

# Increasing model vertical resolution may not necessarily lead to improved atmospheric predictability

Cite as: Chaos **32**, 073120 (2022); <https://doi.org/10.1063/5.0081734>

Submitted: 10 December 2021 • Accepted: 17 June 2022 • Published Online: 19 July 2022

 Sungju Moon,  Jong-Jin Baik, Hyo-Jong Song, et al.



[View Online](#)



[Export Citation](#)



[CrossMark](#)

**APL Machine Learning**  
Open, quality research for the networking communities

## Now Open for Submissions

[LEARN MORE](#)



# Increasing model vertical resolution may not necessarily lead to improved atmospheric predictability

Cite as: Chaos **32**, 073120 (2022); doi: 10.1063/5.0081734

Submitted: 10 December 2021 · Accepted: 17 June 2022 ·

Published Online: 19 July 2022



View Online



Export Citation



CrossMark

Sungju Moon,<sup>1,a)</sup> Jong-Jin Baik,<sup>1,b)</sup> Hyo-Jong Song,<sup>2</sup> and Ji-Young Han<sup>3</sup>

## AFFILIATIONS

<sup>1</sup>School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, South Korea

<sup>2</sup>Department of Environmental Engineering and Energy, Myongji University, Yongin 17058, South Korea

<sup>3</sup>Korea Institute of Atmospheric Prediction Systems, Seoul 07071, South Korea

<sup>a)</sup>**Present address:** Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario L8S 4K1, Canada.

<sup>b)</sup>**Author to whom correspondence should be addressed:** jjbaik@snu.ac.kr

## ABSTRACT

The widely accepted existence of an inherent limit of atmospheric predictability is usually attributed to weather's sensitive dependence on initial conditions. This signature feature of chaos was first discovered in the Lorenz system, initially derived as a simplified model of thermal convection. In a recent study of a high-dimensional generalization of the Lorenz system, it was reported that the predictability of its chaotic solutions exhibits a non-monotonic dimensional dependence. Since raising the dimension of the Lorenz system is analogous to refining the model vertical resolution when viewed as a thermal convection model, it is questioned whether this non-monotonicity is also found in numerical weather prediction models. Predictability in the sense of sensitive dependence on initial conditions can be measured based on deviation time, that is, the time of threshold-exceeding deviations between the solutions with minute differences in initial conditions. Through ensemble experiments involving both the high-dimensional generalizations of the Lorenz system and real-case simulations by a numerical weather prediction model, this study demonstrates that predictability can depend non-monotonically on model vertical resolution. Further analysis shows that the spatial distribution of deviation time strongly contributes to this non-monotonicity. It is suggested that chaos, or sensitive dependence on initial conditions, leads to non-monotonic dependence on model vertical resolution of deviation time and, by extension, atmospheric predictability.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0081734>

A reliable weather forecast hinges on successful utilization of numerical models, requiring tremendous amounts of computing power. Despite the cost, much effort goes into further refining the model resolution. Unfortunately, there is a limit to improving predictability in this way, which is often vaguely blamed on chaos or weather's sensitive dependence on initial conditions. By estimating the time it takes for tiny initial errors to grow significant enough in experiments involving a realistic weather model as well as a simpler chaotic dynamical system, this study seeks to clarify and evaluate the claimed connections between chaos and atmospheric predictability. The analysis demonstrates that chaos theory-based approaches can complement the conventional, feature-specific explanations for atmospheric predictability.

## I. INTRODUCTION

From daily weather forecasts to forewarning of extreme weather events, knowledge about future states of the atmosphere plays an important role in shaping our day-to-day lives. Modern weather forecasting relies on the practice of numerical weather prediction (NWP), which uses numerical models that solve equations governing Earth's atmosphere, parameterize various physical processes, such as clouds and precipitation,<sup>1,2</sup> and incorporate data from a vast network of observations.<sup>3</sup>

