NOTES AND CORRESPONDENCE

Performance of a 250-m Grid Eulerian Dispersion Simulation Evaluated at Two Coastal Monitoring Stations in the Vicinity of the Fukushima Daiichi Nuclear Power Plant

Tsuyoshi Thomas SEKIYAMA

Meteorological Research Institute, Tsukuba, Japan

and

Mizuo KAJINO

Meteorological Research Institute, Tsukuba, Japan

Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

(Manuscript received 16 March 2021, in final form 22 April 2021)

Abstract

Using two horizontal resolutions (3 km and 250 m), this study examined the performance of Eulerian models in simulating dispersion fields at two coastal monitoring stations in the vicinity of a pollutant source (3.2 km and 17.5 km distant) under the situation of the Fukushima 2011 nuclear accident. A 250-m grid simulation was newly conducted for the examination and was able to reproduce the wind and concentration fields in detail over complex terrain. The 3-km grid model could not reproduce the details of the winds and plumes around the Fukushima Daiichi Nuclear Power Plant but occasionally yielded a higher performance with a lower undetected error rate compared with the 250-m grid model due to the large numerical diffusion of the former. A disadvantage of Eulerian dispersion models is expected to be the artificial numerical diffusion in the advection process near emission sources. The artificial numerical diffusion increases the false alarm ratio (number of strikeouts while swinging) but fortunately decreases the undetected error rate (number of strikeouts while looking). This characteristic is appropriate for environmental emergency response (EER) systems. Furthermore, the 250-m grid model result was improved by a plume augmentation (i.e., max pooling) process, which enlarged the plume widths and masked short time lags and small plume drifts. Plume augmentation was advantageous to the high-resolution model for improving statistical scores, which is beneficial for EER systems.

Keywords high resolution modeling; weather modeling; advection modeling; radioactive cesium-137; environmental emergency response

Citation Sekiyama, T. T., and M. Kajino, 2021; Performance of a 250-m grid Eulerian dispersion simulation evaluated at two coastal monitoring stations in the vicinity of the Fukushima Daiichi Nuclear Power Plant. *J. Meteor. Soc. Japan*, **99**, 1089–1098, doi:10.2151/jmsj.2021-052.

Corresponding author: Tsuyoshi Thomas Sekiyama, Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
E-mail: tsekiyam@mri-jma.go.jp
J-stage Advance Published Date: 10 May 2021

1. Introduction

The authors have previously investigated the dependence of dispersion model performance on the horizontal resolution using the data of the Fukushima 2011 nuclear accident (Sekiyama et al. 2015, 2017; Sato et al. 2018, 2020; Sekiyama and Kajino 2020). These studies have essentially found that when using 15-, 5-, 3-, and 1-km grid models, higher-resolution models yield better performance over complex Japanese terrain. Specifically, while the 3-km and 1-km grid models performed comparably over plain regions more than 100 km from the Fukushima Daiichi Nuclear Power Plant (FDNPP), the 1-km grid model was evidently superior to the 3-km grid model over mountainous regions approximately 50 km from the FDNPP. Unfortunately, the 15-km and 5-km grid models exhibited poorer performance compared with the 3-km and 1-km grid models.

The Cs-137 concentration data used for validation in these studies were retrieved by Tsuruta et al. (2014) and Oura et al. (2015) and recorded at approximately 100 monitoring stations. However, most of the stations are located more than 50 km from the FDNPP; among them, the nearest station is 26 km away. Consequently, these studies did not evaluate the model performance in the area immediately surrounding the FDNPP. In principle, Eulerian models, which were utilized in these studies, are generally not expected to be good at simulating plume dispersion in the vicinity of emission sources (cf. Rood 1987). However, it is not theoretically trivial to estimate the poor performance of a Eulerian model, especially when the horizontal resolution is almost comparable to the distance between a monitoring point and an emission source. Therefore, the model performance should be evaluated through observations at not only distant but also close ranges.

In this paper, the close-range performance of Eulerian dispersion simulations is evaluated using two horizontal resolutions, namely, 3 km and 250 m. The 3-km grid simulation is derived from Sekiyama and Kajino (2020), whereas the 250-m grid simulation is newly conducted for the evaluation. Fortunately, Tsuruta et al. (2018) released new Cs-137 concentration data retrieved at two monitoring stations in the vicinity of the FDNPP (3.2 km and 17.5 km distant). This new dataset allows the authors to evaluate the close-range dispersion model performance. The 250-m grid plume dispersion is categorized as a sub-kilometer-scale simulation, which is currently being examined by an increasing number of studies, mainly for complex terrain (e.g., Bao et al. 2018; Wiersema et al. 2020)

and urban pollution (e.g., Lateb et al. 2016; Nakayama et al. 2016; Li et al. 2018; Hamer et al. 2020). The authors would like to contribute to the studies on high-resolution geophysical model by focusing on the rural and coastal locations near the FDNPP.

