Impact of horizontal resolution on the robustness of radiation emulators in a numerical weather prediction model

Hwan-Jin Song[1]

[1]National Institute of Meteorological Sciences, Korea Meteorological Administration

November 23, 2022

Abstract

Developing a radiation emulator based on machine learning in a weather forecasting model is valuable because it can significantly improve the computational speed of forecasting severe weather events. In order to fully replace the radiation parameterization in the weather forecasting model, the universal applicability of radiation emulator is essential, indicating a transition from the research to the operational level. This study addressed the universal issue of radiation emulators associated with horizontal resolutions from the climate simulation scale (100 km) to the cloud-resolving scale (0.25 km). All simulations were performed using an emulator trained at 5 km simulation. In real-case simulations (100–5 km), the forecast errors of radiative fluxes and precipitation were reduced at coarse resolutions. The ideal-case simulations (5–0.25 km) also showed a similar feature with increased errors in heating rates and fluxes at fine resolutions. However, all simulations maintained an appropriate accuracy range compared with observations in real-case simulations or the infrequent use of radiation parameterization in ideal-case simulations. These findings demonstrate the feasibility of a universal radiation emulator associated with different resolutions and models and emphasize the importance of future development directions toward the emulation of high-resolution modeling.

Impact of horizontal resolution on the robustness of radiation emulators in a numerical weather prediction model

Hwan-Jin Song

National Institute of Meteorological Sciences, Korea Meteorological Administration, Jeju-do, Republic of Korea

Submitted to Journal of Geophysical Research – Atmospheres (28 March 2022)

Key Points

- The universal applicability of the radiation emulator was examined at different resolutions using two modeling frameworks.
- In both frameworks, the forecast errors at coarse resolutions were smaller than those at fine resolutions.
- Because the stability of radiation emulator was universally maintained, and speed improvement by the emulator can be highlighted.

---

* *Corresponding author's address*
Hwan-Jin Song
National Institute of Meteorological Sciences,
63568, Seogwipo-si, Jeju-do, Republic of Korea
E-mail: hwanjinsong@gmail.com

Abstract

Developing a radiation emulator based on machine learning in a weather forecasting model is valuable because it can significantly improve the computational speed of forecasting severe weather events. In order to fully replace the radiation parameterization in the weather forecasting model, the universal applicability of radiation emulator is essential, indicating a transition from the research to the operational level. This study addressed the universal issue of radiation emulators associated with horizontal resolutions from the climate simulation scale (100 km) to the cloud-resolving scale (0.25 km). All simulations were performed using an emulator trained at 5 km simulation. In real-case simulations (100–5 km), the forecast errors of radiative fluxes and precipitation were reduced at coarse resolutions. The ideal-case simulations (5–0.25 km) also showed a similar feature with increased errors in heating rates and fluxes at fine resolutions. However, all simulations maintained an appropriate accuracy range compared with observations in real-case simulations or the infrequent use of radiation parameterization in ideal-case simulations. These findings demonstrate the feasibility of a universal radiation emulator associated with different resolutions and models and emphasize the importance of future development directions toward the emulation of high-resolution modeling.

Keywords: WRF, RRTMG, radiation, resolution, neural network, emulator

1. Introduction

The atmospheric radiation process can be accurately expressed using line-by-line models (Tjemkes et al., 2003; Clough et al., 2005; Kratz et al., 2005). However, it bears heavy computation time, which limits its applicability. To overcome this challenge, machine-learning emulators imitating radiative transfer processes have been actively developed. Initially, it was confined to imitating the radiative transfer model (RTM) under specific conditions (Chevallier et al., 1998; Liu et al., 2020; Ukkonen et al., 2020; Lagerquist et al., 2021; Veerman et al., 2021; Meyer et al., 2022; Ukkonen, 2022). For RTM emulation studies, advanced machine-learning techniques besides the common neural network (NN), such as the random forest (Belochitski et al., 2011), the convolutional neural network (CNN; Liu et al., 2020), the recurrent neural network (RNN; Ukkonen, 2022), and the U-net++ model (Lagerquist et al., 2021), have been actively attempted. Emulation study applications have completely replaced radiation parameterization in atmospheric forecasting models (Krasnopolsky et al., 2005, 2008a, 2008b, 2010; 2012; Belochitski et al., 2011; Pal et al., 2019; Roh and Song, 2020; Belochitski and Krasnopolsky, 2021; Song and Roh, 2021; Song et al., 2021, 2022; Song and Kim, 2022). These emulator studies have shown sufficient speed improvements of 10–100 times, compared with theoretical radiation schemes based on discrete bands (Morcrette, 1991; Iacono et al., 2008; Baek, 2017; Hogan and Bozzo, 2018; Pincus et al., 2019). All these emulators have been developed using the common NN with single or multiple hidden layers because most radiation schemes and atmospheric forecasting models are based on the Fortran software, and techniques besides the NN are difficult to implement in the Fortran code. This case is different from the independent RTM case, which is not linked with numerical prediction models such as the global climate model (GCM) and numerical weather prediction (NWP) model.

The speed improvement by the RTM emulator was confined to the radiation process only. In contrast, the radiation emulator in the NWP can further provide speedups for the entire numerical prediction system, benefiting many applications (e.g., typhoon, flood, and heavy snowfall) in which urgent weather forecasting is essential. In addition, because the radiation emulator is linked to many dynamic and physical variables in the NWP or GCM, the emulator can produce tremendous variables and outputs from radiative transfer processes at a faster speed. Therefore, the radiation emulator in numerical prediction models is incomparably valuable for broad applications compared to the RTM. Krasnopolsky et al. (2005) first developed an NN emulator 50–80 times faster than the longwave (LW) parameterization in the Community Atmospheric Model (CAM) with T42 resolution (~300 km). The emulation for shortwave (SW) parameterization, which was 20-fold faster, was further included under the same CAM (Krasnopolsky et al., 2008a). Belochitski et al. (2011) for different machine-learning methods and Krasnopolsky et al. (2008b) for compound parameterization were conducted under the same CAM framework. NN emulators for both LW and SW can improve the computational speed of radiative calculations by approximately 30 times and the total computation time by a maximum of 25% under the Climate Forecast System (CFS) model with T126 resolution (~100 km) (Krasnopolsky et al., 2010). Pal et al. (2019) developed an NN emulator with multiple hidden layers that can accelerate the computational speed of the radiation process by a maximum of 10 times under a super-parameterized energy exascale earth system model (SP-E3SM) with a 1° horizontal resolution (~100 km). Krasnopolsky et al. (2012) further reported a speedup of approximately 37.5 times for the radiation process and a total reduction by a maximum of 18% under the Global Forecast System (GFS) model with T574 resolution (~25 km). Recently, Belochitski and Krasnopolsky (2021) demonstrated the universal performance of a radiation emulator developed in the CFS model (Krasnopolsky et al., 2010) by applying it to the GFS model.

Although the current GFS model has a different horizontal resolution (~13 km) to 100 km of the CFS, they kept the resolution as 100 km in the GFS experiment. In contrast to global models, the radiation emulators for the Korean local model with a 5 km resolution called the Korea Local Analysis and Prediction System (KLAPS; Shin et al., 2022) were developed, showing a significant speedup of the radiation process by 60-fold and 87% reduction in total computation time (Song and Roh, 2021; Song et al. 2021, 2022; Song and Kim, 2022). Uniquely, Roh and Song (2020) developed a radiation emulator with speed improvements for radiation process by 20–100 times and 82–86% reduction in total model time under a cloud-resolving model (CRM) with a 0.25 km resolution. However, their result was limited to an ideal case of a 6 h forecast in the daytime. Because previous studies on radiation emulators were conducted in different modeling frameworks based on different horizontal resolutions (from GCM to CRM), fully interpreting the meaning of the absolute errors found in the literature is difficult. Therefore, the impact of the horizontal resolutions on the forecast accuracy of the radiation emulator needs comprehensive investigation. To evaluate the accuracy of emulator, various aspects need to be validated, such as a common comparison between control simulations (Krasnopolsky et al., 2005, 2008a, 2008b, 2010, 2012; Belochitski et al., 2011; Pal et al., 2019; Belochitski and Krasnopolsky, 2021) and evaluations with the infrequent use of radiation parameterization and observations (Roh and Son, 2020; Song and Roh, 2021; Song et al., 2021; Song et al., 2022; Song and Park, 2022).

