Model Intercomparison of Atmospheric 137Cs From the Fukushima Daiichi Nuclear Power Plant Accident Simulations Based on Identical Input Data

Y. Sato, M. Takigawa, T.T. Sekiyama, M. Kajino, H. Terada, H. Nagai, H. Kondo, J. Uchida, D. Goto, D. Quélo, et al.

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD029144

Key Points:
- A model intercomparison of the atmospheric dispersion of ^{137}Cs using identical input data was conducted
- Deposition and diffusion were key processes responsible for the differences among the models when the meteorological field was reproduced
- A score-weighted ensemble-mean distribution of the atmospheric ^{137}Cs concentration was created

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Sato,
y-sato@energy.nagoya-u.ac.jp

Citation:
Sato, Y., Takigawa, M., Sekiyama, T. T., Kajino, M., Terada, H., Nagai, H., et al (2018). Model intercomparison of atmospheric ^{137}Cs from the Fukushima Daiichi Nuclear Power Plant accident: Simulations based on identical input data. *Journal of Geophysical Research: Atmospheres*, *123*. https://doi.org/10.1029/2018JD029144

Received 12 JUN 2018
Accepted 4 OCT 2018
Accepted article online 12 OCT 2018

Author Contributions:
Conceptualization: Yousuke Sato, Hiromi Yamazawa, Teruyuki Nakajima
Data curation: Yousuke Sato, Masayuki Takigawa
Formal analysis: Yousuke Sato
Funding acquisition: Hiromi Yamazawa, Teruyuki Nakajima
Investigation: Yousuke Sato, Hiroaki Terada, Yu Morino
Methodology: Yousuke Sato, Tsuyoshi Thomas Sekiyama, Mizuo Kajino, Hiroaki Terada, Haruyasu Nagai, Hiroaki Kondo, Junya Uchida, Daisuke Goto, Denis Quélo, Anne Mathieu, Arnaud Quérel, *(continued)*

Model Intercomparison of Atmospheric ^{137}Cs From the Fukushima Daiichi Nuclear Power Plant Accident: Simulations Based on Identical Input Data

Yousuke Sato[1], Masayuki Takigawa[2], Tsuyoshi Thomas Sekiyama[3], Mizuo Kajino[3], Hiroaki Terada[4], Haruyasu Nagai[4], Hiroaki Kondo[5], Junya Uchida[6], Daisuke Goto[7], Denis Quélo[8], Anne Mathieu[8], Arnaud Quérel[8], Sheng Fang[9], Yu Morino[7], Pontus von Schoenberg[10], Håkan Grahn[11], Niklas Brännström[11], Shigekazu Hirao[12], Haruo Tsuruta[13], Hiromi Yamazawa[1], and Teruyuki Nakajima[14]

[1]Department of Applied Energy, Graduate School of Engineering, Nagoya University, Nagoya, Aichi, Japan, [2]Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan, [3]Meteorological Research Institute, Tsukuba, Japan, [4]Japan Atomic Energy Agency, Tokai, Ibaraki, Japan, [5]National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, [6]Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan, [7]National Institute for Environmental Studies, Tsukuba, Japan, [8]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), Fontenay-aux-Roses, France, [9]Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing, China, [10]Department of Environmental Science and Analytical Chemistry (ACES), Stockholm University, Stockholm, Sweden, [11]The Swedish Defense Research Agency (FOI), Umeå, Sweden, [12]Institute of Environmental Radioactivity, Fukushima University, Fukushima, Fukushima, Japan, [13]Remote Sensing Technology Center of Japan, Minato, Tokyo, Japan, [14]Earth Observation Research Center, Japan Aerospace Exploration Agency, Tsukuba, Japan

Abstract A model intercomparison of the atmospheric dispersion of cesium-137 (^{137}Cs) emitted after the Fukushima Daiichi Nuclear Power Plant accident in Japan was conducted to understand the behavior of atmospheric ^{137}Cs in greater detail. The same meteorological data with a fine spatiotemporal resolution and an emission inventory were applied to all models to exclude the differences among the models originating from differences in meteorological and emission data. The meteorological data were used for initial, boundary, and nudging data or offline meteorological field. Furthermore, a horizontal grid with the same resolution as that of the meteorological data was adopted for all models. This setup enabled us to focus on model variability originating from the processes included in each model, for example, physical processes. The multimodel ensemble captured 40% of the atmospheric ^{137}Cs events observed by measurements, and the figure of merit in space for the total deposition of ^{137}Cs exceeded 80. The lower score of the atmospheric ^{137}Cs than that of the deposition originated from the difference in timing between observed and simulated atmospheric ^{137}Cs. Our analyses indicated that meteorological data were most critical for reproducing the atmospheric ^{137}Cs events. The results further revealed that differences in ^{137}Cs concentrations among the models originated from deposition and diffusion processes when the meteorological field was simulated reasonably well. The models with small deposition fluxes produced higher scores for atmospheric ^{137}Cs, and those with strong diffusion succeeded in capturing the high ^{137}Cs concentrations observed; however, they also tended to overestimate the concentrations.

1. Introduction

The accident at the Fukushima Daiichi Nuclear Power Plant (FDNPP) in Japan, triggered by an earthquake in March 2011, released a number of radionuclides, including cesium-137 (^{137}Cs), into the atmosphere. These radionuclides dispersed widely through atmospheric advection and diffusion, and their deposition resulted in contamination over the land and ocean (Science Council of Japan, 2014). Following the accident, the behavior of the released radionuclides was investigated in a number of studies through observation (e.g., Adachi et al., 2013; Igarashi et al., 2015; Kaneyasu et al., 2012; Oura et al., 2015; Science Council of Japan, 2014; Terasaka et al., 2016; Tsuruta et al., 2014) and modeling (e.g., Chino et al., 2011; Draxler et al., 2015; Kajino et al., 2016; Katata et al., 2012; Morino et al., 2011; Stohl et al., 2012; Takemura et al., 2011; Terada et al., 2012; Yasunari et al., 2011). The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT; 2011) estimated the accumulated deposition densities of cesium (^{134}Cs and ^{137}Cs) using aircraft

Sheng Fang, Yu Morino, Pontus von Schoenberg, Håkan Grahn, Niklas Brännström, Shigekazu Hirao, Haruo Tsuruta, Hiromi Yamazawa

Resources: Yousuke Sato, Masayuki Takigawa, Tsuyoshi Thomas Sekiyama, Mizuo Kajino, Hiroaki Terada, Haruyasu Nagai, Hiroaki Kondo, Junya Uchida, Daisuke Goto, Denis Quélo, Anne Mathieu, Arnaud Quérel, Sheng Fang, Yu Morino, Pontus von Schoenberg, Håkan Grahn, Niklas Brännström, Shigekazu Hirao, Hiromi Yamazawa, Teruyuki Nakajima

Supervision: Masayuki Takigawa, Denis Quélo, Anne Mathieu, Hiromi Yamazawa, Teruyuki Nakajima

Validation: Yousuke Sato, Masayuki Takigawa, Tsuyoshi Thomas Sekiyama, Mizuo Kajino, Hiroaki Terada, Haruyasu Nagai, Hiroaki Kondo, Junya Uchida, Daisuke Goto, Denis Quélo, Anne Mathieu, Arnaud Quérel, Sheng Fang, Yu Morino, Pontus von Schoenberg, Håkan Grahn, Niklas Brännström, Shigekazu Hirao, Hiromi Yamazawa

Visualization: Yousuke Sato

Writing - original draft: Yousuke Sato

Writing – review & editing: Masayuki Takigawa, Hiroaki Terada, Haruyasu Nagai, Hiroaki Kondo, Junya Uchida, Daisuke Goto, Denis Quélo, Anne Mathieu, Arnaud Quérel, Sheng Fang, Yu Morino, Pontus von Schoenberg, Shigekazu Hirao, Haruo Tsuruta, Hiromi Yamazawa

measurements acquired over the eastern parts of Japan. The results indicated that over 3,000 kBq/m^2 of cesium was deposited over areas to the northwest of the FDNPP. It was also reported that the deposition of cesium had occurred over the northern parts of the Tokyo metropolitan area (TMA). Tsuruta et al. (2014) and Oura et al. (2015) retrieved hourly atmospheric ^{137}Cs concentrations by analyzing the measurements of the operational filter sampling of the suspended particulate matter (SPM) network, which covers the TMA and Tohoku area. Through analyses of the SPM network data, it was determined that nine high-concentration events of ^{137}Cs (plumes) occurred during March 2011 in the TMA and the Tohoku area. Recent analyses indicated that other plumes were observed during this period (Tsuruta et al., 2017, 2018). Such observational studies have clarified the behavior of ^{137}Cs in the atmosphere. However, the available observational sites were limited over the land, making it difficult to assess the behavior of radionuclides over areas without observation sites.

Numerical simulations have been applied to compensate for the spatiotemporal limitations of observations. For example, Yasunari et al. (2011) determined that the contamination by radionuclides was restricted to the eastern side of the mountains of the Tohoku area because most of the radionuclides were deposited over the east of the mountainous "backbone" of Japan. Terada et al. (2012) revealed that the contamination over the TMA and Tohoku area originated from the plumes emitted on 12, 15, 16, 20, and 21–23 March 2011. These studies indicated that numerical simulations are useful tools for enhancing the understanding of the behavior of radionuclides in the atmosphere. However, they could not reproduce the temporal variation in atmospheric ^{137}Cs concentrations quantitatively because of the uncertainties in the source terms, meteorological fields, and physical processes in the models. For example, Morino et al. (2011) and Nakajima et al. (2017) reported that their numerical models generally reproduced the transportation pathways of plumes over the TMA and Tohoku area, but they found several discrepancies between their models and the observations.

The discrepancies between the models and the observations hint at the causes of the inaccuracies in the models and allow for the acquisition of knowledge needed to improve the models. As there is no natural source of ^{137}Cs, the observed ^{137}Cs must have been released after the FDNPP accident. In other words, ^{137}Cs is an appropriate tracer for the validation of atmospheric models (Kristiansen et al., 2016). A number of studies have been conducted regarding the numerical simulation of the radionuclides released in March 2011, and they have suggested that several elements could be responsible for the uncertainties in the numerical simulations.

These elements can be divided into two groups: external and internal elements. The former comprises external data input into the models, for example, meteorological data, the impact of which has been discussed by Arnold et al. (2015), and the emission inventory, as discussed by both Morino et al. (2011) and Nakajima et al. (2017). In general, reanalysis data are used for the meteorological fields required to reproduce the events with numerical models. However, the spatiotemporal resolutions of the reanalysis data are generally too coarse to successfully reproduce the atmospheric phenomena critical for the regional-scale advection of ^{137}Cs (Sekiyama et al., 2015). The emission inventory has been estimated by various studies (Katata et al., 2015; Saunier et al., 2013; Terada et al., 2012; Winiarek et al., 2014; Yumimoto et al., 2016), but it remains a major cause of the uncertainties in the simulated atmospheric ^{137}Cs concentrations because of the limited acquisition of direct measurements of radionuclide concentrations during their release. Internal elements refer to the components included in the numerical models, such as the physical processes of wet and dry deposition of ^{137}Cs (Morino et al., 2013; Quérel et al., 2015), clouds and rain (Saito et al., 2015), grid resolution (Sekiyama et al., 2015), and horizontal diffusion (Terada et al., 2013). Overall, the results derived from numerical models have large discrepancies.