In addition, in the case of environmental emergency responses (EER; cf. World Meteorological Organization 2006), both the Eulerian and Lagrangian plume dispersion models would be required, especially in the areas very close to emission sources. Hence, we should understand beforehand what types of models, what model resolution, and what model configuration are needed for EER. The authors hope that this paper will provide insight into what model resolution and configuration are suitable for supporting EER.

2. Methodology

2.1 Models

The 3-km grid meteorological analysis was provided at an hourly resolution by Sekiyama et al. (2017) and Sekiyama and Kajino (2020). This simulation was performed using a 4-dimensional data assimilation system consisting of a nonhydrostatic regional weather prediction model (referred to as the NHM; cf. Saito et al. 2006, 2007), a local ensemble transform Kalman filter (LETKF; cf. Miyoshi and Aranami 2006; Kunii 2014), and Japan Meteorological Agency (JMA) operational observation datasets. The NHM was being operationally used by JMA at the time of the Fukushima nuclear accident (March 2011) for daily national weather forecasts with four-dimensional variational assimilation (cf. Honda et al. 2005). The analysis has 60 vertical layers from the surface to a 22-km elevation within the model domain over eastern Japan, as presented in Fig. 1a. The boundary conditions for the model domain were provided by the JMA operational global analysis system. The details of the model and data assimilation settings are described in Sekiyama et al. (2017) and Sekiyama and Kajino (2020). This meteorological analysis has been employed not only by the above studies referenced but also by Sato et al. (2018), Sekiyama and Iwasaki (2018), Iwasaki et al. (2019), Takagi et al. (2020), and Goto et al. (2020) for nuclear accident air pollution modeling.

The 250-m grid meteorological analysis was calculated by the NHM nested by the 1-km meteorological analysis within the model domain, as presented in Fig. 1b. The 1-km analysis was provided by Sekiyama and Kajino (2020), who implemented a one-way nested data assimilation scheme (Kunii 2014) nested by the 3-km meteorological analysis aforementioned. Note that the 3-km, 1-km, and 250-m grid analyses were

calculated using the same model and physical configurations, except for the model domains and horizontal resolution-dependent setups. All the resolution models implemented the improved Mellor–Yamada level 3 closure model (Nakanishi and Niino 2004, 2006) as a turbulence scheme. A cumulus parameterization was not utilized for any models in this study. The 250-m grid NHM calculation lasted 20 h for a 24-h simulation using 96 cores of Fujitsu FX100 with 840×980 grids. On the other hand, the 3-km grid NHM calculation lasted 4 h for a 24-h simulation using one core of Fujitsu FX100 with 215×259 grids.

Cs-137 plume dispersion was calculated by an offline Eulerian regional air quality model, which was driven by either the 3-km or 250-m grid meteorological analysis. This model was previously developed and evaluated by Kajino et al. (2012, 2016, 2018, 2019a, b) and Mathieu et al. (2018). The 3-km and 250-m grid meteorological analyses were input at 1-h (3-km grid) or 10-min (250-m grid) intervals into the offline air quality model, in which the dynamical time step was set to 24 s (3-km grid) or 2 s (250-m grid) using the time-interpolated meteorological analysis. The Cs-137 emission scenario was provided by Katata et al. (2015). In this model, Cs-137 was assumed to be mixed in sulfate-organic mixture aerosol particles, which were injected into a grid cell above the FDNPP at 20–150-m heights, following the time-varying emission scenario. The details of the model settings are described in Sekiyama et al. (2015, 2017) and Sekiyama and Kajino (2020). The 250-m grid offline model calculation lasted 12 h for a 24-h dispersion simulation using 112 cores of Intel Xeon (Haswell) with the same domain as the 250-m grid NHM. On the other hand, the 3-km grid offline model calculation lasted 0.5 h for a 24-h dispersion simulation using eight cores of Intel Xeon (Haswell) with the same domain as the 3-km grid NHM.

2.2 Observations

The observational Cs-137 concentrations were retrieved hourly from filter tapes installed in prefectural governments' suspended particulate matter (SPM) monitors, of which the theoretical detection limit was 0.1 Bq m^{-3} (Tsuruta et al. 2014, 2018). The two SPM monitoring stations, as reported by Tsuruta et al. (2018), were located at Futaba and Naraha in the vicinity of the FDNPP along the east coast of Fukushima Prefecture (Fig. 1c). The Futaba and Naraha monitoring stations were located 3.2-km northwest and 17.5-km south-southwest of the FDNPP, respectively. The plume arrivals were defined with a threshold of 1.5

Fig. 1. Model domains of (a) the 3-km grid model and (b) the 250-m grid model. (c) Detailed locations of the Fukushima Daiichi Nuclear Power Plant (FDNPP), Futaba monitoring station, and Naraha monitoring station with 50-m interval elevation contours.