Precipitation, the most important predictor in the weather forecasting model, is implicitly determined by cumulus parameterization at coarse horizontal resolutions (e.g., above 100 km). However, precipitation is explicitly calculated using cloud microphysics parameterization at convection-permitting scales, typically at resolutions of several kilometers. Owing to the spatial smoothing effect, precipitation forecasting at coarse resolution is generally more accurate compared with that at fine resolution (Robert and Lean, 2008; Clark et al., 2016), while high-resolution forecasting remains important. In contrast, the forecast accuracy at coarse resolution for surface temperature is generally lower than that at fine resolution (Pavlik et al., 2012; Kumar et al., 2016) because the smoothing effect at coarse resolution hinders the realistic prediction of temperature variability. The contrast associated with horizontal resolutions can lead to the conjecture that the radiation emulators developed in climate models with 100–300 km resolutions (Krasnopolsky et al., 2005, 2008a, 2008b, 2010, 2012; Belochitski et al., 2011; Pal et al., 2019; Belochitski and Krasnopolsky, 2021) show different behavior compared with those of the convection-permitting NWP model with 5 km resolution (Song and Roh, 2021; Song et al., 2021, 2022; Song and Kim, 2022) or CRM with 0.25-km resolution (Roh and Song, 2020). Although it is not linked with numerical forecasting models, emulation studies for the RTM used datasets based on horizontal resolutions of 80 km (Liu et al., 2020; Ukkonen et al., 2020; Ukkonen, 2022), 30 km (Meyer et al., 2022), 13 km (Lagerquist et al., 2021), and 1 km (Veerman et al., 2021). However, these studies were developed using different numerical models and machine-learning methods, and it is difficult to conclude which studies show more improved results. Furthermore, the previous radiation emulators were evaluated under the same conditions as the trained horizontal resolution. Because the horizontal resolution of datasets and targeting models can be changed per institutional policy or users' interest from the trained version, the universal robustness of the developed radiation emulator at different horizontal scales should be satisfied for applying the emulator to various modeling systems with horizontal resolutions.

The universal applicability of radiation emulators, partially demonstrated by Belochitski and Krasnopolsky (2021) and Song and Kim (2022), is associated with the changes in numerical models (CFS to GFS) and microphysics parameterizations along with different models (real-case to ideal-case simulations), respectively. However, the effect of the horizontal resolution on the robustness of the radiation emulator remains unknown. Therefore, this study aimed to investigate the universal performance of a radiation emulator with a 5 km resolution developed by Song et al. (2022) when it was applied to different horizontal resolutions for climate and cloud-resolving simulations (100 km to 0.25 km). As in Song and Kim (2022), the trained results from three-dimensional real-case simulations were applied to two-dimensional ideal-case simulations. We expect these quantitative analyses to provide stability and accuracy in the operational NWP model, in association with the potential use of a machine-learning radiation scheme. In addition, it can provide a comprehensive insight into various radiation emulators developed at different resolutions, suggesting a future development direction for radiation emulators.

2. Data and Methods

This study utilized the NN radiation scheme developed by Song et al. (2022) for the KLAPS model over the Korean peninsula. The emulator imitated the RRTMG-K radiation parameterization (Baek, 2017), an updated version of the Rapid Radiative Transfer Model for GCMs (RRTMG; Iacono et al., 2008) with LW of 14 bands and 256 g points and SW of 16 bands and 224 g points, using NN training on 288 million input-output pairs throughout 2009–2019. Stochastic weight averaging (SWA; Izmailov et al., 2018) was applied during the NN training. The LW/SW emulators for a certain month consisted of four categories: land, ocean, clear, and cloudy. The input variables for the NN training were vertical profiles of pressure, temperature, water vapor, ozone, and cloud fraction, in addition to skin temperature and surface emissivity (LW), as well as insolation and surface albedo (SW). The output variables used were heating rate profiles, upward fluxes at the top and bottom of the atmosphere, and downward flux at the bottom. Hereafter, the LW/SW fluxes in this study indicate the average of three fluxes at the top and bottom. The nonlinear relationship between the inputs and outputs was approximated by the NN based on 90 neurons and single hidden layer. As a result of NN training, weight and bias coefficients were obtained, which were linked to the KLAPS model. The radiation emulator showed an approximately 60-fold speedup compared to the RRTMG-K and 84–87% reduction of total computation time (Song and Roh, 2021; Song et al., 2022).

The KLAPS model used in this study was primarily based on Advanced Research of the Weather Research and Forecasting (WRF-ARW) model (Skamarock et al., 2019). The physics suites other than radiation parameterization were the WRF double moment 7-Class microphysics, Simplified Arakawa–Schubert cumulus modified by the Korea Institute of Atmospheric Prediction Systems (KIAPS), Shin and Hong boundary layer, unified Noah land surface model, and revised MM5 Monin–Obukhov surface layer (Skamarock et al., 2019). The European Center for Medium-Range Weather Forecasts Reanalysis v5 (ERA5; Hersbach et al., 2020) reanalysis data with 0.25° horizontal and 37 pressure-level resolutions were used to simulate the WRF model in real cases. The real-case simulation was integrated by 7 d with a 20 s time step over 5 km horizontal grids (234×282) based on the Lambert conformal conic projection and 40 vertical levels. The real-case simulations were initialized from the 1st, 8th, 15th, 22nd days of each month in 2020, consisting of 48 weekly cases. For these cases, the LW and SW fluxes at 3 h intervals were used to evaluate the forecast accuracy of the radiation emulator by comparing the control simulation based on the original radiation parameterization. The forecast accuracy using the emulator was also evaluated by comparing it with surface observations in South Korea. The 2 m air temperature (T_{2m}) data were measured from 713 stations, and precipitation data with 5 km resolution were derived by merging rain gauges and ground-based radar measurements (Fig. 1). In addition, this study considered a two-dimensional squall line experiment within the WRF model for an extreme simulation (Skamarock et al., 2019). Although this is a case study, it corresponds to a highly-unstable situation compared with the one-year average of the real-case simulations. For example, the forecast errors of the LW/SW fluxes were 121% and 185% larger in the ideal-case simulation than in the real case (Song and Kim, 2022). This ideal simulation was forced by the default initial sounding in the WRF model and based on 201 horizontal and 40 vertical levels. Furthermore, it was integrated for 24 h with a 3 s time step. The radiation emulator developed in a real-case simulation (July and land) was applied to the ideal-case simulation to test the robustness associated with the representation error (Song and Kim, 2022).

To analyze the effects of resolution on the universal performance of the radiation emulator, simulations were performed at different spatial grids of 234×282 (5 km), 118×142 (10 km), 48×57 (25 km), 25×29 (50 km), 16×20 (75 km), and 13×15 (100 km), while maintaining similar spatial coverage over the Korean peninsula. The trained results under 5 km resolution were applied to 10–100 km resolutions as a completely independent validation, although the 5 km simulation was also evaluated for the independent period (2020) to the training period (2009–2019). In previous studies (Song and Roh, 2021; Song et al., 2021, 2022; Song and Kim, 2022), the surrounding four points around the lateral boundary were excluded from both training and testing because physically unrealistic data can appear around the lateral boundary. However, because the 100 km simulation was performed in a small domain, excluding the data around the lateral boundary was unfair between 5 km and 100 km in terms of spatial coverage. Thus, this study included data around the lateral boundary by modifying the width of the relaxation zone from four to one. For ideal-case simulation, 5, 3, 2, 1, 0.5, and 0.25 km resolutions were used while maintaining 201 horizontal grids.

3. Results and Discussion

Previous studies under different resolutions

There is a trade-off between accuracy and speedup in the emulator study; thus, the accuracy of emulator should be compared under the same (or similar) computation conditions. Because of this, using many neurons and hidden layers in the NN is a constraint leading to a slowdown. Therefore, it is against the ultimate goal of the radiation emulator in CGMs and

NWPs. Even if a study showed better accuracy using an advanced machine-learning method, an additional slowdown from the method should be carefully considered. In offline testing (not linked to the NWP model) for a two-dimensional ideal simulation under 0.25 km resolution, as an extreme case, the radiation emulator with a 57-fold speedup produced root mean square errors (RMSEs) of 1.54 K day^{-1} for LW heating rate and 1.13 K day^{-1} for SW heating rate (Roh and Song, 2020). For 288 million independent data in real-case simulations, the NN radiation scheme developed under 5 km resolution with a 60-fold speedup (Song and Roh, 2021) showed the RMSEs of 0.59 K day^{-1} for LW heating rate, 0.22 K day^{-1} for SW heating rate, 4.41 W m^{-2} for LW flux, and 20.72 W m^{-2} for SW flux. These RMSEs were further reduced by 0.46 K day^{-1}, 0.18 K day^{-1}, 3.59 W m^{-2}, and 19.13 W m^{-2}, respectively, using SWA during the NN training (Song et al., 2022). In contrast, many previous studies used the common NN (Krasnopolsky, 2014) based on a sequential training (Krasnopolsky et al., 2005, 2008a, 2008b, 2010; Belochitski et al., 2011; Roh and Song, 2020; Belochitski and Krasnopolsky, 2021; Song and Roh, 2021; Song et al., 2021). In CGM studies, the RMSEs for LW and SW heating rates in the offline testing were 0.34 K day^{-1} and 0.19 K day^{-1} in 300 km resolution (CAM), 0.49 K day^{-1} and 0.20 K day^{-1} in 100 km resolution (CFS), and 0.52 K day^{-1} and 0.26 K day^{-1} in 25 km resolution (GFS), respectively (Krasnopolsky et al., 2020, 2012). The radiation emulators for CFS and GFS were 30–40 times faster than that for RRTMG, similar to RRTMG-K for targeting reported by Roh and Song (2020), Song and Roh (2021), Song et al. (2021), Song et al. (2022), and Song and Kim (2022). Because the CGM studies were performed in the same group using the same NN technique and input-output structure, we suspect that the heating rate errors by the emulator reduced when the horizontal resolution became coarse. The results of Krasnopolsky's groups were obtained by using 0.4 million data (both LW and SW) based on individual NN training for each GCM model. Pal et al. (2019) used additional training datasets of 16.2 million. Furthermore, Song and Roh (2021) and Song et al. (2022) used a large number of training sets (288 million data) for a small area (i.e., Korea). Because the representation error can be reduced by using more datasets covering natural variability, despite the 5 km resolution, the RMSEs of the LW and SW heating rates in Song et al. (2022) were smaller than those in Krasnopolsky et al. (2010, 2012).