To understand the reasons for the discrepancies among models, several model intercomparison projects have been conducted, targeting the radionuclides released after the FDNPP accident (Draxler et al., 2015; Kitayama et al., 2018; Kristiansen et al., 2016; Science Council of Japan, 2014). On the basis of an intercomparison, Draxler et al. (2015) reported that models with good performance in terms of accumulated deposition do not always show good performance in terms of the atmospheric concentration of ^{137}Cs. These model intercomparisons are useful for assessing the level of discrepancy between the observations and the models, but it is difficult to estimate the contribution of each element (external or internal) to the overall discrepancy. A model intercomparison using identical external data (i.e., emission inventory and meteorological data) is useful for assessing the relative importance of the contribution from the internal components. In addition,

Table 1
Summary of Models Included in the Intercomparison

Model name	Group/institute	Scientist	E/L	OD/ND	Reference
AIST-MM	AIST[a]	H. Kondo	E	OD	Kondo et al. (2001)
PELLO	FOI[b]	P. Schoenberg	L	ND	von Schoenberg et al. (2014)
HIRAT-LPRM	Fukushima U.	S. Hirao	L	ND	Hirao et al. (2013)
IdX	IRSN[c]	D. Quélo	E	ND	Mathieu et al. (2012)
GEARN	JAEA[d]	H. Terada	L	OD	Katata et al. (2015)
WRF-Chem-J	JAMSTEC[e]	M. Takigawa	E	OD	Grell et al. (2005)
NHM-Chem	MRI[f]	M. Kajino	E	ND	Kajino et al. (2012)
WRF-CMAQ	NIES[g]	Y. Morino	E	OD	Morino et al. (2013)
SCALE[h]	Nagoya U.	Y. Sato	E	OD	Nishizawa et al. (2015) Sato et al. (2015)
Polyphemus	Tsinghua U.	S. Fang	E	ND	Brandt et al. (2002)
WRF-Chem-T	Tsinghua U.	S. Fang	E	OD	Hu et al. (2014)
NICAM	U. Tokyo/NIES	J. Uchida/ D. Goto	E	OD	Satoh et al. (2014) Uchida et al. (2017)

Note. E = Eulerian atmospheric model, L = Lagrangian atmospheric model. Definitions of OD and ND models are described in section 2.1.
[a]AIST means National institute of Advanced Industrial Science and Technology. [b]FOI means Swedish Defense Research Agency. [c]IRSN means Institut de Radioprotection et de Sûreté Nucléaire. [d]JAEA means Japan Atomic Energy Agency. [e]JAMSTEC means Japan Agency for Marine-Earth Science and Technology. [f]MRI means Meteorological Research Institute. [g]NIES means National Institute for Environmental Studies. [h]The original version of SCALE does not have radionuclide component, and it was implemented in this study based on the aerosol component in NICAM as Nakajima et al. (2017)

meteorological data with fine spatiotemporal resolutions are required for reproducing the detailed phenomena that affect the behavior of atmospheric ^{137}Cs (Sekiyama et al., 2015).

In this study, a model intercomparison using identical meteorological data with a fine spatiotemporal resolution and an identical emission inventory was conducted. Through an analysis of the results of the model intercomparison, we aimed to achieve the following objectives: (1) to estimate the uncertainties originating from internal elements (e.g., the physical process of atmospheric ^{137}Cs), (2) to elucidate the key processes for the precise reproduction of the transportation pathways of the ^{137}Cs plumes observed by measurements, (3) to assess the score-weighted multimodel ensemble of the atmospheric ^{137}Cs concentration, and (4) to understand the behavior of atmospheric ^{137}Cs in greater detail.

2. Methods

2.1. Participating Models

Overall, 12 models from 11 groups participated in this model intercomparison; both Lagrangian and Eulerian dispersion models were included. Of the 12 models, seven have their own dynamical cores that calculate meteorological fields (e.g., wind field, temperature, eddy diffusion coefficient, relative humidity, and mixing ratio of hydrometeors), as mentioned in section 2.3. Hereafter, we refer to these models that include their Own Dynamical core as OD models. The remaining models use external meteorological data directly for the advection, diffusion, and deposition processes of ^{137}Cs. Hereafter, we refer to these models, which do Not have their own Dynamical core, as ND models. The names of each model, associated lead scientists, and other information are summarized in Table 1.

2.2. Experimental Setup Strategy

One of the objectives of this intercomparison was to elucidate the uncertainties originating from the internal elements of each model. For this purpose, the external data used with each model needed to be as similar as possible. In addition, using the same grid resolution was desirable because the grid resolution has considerable impact on the representation of the advection and diffusion of ^{137}Cs in the models. In this context, all models used the same meteorological data (Sekiyama et al., 2015, 2017; henceforth we referred to S15 and S17, respectively), emission inventory (Katata et al., 2015), and horizontal grid resolution, as described in section 2.3. The spatial resolution of the simulations was set to the same as that of the meteorological data to reduce errors that could be induced through data interpolation. The calculation domain was set to cover an area of approximately 500 × 600 km by zonal and meridional direction from the TMA to the Tohoku area

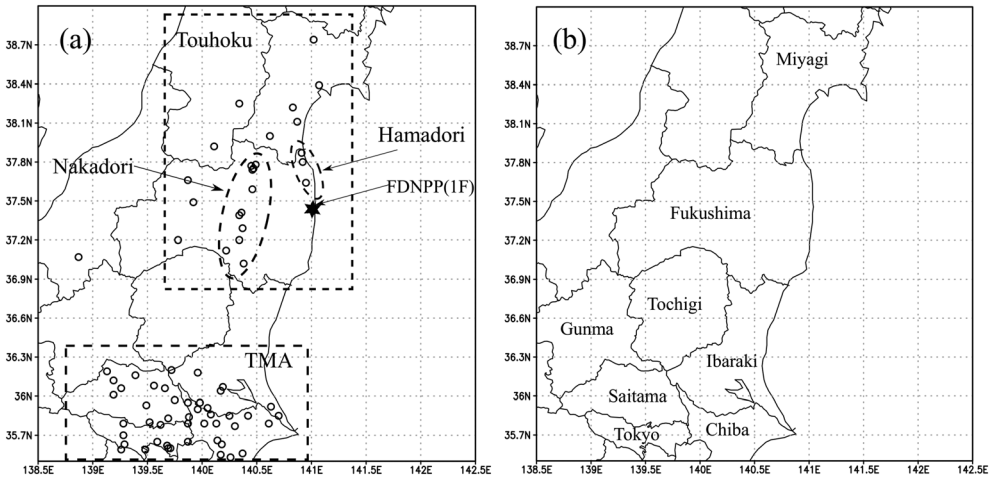

Figure 1. (a) Calculation domain and the locations of the suspended particulate matter observation sites (open circles) and (b) the names of the prefectures in the calculation domain. Arrows, dashed boxes, and dashed circles in (a) correspond to the names of the areas. The map was drawn using GrADS ver 2.1.1a (Institute for Global Environment and Society (IGES), 1989). FDNPP = Fukushima Daiichi Nuclear Power Plant, TMA = Tokyo metropolitan area.

(Figure 1), because most of the observational sites (shown in section 2.4) are encompassed within that region. The same meteorological data (S15 and S17) were used by all models, but the meteorological field in OD models differs from the meteorological data, because OD models calculated meteorological fields by themselves. By using the meteorological data of S15 and S17 for nudging, we attempted to reduce the differences between the meteorological field of OD models and that of S15 and S17 as much as possible.

2.3. Experimental Setup

The meteorological data derived from the operational model for regional weather forecasting around Japan (nonhydrostatic model (NHM); Saito et al., 2006) coupled with a local ensemble Kalman filter (LETKF; NHM-LETKF; Sekiyama et al., 2015, 2017) were used by all the models. The horizontal spatial and temporal resolutions of the NHM-LETKF were 3 km and 10 min, respectively. The temporal resolution (10 min) was sufficiently fine to resolve phenomena whose spatial scale is about 3 km (i.e., the spatial resolution of the NHM-LETKF) on the basis of the typical relationship between the temporal and spatial scales of atmospheric phenomena (cf., Mölders & Kramm, 2014). The horizontal resolution of each model was set to about 3 km. The vertical resolutions of the NHM-LETKF and each model are summarized in Table 2. For the ND models, the meteorological data of the NHM-LETKF were used directly with the temporal interpolation. The OD models calculated the meteorological fields themselves, and they used the meteorological data of the NHM-LETKF for nudging, lateral boundary conditions, and initial conditions of the meteorological fields. The meteorological field simulated by OD models was similar to that of NHM-LETKF, besides some cases like for NICAM and WRF-Chem-T for a high concentration event on 16 March, as discussed in section 3.2.2.

The emission inventory of ^{137}Cs by Katata et al. (2015) was adopted by all the models. The particle size distribution function of ^{137}Cs assumed in the models differed from each other because the emission inventory does not include information on the particle size distribution of ^{137}Cs. We regarded the assumption of the size distribution function as an internal component in this study.

Another important topic in terms of ^{137}Cs is the resuspension. The resuspension rates of radioactive cesium from the contaminated ground surface were estimated by using field experiments together with numerical assessments (Igarashi et al., 2015; Kajino et al., 2016). Based on these assessments, the surface air concentration of ^{137}Cs caused by resuspension was three to four orders of magnitude smaller than that caused by the direct plume from the FDNPP, and therefore, the effect of the resuspension was ignored in the current model intercomparison study.

The calculation was performed from 09:00 JST (00:00 UTC) on 11 March 2011 to 09:00 JST on 31 March 2011. The interval of the model output was 1 hr.

Table 2
Summary of Grid Resolution and Coordinate System of Each Model

Model name	Resolution for dispersion model: Horizontal (m)	Resolution for dispersion model: Vertical[a]	Model top (number of layers)
AIST-MM	3,000	Δz = 10–300 m	5,400 m (35)
PELLO	3,000	—[b]	20,000 m (—[b])
HIRAT-LPRM	3,000	—[b]	10,000 m (—[b])
IdX	3,000	Δz = 40–600 m	5,000 m (17)
GEARN	3,000	$\Delta\eta$ = 0.003–0.019/ Δz = 20–678 m[c]	100 hPa (30)/ 15,000 m (43)[c]
WRF-Chem-J	3,000	$\Delta\sigma$ = 0.003–0.0341	100 hPa (34)
NHM-Chem	3,000	$\Delta\sigma$ = 0.005–0.1	10,000 m (19)
WRF-CMAQ	3,000	$\Delta\sigma$ = 0.0026–0.0332	100 hPa (35)
SCALE	3,000	Δz = 125–1,322 m	16,222 m (30)
Polyphemus	3,000	—[b]	7,523 m (—)
WRF-Chem-T	3,000	$\Delta\sigma$ = 0.0026–0.0332	100 hPa (35)
NICAM	3,000	Δz = 30–2,400 m	22,000 m (40)
NHM-LETKF	3,000	Δz = 40–662 m	20,200 m (60)

[a] z, η, and σ mean height coordinate, η coordinate, and σ coordinate systems. [b]"—" means the models without their own dynamics; meteorological data and the resolution of NHM-LETKF were directly used. [c]GEARN used the η coordinate system for its dynamical core and the z coordinate system for its dispersion model.

2.4. Observational Data and Target Events of Atmospheric ^{137}Cs

Two types of ^{137}Cs observations were used for comparing the results of the models. The first was the cumulative deposition density of ^{137}Cs estimated by aircraft observations (MEXT, 2011). The aircraft observation was conducted from June to November 2011. This period differs from the target period (March 2011) of the simulation, but the decrease of ^{137}Cs from April to June was small (Morino et al., 2013). In addition, the observed deposition had uncertainties that originated from the spatial resolution, land use, and natural radioactive nuclides, among others. Torii et al. (2012) indicated that the error of the airborne estimation was smaller than 50% compared with in situ measurement (as shown in Figures S5a and S5b in the supporting information). In contrast, Morino et al. (2013) indicated that the total deposition of ^{137}Cs varies by one order of magnitude by changing the deposition scheme (as shown in Figures S5c and S5d). Based on the results, the uncertainties of the measured deposition would be much smaller than those in the models. Thus, the cumulative deposition density estimated via aircraft observation can be used as reference data for model evaluation. The second type of observation comprised measurements of the atmospheric concentration of ^{137}Cs obtained from the operational aerosol sampling of the national SPM network (Oura et al., 2015). The hourly averaged atmospheric concentration of ^{137}Cs near the surface was measured at 99 SPM sites. The data from the SPM sites included in the calculation domain (Figure 1) were used for the comparison. The atmospheric concentrations of ^{137}Cs measured by the SPM network were compared with those in the first (lowest) layer of each model because the SPM instruments collect air samples near the surface. The details of the SPM data are described by Tsuruta et al. (2014) and Oura et al. (2015).

