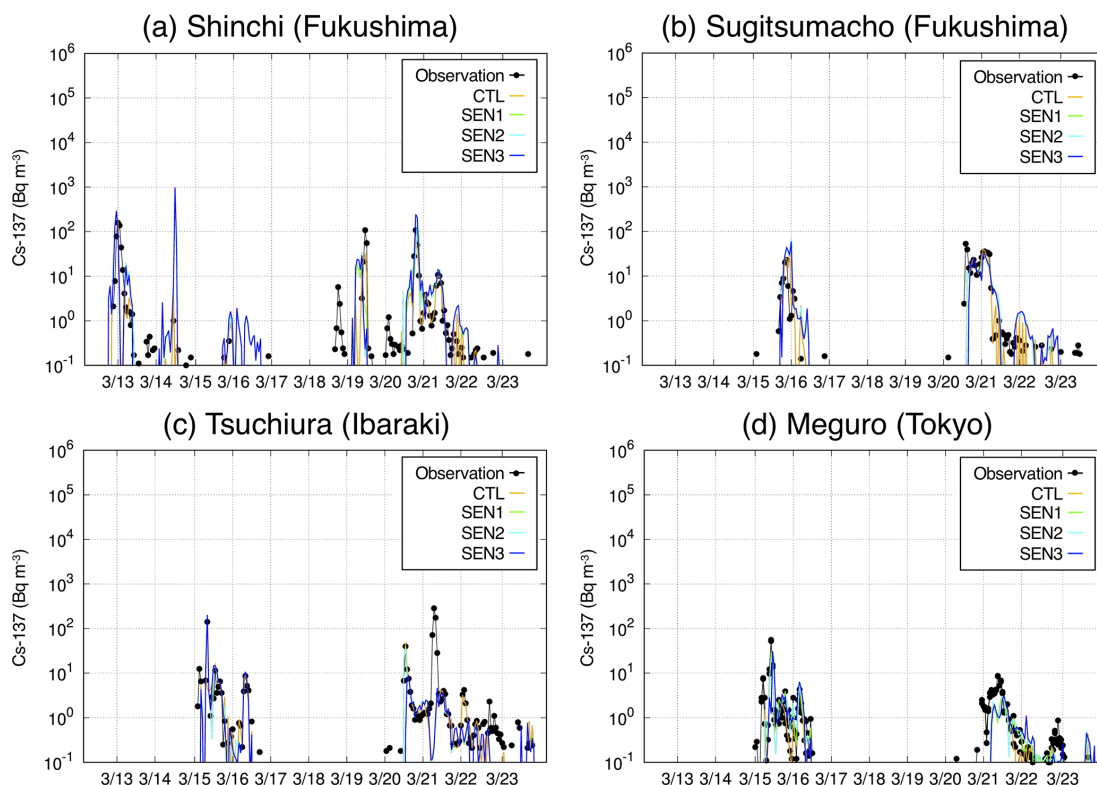
Figure 5. Temporal variations in Cs-137 at the independent sites (not learning but validation sites) using the LMVE ensemble method for CTL, SEN1, SEN2 and SEN3. The time is JST.

are conducted. The test results are evaluated at the sites that are independent (used as validation data) from the other sites (used as learning data) in the LMVE ensemble method (Table 3). At the independent sites, the temporal variations in the simulated Cs-137 are compared at the four sites near FDNPS and in the Kantō region (Fig. 5). The sensitivity depends on the location; the results of all sensitivity tests (SEN1, SEN2 and SEN3) are sometimes far from the observations at Fukushima, as shown in Fig. 5a and b, whereas those of all the sensitivity tests (SEN1, SEN2 and SEN3) are generally close to the observations at the sites in the Kantō region (Fig. 5c and d). This suggests that the interpolation of variance (and thus the Cs-137 concentrations) near the FDNPS is sometimes not applicable, which is probably because the plume of high-density Cs-137 near the FDNPS is very narrow and strongly depends on local winds (Nakajima et al., 2017). The wind, especially low wind speeds, tends to influence the results at the observation sites (Weil et al., 1992).

Figure 6 summarizes the statistical metrics among the sensitivity tests. This figure indicates that the GMB and RMSE values are larger and the PCC and FAC2 values are smaller as the distance between the learning and validation sites increases. The figure also shows that the differences in these metrics between the results using all and independent sites are small and thus clearly show the success of the linear interpolation of variance between the simulation and the obser-

vation in the LMVE ensemble method. These results are consistent with the results shown in the Kantō region of Fig. 5c and d, indicating that the largely spread plumes are generally reproduced by the ensemble method. As shown in Fig. 6, the relationship between the two axes, i.e. distance vs. PCC, can be fitted as a linear line with a slope of 0.18 for the results using all sites and 0.36 for the results using only the validation sites. The slope indicates that the PCC decreases by 0.02–0.04 when the distance from the learning site to the validation site increases by 0.1°. According to the approximate line, the distance is calculated to be 1.03° when all sites are used and 0.68° when the independent sites are used to obtain moderate correlation (PCC > 0.4). Therefore, the proposed interpolation is generally applicable for the best estimation within a distance of 0.7–1.0°, i.e. at least 70 km, from the observation site.

4 Discussion

This section discusses the uncertainties caused by several assumptions in the LMVE ensemble method. As described in Sect. 2.2, the spatial interpolation of variance defined in Eq. (3) assumes IDW. The selection of the sites is also uncertain and is investigated in Sect. 4.1. The time window used in Eq. (3) is assumed to be 1 h, which is discussed in Sect. 4.2.
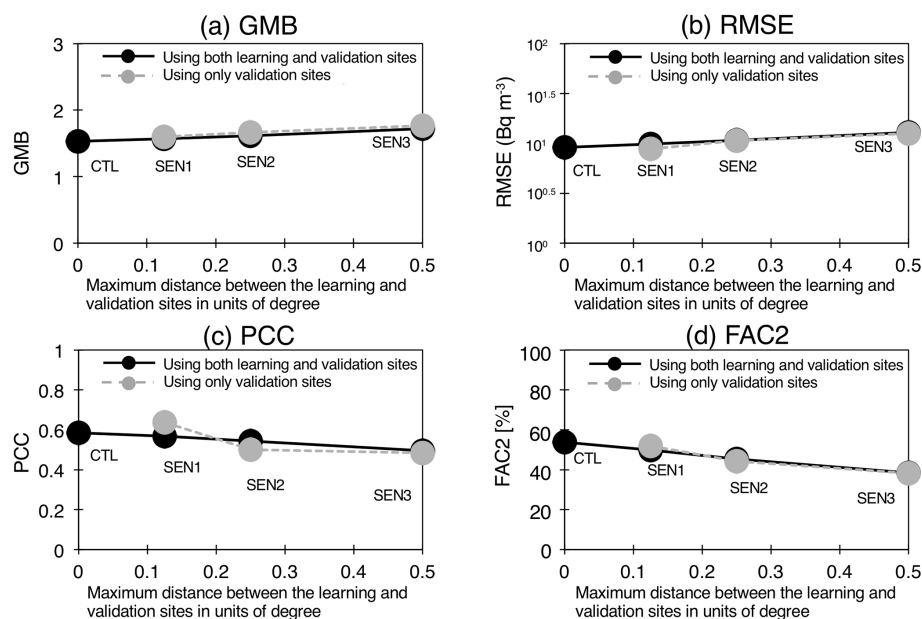
Figure 6. Statistical metrics (GMB, RMSE, PCC and FAC2) at the available sites for CTL, SEN1, SEN2 and SEN3. The statistical metrics are calculated using all sites (in black) and the independent sites (in grey), which are not used in the LMVE ensemble method. The names of the experiments are shown in each panel. The x axis represents a maximum distance between the learning and validation sites in units of degrees.