Numerical errors caused by a radiation emulator can be rapidly amplified during long-term integration into CGMs or NWPs (called online testing). For an ideal-case simulation under 0.25 km resolution, the RMSEs of LW/SW fluxes were amplified by 135% and 72% during 6 h forecasts (7200-time steps) compared with the offline testing results (Roh and Song, 2020). For real-case simulation under 5 km resolution, the RMSEs for LW/SW fluxes during 1 d forecasts (4320-time steps) increased by 84% and 136%, respectively (Song and Roh, 2021). For 7 d forecasts (30240-time steps), the RMSEs for LW/SW fluxes were further increased by 148% and 215%, respectively (Song et al., 2022). From these results, we expect that the numerical errors of the radiation emulator can be amplified more in the case of seasonal or inter-annual predictions based on the GCM. However, because the GCM forecasts are evaluated at monthly or yearly scales in contrast to the hourly scale for the NWP forecasts, error amplification by the long-term integration of the radiation emulator was mostly hidden in previous GCM studies (Krasnopolsky et al., 2005, 2008a, 2008b, 2010, 2012; Belochitski et al., 2011; Pal et al., 2019; Belochitski and Krasnopolsky, 2021). For example, although Belochitski and Krasnopolsky (2021) attempted a universal application of the radiation emulator by applying the training results based on the 100-km CFS into the 100-km GFS, they did not quantitatively evaluate the hourly scale (e.g., RMSE), except for global mean bias showing systematic stability. Under 5 km resolution, the RMSEs of LW/SW heating rates and fluxes for the radiation emulator of Song et al. (2022) can be magnified by 8.6%–41.3% if different microphysics schemes with the trained version are used (Song and

Kim, 2022). The one-year mean bias for LW and SW fluxes (−0.08 W m^{-2} and 0.57 W m^{-2}) in Song and Kim (2022) was smaller than those (−0.26 W m^{-2} and 0.59 W m^{-2}) in Belochitski and Krasnopolsky (2021), despite different resolutions (5 km vs. 100 km). From these previous studies, we can conclude that the radiation emulator developed by Song et al. (2022) is the most mature among the developed radiation emulators in terms of universal robustness.. This emulator also showed stable results when evaluated with surface temperature, precipitation observations (Song et al., 2022), and the changes in 14 microphysics schemes (Song and Kim, 2022).

Real-case simulations

The forecast accuracy of radiation emulator in the NWP model can be evaluated by comparing the control simulations based on the original radiation parameterization or observation data. To evaluate the accuracy of the LW/SW fluxes using a radiation emulator, we utilized the framework used by Song et al. (2022) and Song and Kim (2022). This framework considered 48 weekly cases (approximately one year) with a 3 h interval for 234×282 grids (5 km); thus, the statistics were obtained from 177,375,744 data points. Because different domain sizes, such as 118×142 (10 km), 48×57 (25 km), 25×29 (50 km), 16×20 (75 km), and 13×15 (100 km), were considered for resolution experiments, the data points used were also reduced in proportion to the domain size. The emulator results at different resolutions were compared with each control simulation. T_{2m} and 3-hourly accumulated precipitation simulated using the radiation emulator were compared with surface observation data for the 48 weekly cases.

Figure 2 illustrates the spatial distributions of RMSEs for LW/SW fluxes when the radiation emulator developed at 5 km resolution was applied to 10–100 km resolutions. The WRF simulation results based on the Lambert conformal conic projection were interpolated to the latitude-longitude projection with a regular grid interval to draw the plots. First, the

RMSEs of the LW flux showed a clear contrast between the land and ocean because the skin temperature and surface emissivity, as inputs for the LW radiation process, were separated between land and ocean. Furthermore, mountainous areas can represent a strong variability in skin temperature toward colder temperatures than the surrounding low-latitude areas. Because KLAPS models use a terrain-following vertical coordinate, and vertical profiles (e.g., pressure, temperature, and moisture) around the surface are affected by topographic altitudes. For these reasons, the LW flux shows the largest error above 11 W m^{-2} in the Gaema Plateau area in North Korea, having the highest topographic altitude (Fig. 2a). For the RMSEs of the SW flux, the land–ocean contrast is unclear; it tends to slightly increase toward the southern region (Fig. 2g). The more abundant cloud conditions and slightly larger insolation in the southern region explain the error pattern of SW flux. In contrast, in the Chinese desert areas located in the northwest, the lowest RMSEs for SW flux were found because of the low-cloud condition. When the horizontal resolutions increased to 10, 25, 50, 75, and 100 km, the RMSEs in both LW and SW dropped sharply. Because the radiation emulator used in this study was trained at 5 km resolution, the results at different resolutions should produce outputs with greater uncertainty in terms of representation error. However, the smoothing effects for input-output variables contributed to producing more accurate results with a universal application beyond the representation error at different resolutions. The lower representation error could be due to using large training sets for the small Korean domain. The number of training sets used by Song and Roh (2021) and Song et al. (2022) was 720 times larger than those used by Krasnopolsky et al. (2010, 2012) and Belochitski and Krasnopolsky (2021). Because the GCM domain covers the entire globe, using more training sets is essential to ensure universality and accuracy in the global region, similar to Song et al. (2022). Therefore, we can conclude that the 5 km simulation results of Song and Roh (2021), Song et al. (2021), Song et al. (2022), and Song and Kim (2022) were developed under more difficult condition to secure the universality of radiation emulators compared with 25 km resolution (Krasnopolsky et al., 2012) and resolutions larger than 100 km (Krasnopolsky et al., 2005, 2008a, 2008b, 2010; Belochitski et al., 2011; Pal et al., 2019; Belochitski and Krasnopolsky, 2021). Similarly, we can expect that the development of a radiation emulator at resolutions less than 5 km, such as the 0.25 km resolution in Roh and Song (2020), would be more difficult. Because of the limitation of computational resources, although this study did not consider real-case simulations at resolutions less than 5 km, this issue will be dealt with in the ideal-case simulation framework.

Figure 3 shows the temporal variations in RMSEs for LW/SW fluxes, T_{2m}, and 3-hourly accumulated precipitation during the 7-day forecasts averaged from the total 48 weekly cases and entire domain. The RMSEs of the LW flux increase steadily with forecast time while showing diurnal variation between night and day (Fig. 3a). Because the variability in surface temperature is larger during the daytime, the RMSEs of the LW flux also show an amplified error during the day. The RMSEs of the SW flux are characterized by a strong diurnal variation associated with the evident diurnal cycle of insolation while showing a gradual increase in error with forecast time (Fig. 3b). In both LW and SW fluxes, the RMSEs at coarse resolutions were lower than those at fine resolutions. For resolutions larger than 50 km, the RMSE patterns in the LW flux tended to be saturated (Fig. 3a); similar trends are found in the spatial distribution (Figs. 2d–f). The total RMSEs for the three LW and SW fluxes are listed in Table 1. The RMSEs of LW and SW fluxes for 5, 10, 25, 50, 75, and 100 km were 9.59, 9.16, 8.16, 7.79, 7.78, and 7.99 W m^{-2} and 63.17, 60.34, 52.78, 49.03, 46.89, and 47.40 W m^{-2}, respectively. Accordingly, the 5-km results were more uncertain by 20% (LW) and 33% (SW) compared with the 100-km results. If the results were re-trained at each resolution, not aiming at the universal application of the 5-km radiation emulator, the difference between the 5-km and 100-km results would have been more. These results show that the 5-km simulations in Song and Roh (2021), Song et al. (2021), Song et al. (2022), and Song and Kim (2022) were harsher environments than resolutions larger than 100 km (Krasnopolsky et al., 2005, 2008a, 2008b, 2010; Belochitski et al., 2011; Pal et al., 2019; Belochitski and Krasnopolsky, 2021). The evaluation results compared with the observation data revealed the contrast between T_{2m} and precipitation (Figs. 3c–d). The RMSEs of T_{2m} increased with increasing resolution (2.2619 K to 2.9405 K in Table 1), whereas an opposite trend was found for the 3-hourly precipitation (1.5515 mm to 1.1479 mm in Table 1). When the horizontal resolution is coarse, the spatial variability of the surface temperature in the model simulation can be reduced from actual observations. Thus, overly coarse resolutions can degrade the forecast accuracy of surface temperature. Pavlik et al. (2012) and Kumar et al. (2016) reported similar results, showing lower accuracy at coarse resolution for surface temperature. Because precipitation forecast is highly uncertain, the smoothing effect at a coarse resolution can be more important in determining the forecast error of precipitation. The forecast skill of precipitation in numerical models was rather improved at coarse horizontal scales (Robert and Lean, 2008; Clark et al., 2016). This illusion effect at coarse resolution does not indicate that high-resolution modeling is unnecessary.