On the basis of analyses using the SPM network data, Tsuruta et al. (2014) found that nine notable events (plumes) of atmospheric ^{137}Cs occurred, as summarized in Table 3. The present study targeted Plumes 2, 3, 4, 7, 8, and 9. The other plumes were observed in the Hamadori area, which is about 30 km north of the FDNPP. It is difficult for models with a 3-km grid resolution to simulate the plumes in that area because models cannot simulate phenomena with spatial scales smaller than 6–10 times of the grid resolution (Frehlich & Sharman, 2008; Skamarock, 2004). Lagrangian dispersion models are grid-free and thus in principle are not affected by this problem; however, their performance is conditioned by the grid resolution of the meteorological field (NHM-LETKF) used in input and created from the Eulerian dynamical core (NHM; Saito et al., 2006). Therefore, we excluded Plumes 1, 5, and 6 from our analyses.

Table 3
Summary of the Nine Notable Events (Plumes) of Atmospheric ^{137}Cs Measured by the Suspended Particulate Matter Network (Tsuruta et al., 2014) After the Fukushima Daiichi Nuclear Power Plant Accident

Plume number	Measured time (JST)[a]	Location	Area for analyses
1	12 March 8:00 to 13 March 5:00	Hamadori	—[b]
2	15 March 3:00 to 15 March 23:00	TMA	(35.5°N–36.9°N, 138.5°E–141.0°E)
3	15 March 8:00 to 16 March 1:00	Nakadori	(36.5°N–37.9°N, 138.5°E–142.5°E)
4	16 March 2:00 to 16 March 23:00	TMA	(35.5°N–38.1°N, 140.2°E–141.5°E)
5	18 March 4:00 to 18 March 23:00	Hamadori	—[b]
6	19 March 9:00 to 20 March 4:00	Hamadori	—[b]
7	20 March 12:00 to 20 March 23:00	TMA	(35.5°N–36.5°N, 138.5°E–142.5°E)
8	20 March 13:00 to 21 March 10:00	Nakadori	(36.9°N–38.0°N, 140.0°E–141.5°E)
9	21 March 6:00 to 21 March 20:00	TMA	(35.5°N–36.5°N, 138.5°E–142.5°E)

Note. Locations of the names are shown in Figure 1. JST = Japan Standard Time, which is 9 hr earlier than the Coordinated Universal Time (UTC).
[a]Measured time means the period from plume transit start time and plume transit end time (Kitayama et al., 2018). [b]"–" means that the plume was not the target of the analyses in this study.

In addition to atmospheric ^{137}Cs concentrations, the wind velocity and wind direction, as measured by the Automated Meteorological Data Acquisition System operated by the Japan Meteorological Agency, were used for comparison with the meteorological fields simulated by the models.

2.5. Analyses

In this study, we evaluated the performance of the models through the metrics shown below. Since the grid points of the models were not always defined at the same location as the observation sites, we calculated the modeled values at each observation site as described later in this section. We considered several metrics for the evaluation of the performance of each model. The correlation coefficient (CC), fractional bias (FB), figure of merit in space (FMS), and Kolmogorov-Smirnov parameter (KSP) were introduced for the evaluation of the cumulative amount of ^{137}Cs deposition. The definitions of the CC, FB, FMS, and KSP are as follows:

$$\mathrm{CC} = \frac{\Sigma\left(D_{\mathrm{model}} - \overline{D_{\mathrm{model}}}\right)\left(D_{\mathrm{obs}} - \overline{D_{\mathrm{obs}}}\right)}{\sqrt{\Sigma\left(D_{\mathrm{obs}} - \overline{D_{\mathrm{obs}}}\right)^2 \Sigma\left(D_{\mathrm{model}} - \overline{D_{\mathrm{model}}}\right)^2}}, \tag{1}$$

$$\mathrm{FB} = 2\times\frac{\overline{D_{\mathrm{model}}} - \overline{D_{\mathrm{obs}}}}{\overline{D_{\mathrm{model}}} + \overline{D_{\mathrm{obs}}}}, \tag{2}$$

$$\mathrm{FMS} = 100\times\frac{A_{\mathrm{obs}}\cap A_{\mathrm{model}}}{A_{\mathrm{obs}}\cup A_{\mathrm{model}}}, \tag{3}$$

$$\mathrm{KSP} = \mathrm{Max}\left|P_{\mathrm{obs}}\left(D_{\mathrm{ksp}}\right) - P_{\mathrm{model}}\left(D_{\mathrm{ksp}}\right)\right|, \tag{4}$$

where D_{obs} and D_{model} are the cumulative deposition amounts until the end of March 2011 at each grid point as estimated by aircraft observations and simulated by the models, respectively; A_{xx} is the area in which D_{xx} exceeds 10^4 Bq/m^2 (xx is either model or obs); and P_{xx} (D_{ksp}) is the probability that the occurrence of deposition by xx is not greater than D_{ksp}. In this study, D_{ksp} ranged from 1.5×10^4 to 4.0×10^6 Bq/m^2. The FMS is the ratio of the area common to both the observed and simulated deposition that exceeds 10^4 Bq/m^2 to envelope both areas. The overbar indicates the value averaged over all grid points. In addition, a metric (RANK) combining these four metrics (Draxler et al., 2013; Leadbetter et al., 2015) is defined as follows:

$$\mathrm{RANK} = \mathrm{CC}^2 + \left(1 - \left|\frac{\mathrm{FB}}{2}\right|\right) + \frac{\mathrm{FMS}}{100} + \left(1 - \frac{\mathrm{KSP}}{100}\right), \tag{5}$$

whose value (0–4) was used for the evaluation of the cumulative deposition. These scores were calculated from the results over land where the cumulative deposition amount exceeded 10^4 Bq/m^2. This was because the aircraft observational data were only available over a limited land area and their detection threshold was 10^4 Bq/m^2. The cumulative deposition was calculated from the hourly averaged deposition amount every hour.

In contrast to the observed cumulative deposition of ^{137}Cs, which only has the spatial (geographical) distribution, the atmospheric concentration of ^{137}Cs has spatial and temporal distribution. For the evaluation based on both spatial and temporal distribution, the scores used in the weather prediction were useful. For the evaluation of the atmospheric concentration of ^{137}Cs, we used scores named "CAPTURE," "OVERESTIMATE," "threat score," "factor of two (FA2)," and "factor of five (FA5)." The definitions of these scores are as follows:

$$(\mathrm{CAPTURE}) = \mathrm{MO}/(\mathrm{MO} + \mathrm{XO})\times100, \tag{6}$$

$$(\mathrm{OVERESTIMATE}) = \mathrm{MX}/(\mathrm{MO} + \mathrm{XO})\times100, \tag{7}$$

$$(\mathrm{Threat\ score}) = \mathrm{MO}/(\mathrm{MO} + \mathrm{XO} + \mathrm{MX})\times100, \tag{8}$$

where MO is the number of grids at each time step where both the model and observed ^{137}Cs concentrations exceeded 10 Bq/m^3 and XO (MX) is the number of grids at each time step where only the observation (model) showed ^{137}Cs over 10 Bq/m^3. FA2 and FA5 are defined as the percentage of simulated ^{137}Cs within a factor 2

or 5, respectively, of observed ^{137}Cs. FA2 and FA5 were calculated when the observation showed ^{137}Cs over 10 Bq/m^3. The threat score has been used for the evaluation of weather forecasting models in which the phenomena occurred only a few times a year (Wilks, 2006). Therefore, it was considered to be suitable for the evaluation of the atmospheric ^{137}Cs concentration in this study, because only nine plumes were observed during the study period. High values of all these metrics except for OVERESTIMATE indicate the good performance, and small values for OVERESTIMATE indicate good performance. In addition, a metric (RANK2) combining these three metrics is defined as follows:

$$\mathrm{RANK2} = \frac{\mathrm{FA2}}{100} + \frac{\mathrm{CAPTURE}}{100} + F \times \left(1 - \frac{\mathrm{OVERSESTIMATE}}{100}\right), \quad F = \begin{cases} 0 & (\mathrm{OVERESTIMATE} = 0) \\ 1 & (\mathrm{OVERESTIMATE} \neq 0) \end{cases} \tag{9}$$

We regarded models with FMS $>$ 70, RANK $>$ 2, CAPTURE $>$ 20, threat score $>$ 20, and RANK2 $>$ 1 as showing good performance.

For the evaluation based on the scores shown above, the simulated values were compared with measured value at each observation site. The outputs of the models were defined at grid points for the evaluation, for Eulerian, and Lagrangian models. Since the grid points were not always defined at the same point of the observation site, the simulated values for the comparison were calculated as follows. First, we determined a model's grid point of which the latitude and longitude are nearest to those of the observation site and named the grid as the nearest grid. If the nearest grid is *i*th and *j*th grid of the model for zonal and meridional direction (i, j), the simulated values for the comparison were calculated by averaging the nearest grid (i, j) and eight grids around the nearest grid, that is, ($i-1$, $j-1$), ($i-1$, j), ($i-1$, $j+1$), (i, $j-1$), (i, $j+1$), ($i+1$, $j-1$), ($i+1$, j), and ($i+1$, $j+1$). For the Lagrangian dispersion models, ^{137}Cs at each Eulerian grid was calculated from the distribution of Lagrangian particles through box counting. Each Lagrangian particle is carrying radioactivity of ^{137}Cs, and the sum of this, for all particles in a grid box, determines the concentration in the box. In the comparison with all the models, the center of this grid box is considered the location of the grid box value that is used when the nearest grid was determined as described above.

To evaluate the models for each plume shown in Table 3, we calculated CAPTURE, OVERESTIMATE, the threat score, FA2, FA5, and RANK2 for each plume for each model. The plume area was defined as the grid whose atmospheric ^{137}Cs concentration exceeded 10 Bq/m^3 in the area for the analyses summarized in Table 3.

2.6. Ensemble Mean Evaluation

We calculated multimodel ensembles using two types of weights. The simplest multimodel ensemble was calculated by a simple average with equal weights, which was used for most of analyses in this study. We calculated another multimodel ensemble to evaluate the effectiveness of the ensemble average weighted by the performance of each model. Multimodel ensemble weighted with the product of RANK and the threat score were calculated. The latter ensemble was used to determine the effectiveness of the weighted multimodel ensemble (see section 3.4).

3. Results and Discussion

3.1. Overview of the Simulated ^{137}Cs

First, we statistically evaluated the performance of the models throughout the entire simulation period (i.e., from 11 March to 31 March 2011). Both the atmospheric concentrations and deposition amounts of ^{137}Cs were evaluated following the procedure outlined by Draxler et al. (2015). For the evaluation, a multimodel ensemble was calculated using an equal weighting for each model.