Fig. 2. Time series of the hourly averaged Cs-137 concentrations in March 2011 local time at the (a) Futaba monitoring station and (b) Naraha monitoring station. [A] and [B] indicate the snapshot times for Figs. 3a and 3b, respectively. Closed and open circles are observations defined as plumes (equal to or more than 1.5 Bq m⁻³) and non-plumes (less than 1.5 Bq m⁻³), respectively.

Bq m⁻³ using hourly averaged concentrations. The modeled concentrations were spatially linear-interpolated at each monitoring station. Sekiyama et al. (2017) and Sekiyama and Kajino (2020) utilized a threshold of 1.0 Bq m⁻³; the reason why this study used a threshold of 1.5 Bq m⁻³ is described below.

The major leakage of radioactive substances caused by the Fukushima nuclear accident lasted for 2 or 3 weeks starting on March 12, 2011, the day after the great earthquake. Most of the plumes flowed offshore to the Pacific Ocean with the Siberian winter monsoon. Nakajima et al. (2017) reported that the time windows of onshore plumes were limited to less than 50 h in total. However, the plumes often reached both Futaba and Naraha, as reported by Tsuruta et al. (2018), as these stations are near the FDNPP along the Pacific Ocean coast. Tsuruta et al. (2018) reported that high Cs-137 concentrations were observed at Futaba intermittently at a local time (LT) between March 12 and March 25, 2011, and at Naraha intermittently (LT) between March 14 and March 23, 2011. The observed concentrations are presented in Fig. 2, where the closed (open) circles indicate the existence (non-existence) of a plume defined by a threshold of 1.5 Bq m⁻³.

The background concentration seems to be approximately 1 Bq m⁻³ at Futaba (Fig. 2a), which is probably due to equipment contamination and filter tape cross-contamination (cf. Tsuruta et al. 2014, 2018). The background concentration at Naraha is also higher than the detection limit (0.1 Bq m⁻³) with a large deviation, in which the maximum values seem to exceed 1 Bq m⁻³ (Fig. 2b). This is also probably due to contamination and cross-contamination. Although turbulent diffusion may have partially influenced the background concentrations near the FDNPP, it is unnatural that the values did not fall below 1 Bq m⁻³, even during time slots in which winds are strongly directed toward the ocean at the FDNPP. Therefore, the threshold of plume existence/nonexistence was defined as 1.5 Bq m⁻³ to avoid contamination errors at these two locations. The sensitivity of the statistical scores to the threshold value was small between 1.5 Bq m⁻³ and 3 Bq m⁻³, although the scores evidently deteriorated with the threshold of 1 Bq m⁻³.

In this study, the model performance was evaluated by the following statistical scores, namely, the proportion correct (PC), false alarm ratio (FAR), undetected error rate (UER), bias score (BS), and threat score (TS) (see Appendix). These statistics are based on

Table 1. Statistics (cf. Appendix) of each model at the Futaba monitoring station and the Naraha monitoring station.

		FO	XO	FX	XX	Proportion Correct	False Alarm Ratio	Undetected Error Rate	Bias Score	Threat Score
Futaba	3 km-grid model	140	48	41	122	0.75	0.23	0.26	0.96	0.61
	250 m-grid model	104	84	35	128	0.66	0.25	0.45	0.74	0.47
	Max pooling 250 m-grid model	139	49	67	96	0.67	0.33	0.26	1.10	0.55
Naraha	3 km-grid model	49	11	49	194	0.80	0.50	0.18	1.63	0.45
	250 m-grid model	41	19	23	220	0.86	0.36	0.32	1.07	0.49
	Max pooling 250 m-grid model	44	16	27	216	0.86	0.38	0.27	1.18	0.51

the occurrence of a binary event, in this study, the existence/nonexistence of a Cs-137 plume. Thus, a threshold was required to determine the plume arrival period. Plume arrivals can be detected by not only the concentration but also the radiation dose rate at each station. However, since the equipment for measuring the radiation dose rate is often seriously contaminated in the environment, especially after rainfall, concentration observations are favorable for detecting plumes with high temporal resolution (cf. Tsuruta et al. 2018). Unlike the observational plume arrivals, the modeled plume arrivals were easily detected without the influence of the threshold. This is because the plume edge concentration is increased by more than five orders of magnitude in the model simulations (cf. Iwasaki et al. 2019).