The role of cumulus parameterization in determining precipitation and surface temperature is more important at coarse resolutions than at fine resolutions. Therefore, we need to examine the behavior of radiation emulator results with a turned-off cumulus scheme. To turn off the cumulus scheme, the RMSEs for LW/SW fluxes and T_{2m} in Table 1 were changed by a maximum 0.9% (at 100 km resolution), while the general trends were unchanged. The RMSEs of the 3-hourly precipitation forecasts at 100 km resolution were changed the most by 6.7%, resulting in 0.2472 mm. Nevertheless, these results do not affect the conclusion of this study regarding the universal application of a radiation emulator. Furthermore, readers may suspect that the observational evaluations in Table 1 are a characteristic of the radiation emulator. However, the control simulations based on the original radiation parametrization showed similar RMSEs for T_{2m} (2.2619 K to 2.9895 K) and 3-hourly precipitation (1.5641 mm to 1.1433 mm). Therefore, we can conclude that the radiation emulator developed at 5 km resolution can be universally applied for horizontal resolutions larger than 5 km while maintaining the accuracy and stability. However, the opposite situation (coarse to fine resolutions) cannot be guaranteed because the potential error from the radiation emulator can lead to unstable results in numerical models (Krasnopolsky et al., 2008b; Song et al., 2021; Belochitski and Krasnopolsky, 2021; Song and Kim, 2022). This drawback led to the next analysis based on ideal-case simulations.

Ideal-case simulations

Considering a tremendous computation resource for long-term simulations at fine resolutions less than 5 km, the universal application of a radiation emulator at 5 to 0.25 km resolutions was evaluated in a two-dimensional idealized squall line simulation as an extreme precipitating case. Similar ideal simulations were conducted by Roh and Song (2020), Song et al. (2022), and Song and Kim (2021). Song and Kim (2021) examined the universal application of a radiation emulator with changes in 14 additional microphysics schemes at 5 km resolution. As a follow-up, this study examined the effects of horizontal resolution on a universal radiation emulator. The radiation emulator used in this study was obtained from real-case simulations (July and land). This study considered a 24-hour integral period with a 3-s time step; hence, the emulator was applied four times more than Roh and Song (2020) for a 6-hour forecast and 6.66 times more than Song et al. (2022) and Song and Kim (2022) using a 20 s time step. A 3 s time step was essential for the control simulation at 0.25 km (the use of larger time step led to a blow up of the control simulation). Comparing the representation error of the real-case simulation to the ideal case, the RMSEs of the radiation emulator in Song and Kim (2022) were larger by 23%–48% than the infrequent use of radiation parameterization by 60 times in Song et al. (2022). Because an ideal simulation is an extreme case, the error caused by the emulator can be rapidly amplified. Accordingly, the RMSEs of the radiation emulator in Table 2 appear moderately large. These RMSEs were calculated for 201 horizontal grids and 1440 temporal data points at 10 m intervals. The heating rate and flux in this study represent the average of 39 vertical layers and three flux components (at top and bottom), respectively. To minimize the error associated with universal representation, Song and Kim (2022) attempted compound parameterization (CP), returning the original parameterization when the predicted heating-rate errors exceeded a predefined threshold. They used thresholds of 1.0341 and 0.4820 K day^{-1} for LW and SW heating rates, respectively, to target an approximately 3-fold slowdown to the emulator with 60-fold speedup. When the same concept was applied in this study, the emulator + CP results were 3.23–4.21 times slower than the emulator only. This result implies that the emulator + CP is still 14–19 times faster than the original radiation parameterization. By adding CP, the total RMSEs at 5 km resolution was reduced by 27.3%, 26.7%, 22.3%, and 16.8% for the LW heating rate, SW heating rate, LW flux, and SW flux, respectively. The resulting RMSEs of heating rates (2.61 K day^{-1} and 1.21 K day^{-1}) are comparable to 2.57 K day^{-1} and 1.20 K day^{-1} based on the infrequent use of radiation parameterization by 30 times (Song et al., 2022). These results indicate that using a radiation emulator along with CP can maintain stable accuracy while overcoming the representation error induced by the difference between real and ideal cases, as well as different horizontal resolutions. The successful results of this study associated with universal robustness are novel and worthy of recognition.

Figure 4 displays the temporal and spatial variations of outgoing LW radiation (OLR) and upward LW flux at the top of the atmosphere for 5, 3, 2, 1, 0.5, and 0.25 km resolutions. Each simulation had the same number of horizontal grids but different coverage areas from 1000 km to 50 km. The difference in area coverage induced different evolutionary patterns among the control simulations. A low OLR indicates vigorous deep convention, whereas a high OLR represents a clear condition. While the 5-km control simulation is characterized by widely spread clear areas (expressed by high OLR values) before hour 12 (Fig. 4a), the clear sky portion is rapidly reduced when the horizontal resolution decreases from 5 to 0.25 km. At 0.25 km, there are small clear areas before hour 1. Because cloud simulation is uncertain in the radiation parameterization, the 0.25-km simulation corresponds to a more highly nonlinear situation than the 5-km simulation. The sharp effect at a fine resolution (in contrast to smoothing effect at a coarse resolution) provides a more uncertain situation at 0.25-km resolution. For these reasons, the occurrence frequencies of lower OLRs, regarded as deep convective clouds, are higher at fine resolutions in both real and ideal cases (Fig. 5). In contrast to the long-term (one year) results for the real case showing a stable smoothing effect with resolution (Fig. 5a), the ideal simulations based on one case show great variability in probability density functions (Fig. 5b). Because of the sharp effect and high cloud conditions at fine resolutions, the occurrence frequencies of OLRs less than 180 W m^{-2} were higher at fine resolutions, whereas those around 270 W m^{-2}, regarded as a clear condition, were relatively reduced at fine resolutions. Characteristically, the 0.25-km simulation showed a rare occurrence with OLRs less than 180 W m^{-2}. In contrast, it showed frequent occurrences with OLRs in the range of 180–220 W m^{-2}, compared with other fine resolutions (2 km, 1 km, and 0.5 km). As shown in blue colors in Fig. 4f, the 0.25-km simulation produced medium OLRs of 180–220 W m^{-2} after 12 h. Accordingly, this is the result of nonlinear characteristics in the cloud-resolving simulation.

With these control simulation characteristics, the radiation emulators (along with CP) successfully reproduce similar features to the control simulations until 24 h (Figs. 4g–i). The ideal simulation is sensitive to small initial perturbations and rapidly changed during integration; therefore, obtaining accurate plots for Figs. 4g–i is difficult. The clear area in the initial stage at coarse resolutions and the subsequent widely spread clouds are realistically expressed in the emulator results, despite the results being trained from real-case simulations. The difference between control and emulator results tended to be larger at fine resolutions. For example, OLRs higher than 200 W m^{-2} after 20 h in the control simulation at 0.25 km are not found in the 0.25-km emulator result. Figure 6 illustrates the vertical RMSEs of the heating rates and temporal RMSEs of the fluxes. The RMSEs were calculated for the total temporal-horizontal grids and horizontal grids, respectively. The 1-km and 0.5-km simulations show a larger error in the heating rates at middle levels (Figs. 6a–b). The 0.5-km simulation is also characterized by a large error in the LW heating rate around the surface (Fig. 6a). For the LW flux, no characteristic features were observed at a specific resolution (Fig. 6c). The RMSE of the SW flux is the largest in the 0.25-km simulation, especially around 15–16 h. In contrast, the RMSEs for 1-km and 0.5-km simulations are characteristically lower than those for coarse resolutions (Fig. 6d). The total RMSE statistics are listed in Table 2. Because the ideal simulation is a highly nonlinear process, a consistent tendency with resolutions, such as in the real case, is not found. In general, increasing errors at fine resolutions using an emulator are evident. Compared with the 5-km simulation, for the 0.25–3 km resolutions, the errors of the LW heating rate, SW heating rate, LW flux, and SW flux increased by 19%–128%, 41%–104%, 11%–123%, and 17%–57%, respectively,. These errors increase with horizontal resolutions and are larger than those induced by changes in microphysics parameterization at 5-km resolution in Song and Kim (2022). By using CP, the increased RMSE in the LW and SW heating rates are further reduced by 13%–72% and 9%–45%, respectively, compared with the 5-km simulation. For LW and SW fluxes, the RMSEs were improved by a maximum of 6% and 37%, respectively, in coarse resolutions. These features are closely related to the CP because using CP contributed more to the improvements in LW/SW fluxes (25%–35% and 27%–52%) than LW/SW heating rates (22%–67% and

17%–64%), especially for 1 km and 0.5 km resolutions. Although the RMSEs are higher at fine resolutions than at coarse resolutions, this study is not the blow-up issue of the entire model, such as the unphysical OLR in Belochitski and Krasnopolsky (2021). Therefore, we can conclude that the radiation emulator developed at 5 km resolution can be universally applied for cloud-resolving resolutions less than 5 km while maintaining accuracy at the expense of computational speed by using CP (i.e., 60-fold to 14–19-fold speedup).