The ensemble size defined in Eqs. (4) and (5) is also discussed in Sect. 4.3. In Sect. 4.4, the spatial distribution of the Cs-137 surface concentrations estimated by the LMVE ensemble method is shown as an example for estimating the impact of Cs-137 on inhalation exposure of residents in Japan.

4.1 Sensitivity of spatial interpolation

The spatial interpolation of variance defined in Eq. (3) adopts IDW using two of the nearest sites as denoted by LIP1 in Table 2. Here, four extra methods in Table 2 are used for testing SEN1, SEN2 and SEN3, as shown in Sect. 3.2. Figure 7 illustrates the statistical metrics for 15 tests using the validation sites, which are not used in the LMVE ensemble method as learning data. For SEN1, all metrics in the five interpolation methods are estimated as 1.60 ± 0.01 for GMB, $10^{0.95 \pm 0.00}\,\mathrm{Bq\,m^{-3}}$ for RMSE, 0.64 ± 0.00 for PCC and $52 \pm 0.0\,\%$ for FAC2 by using 1865 data samples. The differences among the interpolation methods are close to zero. In SEN2 and SEN3, however, the differences among the five interpolation methods, especially between the nearest neighbours (named "nearest" in Table 2) and the others, become slightly larger than the others. The largest difference is observed between nearest and the others at 0.01 for GMB, $10^{0.01}\,\mathrm{Bq\,m^{-3}}$ for RMSE, 0.01 for PCC and 1 % for FAC2. Judging from the slight differences and the simplest method, we use the LIP1 method (using the two nearest sites and the weighting coefficient of the inverse distance) in the standard experiment. It should be noted that other interpolations were

considered in the discussion, although they provided much worse results than those shown in Fig. 7 (not shown). The method that used the covariance between the observations at the nearest observation site and the simulation at the target grid was not applicable to this study because the covariance values are generally negative or close to zero at most grids due to the heterogenous distribution of Cs-137. Another method that uses all sites (not two or three grids) for the interpolation was also not applicable to this study because the influence of the ensemble coefficients at the grids far from the target grid cannot be ignored.

4.2 Sensitivity of the time window

To investigate the sensitivity of the time window in Eq. (3), the temporal variations in Cs-137 simulated by the ensemble methods are shown in Fig. 8 using various time windows ranging from 1 to 49 h (not all results are shown in Fig. 8). The difference in the estimated Cs-137 concentrations is generally small but sometimes very large. In Fig. 8a, for example, at Naraha on 20 March, the observed peak is sharp, whereas the sharpness of the estimated peaks depends on the time window values. As the time window increases, the sharpness of the peak becomes weak, i.e. the peak is broadly distributed. At Kawagoe (Fig. 8d) on 20–22 March, the estimated Cs-137 concentrations using the longer time window are far from the observations and estimations using the shorter time window. Such situations are found at the other sites and during other periods (not shown). This also in-
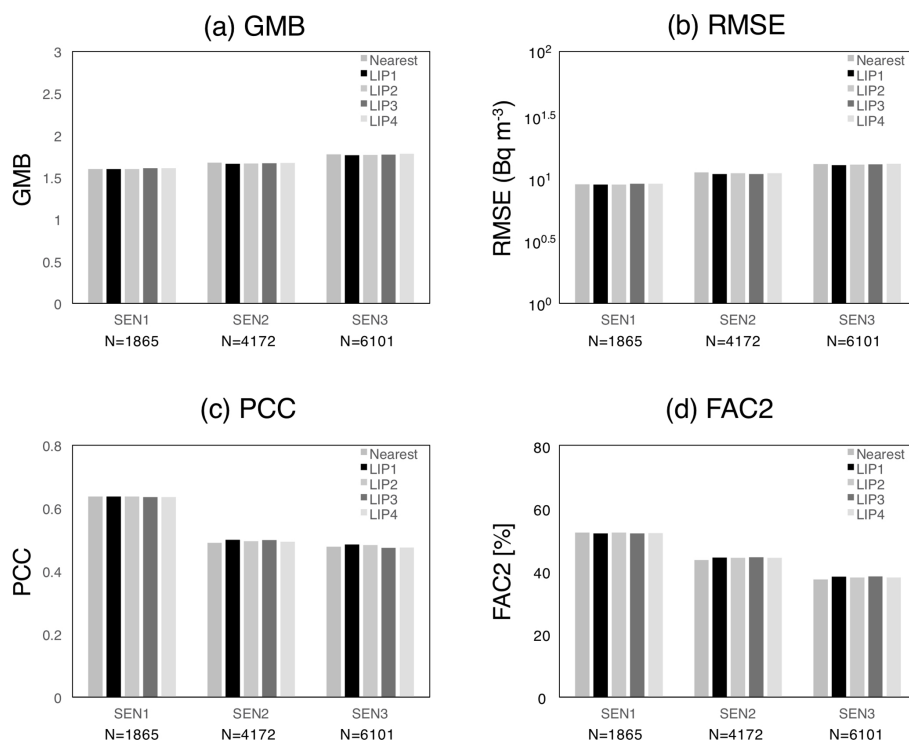
Figure 7. Statistical metrics (GMB, RMSE, PCC and FAC2) for the three sensitivity tests (SEN1, SEN2 and SEN3) as described in Table 3 using the five interpolation methods (nearest, LIP1, LIP2, LIP3 and LIP4) described in Table 2. The x axis represents the sensitivity experiments with the sampling number (N).
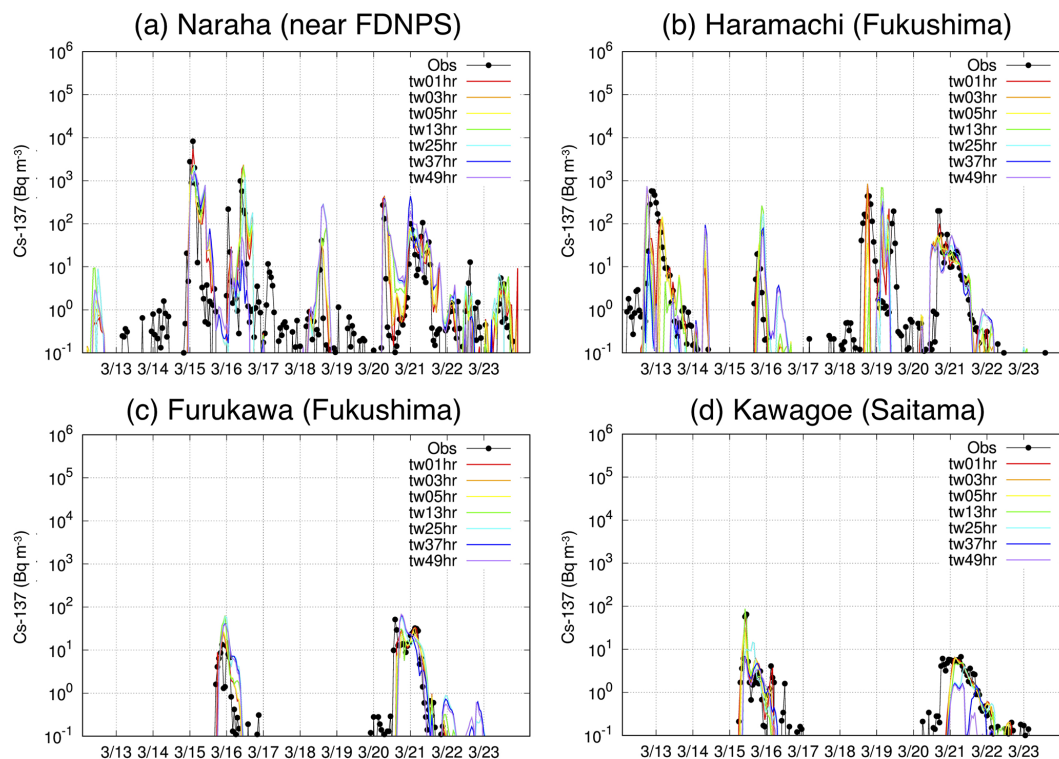
Figure 8. Same as Fig. 2 except for the use of the ensemble results with various time windows ranging from 1 h (tw01hr) to 49 h (tw49hr).