3.1.1. Deposition Amount of ^{137}Cs

The cumulative deposition amounts of ^{137}Cs through March 2011, as estimated by aircraft observations and simulated by the multimodel ensemble, are shown as gray shade in Figures 2a and 2b, respectively. The temporal evolution of the calculated deposition is shown in Figures 2c–2f, and the cumulative wet deposition is also shown. The red contour lines in Figures 2c–2f indicate cumulative deposition originating from the wet deposition, and the solid, dashed, and dot-dashed red lines indicate 10^4, 10^5, and 10^6 Bq/m^2, respectively. The models did not reproduce the large amount of deposition ($>10^5$ Bq/m^2) over Nakadori (shown with a

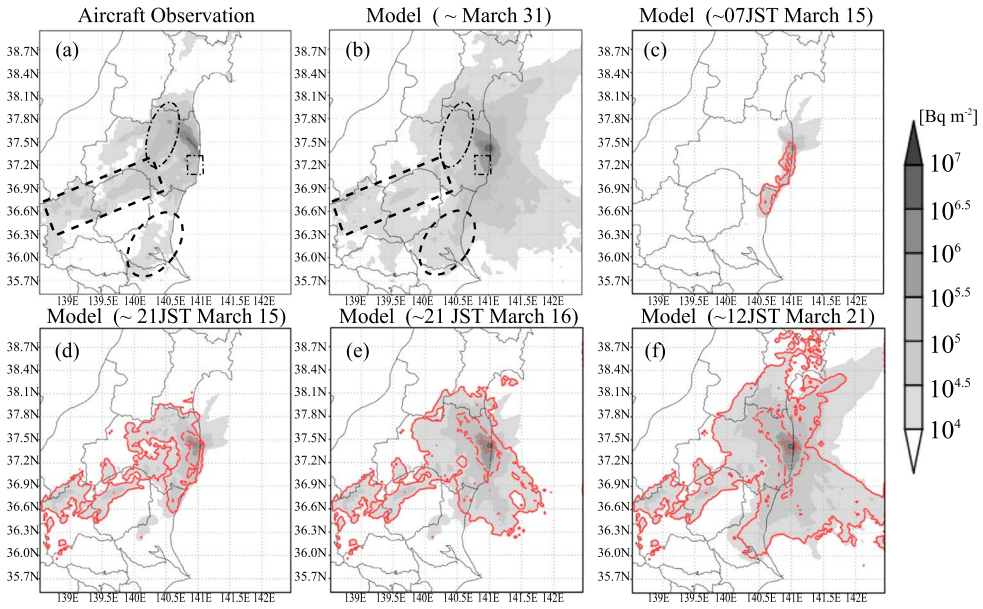

Figure 2. Cumulative deposition amount of ^{137}Cs during March 2011 (a) estimated from aircraft observations (MEXT, 2011) and (b) simulated by the multimodel ensemble. Also shown are the cumulative deposition amounts of ^{137}Cs from 09:00 JST on 11 March to (c) 07:00 JST on 15 March, (d) 21:00 JST on 15 March, (e) 21:00 JST on 16 March, and (f) 12:00 JST on 21 March. Values $<10^4$ Bq/m^2 were ignored for ease of visualization. Red contour lines in (c)–(f) indicate cumulative deposition originating from the wet deposition, and the solid, dashed, and dot-dashed red lines indicate 10^4, 10^5, and 10^6 Bq/m^2, respectively. The dashed and dot-dashed rectangle and circles in (a) and (b) identify areas described in section 3.1.1. The mapping was created using GrADS 2.1.1a (Institute for Global Environment and Society (IGES), 1989).

dotted-dashed circle in Figures 2a and 2b). In addition, the models overestimated the deposition over the south of the FDNPP (shown with a dotted-dashed square in Figures 2a and 2b).

In contrast to the poor performance of the models over the Nakadori and the south of the FDNPP, the model ensemble successfully reproduced the large deposition amount ($>10^5$ Bq/m^2) to the northwest of the FDNPP. The model ensemble also reproduced the deposition in the northern parts of Gunma and Tochigi Prefecture (shown as a dashed rectangle in Figures 2a and 2b) and in Ibaraki Prefecture (shown as a dashed circle in Figures 2a and 2b), the value of which is in the order of 10^4 Bq/m^2. The large deposition amount over the northwest of the FDNPP occurred until 16 March (Figure 2e). The red contour lines in Figure 2e are mostly overlapping edges of grayscale shade. This indicates that the large deposition amount over this area mainly originated from wet deposition. Most of the deposition in Gunma and Tochigi Prefecture (shown as a dashed rectangle in Figures 2a and 2b) was also reproduced to have occurred until 16 March, and most of the edges of grayscale shade in this area until 16 March (Figure 2e) overlap the red contour lines. These results indicate that the deposition of ^{137}Cs over these areas originated from the wet deposition caused by a precipitation event that occurred from the TMA to Tohoku area in the period from late hours on 15 to 16 March (the distributions of the precipitation areas are shown in Figure 6 of Nakajima et al., 2017). These results are consistent with the suggestions of Terada et al. (2012), who investigated the timing of the deposition event using their model, and those of Mathieu et al. (2018), who proposed that the timing of the deposition was supported by the observations and model simulations.

The scores regarding the cumulative deposition of ^{137}Cs are summarized in Table 4. The CC, FMS, FB, and KSP values of the model ensemble mean are 0.69, 80.7, −0.074, and 3.41, respectively. In terms of the CC, WRF-Chem-T and WRF-Chem-J produced high scores. With respect to the FMS, GEARN, WRF-Chem-J, and AIST-MM had good performances (FMS $>$ 70). The value of the FB of AIST-MM was smaller than that of the other models. The total evaluation by RANK indicates that WRF-Chem-J had the best performance, followed by IdX, AIST-MM, HIRAT-LPRM, GEARN, and WRF-Chem-T, which also showed good performance. RANK values ranged from 0.73 to 2.98, with most models exceeding 2. The averaged RANK values for Lagrangian and Eulerian dispersion models were 2.73 and 2.41, respectively.

Table 4
Scores of ^{137}Cs Deposition Amounts Accumulated During March 2011

Model (L/E)	CC	FMS	FB	KSP	RANK
AIST-MM (E)	0.473	71.0	0.068	11.8	2.78
PELLO (L)	0.563	51.7	0.099	25.6	2.51
HIRAT-LPRM (L)	0.609	67.8	0.267	8.48	2.83
IdX (E)	0.589	63.3	0.122	17.2	2.75
GEARN (L)	0.602	76.7	0.317	11.5	2.86
WRF-Chem-J[a] (E)	0.707	73.2	0.294	10.5	2.98
NHM-Chem (E)	0.502	38.8	−0.657	41.4	1.90
WRF-CMAQ (E)	0.525	54.3	−0.452	21.6	2.38
SCALE (E)	0.422	8.1	−1.80	62.5	0.73
Polyphemus (E)	0.636	52.9	−0.133	21.8	2.65
WRF-Chem-T[a] (E)	0.866	61.0	−0.472	19.3	2.93
NICAM (E)	0.562	52.9	0.118	21.9	2.57
Ensemble mean	0.690	80.7	−0.074	3.41	3.21

Note. CC = correlation coefficient, FB = fractional bias, FMS = figure of merit in space, KSP = Kolmogorov-Smirnov parameter.
[a]WRF-Chem-J and WRF-Chem-T indicates the WRF-Chem by JAMSTEC and WRF-Chem by Tsinghua University, respectively

The total deposition amount accumulated over the entire land area (Table 5), and the horizontal distribution of the deposition (Figure S4) elucidated the reason for the differences in the scores of each model. In the results of the models that did not show good performance, defined as models with RANK smaller than 2 (i.e., the NHM-Chem and SCALE), the area with a total deposition amount exceeding 10^4 Bq/m^2 was significantly smaller than that of the observations over land. SCALE could not reproduce the horizontal distribution of the cumulative deposition and considerably underestimated the deposition amount over land (Table 5). NHM-Chem reproduced the horizontal distribution but underestimated its amount. The small deposition amounts in these models resulted in their low scores (Table 4). To further evaluate the differences among the models, detailed analyses for each plume summarized in Table 3 were required. This is addressed in sections 3.1.2 and 3.2.

We should also note that even though the fraction of dry deposition in relation to the total deposition has considerable variability among the models (Table 5), most of the models showed good performance (RANK $>$ 2) for the cumulative deposition. Dry deposition was dominant in the PELLO, Polyphemus, and SCALE models, whereas wet deposition was dominant in the other models. Variability in this fraction has been reported by a previous intercomparison study (Science Council of Japan, 2014), but that study was unable to conclude whether the variability originated from uncertainties in the deposition process itself. This is because the external data and grid resolution differed among the models in the previous study. In contrast, the same external data and grid resolution were used in the current study, and therefore, our results showed that the variability in the fraction originated from the differences in the deposition processes of the various models.

3.1.2. Atmospheric Concentration of ^{137}Cs

The composite of the temporal evolutions of the atmospheric ^{137}Cs concentration around noticeable peaks of the SPM observations and the model simulations were compared (Figure 3). To calculate the composite of the SPM observation, peaks of the atmospheric ^{137}Cs concentration were determined as the time when ^{137}Cs exceeded 100 Bq/m^3 and temporal variation of ^{137}Cs was converted from *increase* to *decrease*. The peaks were determined at all SPM sites in the calculation domain during March 2011. Then, ^{137}Cs concentration during 10 hr before and after the peak time was normalized by the peak ^{137}Cs concentration. The composite was calculated by averaging the normalized ^{137}Cs concentration for all peaks in March 2011 over all the SPM sites (red line in Figure 3). Peaks separated by more than 1 hr were regarded as independent peaks. The simulation results were normalized by the observed peak concentrations and averaged over the peaks for the same time windows defined by the observed peaks. The statistics of the composite trend simulated by all the models are also shown in Figure 3.

The models reproduced the timing of the peak with a time lag as small as 1–2 hr averaged over all SPM sites (Figure 3a). The time lag of the peak was large (2–5 hr) in the TMA (Figure 3b), whereas the time lag was as small as ~1 hr in the Tohoku area (Figure 3c). The variation in the time lag was considered to have mainly originated from differences in the distances between the observation sites and FDNPP. Atmospheric ^{137}Cs concentrations observed in the TMA were affected by the uncertainties of the meteorological field and those that originated from physical processes in the atmosphere for a longer time than were those in the Tohoku area. Thus, the models were able to better reproduce the ^{137}Cs observed at sites nearer to the emission source, that is, in the Tohoku area.

Table 5
Cumulative ^{137}Cs Deposition Amount Over Land During March 2011 and Ratio of Dry Deposition to Total Deposition Over Land

Model (L/E)	Total deposition (PBq)	Area (km^2) where total deposition exceeding 10^4 Bq/m^2	Dry deposition fraction (%)
AIST-MM (E)	3.58	4.26×10^4	1.7
PELLO (L)	3.07	2.52×10^4	79.1
HIRAT-LPRM (L)	3.62	3.29×10^4	14.2
IdX (E)	3.08	2.57×10^4	24.4
GEARN (L)	4.03	4.09×10^4	23.4
WRF-Chem-J (E)	3.76	3.57×10^4	6.7
NHM-Chem (E)	1.45	1.69×10^4	18.5
WRF-CMAQ (E)	1.79	2.11×10^4	3.6
SCALE (E)	0.18	4.49×10^3	79.6
Polyphemus (E)	2.48	2.65×10^4	61.9
WRF-Chem-T (E)	2.04	3.84×10^4	6.6
NICAM (E)	3.17	2.57×10^4	9.0
Observation (MEXT, 2011)	2.65	3.28×10^4	—

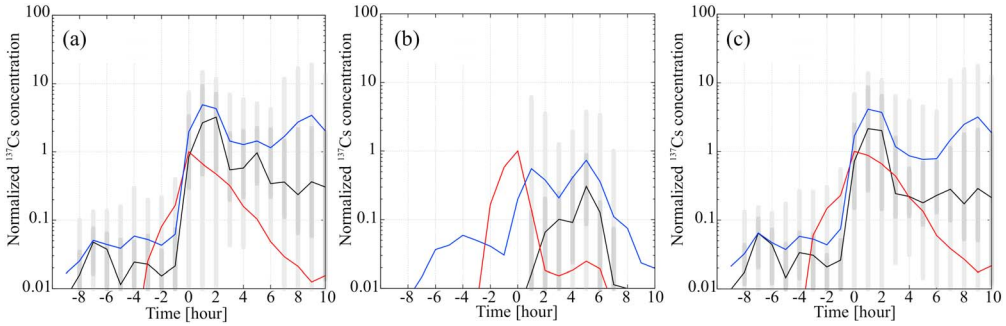

Figure 3. Temporal evolution of atmospheric ^{137}Cs normalized by the observed value averaged during the entire calculation time and averaged over (a) the suspended particulate matter (SPM) sites in the calculation domain, (b) the SPM sites in Tokyo metropolitan area, and (c) those in the Tohoku area. Time (t) = 0 is corresponding to the time when the high concentration event of ^{137}Cs (>100 Bq/m^3) was observed. Red, blue, and black lines show the observed values, model ensemble means, and medians of the models, respectively. Thick and thin gray bars indicate the ranges of the 25th percentile to 75th percentile and from minimum to maximum, respectively.