Despite the great strides made in the 100-year history of NWP,<sup>4,5</sup> it is widely accepted that there exists an inherent limit on how far ahead Earth's atmosphere can be predicted.<sup>6</sup> This limit is thought to exist regardless of improvements in model or observational precision and irrespective of any anticipated growth in

computational resources. Reasons for this predictability limit are usually ascribed to weather's sensitive dependence on initial conditions as a chaotic system, citing Lorenz's 1963 paper, hereafter L63,<sup>7</sup> which introduced a three-dimensional system of ordinary differential equations commonly referred to as the Lorenz system. Indeed, the discovery of deterministic chaos in the celebrated Lorenz system<sup>7</sup> played a crucial role in reshaping our understanding of atmospheric predictability; however, it was the ideas contained in his later paper from 1969, hereafter L69,<sup>8</sup> that had a greater practical impact on the study of atmospheric predictability.<sup>9</sup> In L69, Lorenz found the source of unpredictability in the loss of marginal predictability gain in a model with many scales of motion. These two different works of Lorenz on related subjects, as well as the subsequent studies they spawned, have resulted in a subtle yet unmistakable disconnect<sup>9</sup> between the view of weather as a chaotic system based on L63 and the theoretical understanding of atmospheric predictability according to L69.

This study is an exploratory attempt at reconciliation through a rather intuitively defined measure of atmospheric predictability which we call deviation time. In a series of ensemble experiments involving both a realistic NWP model and a high-dimensional generalization of the Lorenz system, we attempt to measure through deviation time the limit on atmospheric predictability that perhaps offers more insight on the ideas presented in L63, that is, loss of predictability due to weather's sensitive dependence on initial conditions. We find that atmospheric predictability measured in this way depends non-monotonically on the model's vertical resolution; in other words, not only there are diminishing marginal returns in improved predictability coming from refinements in the model vertical resolution, refining the vertical resolution beyond the saturation point can even have negative impacts on predictability. That this behavior is observed in both the realistic NWP model and the generalized Lorenz systems suggests that the culprit behind this non-monotonicity, regardless of the specific reasons relevant to each simulation case, likely lies with the underlying chaotic dynamics of these models. The subsequent analysis of the experimental results aims to find evidence in support of this claim.

## II. BACKGROUND

### A. Lorenz's ideas about atmospheric predictability

In L63, the idea that chaotic solutions exhibit sensitive dependence on initial conditions is demonstrated using the Lorenz system through so-called twin experiments comparing two trajectories, or "analogs,"<sup>10</sup> with slightly perturbed initial conditions. It is easy to see that for the Lorenz system, there appears to be a relatively clear point in time beyond which the deviations among analog trajectories are noticeably pronounced, by which point it can be said that predictability loss in the sense of L63 has occurred.<sup>11</sup>

Meanwhile, in L69, Lorenz approaches the problem from the turbulence theory perspective,<sup>12–14</sup> focusing on marginal predictability gain over many scales of motion. It is argued that as one attempts to resolve motions at finer scales, error growths at these scales start to overwhelm the overall predictability, rendering any such attempt futile. According to Lorenz,<sup>8</sup> the error growth rate through the scales

depends on the prescribed spectral slopes of kinetic energy, whose existence and precise values are backed up by a slew of observational evidence, including the well-received aircraft measurements by Nastrom and Gage.<sup>15</sup> Mathematically, the infinite series representing the marginal predictability gain is found to be convergent, making it possible to theoretically estimate the predictability limit in this sense. A great majority of theoretical progress in the field has thus far focused on adopting<sup>16,17</sup> or expanding upon<sup>14,18–23</sup> the arguments made in L69.

The Lorenz system introduced in L63 was originally conceived as a simplified model of Rayleigh–Bénard convection, a classic thermal convection problem in fluid dynamics. The derivation involves severe truncation of high-order spectral modes. What is surprising about the Lorenz system is, therefore, not just that it has chaotic dynamics, but that it does so despite being low-order by design. It follows that the nonlinear dynamical explanation for limited predictability in L63 based on the chaotic behavior of the Lorenz system should be distinct<sup>9</sup> from the scale-interaction argument made in L69.

When viewed as a thermal convection problem, the Lorenz system can be extended to incorporate additional higher-order spectral modes, and higher-dimensional systems obtained in this way can thus be said to resolve motions at smaller scales than does the original Lorenz system.<sup>24–26</sup> One recent example is the  $(3N)$ - and  $(3N + 2)$ -dimensional generalization<sup>11</sup> of the Lorenz system for any positive integer  $N$ , which allows incorporation of any number of higher-order modes in the vertical direction through controlling the dimension of the system. This approach to resolving smaller-scale motions in the vertical direction by the generalizations is reminiscent of the vertical resolution sensitivity analysis carried out by Adamec,<sup>27</sup> in which two- and three-mode quasi-geostrophic simulations of an ocean flow were compared. In order to see how predictability in the sense of L63 is affected by the high-dimensional extensions, *deviation time*, or the time of significant deviation between the analogs, has been measured in a set of twin experiments run at different dimensions in a relevant parameter space. While the applicability of deviation time in the context of NWP remains uncertain, it can serve as a step toward a reconciled understanding of atmospheric predictability.