dicates that the peak in Cs-137 is very sharp temporally and spatially, so the time window must be shortened. The dependency of the time window on the results is investigated using the statistical metrics at all sites used in the LMVE ensemble method, as shown in Fig. 9. The dependency of the time window on the GMB, RMSE, PCC and FAC2 was found to be strong; moreover a shorter time window tends to provide higher PCC and FAC2 values and lower GMB and RMSE values. Therefore, the time window in the standard experiment is set to the shortest time, i.e. 1 h.

4.3 Sensitivity of the ensemble size

The previous study of Nakajima et al. (2017) used only two members for the LMVE ensemble method, and the ensemble results were better than the original results for each member, but the difference in the PCC was very small (0.03–0.05; Table 1 in Nakajima et al., 2017). Therefore, this study increases the number of LMVE ensembles to six members and investigates the sensitivity of the ensemble size to the results. Figure 10 shows the relationship between the ensemble size and the statistical metrics (GMB, RMSE, PCC and FAC2). The results clearly show that as the ensemble size increases, the GMB and RMSE decrease, and the PCC and FAC2 increase. This tendency can also be found in previous studies (e.g. Pennell and Reichler, 2010; Kioutsioukis and Galmarini, 2014; Solazzo and Galmarini, 2015a). Using two members, the average GMB is calculated to be 1.95, which is smaller than that obtained using a single member by 0.76; the average RMSE is calculated to be $10^{1.208}$ ($= 16.2$) $\mathrm{Bq\,m^{-3}}$, which is smaller than that obtained using a single member by $10^{1.155}$ ($= 14.3$) $\mathrm{Bq\,m^{-3}}$; the average PCC is calculated to be 0.46, which is larger than that obtained using a single member by 0.08; and the average FAC2 is calculated to be 34 %, which is larger than that obtained using a single member by 14 %. Using more than two members, the PCC is calculated to be more than 0.4, i.e. a moderate correlation. Using five members, the average GMB is calculated to be 1.56, which is larger than the value of 1.53 obtained using six members; the average RMSE is calculated to be $10^{0.984}$ ($= 9.63$) $\mathrm{Bq\,m^{-3}}$, which is larger than the value of $10^{0.960}$ ($= 9.12$) $\mathrm{Bq\,m^{-3}}$ obtained using six members; the average PCC is calculated to be 0.57, which is smaller than the value of 0.59 obtained using six members; and the average FAC2 is calculated to be 51 %, smaller than the value of 54 % obtained using six members. However, the best ensemble results obtained using five members are close to those obtained using six members, with differences of +0.1 % for GMB, +1.5 % for PCC, +0.0 % for RMSE and +0.2 % for FAC2. By contrast, the best ensemble results obtained using four members are worse than those obtained using six members, with the differences of −1.6 % for GMB, −0.9 % for PCC, +0.8 % for RMSE and −4.0 % for FAC2. Therefore, we conclude that the minimum number in the LMVE ensemble in this study is five but only when the members are effectively selected. When the members cannot be selected, the best results can be obtained by reducing the weighting coefficients of the members through the calculation of the LMVE method, since the difference in the statistical metrics between six- and five-member ensembles is very small.

Even in the best estimate using selected five or six members, the PCC value is less than 0.7, which means the ensemble results are moderately (not strongly) correlated with the observations. Therefore, to obtain values much closer to the observations, a new ensemble member is required. As explained in Sect. 3.1, when one of the members provides results close to the observations even at 1 h, the ensemble results proposed in this study become closer to the observations.

For the median and average values using six members, the PCC is calculated to be 0.42 and 0.46, respectively, similar to the ensemble results obtained using two members. By contrast, the GMB, RMSE and FAC2 values for the median and average results obtained using six members are close to the results for the single members. Since the median and average values obtained using many members are generally closer to the best estimate compared to the original members, the original members used in this study are not independent of each other. Therefore, these results indicate that the LMVE ensemble method is applicable even when the ensemble size is only two and even when they are not independent, although the bias, uncertainty, correlation and precision dramatically decrease as the ensemble size increases. The proposed ensemble method is very useful for properly estimating Cs-137 concentrations, even under a limited ensemble size.