The scores of the models for atmospheric ^{137}Cs concentrations are shown in Table 6. In terms of CAPTURE, the models captured 5–50% of the ^{137}Cs events in the atmosphere, and WRF-Chem-T, HIRAT-LPRM, PELLO, GEARN, and WRF-CMAQ had better scores (CAPTURE > 20). The ensemble means of the models captured 40% of the events. WRF-CMAQ, WRF-Chem-T, HIRAT-LPRM, and GEARN had good threat scores (>20) and that of the model ensemble mean was 24. In terms of the unified score (RANK2), WRF-CMAQ and WRF-Chem-T showed good performance (RANK2 > 1), and RANK2 of the ensemble mean was 1.01. The averages of RANK2 and threat score for Lagrangian and Eulerian dispersion models were 0.86 and 18.73, and 0.79 and 17.04, respectively.

The lower score of the atmospheric ^{137}Cs by ensemble means (CAPTURE = 40%) than total deposition (FMS = 80) would be due to the difference in timing between observed and modeled atmospheric ^{137}Cs. The scores for deposition (e.g., FMS) are only dependent upon the spatial distribution. In contrast, the scores for atmospheric ^{137}Cs are dependent on differences in timing. Based on this point, evaluation of the atmospheric ^{137}Cs was stricter than that of the deposition.

3.2. Evaluation of each Plume

The scores in terms of the atmospheric concentration of ^{137}Cs for each plume are shown in Table 7, and the scores of each model for each plume are summarized in Tables S1 and S2. Plumes 2 and 8 were reproduced well by the models (RANK2 > 1.5), but Plumes 7 and 9 were not captured (threat score < 10 and RANK2 < 0.8). This trend in scores was common to all models (Tables S1 and S2). The poor performances of the models for Plumes 7 and 9 were attributed to the external data. For these plumes, the wind field was not reproduced well, as discussed in supporting information S2.2 and S2.3. These results indicate that the performance was not good, regardless of the inherent potential of the model, if the key meteorological phenomena were not reproduced well in the external input data. In contrast, the meteorological fields properly represented the meteorological conditions for Plumes 2 and 8, which were reproduced well by the models. These results suggest that the quality of the external data is crucial for reproducing plumes.

The meteorological fields were also properly reproduced for Plume 4 in most of the models (except for two models) and, therefore, was well reproduced in ensemble mean (Table 7). However, half of the models showed poor performance (both threat score and CAPTURE are 0 as shown in Table S1). The poor performance of most of these

Table 6
Scores of Atmospheric ^{137}Cs Concentration Averaged During March 2011

Model (L/E)	C	T	O	FA2	FA5	RANK2
AIST-MM (E)	14.6	11.0	69.4	8.4	22.5	0.54
PELLO (L)	36.1	10.6	86.9	21.0	44.4	0.70
HIRAT-LPRM (L)	38.2	23.3	62.6	17.6	41.0	0.93
IdX (E)	17.2	13.9	58.0	11.6	28.8	0.71
GEARN (L)	20.4	22.3	63.8	21.5	47.0	0.94
WRF-Chem-J (E)	6.4	19.4	53.3	13.7	31.5	0.85
NHM-Chem (E)	13.3	17.2	69.2	18.0	36.3	0.77
WRF-CMAQ (E)	25.7	26.9	42.3	20.4	39.5	1.11
SCALE (E)	7.7	9.6	46.7	6.4	12.4	0.70
Polyphemus (E)	25.2	14.6	71.5	13.3	30.5	0.65
WRF-Chem-T (E)	49.1	26.2	64.0	25.7	50.0	1.11
NICAM (E)	18.4	14.5	58.8	7.68	21.3	0.67
Ensemble Mean	40.3	23.8	63.1	25.2	51.3	1.01

Note. C = CAPTURE, T = threat score, O = OVERESTIMATE, FA2 = factor of 2, FA5 = factor of 5.

Table 7
Scores of Atmospheric ^{137}Cs Concentration for Each Plume by Multimodel Ensemble

	C	T	O	FA2	FA5	RANK2
Plume 2	63.4	44.8	39.6	37.6	72.3	1.61
Plume 3	48.6	19.1	76.1	34.3	51.4	1.07
Plume 4	41.7	30.3	47.4	30.0	60.0	1.24
Plume 7	5.0	4.9	33.3	6.3	20.0	0.78
Plume 8	48.1	44.0	16.2	24.8	65.1	1.57
Plume 9	7.14	2.3	96.7	3.6	14.3	0.14

Note. C = CAPTURE, T = threat score, O = OVERESTIMATE, FA2 = factor of 2, FA5 = factor of 5.

models are originated from the physical process as discussed in section 3.2.2, and therefore, it is worthwhile to discuss the performance and internal components of the model for Plume 4.

The RANK2 of the Plume 3 was higher than that for Plumes 7 and 9, but the wind direction was changed during the period of Plume 3 (as discussed in supporting information S2.1). Some models could not reproduce the change of the wind direction correctly. In addition, the uncertainties in the source term of ^{137}Cs would be another source of error for Plume 3. Based on these results, the uncertainties of external data (meteorological field and emission inventory) had large impact on the performance of the models for Plume 3.

In the following subsections, we discuss the differences in the internal components of the models through analyses of Plumes 2, 4, and 8. For the evaluation of the models for the atmospheric ^{137}Cs concentration, we mainly used threat score and CAPTURE values. We focused on these metrics because miss-detection, which is defined as when ^{137}Cs is measured but not predicted by the model, should be reduced as much as possible for the prediction of radionuclides from the accident of nuclear power plants. The threat score and CAPTURE are suitable for the purpose.

3.2.1. Plume 2: Horizontal Diffusion Was Critical

Plume 2 was observed in the TMA on 15 March 2011. The ^{137}Cs contributing to Plume 2 was released at around 00:00 JST on 15 March, and it was transported by the northeast wind toward the TMA (Figure 4a). During this period, the northeast wind persisted, and there was no precipitation over the area between the FDNPP and TMA. Under such simple meteorological conditions, irrespective of the model used, ^{137}Cs was simply transported in a southwestward direction until it reached the TMA. Thus, the performance of all models for Plume 2 was relatively good compared with the other plumes (Tables 7 and S1).

The differences in the scores of the models originated from the internal components of each model, that is, the physical and diffusion processes. For Plume 2, the wet deposition process made no contribution to the atmospheric ^{137}Cs because there was no precipitation. Thus, the diffusion and dry deposition processes were the main factors that determined the model's score. To understand the reason for the differences between the models, it is useful to consider the horizontal distribution of ^{137}Cs calculated by each model. ^{137}Cs simulated by the models with good scores in terms of CAPTURE and threat score values were distributed widely across the TMA (Figure 4b). In these models, the high-concentration events measured at the SPM sites in the TMA were easily captured. However, the width of the plume may have been overestimated, as shown in

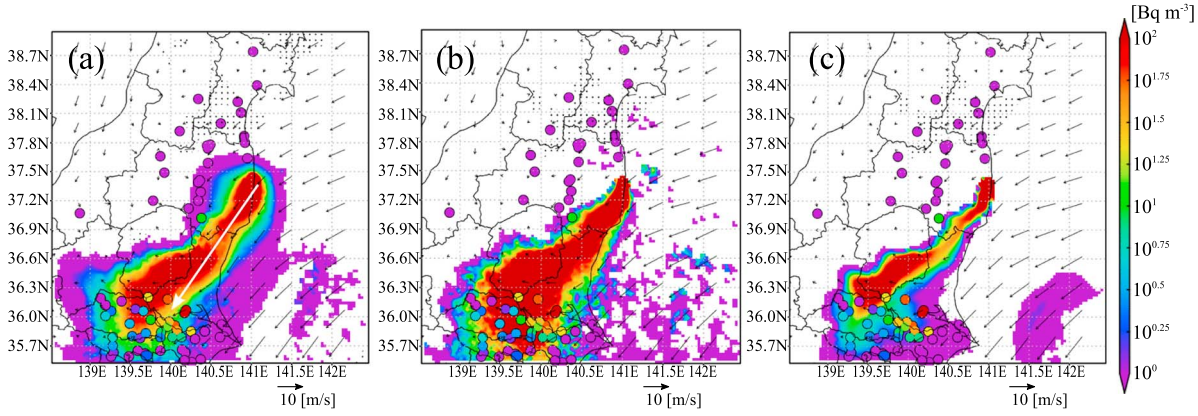

Figure 4. Geographical distribution of ^{137}Cs (shaded area) at 09:00 JST on 15 March 2011 by (a) the model ensemble mean and simulated by (b) HIRAT-LPRM and (c) NHM-Chem, and ^{137}Cs measured by the suspended particulate matter network (circle). Black arrows indicate the wind field at a 10-m height simulated by the OD model or given by the NHM-LETKF data with interpolation for the ND model. Black arrows in (a) represents the multimodel ensemble of the wind field simulated by the OD models and interpolated from the NHM-LETKF for ND models. Dotted area shows areas where precipitation exceeded 0.1 mm/hr. Thick white arrow in (a) indicates the movement of ^{137}Cs in Plume 2. The ^{137}Cs concentrations of <1 Bq/m^3 were ignored for ease of visualization. Definitions of the OD and ND models are described in section 2.1. The mapping was created using GrADS 2.1.1a (Institute for Global Environment and Society (IGES), 1989).

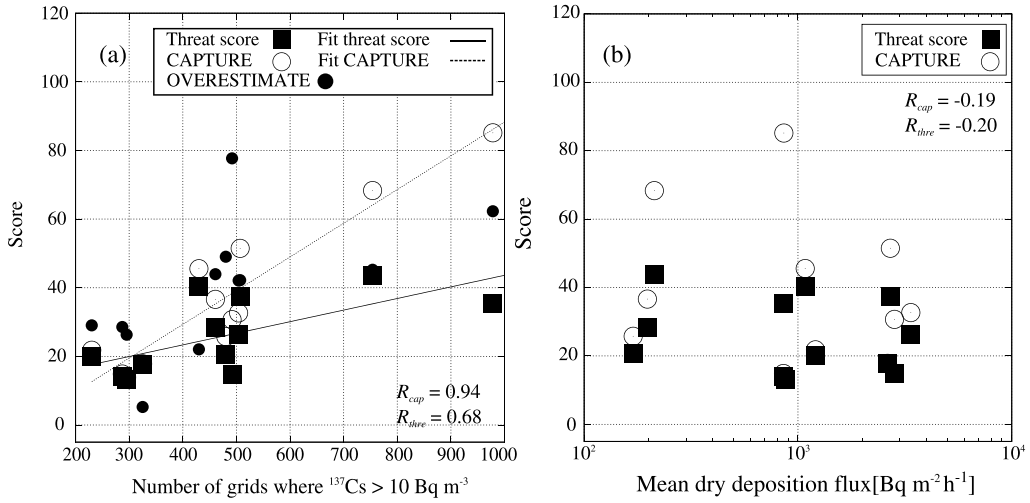

Figure 5. Relationship between the score and (a) an area of high atmospheric ^{137}Cs concentration (>10 Bq/m^3) and (b) the mean dry deposition flux during Plume 2 over the path where the atmospheric ^{137}Cs concentration exceeded 10 Bq/m^3. Closed squares, open circles, and closed circles show the threat score, CAPTURE, and OVERESTIMATE, respectively. R_{cap} (R_{thre}) shows the correlation coefficient between (a) the number of the grid and (b) the dry deposition flux and CAPTURE (threat score).