3. Results and discussion

3.1 Model simulations

Figure 2a presents the modeled concentrations at Futaba, which is located in the grid immediately next to the grid of the FDNPP in the 3-km grid model. The 3-km grid model (blue lines) and 250-m grid model (red lines) are often synchronized with each other. However, the concentration spikes in the 3-km grid model tend to be broader than those in the 250-m grid model. In other words, the plumes tend to stay for a longer time in the 3-km grid model than in the 250-m grid model upon their arrival. This tendency is expected as low-resolution Eulerian models experience large numerical diffusion near emission sources. Moreover, at Naraha (five times farther from the FDNPP than Futaba), the synchronization between the 3-km and 250-m grid models is less prominent, as presented in Fig. 2b. The plumes in the 3-km grid model evidently arrive more often at Futaba than at Naraha. This is mainly due to the large numerical diffusion of the 3-km grid model. Note that both models tend to

overestimate the concentration at Futaba (nearer to the FDNPP), which is probably due to the model's dynamic errors rather than the emission scenario error, as the models are less likely to overestimate the concentration at Naraha.

Within the time windows presented in Figs. 2a and 2b, the statistical scores were calculated for Futaba and Naraha (Table 1). At Futaba, the PC is higher for the 3-km grid model as the UER (the number of strikeouts while looking, if using a baseball analogy) is much lower. In this case, since the FARs (the number of strikeouts while swinging, if using a baseball analogy) are similar between the 3-km and 250-m grid models, a lower UER results in a higher PC. The lower UER is due to the larger numerical diffusion of the 3-km grid model. Since the BS is smaller than 1 for the 250-m grid model, the horizontal diffusion in the models might be weaker than that in reality. This is because the model probably underestimates the frequency of high-concentration events when the BS is much smaller than 1.

In contrast, at Naraha, the 250-m grid model has a slightly higher PC than the 3-km grid model. In this case, the 3-km grid model has a much lower UER but a much higher FAR due to large numerical diffusion, as half of the alarms are false ("the boy who cried wolf" events). Consequently, the BS value is inflated, and the TS value deteriorates in the 3-km grid model. The combination of a high BS and a low TS leads to people adopting a normalcy bias, which is not appropriate for an EER model. Conversely, super high-resolution models generally tend to have a high UER due to the narrowness of plumes, which is also not appropriate for an EER model. Therefore, plume augmentation, or image processing for model results, might raise the statistical scores when using super high-resolution models. Such processing is described and evaluated in the next section.

3.2 Max pooling model

To augment the plume widths, a max pooling process was applied to the 250-m grid model result. At each grid point, an alternative value was sampled by taking the maximum value inside of a circle with a radius of 1.5 km from the grid point. This process is equivalent to the max-pooling layer of an artificial neural network with a 1-grid stride and a 12-grid pooling size. In the output distributions, the plumes were broadened to be comparable to the 3-km grid plumes. The time series of the concentrations and statistical scores for the max-pooling 250-m grid model are presented in Fig. 2 indicated by green lines and listed in Table 1, respectively.

As presented in Fig. 2a, compared with the original 250-m grid model, the plumes arrive more frequently at Futaba (nearer to the FDNPP) in the max-pooling model, and the concentration spikes are evidently broader and higher for the max-pooling model. Consequently, the BS value is improved, and the UER value is successfully decreased to the level of the 3-km grid model (see the rows for Futaba in Table 1). In the case of the original 250-m grid model, the BS value was much smaller than 1 as FO was small while (FO + XO) was large. Therefore, the improvement in the BS indicates the increase in FO, which results in the improvement in the TS. Although the improvement in the PC is small because of the deterioration in the FAR, the TS value is significantly improved from 0.47 to 0.55. In contrast, the difference between the original and max-pooling 250-m grid models is very small at Naraha, as presented in Fig. 2b. This is probably because Naraha is farther from the FDNPP than Futaba, and thus, the pooling size (1.5-km radius) is relatively small in comparison with the plume width or distance between the plumes and Naraha. Consequently, there is slight improvement in the PC and FAR (see the rows for Naraha in Table 1). However, since the UER value is slightly improved, the TS value increases from 0.49 to 0.51. Note that the TS is higher than 0.5 for the max-pooling 250-m grid model at both stations, which is relatively good in comparison with the operational scores of heavy rain ($> 10 \text{ mm h}^{-1}$) weather forecasts (cf. Appendix).