4.4 Cs-137 spatial distribution

The above discussion indicates that the LMVE ensemble method can better estimate the Cs-137 distribution; the spatial distributions of Cs-137 concentrations that are integrated daily on 15 March 2011 are also shown (Fig. 11). In the Fukushima prefecture, including the FDNPS, the results of Cs-137 simulated by each member are largely spread, so the ensemble results, especially in the area far from the observation sites, are very important. Figure 5 suggests that the ensemble results are moderately correlated with the observations around the area where the distance from the observation site is approximately 20–30 km. Therefore, in the Fukushima prefecture, which presents a complex terrain, parts of the results of Cs-137 in this area are still uncertain, even when using models with a 3 km horizontal grid, because part of the area is far (30 km away the coast of Fukushima, which is called Hama-dori in Japan), and the inner area is the location of many observation sites (called Naka-dori in Japan). By contrast, although the difference in the simulated plumes among each member is very large over the Kantō region, the conclusion from Sect. 3.2 supports the results that the ensemble Cs-137 results are closer to the observations, with PCC > 0.4, compared to the results of each member. This
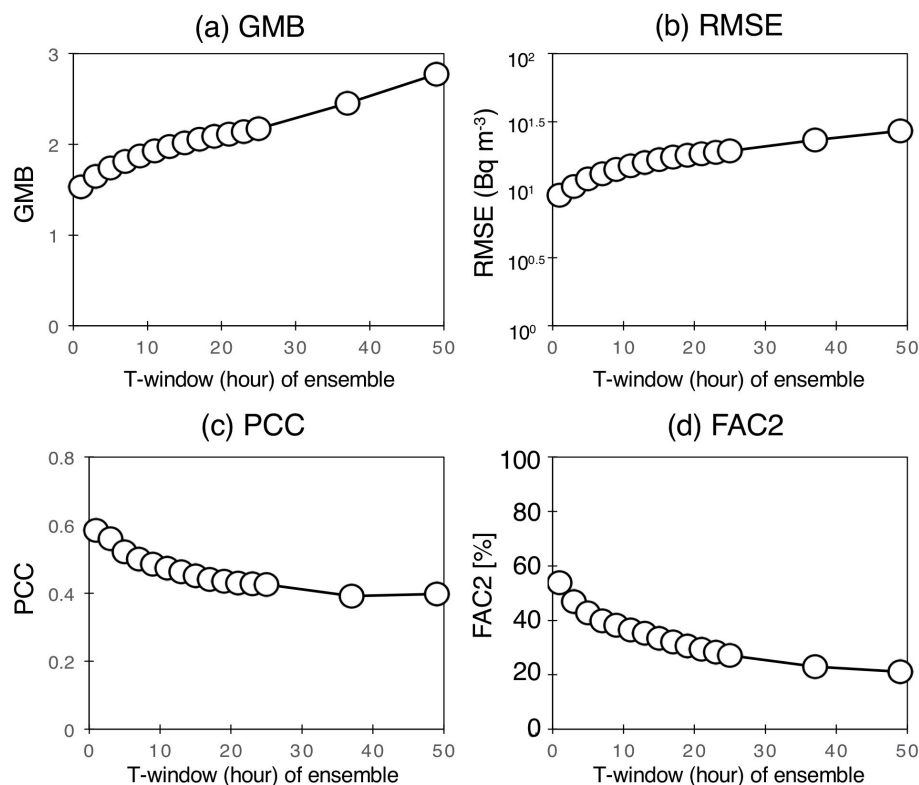
Figure 9. Statistical metrics (GMB, RMSE, PCC and FAC2) at the 101 available sites against various time windows (*x* axis) ranging from 1 to 49 h.

result is obtained because in the Kantō region, most areas are within 70 km of any observation site. However, some of the prefectures in the Kantō region do not have measurement sites, so in these prefectures, the ensemble results are still uncertain. This suggests that in the future, it should be required to observe Cs-137 at distance intervals of 20–30 km distance near the source region and 70 km in other areas to properly estimate the best results of the Cs-137 spatial distribution.

5 Conclusions

The LMVE ensemble method is based on a classical idea but is still useful for estimating the best results using MMEs and observations without requiring a large amount of computer resources for high-resolution models. This method was first applied to estimate the Cs-137 distribution by Nakajima et al. (2017) and is extended in this study. The uniqueness of this approach compared with other MMEs for other species is based on the following: (1) the availability of observed Cs-137 concentrations near the surface at approximately 100 sites, thus providing dense coverage over eastern Japan; (2) the simplicity of identifying the emission source of Cs-137 associated with the point source of FDNPS; (3) the novelty of implementing the MME approach with a high-resolution model over complex terrain in eastern Japan; and

(4) the strong need to better estimate the Cs-137 distribution due to its inhalation exposure risk among residents in Japan. However, Nakajima et al. (2017) did not thoroughly discuss the availability of this method in depth, show the biases, uncertainties, precision and generalizability of this method under varying time windows, space windows and ensemble sizes. Radioactive Cs-137 was released from the FDNPS in March 2011, and many studies have investigated the distribution of Cs-137, but the proper estimations of Cs-137 are not still adequate. Therefore, this study first extended the LMVE ensemble method to an ensemble size of six for simulating Cs-137, including two models, the WRF-CMAQ and NICAM models, and observations and then investigated their uncertainties to confirm their performances to generalize this method and attempt to give the best estimate for the estimation of their inhalation impacts on humans, which is a companion study by Takagi et al. (2020). The results of the ensemble members are also updated from Nakajima et al. (2017) by using a finer horizontal resolution (3 km grid) and by nudging an improved meteorological field provided by Sekiyama et al. (2017).

The proposed LMVE ensemble method provides the best results among the single members of the ensemble. This shows that the MME-estimated Cs-137 concentrations at all available 101 sites have the lowest bias against the observations, with GMB = 1.53; the lowest uncertainties, with
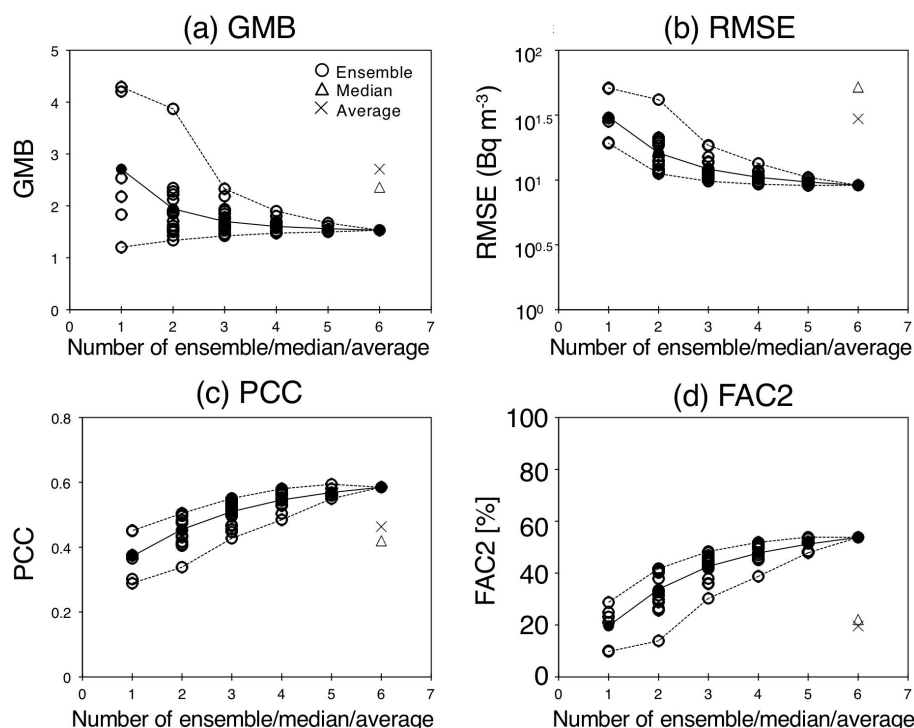
Figure 10. Statistical metrics (GMB, RMSE, PCC and FAC2) at the available sites against the number of the ensemble members, the median and the average (x axis). The black line indicates an average of the ensemble results for each number, whereas the dashed line indicates the maximum and minimum results of the ensemble results for each number.