Figure 4b, compared with the measured area with a high atmospheric ^{137}Cs concentration (>10 Bq/m^3; i.e., the models had poor scores in terms of OVERESTIMATE). In contrast, the ^{137}Cs simulated by the models with low scores was distributed within a narrower area (Figure 4c). The narrower horizontal distribution of the ^{137}Cs plume reproduced the width of the area with a high measured ^{137}Cs concentration and resulted in good performance in terms of OVERESTIMATE. However, the center of the plume was simulated slightly to the northwest of the high-concentration area measured. This displacement of the simulated plume resulted in lower scores in terms of CAPTURE and threat score values for the models.

The clear correlation between the score and the area with a high ^{137}Cs concentration supports this suggestion (Figure 5a). In contrast, dry deposition had a weak correlation with the scores (Figure 5b). Additionally, the meteorological field in each model had small differences in simple meteorological conditions for Plume 2. In this case, the size of the area with a high ^{137}Cs concentration would be mainly determined by the strength of horizontal and vertical diffusion, which is expressed by the eddy diffusion coefficient, the numerical diffusion added for numerical stability, and the diffusion that originated from the spatiotemporal variability of the wind field. We suggest from these results that the geographical distribution of ^{137}Cs simulated by each model for Plume 2, which traveled under simple meteorological conditions, was critically affected by the strength of horizontal and vertical diffusion. The models with strong diffusion resulted in a wide plume, and the CAPTURE and threat score (OVERESTIMATE) for the models were high (low) and low (high) with weak diffusion.

3.2.2. Plume 4: Wet Deposition Was Critical

Plume 4 was observed in Chiba Prefecture from 06:00 to 12:00 JST on 16 March 2011. The ^{137}Cs contributing to the plume was released late in the night of 15 March to early the next morning. During this period, precipitation was simulated in the region from the FDNPP to TMA (Figure 6a), and wet deposition occurred in the models. The plume emitted during this period was transported southward by a north-northeasterly wind across Chiba Prefecture (Figure 6a).

Due to the inaccuracy of the wind field simulated by WRF-Chem-T and NICAM, their simulated plumes were located over the ocean east of Chiba Prefecture (i.e., more east than the plumes simulated by the models with good performance). In SCALE, heavy precipitation occurred around the FDNPP during this period. Most of the atmospheric ^{137}Cs emitted from the FDNPP was removed by wet deposition to reduce the concentration below the threshold (10 Bq/m^3) for the analyses, and the plume did not arrive at Chiba Prefecture in SCALE. The plume simulated by Polyphemus also did not arrive at this prefecture, which was probably caused by the strong vertical diffusion. The scores for Plume 4 obtained by these four models were low (Table S2).

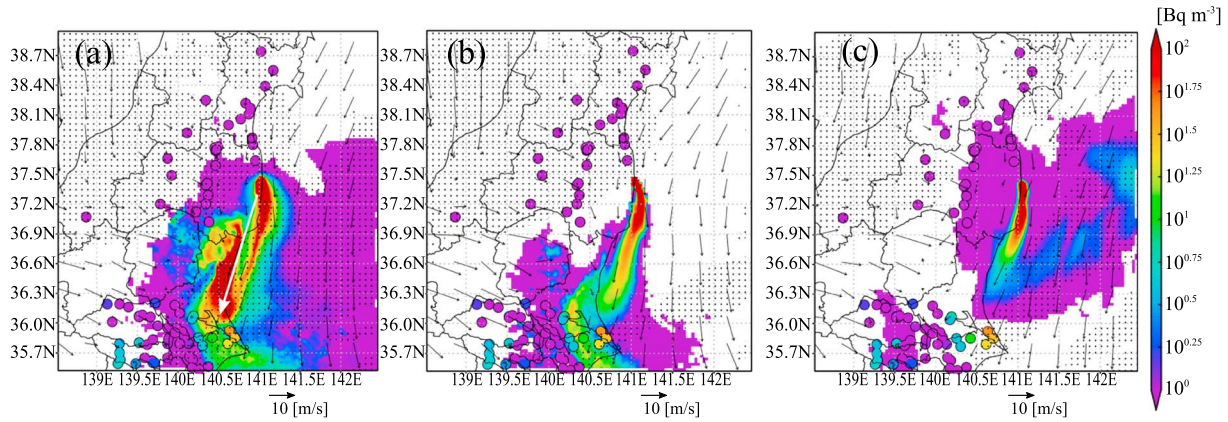

Figure 6. Same as Figure 4 but at 08:00 JST on 16 March 2011 by (a) the model ensemble mean and simulated by (b) GEARN and (c) WRF-CMAQ.

Except for the abovementioned four models, the results of the models with a small wet deposition flux showed that ^{137}Cs was transported by the northerly wind and reached Chiba Prefecture (Figure 6b). In contrast, the models with a large wet deposition flux calculated a significant reduction in the atmospheric ^{137}Cs concentration, and the plume did not reach Chiba Prefecture (Figure 6c). The scatterplot of the score and mean wet deposition flux shows a negative correlation (Figure 7a), whereas the correlation between the dry deposition flux and the score is weak (Figure 7b). Atmospheric ^{137}Cs in Plume 4 was favorably simulated by models with a relatively small wet deposition flux. These results suggest that the performance of the models for Plume 4 was determined by the wet deposition process.

We should note that the wet deposition flux was not only dependent upon the wet deposition itself but also precipitation (e.g., the intensity, duration, and vertical structure of the precipitation). Thus, the results do not directly lead to the conclusion that models with small wet deposition flux always have good performance.

3.2.3. Plume 8: Deposition Was Critical

Plume 8 was observed in the Nakadori area from 18:00 JST on 20 March to 06:00 JST on 21 March 2011. The ^{137}Cs contributing to the plume was released on the afternoon of 20 March. It was transported in a

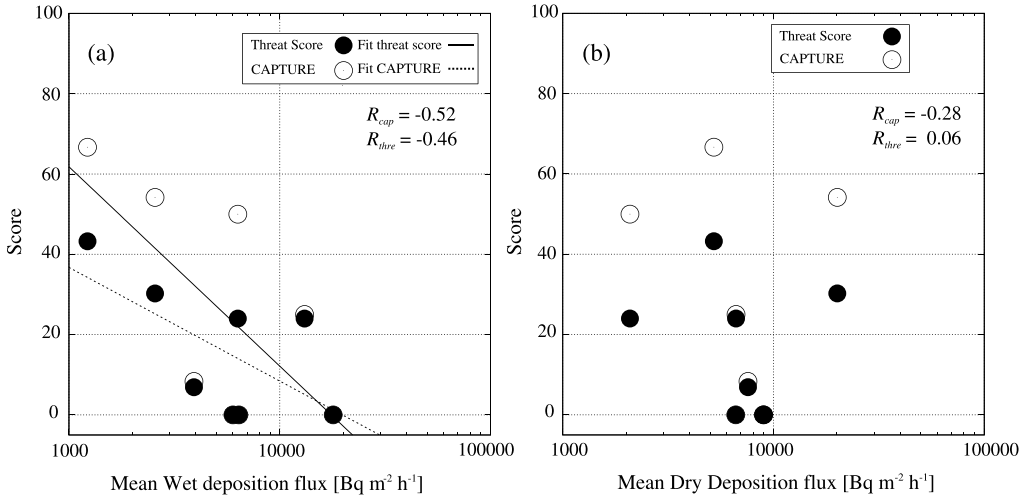

Figure 7. Relationship between the scores and the (a) mean wet deposition flux and (b) mean dry deposition flux averaged during Plume 4 over the path where the atmospheric ^{137}Cs concentration exceeded 10 Bq/m^3. The closed and open symbols show threat score and CAPTURE, respectively. Results of SCALE, WRF-Chem-T, Polyphemus, and NICAM are not shown. R_{cap} (R_{thre}) in (a) and (b) respectively shows the correlation coefficient between (a) the wet deposition flux and (b) the dry deposition flux and CAPTURE (threat score).

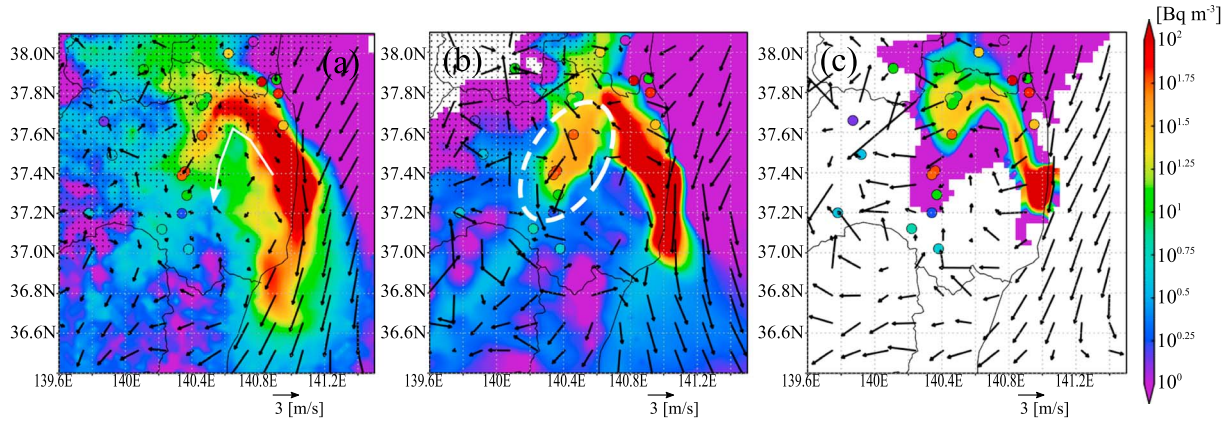

Figure 8. Same as Figure 4 but at 22:00 JST 20 March 2011 by (a) the model ensemble mean and simulated by (b) WRF-CMAQ and (c) SCALE.

northwestward direction under the influence of a southeasterly wind, and the lower part of the plume was then transported southward over the Nakadori area after 21:00 JST on 20 March (Figure 8a).

In the results of the models with good performance, high ^{137}Cs concentrations were distributed across the Nakadori area by a north-northeasterly wind (white dashed circle in Figure 8b). Conversely, the ^{137}Cs simulated by the models with poor performance was not transported southward toward the Nakadori area by the abovementioned wind (Figure 8c). These trends were commonly seen in the results with poor performance. These results suggest that atmospheric ^{137}Cs was excessively removed by the deposition process in the models with poor performance. A scatterplot of the mean total (dry + wet) deposition flux and the scores support this suggestion (Figure 9). The scores were negatively correlated with the mean total deposition flux, which indicated that the performance of the models for Plume 8 was determined by the deposition process. The horizontal distribution of atmospheric ^{137}Cs was more realistically reproduced by the models with a small deposition flux.