4. Summary and Conclusions

In this study, we considered different horizontal resolutions under two simulation frameworks: KLAPS over the Korean peninsula (real case) and two-dimensional squall line simulation (ideal case) to examine the impact of horizontal resolution on the universal applicability of the radiation emulator in NWP models. The real-case simulation was performed for approximately one year with a 7 d forecast time, whereas the ideal-case simulation was an extreme squall line case with a 1 d forecast time. The horizontal resolutions were 5, 10, 25, 50, 75, and 100 km (convection-permitting scale to climate simulation scale) for the real-case simulation and 5, 3, 2, 1, 0.5, and 0.25 km (convection-permitting scale to cloud-resolving scale) for the ideal-case simulations. All emulator simulations were based on the NN radiation scheme developed under the real-case at 5-km simulations by Song et al. (2022). This emulator was 60-fold faster than the RRTMG-K radiation parameterization. Here, all simulations were tested in an independent period using training sets (2009–2019). The real-case simulation focused only on the impact of horizontal resolutions on the universal applicability of the 5-km radiation emulator. In contrast, the ideal-case simulation further considered the universal robustness arising from the difference between the real and ideal cases. Despite the different horizontal resolutions with the trained 5-km resolution, the forecast error of LW/SW fluxes was significantly reduced from fine to coarse resolutions (9.59 to 7.79 W m^{-2} and 63.17 to 46.89 W m^{-2}). In addition, the RMSEs of

T_{2m} and precipitation, compared with the observations, increased and decreased from fine to coarse resolutions (2.2619 K to 2.9405 K and 1.5515 mm to 1.1479 mm), respectively. Because control simulations also showed the same error characteristic for T_{2m}, these results suggested that the radiation emulator developed at a 5 km resolution is universally applied for horizontal resolutions larger than 5 km while maintaining accuracy and stability. For the ideal-case simulation, the temporal and spatial evolutions of the OLRs were examined for different horizontal resolutions (5, 3, 2, 1, 0.5, and 0.25 km). Each control simulation showed a large difference in the temporal and spatial cloud patterns owing to the smoothing effect at coarse resolutions and different cloud conditions. Using radiation emulator successfully reproduces features similar to those of the control simulations. For 0.25–3 km resolutions, the forecast errors of the LW heating rate, SW heating rate, LW flux, and SW flux increased by 19–128%, 41–104%, 11–123%, and 17–57%, respectively, compared with the 5-km simulation. To minimize these errors, the compound parameterization 3.23–4.21 times slower than the emulator was further used (i.e., 14–19 times faster than the original radiation parameterization). By adding CP, the total RMSEs at 5 km resolution was reduced by 27.3%, 26.7%, 22.3%, and 16.8% for the LW heating rate, SW heating rate, LW flux, and SW flux, respectively. The resulting RMSEs of LW heating rate, SW heating rate, LW flux, and SW flux at 5–0.25 km resolutions were 2.61 to 4.49 W m^{-2}, 1.21 to 1.75 W m^{-2}, 15.60 to 17.65 W m^{-2}, and 101.52 to 174.04 W m^{-2}, respectively. Here, the resulting RMSEs of heating rates at 5 km resolution (2.61 and 1.21 K day^{-1}) were comparable to 2.57 and 1.20 K day^{-1} based on the infrequent use of original radiation parameterization by 30 times (Song et al., 2022).

This study provides a comprehensive insight into radiation emulator studies using numerical prediction models at different resolutions. From this study, it was found that coarse-resolution modeling was easier than fine-resolution simulation to ensure the accuracy and stability of the radiation emulator. Therefore, previous emulator studies at convection-permitting and cloud-resolving scales were regarded as more valuable than low-resolution studies based on climate models. In addition, these results provide important information on the universal applicability of radiation emulators associated with using different horizontal resolutions and modeling platforms. Although the universal robustness of the radiation emulator has already been examined for changes in numerical models and microphysics parameterization (Belochitski and Krasnopolsky, 2021; Song and Kim, 2022), no experiments at a different resolution from the trained resolution have been conducted. The efforts in this study are particularly variable as this study is the first to show the universal applicability of radiation emulators at different resolutions. Therefore, the complete replacement of radiation parameterization by a machine-learning emulator with a significant speedup is nearing. This study can also accelerate the computational speed of regional climate simulations or high-resolution modeling in terms of a faster radiation scheme. The findings in this study also suggest an evident direction for developing the universal radiation emulator in the future that it should be developed at a resolution as high as possible. The emulator trained at low resolution has a great uncertainty when it is applied to high-resolution model, because the occurrence frequency of extreme far corner events is underestimated at the low-resolution modeling. In terms of universal application, one drawback of this study confined to the Korean peninsula can be improved by expanding the training sets for global regions.

Acknowledgements

This work was funded by the KMA Research and Development Program "Developing of AI technology for weather forecasting" under Grant (KMA2021-00120).

Data Availability Statement

The radiation emulators used in this study and associated WRF modeling systems are available in https://doi.org/10.5281/zenodo.5638436 and https://doi.org/10.5281/zenodo.6033618, respectively.

References

Baek, S. (2017). A revised radiation package of G-packed McICA and two-stream approximation: Performance evaluation in a global weather forecasting model. *Journal of Advances in Modeling Earth Systems,* 9, 1628–1640. https://doi.org/10.1002/2017MS000994.

Belochitski, A., Binev, P., DeVore, R., Fox-Rabinovitz, M., Krasnopolsky, V., & Lamby, P. (2011). Tree approximation of the long wave radiation parameterization in the NCAR CAM global climate model. *Journal of Computational and Applied Mathematics*, 236, 447–460. https://doi.org/10.1016/j.cam.2011.07.013.

Belochitski, A., & Krasnopolsky, V. (2021). Robustness of neural network emulations of radiative transfer parameterizations in a state-of-the-art General Circulation Model. *Geoscientific Model Development,* 14, 7425–7437. https://doi.org/10.5194/gmd-2021-114.

Chevallier, F., Chéruy. F., Scott, N. A., & Chédin, A. (1998). A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology,* 37, 1385–1397. https://doi.org/10.1175/1520-0450(1998)037.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications,* 23, 165–181. https://doi.org/10.1002/met.1538.

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., & Brown, P. D. (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer,* 91, 233–244. https://doi.org/10.1016/j.jqsrt.2004.05.058.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society,* 146, 1999– 2049. https://doi.org/10.1002/qj.3803.

Hogan, R. J., & Bozzo, A. (2018). A flexible and efficient radiation scheme for the ECMWF model. *Journal of Advances in Modeling Earth Systems,* 10(8), 1990–2008. https://doi.org/10.1029/2018ms001364.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research,* 113, D13103. https://doi.org/10.1029/2008JD009944.

Izmailov, P. Podoprikhin, D., Garipov, T., Vetrov, D., and Wilson, A. G. (2018). Averaging weights leads to wider optima and better generalization. *Conference on Uncertainty in Artificial Intelligence (UAI) 2018.* https://arxiv.org/abs/1803.05407.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (2005). New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model. *Monthly Weather Review*, 133, 1370–1383. https://doi.org/10.1175/MWR2923.1.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Belochitski, A. A. (2008a). Decadal climate simulations using accurate and fast neural network emulation of full, longwave and shortwave, radiation. *Monthly Weather Review,* 136, 3683–3695. https://doi.org/10.1175/2008mwr2385.1.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Tolman, H. L., & Belochitski, A. A. (2008b). Neural network approach for robust and fast calculation of physical processes in numerical environmental models: Compound parameterization with a quality control of larger errors. *Neural Networks,* 21, 535–543. https://doi.org/10.1016/j.neunet.2007.12.019.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Hou, Y. T., Lord, S. J., & Belochitski, A. A. (2010). Accurate and fast neural network emulations of model radiation for the NCEP coupled Climate Forecast System: Climate simulations and seasonal predictions. *Monthly Weather Review*, 138, 1822–1842. https://doi.org/10.1175/2009MWR3149.1.

Krasnopolsky, V. M., Belochitski, A. A., Hou, Y. T., Lord, S. J., & Yang, F. (2012). Accurate and fast neural network emulations of long and short wave radiation for the NCEP Global Forecast System model. NCEP/NWS, NOAA, office note 471. https://www.emc.ncep.noaa.gov/officenotes/newernotes/on471.pdf.

Krasnopolsky, V. M. (2014). NCEP neural network training and validation system: Brief description of NN background and training software. Environment Modeling Center, NCEP/NWS, NOAA. https://doi.org/10.7289/v5qr4v2z.

Kratz, D. P., Mlynczak, M. G., Mertens, C. J., Brindley, H., Gordley, L. L., Martin-Torres, J., Ferenc, M. M., & Turner, D. D. (2005). An inter-comparison of far-infrared line-by-line radiative transfer models. *Journal of Quantitative Spectroscopy and Radiative Transfer, 90(3-4),* 323–341. https://doi.org/10.1016/j.jqsrt.2004.04.006.

Kumar, P., Ojha, S. P., Singh, R., Kishtawal, C. M., & Pal, P. K. (2016). Performance of weather research and forecasting model with variable horizontal resolution. *Theoretical and Applied Climatology,* 126, 705–713. https://doi.org/10.1007/s00704-015-1607-7.

Lagerquist, R., Turner, D., Ebert-Uphoff, I., Stewart, J., & Hagerty, V. (2021). Using deep learning to emulate and accelerate a radiative transfer model. *Journal of Atmospheric and Oceanic Technology,* 38, 1673–1696. https://doi.org/10.1175/JTECH-D-21-0007.1.

Liu, Y., Caballero, R., & Monteiro, J. M. (2020). RadNet 1.0: exploring deep learning architectures for longwave radiative transfer. *Geoscientific Model Development*, 13, 4399–4412. https://doi.org/10.5194/gmd-13-4399-2020.

Meyer, D., Hogan, R. J, Dueben, P. D, & Mason, S. L. (2022). Machine learning emulation of 3D cloud radiative effects. *Journal of Advances in Modeling Earth Systems,* e2021MS002550, https://doi.org/10.1029/2021MS002550.