The scores were improved by the max-pooling process, especially at Futaba (nearer to the FDNPP). These improvements are presented in Fig. 3 for two time slots, [A] and [B], indicated in Fig. 2a. At time [A], the 3-km grid model successfully simulates the observed high concentration, but the original 250-m grid model fails with a sharp drop in its concentration (see Fig. 2a). In this case, as presented in Fig. 3a,

while the 3-km grid plume spreads over the monitoring station with large numerical diffusion, the original 250-m grid plume slightly misses covering the monitoring station. Since the difference between the 3-km and 250-m grid surface wind fields is small around the FDNPP, the plume coverage mainly depends on the magnitude of numerical diffusion. In contrast, the edge of the max-pooling 250-m grid plume successfully covers the monitoring station.

At time [B], while the 3-km grid model completely fails, the original 250-m grid model almost successfully simulates the observed high concentration (see Fig. 2a). However, since the observed concentration rapidly fluctuates, the original 250-m grid model regrettably fails with a very small time lag. In this case, as presented in Fig. 3b, the 3-km grid plume misses covering the monitoring station as the surface wind around the FDNPP continuously flows in one direction opposite to the monitoring station. However, in the 250-m grid model, the surface wind around the FDNPP is not homogeneous and instead follows the complicated terrain near the FDNPP. Consequently, while the prevailing wind is in the opposite direction to the monitoring station, the fluctuating edge of the plume sometimes covers the monitoring station. Even if the timing of plume intrusion at Futaba is slightly lagged in the original 250-m grid model, the augmented plume constantly covers Futaba in the max-pooling 250-m grid model, which provides a lower UER (but a higher FAR) and a better TS.

This study examined the concentration data from two monitoring stations located along the Pacific coast (not in a mountainous region) in the vicinity of the FDNPP (only 3.2 km and 17.5 km distant). The topographical difference was one of the reasons why the higher-resolution model did not perform overwhelmingly better than the lower-resolution model in this study. Previous studies (Sekiyama et al. 2015; Sekiyama and Kajino 2020) mainly focused on inland complex terrain. Although the 3-km grid model could not reproduce the details of the wind direction and plume dispersion near the FDNPP (e.g., at time [B]), the low-resolution model was superior to the original 250-m grid model at Futaba due to the large numerical diffusion therein. This empirical knowledge is not trivial for the construction of EER systems. In addition, a max-pooling (or plume augmentation) process is probably beneficial to high-resolution EER systems. However, the optimal pooling size (or plume augmentation width) depends on the situation. In this study, 4-grid (500-m radius) and 40-grid (5-km radius) pooling sizes were also tested (not shown), but the 12-

Fig. 3. Hourly averaged surface Cs-137 concentrations (green shading) and lowermost layer winds (red arrows) near Futaba and the FDNPP calculated using the models at (a) 0800–0900 local time on March 14, 2011, and (b) 2100–2200 local time on March 16, 2011. The 250-m grid winds are plotted every 6 grids (1.5 km). Gray shading indicates the elevation used in the 250-m grid models. The open triangle indicates the FDNPP. The open circle indicates the Futaba monitoring station.

grid (1.5-km radius) pooling size employed above was superior to the other options.

4. Conclusion

Generally, large numerical diffusion near an emission source is a disadvantage of Eulerian dispersion models, as increasing numerical diffusion increases the FAR (the number of strikeouts while swinging). Fortunately, this drawback, however, will decrease the UER (the number of strikeouts while looking). This characteristic is suitable for EER systems. We demonstrated that the performance of the 3-km grid model at Futaba (very close to the FDNPP) is better than that of the original 250-m grid model. On the one hand, it is scientifically important to determine the causes of plume errors on a case-by-case basis. On the other hand, it is technically troublesome that such a tiny plume location error can cause the pollutant concentration to not exceed a warning threshold. Therefore, it

would be ideal for EER systems to use high-resolution models with augmented plumes, i.e., the max-pooling process. Note that high-resolution models are able to simulate the wind and dispersion fields affected by complex terrain in detail, as presented in Fig. 3b. In addition, plume augmentation effectively masks short time lags and small plume drifts in sub-kilometer-scale high-resolution models. Unfortunately, the computational burden is, however, theoretically $12 \times 12 \times 12 = 1728$ times different between the 3-km and 250-m grid models when compared in the same area according to the Courant–Friedrichs–Lewy (CFL) condition. Therefore, enhancing the model resolution to the sub-kilometer-scale is not cost-effective, considering the available computational resources as of 2021. The operationalization of sub-kilometer-scale dispersion models will be a future issue in the construction of EER systems.

Data availability statement

The 3-km and 250-m grid model data for both the meteorological and dispersion simulations are available upon request to the corresponding author. The authors obtained the Cs-137 observation data from Tsuruta et al. (2018) in text-file format upon request. The numerical model source codes used in this study are available under a collaborative framework between MRI and related institutes/universities.