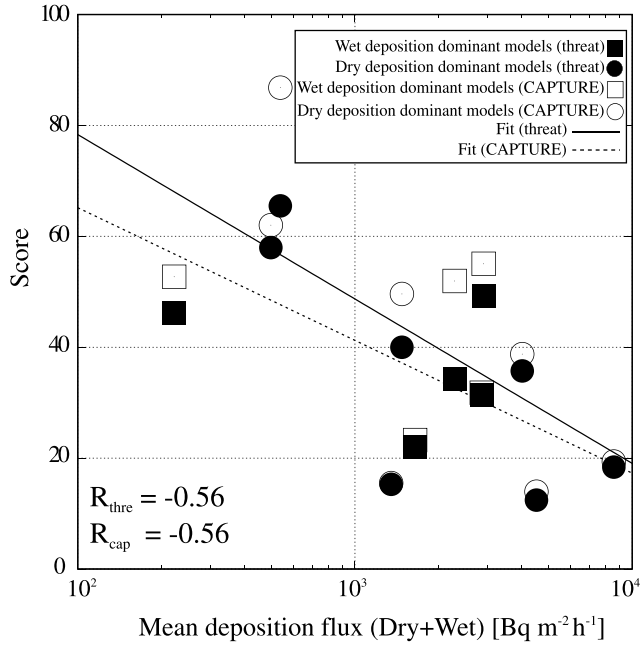

Figure 9. Relationship between the scores and total deposition (dry + wet) flux averaged during Plume 8 over the path where atmospheric ^{137}Cs exceeded 10 Bq/m^3. Squares and circles show the results of models in which wet and dry deposition were respectively dominant. Closed and open symbols show the threat score and CAPTURE, respectively. R_{cap} (R_{thre}) shows the correlation coefficient between the wet deposition flux and the dry deposition flux and CAPTURE (threat score).

The fraction of wet deposition to the total (dry + wet) deposition varied considerably among the models. Wet deposition was dominant in some models (AIST-MM, HIRAT-LPRM, WRF-Chem-J, WRF-CMAQ, WRF-Chem-T, and NICAM, shown as "wet deposition dominant models" in Figure 9), whereas dry deposition was dominant in the other models (PELLO, IdX, GEARN, NHM-Chem, and SCALE, shown as "dry deposition dominant models" in Figure 9) during the Plume 8 period.

3.3. Summary of the Atmospheric ^{137}Cs Score

On the basis of the results above, we concluded that the wind field was most critical for properly reproducing atmospheric ^{137}Cs. The results further showed that the deposition process was one of the main reasons for the variability in the results of the models. We found that a large deposition flux tended to result in poor scores, and the ratio of the contribution of wet deposition to the total (dry + wet) deposition differed noticeably among the models. Improvements in the modeling of the deposition processes would considerably reduce the uncertainties in the atmospheric dispersion simulation. In addition, the strength of the horizontal diffusion was also found to be critical to the results of the models under simple meteorological conditions where the deposition was less important. These conclusions are considered to be robust and pertinent because all models used the same external data, and the uncertainties originating from these data were relatively small compared with those in previous studies.

Table 8
Scores of Each Ensemble Mean

Ensemble mean	CC	FMS	FB	KSP	RANK	C	T
Equal weight	0.690	80.7	−0.074	3.41	3.21	40.3	23.8
Weighted by RANK × (threat score)	0.702	84.1	−0.034	3.49	3.30	46.1	27.2

Note. C and T represent CAPTURE and threat score, respectively; CC = correlation coefficient, FB = fractional bias, FMS = figure of merit in space, KSP = Kolmogorov-Smirnov parameter.

3.4. Weighting for Ensemble Mean Calculation

As reported by a previous model intercomparison study (Kitayama et al., 2018), the model ensemble mean is useful for producing results that properly reproduce the deposition and atmospheric concentration of ^{137}Cs. However, a simple ensemble average with equal weights may include erroneous information from all models regardless of their performance; thus, the performance of a multimodel ensemble could be poorer than that of individual well-performing models (Tables 4 and 6). An ensemble average reflecting the performance of each model enables the improvement of the performance of the multimodel ensemble (Draxler et al., 2015; Solazzo & Galmarini, 2015). In this study, we introduced a score defined as the product of the threat score and RANK. The multimodel ensemble was created by using the weight calculated from the score. Henceforth, we refer to the ensemble created in this way as the weighted multimodel ensemble.

The performance of the weighted multimodel ensemble is summarized in Table 8. The scores of the weighted multimodel ensemble are better than that of the simple (nonweighted) multimodel ensemble. CAPTURE, FMS, and RANK of the multimodel ensemble were about 46, 84, and 3.3, respectively.

4. Conclusions and Future Studies

This study conducted a model intercomparison of atmospheric dispersion models targeting ^{137}Cs released from the FDNPP in March 2011. For the intercomparison, all participating models used the same emission inventory of Katata et al. (2015) and the same meteorological data with a fine spatiotemporal resolution, which were produced by the Japanese operational weather forecast model coupled with an assimilation system (S15 and S17). These setups enabled us to focus on the internal elements of the models (i.e., physical processes and strength of diffusion) that may contribute to the model-to-model variability in performance. Overall, 12 models from 11 groups participated in the intercomparison, including Lagrangian and Eulerian dispersion models.

The results of the intercomparison indicated that the multimodel ensemble captured about 40% of the high-concentration event of ^{137}Cs measured by the SPM network during the study period. Through analyses of the intercomparison results, we revealed that the meteorological fields (i.e., wind direction and velocity) were the most critical factors for reproducing of the atmospheric ^{137}Cs events measured by the SPM network. The model performance was not good for any model when the meteorological fields did not reproduce the measured meteorological conditions. In such cases, the multimodel ensemble captured less than 10% of the high-concentration ^{137}Cs event (e.g., Plumes 7 and 9). Our analyses further revealed that the magnitude of the deposition flux was critical to the reproduction of atmospheric ^{137}Cs events when the uncertainties in the meteorological fields and source term are insignificant. With the source term and meteorological data, models with a small deposition flux tended to have good scores, and vice versa.

The diffusion strength has a large impact on the geographical distribution of plumes under simple meteorological conditions (Plume 2), in which precipitation or changes in wind direction did not occur. However, the reasonable strength of the diffusion was not determined in this study. To determine the reasonable strength of the diffusion, it is necessary to clarify the width and the location of the plumes accurately from the observation. The observations with fine spatiotemporal resolution along the pathway of the ^{137}Cs plumes enable us to elucidate the width and the location of the plume. The measurement of ^{137}Cs over the north part of Ibaraki Prefecture, where SPM measurement could not be used in this study, is also required for addressing the uncertainties of the diffusion. Along with the horizontal diffusion, the vertical diffusion and the three-dimensional structure of simulated ^{137}Cs should be investigated for improved understanding of the behavior of ^{137}Cs in the atmosphere.

In addition, we evaluated the model in detail for each plume based on the threat score, CAPTURE, and OVERESTIMATE (section 3.2). The evaluation based on the magnitude of the difference between models and observation enables us to elucidate the differences among the models in detail. Such analyses should be conducted in the future.

The deposition process was calculated by a variety of methods in each model in this study. For example, some models calculated the wet deposition flux based on the size of the aerosol attached to ^{137}Cs, but the size of the aerosol was not considered in other models. Such variability in the methods used to calculate the deposition would result in large intermodel spread. In addition to the deposition process itself, the uncertainties of precipitation directly resulted in large intermodel spread. Further analyses exploring the uncertainties of both precipitation and deposition processes are required for understanding the uncertainties of the deposition process in detail.

The results of the intercomparison also showed that the multimodel ensemble weighted by the score in terms of both atmospheric ^{137}Cs and deposition shows better performance than does each model individually.

Acknowledgments
This study was supported by the advancement of meteorological and global environmental predictions utilizing observational "Big Data" of the social and scientific priority issues (Theme 4) to be tackled using the post K computer of the FLAGSHIP2020 Project and by the Integrated Research Program for Advancing Climate Models of MEXT. Some of the authors were supported by the Environment Research and Technology Development Fund (5-1501 and 1-1802) of the Ministry of the Environment, Japan (MOEJ), JSPS KAKENHI (Grant Number 24110002, 17K00533), the National Natural Science Foundations of China (Grant Number 11475100). We also thank all the local governments for allowing us to use the radionuclides in SPM collected on the used filter tapes in the SPM monitors. The simulated data are downloadable for research purposes from the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) server (https://ebcrpa.jamstec.go.jp/sato_etal_jgr2018_data). The data of SPM observation was obtained from Appendix A of Oura et al. (2015, http://www.radiochem.org/en/j-online152.html) and the data of airborne measurement was obtained from a website (https://emdb.jaea.go.jp/emdb/en/portals/b1010301/).

References

Adachi, K., Kajino, M., Zaizen, Y., & Igarashi, Y. (2013). Emission of spherical cesium-bearing particles from an early stage of the Fukushima nuclear accident. *Scientific Reports*, *3*(1), 2554. https://doi.org/10.1038/srep02554.

Arnold, D., Maurer, C., Wotawa, G., Draxler, R., Saito, K., & Seibert, P. (2015). Influence of the meteorological input on the atmospheric transport modelling with FLEXPART of radionuclides from the Fukushima Daiichi nuclear accident. *Journal of Environmental Radioactivity*, *139*, 212–225. https://doi.org/10.1016/j.jenvrad.2014.02.013

Brandt, J., Christensen, J. H., & Frohn, L. M. (2002). Modelling transport and deposition of caesium and iodine from the Chernobyl accident using the DREAM model. *Atmospheric Chemistry and Physics*, *2*(5), 397–417. https://doi.org/10.5194/acp-2-397-2002

Chino, M., Nakayama, H., Nagai, H., Terada, H., Katata, G., & Yamazawa, H. (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. https://doi.org/10.1080/18811248.2011.9711799

Draxler, R., Arnold, D., Chino, M., Galmarini, S., Hort, M., Jones, A., et al. (2015). World Meteorological Organization's model simulations of the radionuclide dispersion and deposition from the Fukushima Daiichi Nuclear Power Plant accident. *Journal of Environmental Radioactivity*, *139*, 172–184. https://doi.org/10.1016/j.jenvrad.2013.09.014

Draxler, R., Arnold, D., Galmarini, S., Hort, M., Jones, A., Leadbetter, S., et al. (2013). The World Meteorological Organization's evaluation of meteorological analyses for the radionuclide dispersion and deposition from the Fukushima Daiichi Nuclear Power Plant accident. *Journal of Environmental Radioactivity*, *139*, 172–184. https://doi.org/10.1016/j.jenvrad.2013.09.014

Frehlich, R., & Sharman, R. (2008). The use of structure functions and spectra from numerical model output to determine effective model resolution. *Monthly Weather Review*, *136*(4), 1537–1553. https://doi.org/10.1175/2007MWR2250.1

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Hirao, S., Yamazawa, H., & Nagae, T. (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. https://doi.org/10.1080/00223131.2013.757454

Hu, X., Li, D., Huang, H., Shen, S., & Bou-Zeid, E. (2014). Modeling and sensitivity analysis of transport and deposition of radionuclides from the Fukushima Dai-ichi accident. *Atmospheric Chemistry and Physics*, *14*(20), 11,065–11,092. https://doi.org/10.5194/acp-14-11065-2014

Igarashi, Y., Kajino, M., Zaizen, Y., Adachi, K., & Mikami, M. (2015). Atmospheric radioactivity over Tsukuba, Japan: A summary of three years of observations after the FDNPP accident. *Progress in Earth and Planetary Science*, *2*(1), 44. https://doi.org/10.1186/s40645-015-0066-1.

Institute for Global Environment and Society (IGES) (1989). Grid analysis and display system (GrADS), Retrieved from http://cola.gmu.edu/grads/downloads.php (Accessed 15 Oct, 2018)

Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2011). Results of the fourth airborne monitoring survey by MEXT. http://radioactivity.nsr.go.jp/en/contents/4000/3179/24/1270_1216.pdf.

Kajino, M., Inomata, Y., Sato, K., Ueda, H., Han, Z., An, J., et al. (2012). Development of the RAQM2 aerosol chemical transport model and predictions of the northeast Asian aerosol mass, size, chemistry, and mixing type. *Atmospheric Chemistry and Physics*, *12*(24), 11833–11856. https://doi.org/10.5194/acp-12-11833-2012

Kajino, M., Ishizuka, M., Igarashi, Y., Kita, K., Yoshikawa, C., & Inatsu, M. (2016). Long-term assessment of airborne radiocesium after the Fukushima nuclear accident: Re-suspension from bare soil and forest ecosystems. *Atmospheric Chemistry and Physics*, *16*(20), 13149–13172. https://doi.org/10.5194/acp-16-13149-2016

Kaneyasu, N., Ohashi, H., Suzuki, F., Okuda, T., & Ikemori, F. (2012). Sulfate aerosol as a potential transport medium of radiocesium from the Fukushima nuclear accident. *Environmental Science & Technology*, *46*(11), 5720–5726. https://doi.org/10.1021/es204667h.