RMSE $= 9.42\,\mathrm{Bq\,m^{-3}}$; the highest correlation against the observations, with PCC $= 0.58$; and the highest precision against the observations, with FAC2 $= 54\,\%$. Moreover, the single-model members provided higher biases (GMB $=$ 1.83–4.29, except for 1.20 obtained from one member), higher uncertainties (RMSE $= 19.2$–$51.2\,\mathrm{Bq\,m^{-3}}$), lower correlation coefficients (PCC $= 0.29$–0.45) and lower precision (FAC2 $= 10\,\%$–$29\,\%$). In the model grid excluding the observations, Cs-137 is estimated by a spatial interpolation of variance in the formulation of the LMVE using the inverse distance weighting between the nearest two sites. For the test of this assumption, the available measurements are divided into two sets, learning and validation data, and thus the test finds that the assumption for linear interpolation promises a moderate PCC value (>0.4) within a distance of 0.7–1.0°, i.e. at least 70 km. Extra sensitivity tests for several parameters, i.e. the site number and the weighting coefficients in the spatial interpolation, the time window in the LMVE and the ensemble size, are determined. As a result, the findings for the uncertainty in the proposed LMVE ensemble method are shown. (1) The LIP1 method (using two sites and IDW) is the simplest and provides better results than the other interpolations. (2) The time window in the LMVE ensemble method can be set to 1 h. (3) A larger number of ensemble members (ensemble size) remarkably yielded better results, and having more than two members generates better results

than each member alone, even when the members are not completely independent. In this study, the minimum ensemble size is found to be five, but only when the members are effectively selected. The best ensemble size can be six if the weighting coefficient of the member is minimized through the LMVE calculation without selecting any members.

It should be noted, however, that the LMVE ensemble method presents certain limitations. When Cs-137 simulated by all members is too underestimated compared to the observations, the variance in Cs-137 between the observations and the simulations, as defined in Eq. (3), must be too large, and thus the weighted coefficient of the members, as defined in Eq. (5), becomes very small. Because the cross terms of the Cs-137 concentration and the weighted coefficient are small, the Cs-137 concentrations estimated by the ensemble must be underestimated. By contrast, when Cs-137 is overestimated by some members, the weighted coefficient becomes very small. However, because the cross terms of the Cs-137 concentration and the weighted coefficient are not small, the Cs-137 concentrations estimated by the ensemble are overestimated.

In addition, the spatial interpolation used in this study does not obtain a moderate PCC value (>0.4) in the areas where the distance from the observation sites exceeds approximately 70 km. Therefore, the estimated results over the area with very sporadic site locations are very uncertain, es-
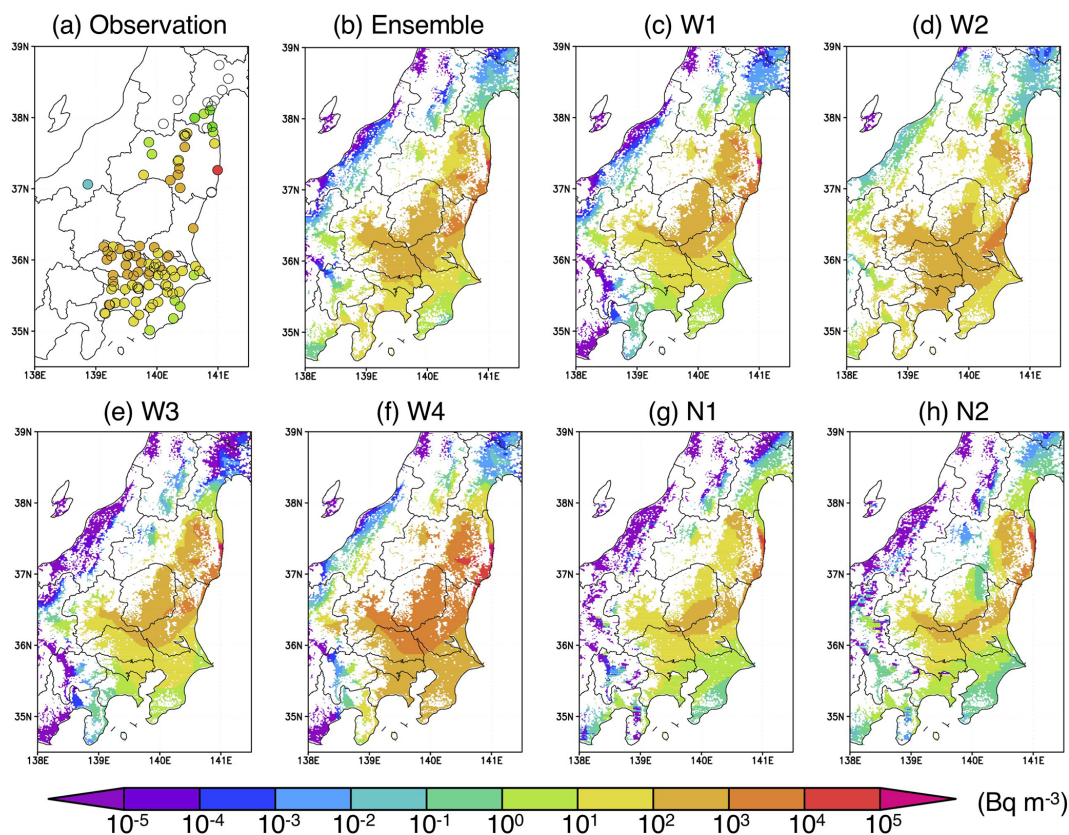
Figure 11. Spatial distribution of the **(a)** observed and **(b–h)** simulated daily integrated Cs-137 concentrations on 15 March 2011 (JST).

pecially for the inner areas of the Kantō region, e.g. Gunma and Tochigi prefectures. The assumption of the spatial interpolation using IDW is difficult to apply to broadly distributed materials, such as Cs-137 emitted from the FDNPS, which is spatially and temporally distributed very heterogeneously (Nakajima et al., 2017). It can be said that it is difficult to use any spatial interpolation, which basically assumes the spatial smoothness of the target's concentrations based on the best estimation of the Cs-137 distribution. In the future, Cs-137 should be observed at distance intervals of 20–30 km near the source region, including in complex terrain, and at intervals of at least 70 km in other areas to properly estimate the best results of the Cs-137 spatial distribution.

This study only applies the LMVE ensemble method to radioactive Cs-137 in the atmosphere, but this method can be applied to Cs-137 deposition and atmospheric pollutants, such as PM$_{2.5}$. However, the results obtained in this study are not directly used for estimation of the best estimate of the Cs-137 deposition because the simulated Cs-137 concentration and the simulated Cs-137 deposition flux are not generally correlated with each other, as suggested by previous model comparison studies (Kitayama et al., 2018). Recently, large areal coverage of surface PM$_{2.5}$ measurements has become available for most countries (https://aqicn.org/map/world/, (last access: 6 March 2020). In addition, since the PM$_{2.5}$

distribution does not vary abruptly, the LMVE ensemble method is easily applied for PM$_{2.5}$ estimations. Therefore, this method will be applied in our future study to estimate the PM$_{2.5}$ distribution for better air quality prediction.