Morcrette, J.-J. (1991). Radiation and cloud radiative properties in the European Centre for Medium Range Weather Forecasts forecasting system. *Journal of Geophysical Research,* 96(D5), 9121. https://doi.org/10.1029/89jd01597.

Pal, A., Mahajan, S., & Norman, M. R. (2019), Using deep neural networks as cost-effective surrogate models for Super-Parameterized E3SM radiative transfer. *Geophysical Research Letters,* 46, 6069–6079. https://doi.org/10.1029/2018GL081646.

Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems,* 11, 3087–3089. https://doi.org/10.1029/2019MS001621.

Pavlik, D., Söhl, D., Pluntke, T., Mykhnovych, A., & Bernhofer, C. (2012). Dynamic downscaling of global climate projections for eastern Europe with a horizontal resolution of 7 km. *Environmental Earth Sciences,* 65, 1475–1482. https://doi.org/10.1007/s12665-011-1081-1.

Roberts, N. M., & Lean, H. W. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97. https://doi.org/10.1175/2007MWR2123.1.

Roh, S., & Song, H.-J. (2020). Evaluation of neural network emulations for radiation parameterization in cloud resolving model. *Geophysical Research Letters,* 47, e2020GL089444. https://doi.org/10.1029/2020GL089444.

Shin, H.-C., Ha, J.-H., Ahn, K. D., Lee. E. H., Kim, C. H., Lee. Y. H., & Clayton, A. (2022). An overview of KMA's operational NWP data assimilation system. In: Park, S. K. & Xu. L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV). Springer, Cham. https://doi.org/10.1007/978-3-030-77722-7_26.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., & Huang, X.-Y. (2019). A description of the Advanced Research WRF model version 4. *NCAR Technical Notes.* https://doi.org/10.5065/1DFH-6P97.

Song, H.-J., & Roh, S. (2021). Improved weather forecasting using neural network emulation for radiation parameterization. *Journal of Advances in Modeling Earth Systems,* 13, e2021MS002609. https://doi.org/10.1029/2021MS002609.

Song, H.-J., Roh, S., & Park, H. (2021). Compound parameterization to improve the accuracy of radiation emulator in a numerical weather prediction model. *Geophysical Research Letters,* 48, e2021GL095043. https://doi.org/10.1029/2021GL095043.

Song, H.-J., Roh, S., Lee, J., Nam, G., Yun, E., Yoon, J., & Kim, P. S., (2022). Benefits of stochastic weight averaging in developing neural network radiation scheme for numerical weather prediction. *Journal of Advances in Modeling Earth Systems,* https://doi.org/10.1002/essoar.10508964.1. (in minor revision)

Song, H.-J., & Kim, P. S. (2022). Effects of cloud microphysics on the universal performance of neural network radiation scheme. *Geophysical Research Letters,* https://doi.org/10.1002/essoar.10510634.1. (in revision)

Tjemkes, S. ., Patterson, T., Rizzi, R., Shephard, M., Clough, S., Matricardi, M., Haigh, J. D., Höpfner, M., Payan, S., Trotsenko, A., Scott, N., Rayer, P., Taylor, J. P., Clerbaux, C., Strowl L. L., DeSouza-Machado, S., Tobin, D., & Knuteson, R. (2003). The ISSWG line-by-line inter-comparison experiment. *Journal of Quantitative Spectroscopy and Radiative Transfer,* 77(4), 433–453. https://doi.org/10.1016/s0022-4073(02)00174-7.

Ukkonen, P., Pincus, R., Hogan, R. J., Nielsen, K. P., & Kaas, E. (2020). Accelerating radiation computations for dynamical models with targeted machine learning and code optimization. *Journal of Advances in Modeling Earth Systems,* 12, e2020MS002226. https://doi.org/10.1029/2020MS002226.

Ukkonen, P. (2022). Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *Journal of Advances in Modeling Earth Systems, e2021MS002875,* https://doi.org/10.1029/2021MS002875.

Veerman M. A., Pincus, R., Stoffer, R., van Leeuwen, C. M., Podareanu, D., & van Heerwaarden, C. C. (2021). Predicting atmospheric optical properties for radiative transfer computations using neural networks. *Philosophical Transactions of the Royal Society A,* 379, 20200095. https://doi.org/10.1098/rsta.2020.0095.

Table 1. Total root mean square error (RMSE) statistics for real-case simulations. The LW and SW fluxes [W m^{-2}], along with three upward (↑) and downward (↓) fluxes at the top and bottom, were compared with the control model simulations, whereas 2-m temperature (T_{2m}) [K] and 3-hourly precipitation [mm] were compared with surface observations.

	5 km	10 km	25 km	50 km	75 km	100 km
LW flux	9.5888	9.1575	8.1587	7.7853	7.7776	7.9902
- top ↑	11.2884	10.6651	9.2559	8.8752	9.2227	9.8026
- bottom ↑	3.8179	3.7173	3.4111	3.3648	3.2523	3.3526
- bottom ↓	13.6601	13.0901	11.8089	11.1159	10.8579	10.8154
SW flux	63.1709	60.3422	52.7777	49.0250	46.8944	47.3956
- top ↑	79.3886	75.9043	62.3987	61.6078	58.9609	59.6203
- bottom ↑	13.6354	12.9018	11.3199	10.4635	10.0865	10.0846
- bottom ↓	96.4886	92.2204	80.6146	75.0035	71.6357	72.4819
T_{2m}	2.2619	2.4536	2.6596	2.7026	2.8249	2.9405
Precipitation	1.5515	1.5170	1.3788	1.2800	1.1747	1.1479

Table 2. Total root mean square error (RMSE) statistics for idealized squalline simulations. The LW and SW heating rates [K day^{-1}], as well as the LW and SW fluxes [W m^{-2}], were compared with the control simulations. The numbers before and after arrows indicate the emulator only and the emulator with compound parameterization.

	LW heating rate	SW heating rate	LW flux	SW flux
5 km	3.59 → 2.61	1.65 → 1.21	21.33 → 16.57	193.19 → 160.76
3 km	4.42 → 2.96	2.48 → 1.64	34.02 → 16.06	260.52 → 154.19
2 km	4.29 → 2.98	2.32 → 1.33	23.69 → 17.65	225.13 → 147.67
1 km	5.19 → 3.87	3.37 → 1.75	37.39 → 16.09	291.27 → 104.46
0.5 km	5.90 → 4.49	3.30 → 1.75	47.59 → 15.60	269.50 → 101.52
0.25 km	5.29 → 3.44	2.99 → 1.43	43.36 → 16.01	245.47 → 174.04

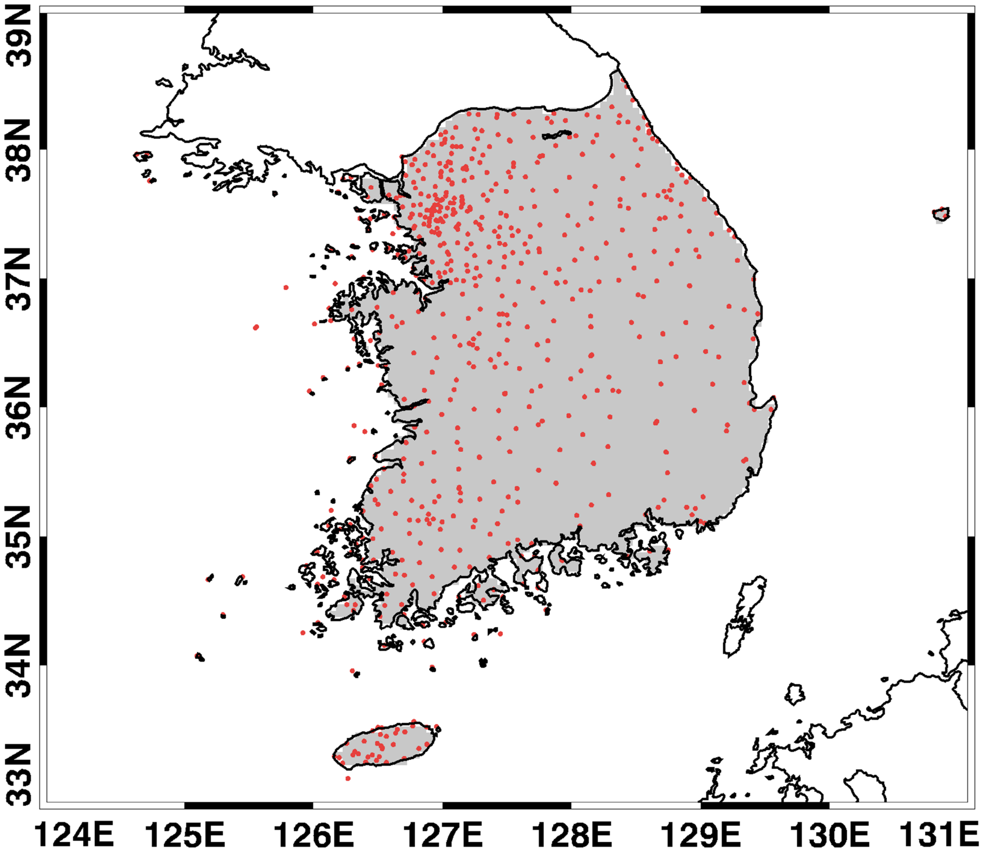

Figure 1. Locations of surface 713 stations (red circles) used for 2-m temperature along with gauge-radar merged precipitation data with 5-km resolution (grey colors).