Acknowledgments

This work was supported by the Environment Research and Technology Development Fund (1-1802) at the Environmental Restoration and Conservation Agency and by JSPS KAKENHI Grant Numbers JP17K00533, JP19H01155, and 21H03593.

Appendix

This appendix highlights the statistical indices referenced in this study. First, we define the following:

FO is the number of positive simulations and positive observations (correct hits),

XO is the number of negative simulations but positive observations (misses),

FX is the number of positive simulations but negative observations (false alarms), and

XX is the number of negative simulations and negative observations (correct rejections),

where positive simulations/observations indicate high Cs-137 concentrations (equal to or more than 1.5 Bq m⁻³) in this study.

The proportion correct (PC) is the ratio of the number of correct events (hits and rejections) to the number of total events, defined by

$$PC \equiv \frac{FO + XX}{FO + XO + FX + XX}, \quad (0 \leq PC \leq 1).$$

A higher PC indicates higher model performance, especially when the number of positive observations is comparable to the number of negative observations.

The false alarm ratio (FAR) is the ratio of the number of false alarm events to the number of positive simulations, defined by

$$FAR \equiv \frac{FX}{FO + FX}, \quad (0 \leq FAR \leq 1).$$

Cases with a high FAR are likened to “the boy who cried wolf” (or the number of strikeouts while swinging). Therefore, a lower FAR is better to prevent people from adopting a normalcy bias.

The undetected error rate (UER) is the ratio of the

number of missed events to the number of positive observations, defined by

$$UER \equiv \frac{XO}{FO + XO}, \quad (0 \leq UER \leq 1).$$

High-UER forecast models are apt to miss disasters (strikeout while looking), which is not desirable for an emergency management system.

The bias score (BS) is the ratio of the number of positive simulations to the number of positive observations, defined by

$$BS \equiv \frac{FO + FX}{FO + XO}, \quad (0 \leq BS).$$

If the frequency of positive simulations is equal to that of positive observations, the BS is unity, which is the best score. The models with a high FAR (the models that “cry wolf”) are more apt to yield a much higher BS than 1.

The threat score (TS) is the ratio of the number of correct hit events to the number of events other than correct rejections, defined by

$$TS \equiv \frac{FO}{FO + FX + XO}, \quad (0 \leq TS \leq 1).$$

The TS is often referred to as the critical success index (CSI). When the number of negative observations is increased, such as no rain or no tornado, XX (the number of correct rejections) tends to be large. In that case, the PC approaches unity and is less affected by the informative values, namely, FO, FX, and XO. In contrast, the TS excludes correct rejection events and thus is applicable to validation with a large number of negative observations. While the best value of the TS is 1, the TS rarely approaches the best value. For example, operational TS values vary between 0.1 and 0.5 for heavy rain (> 10 mm h⁻¹) forecasts in Japan with 0-h or 3-h lead times (Japan Meteorological Agency 2019).

References

- Bao, J., F. K. Chow, and K. A. Lundquist, 2018: Large-eddy simulation over complex terrain using an improved immersed boundary method in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **146**, 2781–2797.
- Goto, D., Y. Morino, T. Ohara, T. T. Sekiyama, J. Uchida, and T. Nakajima, 2020: Application of linear minimum variance estimation to the multi-model ensemble of atmospheric radioactive Cs-137 with observations. *Atmos. Chem. Phys.*, **20**, 3589–3607.
- Hamer, P. D., S.-E. Walker, G. Sousa-Santos, M. Vogt, D.