Data availability. The WRF-CMAQ and NICAM model results used to support this article can be obtained from the corresponding author upon request (goto.daisuke@nies.go.jp). The observational data of Cs-137 are freely accessible in Appendix A of Oura et al. (2015) at http://www.radiochem.org/paper/JN152/jn15201_Appendix_A_rev.pdf (last access: 6 March 2020).

Author contributions. DG designed the experiments, conducted the ensemble calculation and drafted the manuscript. YM and JU conducted model simulations. TO, TTS and TN contributed to the discussion of the ensemble results. All authors contributed to writing the manuscript and all of the discussions.

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

Acknowledgements. We acknowledge Haruo Tsuruta and Ya-suji Oura for measuring the Cs-137 concentrations at the monitoring sites and Mai Takagi for checking our ensemble results to estimate the inhalation exposure.

Financial support. This research has been supported by the Environmental Research and Technology Development Fund (fund nos. 5-1501 and 1-1802) of the Environmental Restoration and Conservation Agency, Japan; the Ministry of the Environment, Japan; Japan Society for the Promotion of Science (JSPS) KAKANHI (grant no. JP17H04711); and the Japan Aerospace Exploration Agency (JAXA)/Earth Observation Research Center/GCOM-C.

Review statement. This paper was edited by Jianping Huang and reviewed by two anonymous referees.

References

Amante, C. and Eakins, B. W.: ETOPO1 Global relief model converted to panmap layer format. NOAA-National geophysical Data Center, PANGAEA, https://doi.org/10.1594/PANGAEA.769615, 2009.

Byun, D. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Model-3 Community Multiscale Air Quality (CMAQ) modeling system, App. Mech. Rev., 59, 51–77, 2006.

Casanova, S. and Ahrens, B.: On the weighting of multi-model ensembles in seasonal and short-range weather forecasting, Mon. Weather Rev., 137, 3811–3822, https://doi.org/10.1175/2009MWR2893.1, 2009.

Chang, J. C., and Hanna, S. R.: Air quality model performance evaluation, Meteorol. Atmos. Phys., 87, 167–196, https://doi.org/10.1007/s00703-003-0070-7, 2004.

Chino, M., Nakayama, H., Nagai, H., Terada, H., Katata, G., and Yamazawa, H.: Preliminary estimation of release amounts of 131I and 137 Cs accidentally discharged from the Fukushima Daiichi Nuclear Power Plant into the atmosphere, J. Nucl. Sci. Technol., 48, 1129–1134, 2011.

Dai, T., Goto, D., Schutgens, N. A. J., Dong, X., Shi, G., and Nakajima, T.: Simulated aerosol key optical properties over global scale using an aerosol transport model coupled with a new type of dynamic core, Atmos. Environ., 82, 71–82, https://doi.org/10.1016/j.atmosenv.2013.10.018, 2014.

Draxler, R., Arnold, D., Chino, M., Galmarini, S., Hort, M., Jones, A., Leadbetter, S., Malo, A., Maurer, C., Rolph, G., Saito, K., Servranckx, R., Shimbori, T., Solazzo, E., and Wotawa, G.: World Meteorological Organization's model simulations of the radionuclide dispersion and deposition from the Fukushima Daiichi Nuclear Power Plant accident, J. Environ. Radioactiv., 139, 172–184, https://doi.org/10.1016/j.jenvrad.2013.09.014, 2015.

Furuta, S., Sumiya, S., Watanabe, H., Nakano, M., Imaizumi, K., Takeyasu, M., Nakada, A., Fujita, H., Mizutani, T., Morisawa, M., Kokubun, Y., Kono, T., Nagaoka, M., Yokoyama, H., Hokama, Y., Isozaki, T., Nemoto, M., Hiyama, Y., Onuma, T., Kato, C., and Kurachi, T.: Results of the environmental radiation monitoring following the accident at the Fukushima Daiichi Nuclear Power Plant – Interim Report (Ambient Radiation Dose Rate, Radioactivity Concentration in the Air and Radioactivity Concentration in the fallout), JAEA-Review, 2011-035, 2011.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y., Schaap, M., Slowik, J. G., Spracklen, D. V., Vignati, E., Wild, M., Williams, M., and Gilardoni, S.: Particulate matter, air quality and climate: lessons learned and future needs, Atmos. Chem. Phys., 15, 8217–8299, https://doi.org/10.5194/acp-15-8217-2015, 2015.

Gneiting T., Raftery, A. E., Westveld, A. H., and Goldman, T.: Calibrated Probabilistic Forecasting Using Ensemble Model Out- put Statistics and Minimum CRPS Estimation, Mon. Weather Rev., 133, 1098–1118, 2005.

Goto, D., Nakajima, T., Takemura, T., and Sudo, K.: A study of uncertainties in the sulfate distribution and its radiative forcing associated with sulfur chemistry in a global aerosol model, Atmos. Chem. Phys., 11, 10889–10910, https://doi.org/10.5194/acp-11-10889-2011, 2011.

Goto, D., Dai, T., Satoh, M., Tomita, H., Uchida, J., Misawa, S., Inoue, T., Tsuruta, H., Ueda, K., Ng, C. F. S., Takami, A., Sugimoto, N., Shimizu, A., Ohara, T., and Nakajima, T.: Application of a global nonhydrostatic model with a stretched-grid system to regional aerosol simulations around Japan, Geosci. Model Dev., 8, 235–259, https://doi.org/10.5194/gmd-8-235-2015, 2015.

Goto, D., Nakajima, T., Dai, T., Yashiro, H., Sato, Y., Suzuki, K., Uchida, J., Misawa, S., Yonemoto, R., Trieu, T. T. N., Tomita, H., and Satoh, M.: Multi-scale Simulations of Atmospheric Pollutants Using a Non-hydrostatic Icosahedral Atmospheric Model, in: Land-Atmospheric Research Applications in South and Southeast Asia, edited by: Vadrevu, K., Ohara, T., and Justice, C., Springer Remote Sensing/Photogrammetry, Springer, Cham, 2018.

Goto, D., Kikuchi, M., Suzuki, K., Hayasaki, M.,Yoshida, M., Nagao, T. M., Choi, M., Kim, J., Sugimoto, N., Shimizu, A., Oikawa, E., and Nakajima, T.: Aerosol model evaluation using two geostationary satellites over East Asia in May 2016, Atmos. Res., 217, 93–113, https://doi.org/10.1016/j.atmosres.2018.10.016, 2019.

Gotway, C. A. and Young, L. J.: Combining incompatible spatial data, J. Am. Stat. Assoc., 97, 632–648, https://doi.org/10.1198/016214502760047140, 2002.

Haughton, N., Abramowitz, G., Pitman, A., and Phipps, S. J.: Weighting climate model ensembles for mean and variance estimates, Clim. Dynam., 45, 3169–3181, https://doi.org/10.1007/s00382-015-2531-3, 2015.

Hollingsworth, A. and Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus, 38A, 111–136, 1986.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, 341 pp., 2003 (reprinted with corrections 2004).