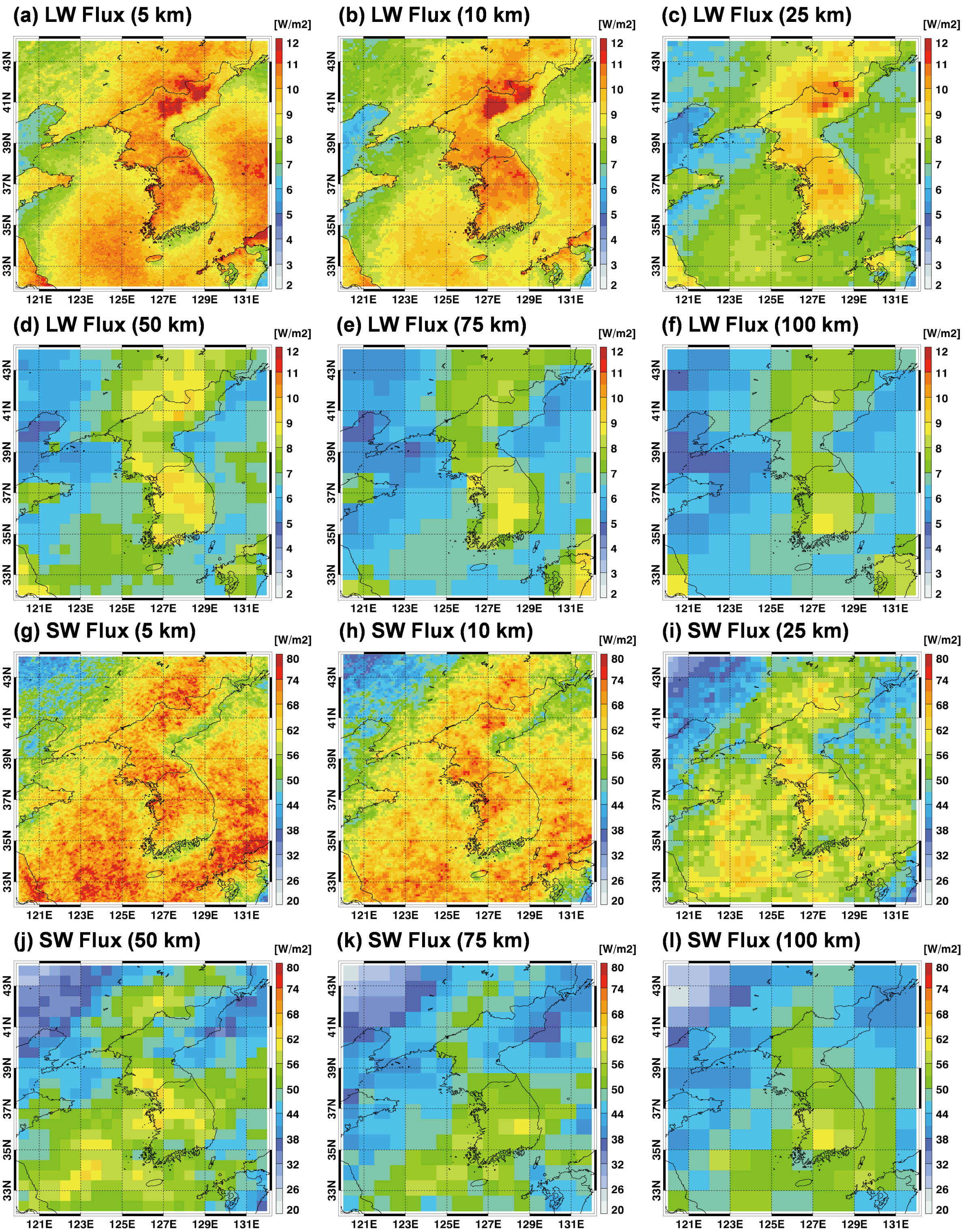

Figure 2. Spatial distributions of RMSEs for LW and SW fluxes compared to the control runs with different horizontal resolutions in real-case simulations.

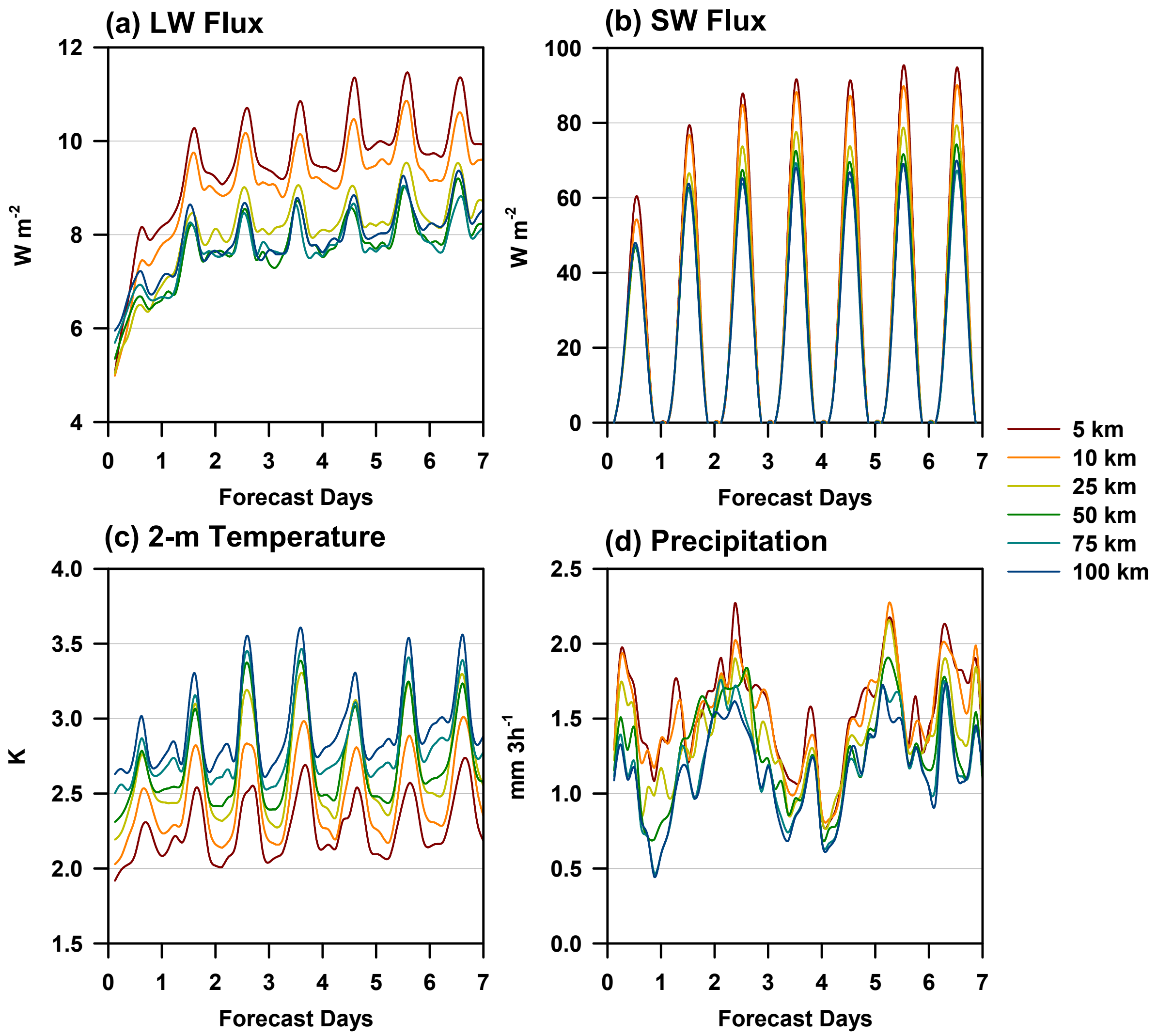

Figure 3. Time series of RMSEs for (a) LW and (b) SW fluxes compared to the control runs with different horizontal resolutions in real-case simulations. The results of (c) 2-m temperature and (d) 3 hourly precipitation compared with surface observations were also given.