- Vo-Thanh, S. Lopez-Aparicio, P. Schneider, M. O. P. Ramacher, and M. Karl, 2020: The urban dispersion model EPISODE v10.0 – Part 1: An Eulerian and sub-grid-scale air quality model and its application in Nordic winter conditions. *Geosci. Model Dev.*, **13**, 4323–4353.
- Honda, Y., M. Nishijima, K. Koizumi, Y. Ohta, K. Tamiya, T. Kawabata, and T. Tsuyuki, 2005: A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, **131**, 3465–3475.
- Iwasaki, T., T. T. Sekiyama, T. Nakajima, A. Watanabe, Y. Suzuki, H. Kondo, Y. Morino, H. Terada, H. Nagai, M. Takigawa, H. Yamazawa, D. Qulo, and A. Mathieu, 2019: Intercomparison of numerical atmospheric dispersion prediction models for emergency response to emissions of radionuclides with limited source information in the Fukushima Dai-ichi Nuclear Power Plant accident. *Atmos. Environ.*, **214**, 116830, doi:10.1016/j.atmosenv.2019.116830.
- Japan Meteorological Agency, 2019: *Outline of the operational numerical weather prediction at the Japan Meteorological Agency, March 2019. Appendix to WMO technical progress report on the global data-processing and forecasting system and numerical weather prediction*. Japan Meteorological Agency, 229 pp. [Available at <https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2019-nwp/index.htm>.]
- Kajino, M., Y. Inomata, K. Sato, H. Ueda, Z. Han, J. An, G. Katata, M. Deushi, T. Maki, N. Oshima, J. Kurokawa, T. Ohara, A. Takami, and S. Hatakeyama, 2012: Development of the RAQM2 aerosol chemical transport model and predictions of the Northeast Asian aerosol mass, size, chemistry, and mixing type. *Atmos. Chem. Phys.*, **12**, 11833–11856.
- Kajino, M., M. Ishizuka, Y. Igarashi, K. Kita, C. Yoshikawa, and M. Inatsu, 2016: Long-term assessment of airborne radiocesium after the Fukushima nuclear accident: Re-suspension from bare soil and forest ecosystems. *Atmos. Chem. Phys.*, **16**, 13149–13172.
- Kajino, M., T. T. Sekiyama, A. Mathieu, I. Korsakissok, R. Prillat, D. Qulo, A. Qurel, O. Saunier, K. Adachi, S. Girard, T. Maki, K. Yumimoto, D. Didier, O. Masson, and Y. Igarashi, 2018: Lessons learned from atmospheric modeling studies after the Fukushima nuclear accident: Ensemble simulations, data assimilation, elemental process modeling, and inverse modeling. *Geochem. J.*, **52**, 85–101.
- Kajino, M., M. Deushi, T. T. Sekiyama, N. Oshima, K. Yumimoto, T. Tanaka, J. Ching, A. Hashimoto, T. Yamamoto, M. Ikegami, A. Kamada, M. Miyashita, Y. Inomata, S. Shima, A. Takami, A. Shimizu, S. Hatakeyama, Y. Sadanaga, H. Irie, K. Adachi, Y. Zaizen, Y. Igarashi, H. Ueda, T. Maki, and M. Mikami, 2019a: NHM-Chem, the Japan Meteorological Agency’s regional meteorology-chemistry model: Model evaluations toward the consistent predictions of the chemical, physical, and optical properties of aerosols. *J. Meteor. Soc. Japan*, **97**, 337–374.
- Kajino, M., T. T. Sekiyama, Y. Igarashi, G. Katata, M. Sawada, K. Adachi, Y. Zaizen, H. Tsuruta, and T. Nakajima, 2019b: Deposition and dispersion of radiocesium released due to the Fukushima nuclear accident: Sensitivity to meteorological models and physical modules. *J. Geophys. Res.: Atmos.*, **124**, 1823–1845.
- Katata, G., M. Chino, T. Kobayashi, H. Terada, M. Ota, H. Nagai, M. Kajino, R. Draxler, M. C. Hort, A. Malo, T. Torii, and Y. Sanada, 2015: Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and oceanic dispersion model. *Atmos. Chem. Phys.*, **15**, 1029–1070.
- Kunii, M., 2014: Mesoscale data assimilation for a local severe rainfall event with the NHM–LETKF system. *Wea. Forecasting*, **29**, 1093–1105.
- Lateb, M., R. N. Meroney, M. Yataghene, H. Fellouah, F. Saleh, and M. C. Boufadel, 2016: On the use of numerical modelling for near-field pollutant dispersion in urban environments – A review. *Environ. Pollut.*, **208**, 271–283.
- Li, H., G. Cui, and Z. Zhang, 2018: A new scheme for the simulation of microscale flow and dispersion in urban areas by coupling large-eddy simulation with mesoscale models. *Bound.-Layer Meteor.*, **167**, 145–170.
- Mathieu, A., M. Kajino, I. Korsakissok, R. Prillat, D. Qulo, A. Qurel, O. Saunier, T. T. Sekiyama, Y. Igarashi, and D. Didier, 2018: Fukushima Daiichi-derived radionuclides in the atmosphere, transport and deposition in Japan: A review. *Appl. Geochem.*, **91**, 122–139.
- Miyoshi, T., and K. Aranami, 2006: Applying a four-dimensional local ensemble transform Kalman filter (4D-LETKF) to the JMA nonhydrostatic model (NHM). *SOLA*, **2**, 128–131.
- Nakajima, T., S. Misawa, Y. Morino, H. Tsuruta, D. Goto, J. Uchida, T. Takemura, T. Ohara, Y. Oura, M. Ebihara, and M. Satoh, 2017: Model depiction of the atmospheric flows of radioactive cesium emitted from the Fukushima Daiichi Nuclear Power Station accident. *Prog. Earth Planet. Sci.*, **4**, 2, doi:10.1186/s40645-017-0117-x.
- Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31.
- Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407.
- Nakayama, H., T. Takemi, and H. Nagai, 2016: Devel-