Katata, G., Chino, M., Kobayashi, T., Terada, H., Ota, M., Nagai, H., Kajino, M., Draxler, R., Hort, M. C., Malo, A., Torii, T., and Sanada, Y.: Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and

oceanic dispersion model, Atmos. Chem. Phys., 15, 1029–1070, https://doi.org/10.5194/acp-15-1029-2015, 2015.

Kioutsioukis, I. and Galmarini, S.: *De praeceptis ferendis*: good practice in multi-model ensembles, Atmos. Chem. Phys., 14, 11791–11815, https://doi.org/10.5194/acp-14-11791-2014, 2014.

Kioutsioukis, I., Im, U., Solazzo, E., Bianconi, R., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., van der Gon, H. D., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Jorba, O., Manders-Groot, A., Neal, L., Pérez, J. L., Pirovano, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Tuccella, P., Werhahn, J., Wolke, R., Hogrefe, C., and Galmarini, S.: Insights into the deterministic skill of air quality ensembles from the analysis of AQMEII data, Atmos. Chem. Phys., 16, 15629–15652, https://doi.org/10.5194/acp-16-15629-2016, 2016.

Kitayama, K., Morino, Y., Takigawa, M., Nakajima, T., Hayami, H., Nagai, H., Terada, H., Saito, K., Shimbori, T., Kajino, M., Sekiyama, T. T., Didier, D., Mathieu, A., Quélo, D., Ohara, T., Tsuruta, H., Oura, Y., Ebihara, M., Moriguchi, Y., and Shibata, T.: Atmospheric modeling of ^{137}Cs plumes from the Fukushima Daiichi Nuclear Power Plant–Evaluation of the model intercomparison data of the science council of Japan, J. Geophys. Res.-Atmos., 123, 7754–7770, https://doi.org/10.1029/2017JD028230, 2018.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in combing projections from multiple climate models, J. Climate, 23, 2739–2758, https://doi.org/10.1175/2009JCLI3361.1, 2010.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Improved weather and seasonal climate forecasts from multimodel superensemble, Science, 285, 1548–1550, https://doi.org/10.1126/science.285.5433.1548, 1999.

Lamarque, J.-F., Dentener, F., McConnell, J., Ro, C.-U., Shaw, M., Vet, R., Bergmann, D., Cameron-Smith, P., Dalsoren, S., Doherty, R., Faluvegi, G., Ghan, S. J., Josse, B., Lee, Y. H., MacKenzie, I. A., Plummer, D., Shindell, D. T., Skeie, R. B., Stevenson, D. S., Strode, S., Zeng, G., Curran, M., Dahl-Jensen, D., Das, S., Fritzsche, D., and Nolan, M.: Multi-model mean nitrogen and sulfur deposition from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): evaluation of historical and projected future changes, Atmos. Chem. Phys., 13, 7997–8018, https://doi.org/10.5194/acp-13-7997-2013, 2013

McKeen, S., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Hsie, E.-Y., Gong, W., Bouchet, V., Menard, S., Moffet, R., McHenry, J., McQueen, J., Tang, Y., Carmichael, G. R., Pagowski, M., Chan, A., Dye, T., Frost, G., Lee, P., and Mathur, R.: Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, J. Geophys. Res., 110, D21307, https://doi.org/10.1029/2005JD005858, 2005.

Morino, Y., Ohara, T., and Nishizawa, M.: Atmospheric behavior, deposition, and budget of radioactive materials from the Fukushima Daiichi nuclear power plant in March 2011, Geophys. Res. Lett., 38, L00G11, https://doi.org/10.1029/2011GL048689, 2011.

Morino, Y., Ohara, T., Watanabe, M., Hayashi, S., and Nishizawa, M.: Episode analysis of deposition of radiocesium from the Fukushima Daiichi Nuclear Power Plant accident, Environ. Sci. Technol., 47, 2314–2322, 2013.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853–1877, https://doi.org/10.5194/acp-13-1853-2013, 2013.

Nakajima, T., Misawa, S., Morino, Y., Tsuruta, H., Goto, D., Uchida, J., Takemura, T., Ohara, T., Oura, Y., Ebihara, M., and Satoh, M.: Model depiction of the atmospheric flows of radioactive cesium emitted from the Fukushima Daiichi Nuclear Power Station accident, Progress in Earth and Planetary Science, 4, 2, https://doi.org/10.1186/s40645-017-0117-x, 2017.

Oura, Y., Ebihara, M., Tsuruta, H., Nakajima, T., Ohara, T., Ishimoto, M., Sawahata, H., Katsumura, Y., and Nitta, W.: A A Database of Hourly Atmospheric Concentrations of Radiocesium (^{134}Cs and ^{137}Cs) in Suspended Particulate Matter Collected in March 2011 at 99 Air Pollution Monitoring Stations in Eastern Japan, J. Nuclear Radiochem. Sci., 15, 15–26, 2015.

Pennell, C. and Reichler, T.: On the Effective Number of Climate Models, J. Climate, 24, 2358–2367, https://doi.org/10.1175/2010JCLI3814.1, 2010.

Potempski, S. and Galmarini, S.: Est modus in rebus: analytical properties of multi-model ensembles, Atmos. Chem. Phys., 9, 9471–9489, https://doi.org/10.5194/acp-9-9471-2009, 2009.

Riccio, A., Ciaramella, A., Giunta, G., Galmarini, S., Solazzo, E., and Potempski, S.: On the systematic reduction of data complexity in multimodel atmospheric dispersion ensemble modeling, J. Geophys, Res., 117, D05314, https://doi.org/10.1029/2011JD016503, 2012.

Rutherford, I. D.: Data assimilation by statistical interpolation of forecast error fields, J. Atmos. Sci., 29, 809–815, 1972.

Robichaud, A. and Ménard, R.: Multi-year objective analyses of warm season ground-level ozone and PM$_{2.5}$ over North America using real-time observations and Canadian operational air quality models, Atmos. Chem. Phys., 14, 1769–1800, https://doi.org/10.5194/acp-14-1769-2014, 2014.

Saito, K., Fujita, T., Yamada, Y., Ishida, J., Kumagai, Y., Aranami, K., Ohmori, S., Nagasawa, R., Kumagai, S., Muroi, C., Kato, T., Eito, H., and Yamazaki, Y.: The operational JMS nonhydrostatic model, Mon. Weather Rev., 134, 1266–1298, 2006.

Sato, Y., Takigawa, M., Sekiyama, T.T., Kajino, M., Terada, H., Nagai, H., Kondo, H., Uchida, J., Goto, D., Quélo, D., Mathieu, A., Querel, A., Fang, S., Morino, Y., von Schoenberg, P., Grahn, H., Brannström, N., Hirao, S., Tsuruta, H., Yamazawa, H., and Nakajima T.: Model intercomparison of atmospheric ^{137}Cs from the Fukushima Daiichi Nuclear Power Plant Accident: Simulations based on identical input data. J. Geophys. Res.-Atmos., 123, 11748–11765, https://doi.org/10.1029/2018JD029144, 2018.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.: Nonhydrostatic Icosaedral Atmospheric Model (NICAM) for global cloud resolving simulations, J. Comput. Phys., 227, 3486–3514, https://doi.org/10.1016/j.jcp.2007.02.006, 2008.

Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., Noda, A., Yamada, T., Goto, D., Sawada, M., Miyoshi, T., Niwa, Y., Hara, M., Ohno, T., Iga, S., Arakawa, T., Inoue, T., and Kubokawa, H.: The Non-hydrostatic icosahedral atmospheric model: description and development, Prog. Earth Planet. Sci., 1, 18–49, https://doi.org/10.1186/s40645-014-0018-1, 2014.

Science Council of Japan: A review of the model comparison of transportation and deposition of radioactive materials released to the environment as a result of the Tokyo Electric power company's Fukushima Daiichi Nuclear Power Plant accident, Tokyo, Japan, 2014.

Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, Atmos. Chem. Phys., 10, 39–49, https://doi.org/10.5194/acp-10-39-2010, 2010.

Sekiyama, T. T., Kunii, M., Kajino, M., and Shimbori T.: Horizontal resolution dependence of atmospheric simulations of the Fukushima Nuclear Accident using 15 km, 3 km, and 500-m grid models, J. Meteorol. Soc. Jpn., 93, 49–64, https://doi.org/10.2151/jmsj.2015-002, 2017.

Sessions, W. R., Reid, J. S., Benedetti, A., Colarco, P. R., da Silva, A., Lu, S., Sekiyama, T., Tanaka, T. Y., Baldasano, J. M., Basart, S., Brooks, M. E., Eck, T. F., Iredell, M., Hansen, J. A., Jorba, O. C., Juang, H.-M. H., Lynch, P., Morcrette, J.-J., Moorthi, S., Mulcahy, J., Pradhan, Y., Razinger, M., Sampson, C. B., Wang, J., and Westphal, D. L.: Development towards a global operational aerosol consensus: basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME), Atmos. Chem. Phys., 15, 335–362, https://doi.org/10.5194/acp-15-335-2015, 2015.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3. NCAR/TN.475+STR, National Center for Atmospheric Research, Boulder, Colorado, USA, 2008.

Solazzo, E. and Galmarini, S.: A science-based use of ensembles of opportunities for assessment and scenario studies, Atmos. Chem. Phys., 15, 2535–2544, https://doi.org/10.5194/acp-15-2535-2015, 2015a.

Solazzo, E. and Galmarini, S.: The Fukushima-^{137}Cs deposition case study: properties of the multi-model ensemble, J. Environ. Radioactiv., 139, 226–233, https://doi.org/10.1016/j.jenvrad.2014.02.017, 2015b.

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., van der Gon, H. D., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jericevic, A., Kraljevic, L., Miranda, A. I., Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhim, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 60–74, https://doi.org/10.1016/j.atmosenv.2012.01.003, 2012.

Solazzo, E., Riccio, A., Kioutsioukis, I., and Galmarini, S.: Pauci ex tanto numero: reduce redundancy in multi-model ensembles, Atmos. Chem. Phys., 13, 8315–8333, https://doi.org/10.5194/acp-13-8315-2013, 2013.

Stensrud, D. J., Bao, J., and Warner, T. T.: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems, Mon. Weather Rev., 128, 2077–2107, 2000.

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., Tapia, C., Vargas, A., and Yasunari, T. J.: 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, Atmos. Chem. Phys., 12, 2313–2343, https://doi.org/10.5194/acp-12-2313-2012, 2012.

Suzuki, K., Nakajima, T., Satoh, M., Tomita, H., Takemura, T., Nakajima, T. Y., and Stephens, G. L.: Global cloud-system resolving simulation of aerosol effect on warm clouds, Geophys. Res. Lett., 35, L19817, https://doi.org/10.1029/2008GL035449, 2008.

Takagi, M., Ohara, T., Goto, D., Morino, Y., Uchida, J., Sekiyama, T., Nakayama, S., Ebihara, M., Oura, Y., Nakajima, T., Tsuruta, H., and Moriguchi, Y.: Reassessment of early ^{131}I inhalation doses by the Fukushima nuclear accident based on atmospheric ^{137}Cs and ^{131}I/^{137}Cs observation data and multi-ensemble of atmospheric transport and deposition models, J. Environ. Radioact., accepted, 2020.

Takemura, T., Nozawa, T., Emori, S., Nakajima, T. Y., and Nakajima, T.: Simulation of climate response to aerosol direct and indirect effects with aerosol transport-radiation model, J. Geophys. Res., 110, D02202, https://doi.org/10.1029/2004JD005029, 2005.

Talagrand, O.: Assimilation of observations, an introduction, J. Meteorol. Soc. Jpn., 75, 191–209, 1997.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Terada, H., Nagai, H., Furuno, A., Kakefuda, T., Harayama, T., and Chino, M.: Development of worldwide version of system for prediction of environmental emergency dose information: WSPEEDI 2nd version, Transactions of the Atomic Energy Society of Japan, 7, 257–267, 2008 (in Japanese).

Terada, H., Katata, G., Chino, M., and Nagai, H.: Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi nuclear power plant accident. Part II: verification of the source term and analysis of regional-scale atmospheric dispersion, J. Environ. Radioactiv., 112, 141–154, 2012.

Tomita, H. and Satoh, M.: A new dynamical framework of non-hydrostatic global model using the icosahedral grid, Fluid Dyn. Res., 34, 357–400, 2004.

Tsuruta, H., Oura, Y., Ebihara, M., Ohara, T., and Nakajima, T.: First retrieval of hourly atmospheric radionuclides just after the Fukushima accident by analyzing filter-tapes of operational air pollution monitoring stations, Sci. Rep., 4, 6717, https://doi.org/10.1038/srep06717, 2014.

Tsuruta, H., Oura, Y., Ebihara, M., Moriguchi, Y., Ohara, T., and Nakajima, T.: Time- series analysis of atmospheric radiocesium at two SPM monitoring sites near the Fukushima Daiichi Nuclear Power Plant just after the Fukushima accident on March 11, 2011, Geochem. J., 52, 103–121, 2018.

Uchida J., Mori M., Hara M., Satoh M., Goto D., Kataoka T., Suzuki K., and Nakajima T.: Impact of lateral boundary errors on the simulation of clouds with a non-hydrostatic regional climate model, Mon. Weather Rev., https://doi.org/10.1175/MWR-D-17-0158.1, 145, 12, 5059–5082, 2017.

van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., graff, A., Jonson, J. E., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tattasón, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, Atmos. Environ., 41, 2083–2907, 2007.

Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C.: Risks of model weighting in multimodel climate projections, J. Climate, 23, 4175–4191, https://doi.org/10.1175/2010JCLI3594.1, 2010.

Weil, J. C., Sykes, R. I., and Venkatran, A.: Evaluating air-quality models: review and outlook, J. Appl. Meteorol., 31, 1121–1145, 1992.

Remarks from the typesetter

TS1   Please give an explanation of why this needs to be changed. We have to ask the handling editor for approval. Thanks.

TS2   Please give an explanation of why this needs to be changed. We have to ask the handling editor for approval. Thanks.