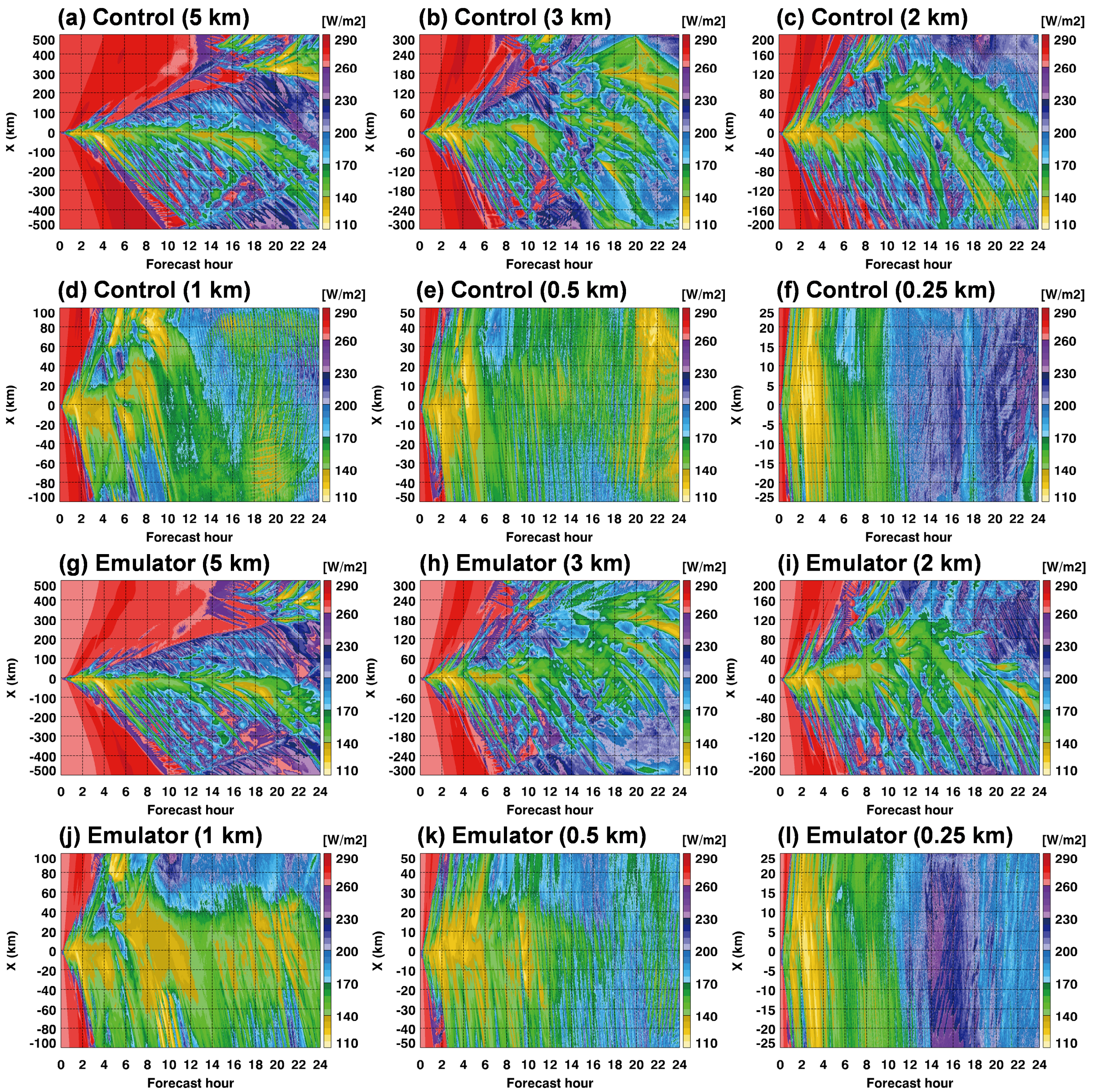

Figure 4. Spatio-temporal evolutions of outgoing longwave radiation (OLR) with different horizontal resolutions between control and emulator results in ideal-case simulations.

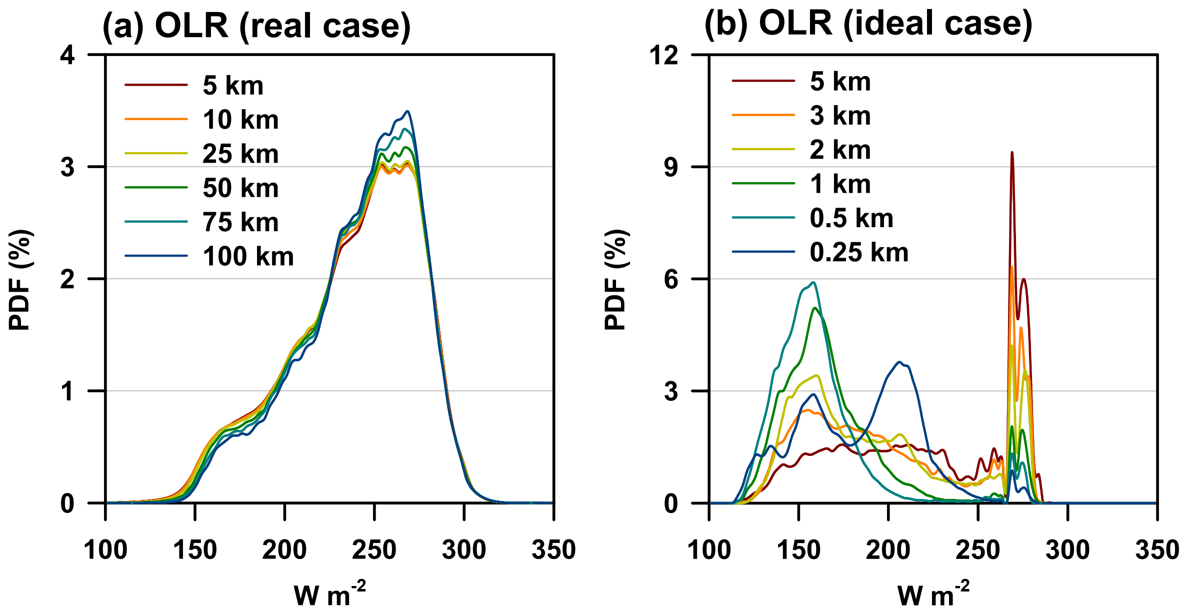

Figure 5. Probability density functions of outgoing longwave radiation (OLR) between (a) real-case and (b) ideal-case simulations at different horizontal resolutions.

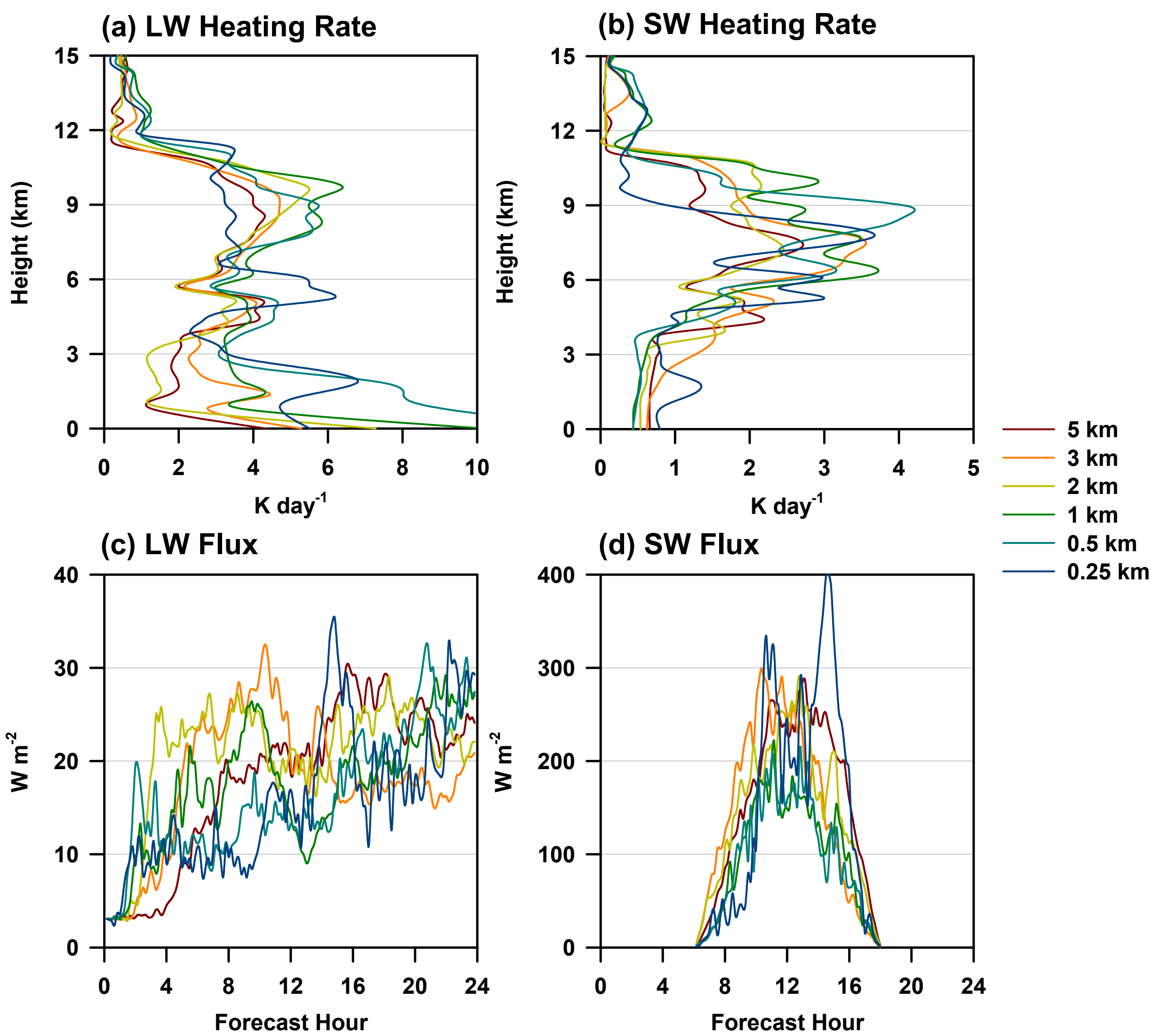

Figure 6. Vertical RMSEs of (a) LW and (b) SW heating rates, as well as temporal RMSEs of (c) LW and (d) SW fluxes, compared to the control runs with different horizontal resolutions in ideal-case simulations.