- opment of local-scale high-resolution atmospheric dispersion model using large-eddy simulation. Part 5: Detailed simulation of turbulent flows and plume dispersion in an actual urban area under real meteorological conditions. *J. Nucl. Sci. Tech.*, **53**, 887–908.
- Oura, Y., M. Ebihara, H. Tsuruta, T. Nakajima, T. Ohara, M. Ishimoto, H. Sawahata, Y. Katsumura, and W. Nitta, 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. *J. Nucl. Radiochem. Sci.*, **15**, 15–26.
- Rood, R. B., 1987: Numerical advection algorithms and their role in atmospheric transport and chemistry models. *Rev. Geophys.*, **25**, 71–100.
- Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito, and Y. Yamazaki, 2006: The operational JMA nonhydrostatic mesoscale model. *Mon. Wea. Rev.*, **134**, 1266–1298.
- Saito, K., J. Ishida, K. Aranami, T. Hara, T. Segawa, M. Narita, and Y. Honda, 2007: Nonhydrostatic atmospheric models and operational development at JMA. *J. Meteor. Soc. Japan*, **85B**, 271–304.
- Sato, Y., M. Takigawa, T. T. Sekiyama, M. Kajino, H. Terada, H. Nagai, H. Kondo, J. Uchida, D. Goto, D. Quélo, A. Mathieu, A. Quérel, S. Fang, Y. Morino, P. von Schoenberg, H. Grahm, N. Brännström, S. Hirao, H. Tsuruta, H. Yamazawa, and T. Nakajima, 2018: 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.
- Sato, Y., T. T. Sekiyama, S. Fang, M. Kajino, A. Quérel, D. Quélo, H. Kondo, H. Terada, M. Kadowaki, M. Takigawa, Y. Morino, J. Uchida, D. Goto, and H. Yamazawa, 2020: A model intercomparison of atmospheric ^{137}Cs concentrations from the Fukushima Daiichi Nuclear Power Plant accident, phase III: Simulation with an identical source term and meteorological field at 1-km resolution. *Atmos. Environ.: X*, **7**, 100086, doi:10.1016/j.aeoa.2020.100086.
- Sekiyama, T. T., and T. Iwasaki, 2018: Mass flux analysis of ^{137}Cs plumes emitted from the Fukushima Daiichi nuclear power plant. *Tellus B*, **70**, 1–11.
- Sekiyama, T. T., and M. Kajino, 2020: Reproducibility of surface wind and tracer transport simulations over complex terrain using 5-, 3-, and 1-km-grid models. *J. Appl. Meteor. Climatol.*, **59**, 937–952.
- Sekiyama, T. T., M. Kunii, M. Kajino, and T. Shimbori, 2015: Horizontal resolution dependence of atmospheric simulations of the Fukushima nuclear accident using 15-km, 3-km, and 500-m grid models. *J. Meteor. Soc. Japan*, **93**, 49–64.
- Sekiyama, T. T., M. Kunii, and M. Kajino, 2017: The impact of surface wind data assimilation on the predictability of near-surface plume advection in the case of the Fukushima nuclear accident. *J. Meteor. Soc. Japan*, **95**, 447–454.
- Takagi, M., T. Ohara, D. Goto, Y. Morino, J. Uchida, T. T. Sekiyama, S. F. Nakayama, M. Ebihara, Y. Oura, T. Nakajima, H. Tsuruta, and Y. Moriguchi, 2020: Re-assessment of early ^{131}I inhalation doses by the Fukushima nuclear accident based on atmospheric ^{137}Cs and $^{131}\text{I}/^{137}\text{Cs}$ observation data and multi-ensemble of atmospheric transport and deposition models. *J. Environ. Radioact.*, **218**, 106233, doi:10.1016/j.jenvrad.2020.106233.
- Tsuruta, H., Y. Oura, M. Ebihara, T. Ohara, and T. Nakajima, 2014: First retrieval of hourly atmospheric radionuclides just after the Fukushima accident by analyzing filter-tapes of operational air pollution monitoring stations. *Sci. Rep.*, **4**, 6717, doi:10.1038/srep06717.
- Tsuruta, H., Y. Oura, M. Ebihara, Y. Moriguchi, T. Ohara, and T. Nakajima, 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. *Geochem. J.*, **52**, 103–121.
- Wiersema, D. J., K. A. Lundquist, and F. K. Chow, 2020: Mesoscale to microscale simulations over complex terrain with the immersed boundary method in the Weather Research and Forecasting Model. *Mon. Wea. Rev.*, **148**, 577–595.
- World Meteorological Organization, 2006: Environmental emergency response: WMO activities. *WMO Bull.*, **55**, 37–41.