Katata, G., Chino, M., Kobayashi, T., Terada, H., Ota, M., Nagai, H., et al. (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. *Atmospheric Chemistry and Physics*, *15*(2), 1029–1070. https://doi.org/10.5194/acp-15-1029-2015

Katata, G., Ota, M., Terada, H., Chino, M., & Nagai, H. (2012). Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source term estimation and local-scale atmospheric dispersion in early phase of the accident. *Journal of Environmental Radioactivity*, *109*, 103–113. https://doi.org/10.1016/j.jenvrad.2012.02.006

Kitayama, K., Morino, Y., Takigawa, M., Nakajima, T., Hayami, H., Nagai, H., et al. (2018). 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. *Journal of Geophysical Research: Atmospheres*, *123*, 7754–7770. https://doi.org/10.1029/2017JD028230

Kondo, H., Saigusa, N., Murayama, S., Yamamoto, S., & Kannari, A. (2001). A numerical simulation of the daily variation of CO_2 in the central part of Japan-summer case. *Journal of the Meteorological Society of Japan*, *79*(1), 11–21. https://doi.org/10.2151/jmsj.79.11

Kristiansen, N. I., Stohl, A., Olivié, D. J. L., Croft, B., Søvde, O. A., Klein, H., et al. (2016). Evaluation of observed and modelled aerosol lifetimes using radioactive tracers of opportunity and an ensemble of 19 global models. *Atmospheric Chemistry and Physics*, *16*(5), 3525–3561. https://doi.org/10.5194/acp-16-3525-2016

Leadbetter, S. J., Hort, M. C., Jones, A. R., Webster, H. N., & Draxler, R. R. (2015). Sensitivity of the modelled deposition of caesium-137 from the Fukushima Dai-ichi Nuclear Power Plant to the wet deposition parameterisation in NAME. *Journal of Environmental Radioactivity*, *139*, 200–211. https://doi.org/10.1016/j.jenvrad.2014.03.018

Mathieu, A., Kajino, M., Korsakissok, I., Périllat, R., Quélo, D., Quérel, A., et al. (2018). Fukushima Daiichi-derived radionuclides in the atmosphere, transport and deposition in Japan: A review. *Applied Geochemistry*, *91*(June 2017), 122–139. https://doi.org/10.1016/j.apgeochem.2018.01.002

Mathieu, A., Korsakissok, I., Quelo, D., Groell, J., Tombette, M., Didier, D., et al. (2012). Atmospheric dispersion and deposition of radionuclides from the Fukushima Daiichi Nuclear Power Plant accident. *Elements*, *8*(3), 195–200. https://doi.org/10.2113/gselements.8.3.195

Mölders, N., & Kramm, G. (2014). *Lectures in meteorology*. Cham: Springer Atmospheric Sciences, Springer International Publishing.

Morino, Y., Ohara, T., & Nishizawa, M. (2011). Atmospheric behavior, deposition, and budget of radioactive materials from the Fukushima Daiichi Nuclear Power Plant in March 2011. *Geophysical Research Letters*, *38*, LOOG11. https://doi.org/10.1029/2011GL048689.

Morino, Y., Ohara, T., Watanabe, M., Hayashi, S., & Nishizawa, M. (2013). Episode analysis of deposition of radiocesium from the Fukushima Daiichi Nuclear Power Plant accident. *Environmental Science & Technology*, *47*(5), 2314–2322. https://doi.org/10.1021/es304620x

Nakajima, T., Misawa, S., Morino, Y., Tsuruta, H., Goto, D., Uchida, J., et al. (2017). 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*(1), 2. https://doi.org/10.1186/s40645-017-0117-x.

Nishizawa, S., Yashiro, H., Sato, Y., Miyamoto, Y., & Tomita, H. (2015). Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations. *Geoscientific Model Development*, *8*(10), 3393–3419. https://doi.org/10.5194/gmd-8-3393-2015

Oura, Y., Ebihara, M., Tsuruta, H., Nakajima, T., Ohara, T., Ishimoto, M., et al. (2015). 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. *Journal of Nuclear and Radiochemical Sciences*, *15*(2), 2_1–2_12. https://doi.org/10.14494/jnrs.15.2_1

Quérel, A., Roustan, Y., Quélo, D., & Benoit, J. P. (2015). Hints to discriminate the choice of wet deposition models applied to an accidental radioactive release. *International Journal of Environment and Pollution*, *58*(4), 268. https://doi.org/10.1504/IJEP.2015.077457

Saito, K., Fujita, T., Yamada, Y., Ishida, J.-I., Kumagai, Y., Aramani, K., et al. (2006). The operational JMA nonhydrostatic mesoscale model. *Monthly Weather Review*, *134*(4), 1266–1298. https://doi.org/10.1175/MWR3120.1

Saito, K., Shimbori, T., & Draxler, R. (2015). JMA's regional atmospheric transport model calculations for the WMO technical task team on meteorological analyses for Fukushima Daiichi Nuclear Power Plant accident. *Journal of Environmental Radioactivity*, *139*, 185–199. https://doi.org/10.1016/j.jenvrad.2014.02.007

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., Kajikawa, Y., & Tomita, H. (2015). Impacts of cloud microphysics on trade wind cumulus: Which cloud microphysics processes contribute to the diversity in a large eddy simulation? *Progress in Earth and Planetary Science*, *2*(1), 23. https://doi.org/10.1186/s40645-015-0053-6

Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., et al. (2014). The non-hydrostatic icosahedral atmospheric model: Description and development. *Progress in Earth and Planetary Science*, *1*(1), 18. https://doi.org/10.1186/s40645-014-0018-1

Saunier, O., Mathieu, A., Didier, D., Tombette, M., Quélo, D., Winiarek, V., & Bocquet, M. (2013). An inverse modeling method to assess the source term of the Fukushima Nuclear Power Plant accident using gamma dose rate observations. *Atmospheric Chemistry and Physics*, *13*(22), 11403–11421. https://doi.org/10.5194/acp-13-11403-2013

Science Council of Japan (2014). 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.

Sekiyama, T. T., Kajino, M., & Kunii, M. (2017). The impact of surface wind data assimilation on the predictability of near-surface plume advection in the case of the Fukushima nuclear accident. *Journal of the Meteorological Society of Japan Series II*, *95*(6), 447–454. https://doi.org/10.2151/jmsj.2017-025

Sekiyama, T. T., Kunii, M., Kajino, M., & Shimbori, T. (2015). Horizontal resolution dependence of atmospheric simulations of the Fukushima nuclear accident using 15-km, 3-km, and 500-m grid models. *Journal of the Meteorological Society of Japan Series II*, *93*(1), 49–64. https://doi.org/10.2151/jmsj.2015-002

Skamarock, W. C. (2004). Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, *132*(12), 3019–3032. https://doi.org/10.1175/MWR2830.1

Solazzo, E., & Galmarini, S. (2015). The Fukushima-^{137}Cs deposition case study: Properties of the multi-model ensemble. *Journal of Environmental Radioactivity*, *139*, 226–233. https://doi.org/10.1016/j.jenvrad.2014.02.017

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., et al. (2012). 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*, *12*(5), 2313–2343. https://doi.org/10.5194/acp-12-2313-2012

Takemura, T., Nakamura, H., Takigawa, M., Kondo, H., Satomura, T., Miyasaka, T., & Nakajima, T. (2011). A numerical simulation of global transport of atmospheric particles emitted from the Fukushima Daiichi Nuclear Power Plant. *SOLA*, *7*, 101–104. https://doi.org/10.2151/sola.2011-026

Terada, H., Katata, G., Chino, M., & Nagai, H. (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. https://doi.org/10.1016/j.jenvrad.2012.05.023

Terada, H., Nagai, H., & Yamazawa, H. (2013). Validation of a Lagrangian atmospheric dispersion model against middle-range scale measurements of ^{85}Kr concentration in Japan. *Journal of Nuclear Science and Technology*, *50*(12), 1198–1212. https://doi.org/10.1080/00223131.2013.840545

Terasaka, Y., Yamazawa, H., Hirouchi, J., Hirao, S., Sugiura, H., Moriizumi, J., & Kuwahara, Y. (2016). Air concentration estimation of radionuclides discharged from Fukushima Daiichi Nuclear Power Station using NaI (Tl) detector pulse height distribution measured in Ibaraki Prefecture. *Journal of Nuclear Science and Technology*, *53*(12), 1919–1932. https://doi.org/10.1080/00223131.2016.1193453

Torii, T., Sanda, Y., Sugita, T., Kondo, A., Shikaze, Y., Takahashi, M., et al. (2012). Investigation of radionuclide distribution using aircraft for surrounding environmental survey from Fukushima Dai-ichi Nuclear Power Plant (in Japanese), JAEA-Technology, 2012-036, Japan Atomic Energy Agency, Ibaraki, Japan, https://jopss.jaea.go.jp/pdfdata/JAEA-Technology-2012-036.pdf

Tsuruta, H., Oura, Y., Ebihara, M., Moriguchi, Y., Ohara, T., & Nakajima, T. (2017). Spatio-temporal distribution of atmospheric radiocesium in eastern Japan just after the TEPCO Fukushima Daiichi Nuclear Power Plant accident—Analysis of used filter-tapes of SPM monitors in air quality monitoring stations (in Japanese). *Earosoru Kenkyu*, *32*, 244–255. https://doi.org/10.11203/jar.32.244

Tsuruta, H., Oura, Y., Ebihara, M., Moriguchi, Y., Ohara, T., & Nakajima, T. (2018). 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. *Geochemical Journal*, *52*(2), 103–121. https://doi.org/10.2343/geochemj.2.0520

Tsuruta, H., Oura, Y., Ebihara, M., Ohara, T., & Nakajima, T. (2014). First retrieval of hourly atmospheric radionuclides just after the Fukushima accident by analyzing filter-tapes of operational air pollution monitoring stations. *Scientific Reports*, *4*(1), 6717. https://doi.org/10.1038/srep06717

Uchida, J., Mori, M., Hara, M., Satoh, M., Goto, D., Kataoka, T., et al. (2017). Impact of lateral boundary errors on the simulation of clouds with a nonhydrostatic regional climate model. *Monthly Weather Review*, *145*(12), 5059–5082. https://doi.org/10.1175/MWR-D-17-0158.1

von Schoenberg, P., Boson, J., Grahn, H., Nylén, T., Ramebäck, H., & Thaning, L. (2014). Atmospheric dispersion of radioactive material from the Fukushima Daiichi Nuclear Power Plant. In R. M. A. T. D. G. Steyn & P. J. H. Builtjes (Eds.), *Air pollution modeling and its application XXII*, edited by, (pp. 345–349). Netherlands: Springer Netherlands.

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed.). Burlington, MA: Academic Press.

Winiarek, V., Bocquet, M., Duhanyan, N., Roustan, Y., Saunier, O., & Mathieu, A. (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*, 268–279. https://doi.org/10.1016/j.atmosenv.2013.10.017

Yasunari, T. J., Stohl, A., Hayano, R. S., Burkhart, J. F., Eckhardt, S., & Yasunari, T. (2011). Cesium-137 deposition and contamination of Japanese soils due to the Fukushima nuclear accident. *Proceedings of the National Academy of Sciences*, *108*(49), 19,530–19,534. https://doi.org/10.1073/pnas.1112058108

Yumimoto, K., Morino, Y., Ohara, T., Oura, Y., Ebihara, M., Tsuruta, H., & Nakajima, T. (2016). Inverse modeling of the ^{137}Cs 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. https://doi.org/10.1016/j.jenvrad.2016.06.018.