The twin experiments performed on the generalized Lorenz systems<sup>11</sup> with a rather crudely chosen threshold deviation revealed a few surprisingly rich facts about predictability in these high-dimensional extensions of the Lorenz system: (1) Deviation time does not rise or fall monotonically with increasing dimension. (2) Deviation time tends to be longer around the border between chaotic and non-chaotic regions of the dimension–parameter space, forming a sort of depression—we can call it the *valley of uncertainty*—at parameters that are deeply embedded within the chaotic region. (3) The thalweg, a line connecting the lowest points of the valley, is not a straight line aligned with the dimensional axis, which explains why changing the dimension under fixed parameters leads to ups and downs in deviation time, resulting in non-monotonicity of predictability with respect to dimension. Furthermore, deviation time was not found to be sensitive to the choice of variable for computing deviations between the analogs.<sup>11</sup>

## B. Model vertical resolution and predictability in NWP

The most promising results toward measuring atmospheric predictability in the sense of L63 are found in studies about ensemble error growths in operational NWP models. Here, an ensemble can be thought of as an extension of analogs but with its members coming in greater numbers than just two in twin experiments. Of course, in realistic NWP models, atmospheric predictability involves more complex processes than what goes on in the simplified models,<sup>28,29</sup> but the reasoning goes that measuring predictability should ultimately come down to measuring the growth of deviations among the ensemble members, more commonly referred to as ensemble error or ensemble dispersion. Fittingly, sensitive dependence on initial conditions from L63 had been the *de facto* justification for bringing ensemble approaches to NWP.<sup>30,31</sup> In this context, some reconciliation efforts between L63 and L69 were made by adopting a scale-dependent measure of ensemble error or comparing multiple model runs with different model resolutions.<sup>17,32</sup> For example, in a series of global forecasts with different horizontal resolutions, Buizza<sup>33</sup> found that predictability is sensitive to model resolution in the short range but not in the long range.

It is still unclear, however, whether the problems considered by many of these ensemble-based error growth studies are representative of sensitive dependence on initial conditions in the strict sense of L63. Instead of the time of significant deviation, the usual focus of the ensemble error growth research is the time of error saturation<sup>23,34</sup> beyond which any marginal advantage of using an ensemble becomes nil. It is in this sense the oft-quoted *intrinsic* predictability limit of around two weeks came to be, estimated based on error saturation comparisons against simulations assuming nearly perfect initial conditions.<sup>6</sup> Moreover, rather than taking the ensemble spread problem as given or trying to mitigate the spread, the end-goal for some of these studies is the opposite, that is, to avoid underdispersion of ensemble members in order to cover all potential trajectories. Understandably, ensemble generation in operational NWP generally involves some form of artificial cultivation of dispersion among the members,<sup>16,35–37</sup> which may contaminate the outcome if one's aim is to detect the predictability loss strictly caused by sensitive dependence on initial conditions following L63. Evidence suggests that Lorenz's idea about longer-term statistical behavior of weather, called the ergodicity principle, is also better captured by an ensemble of simple analogs than by purely random noise or trajectories infused with artificial error enhancers.<sup>38</sup>

In the context of NWP, a situation analogous to raising the dimension of the Lorenz system<sup>11</sup> can be arranged by refining the model vertical resolution. The relationship between a model's resolution and predictability is a delicate matter in operational NWP partly because it also involves the economic problem of optimal distribution of computational resources.<sup>39,40</sup> Studies so far have also focused more on the impact of refining the horizontal resolution than the vertical resolution, but in recent years, the modeling community started to pay more attention to the importance of also having a sufficiently high vertical resolution for its cost-effectiveness and the potentially deleterious effects without it.<sup>21,39,41</sup> The effect of resolution refinements on horizontal dynamics can be very different from that on vertical dynamics and boundary conditions.

Note that while this study is concerned with inherent predictability of the atmosphere, the majority of discussions in the literature surrounding the relationship between model resolution and predictability is concerned with *predictive skill*, which is related to but distinct from the upper limit of inherent predictability.<sup>42,43</sup> Unlike horizontal resolution refinements, it is well-documented that refining the vertical resolution does not always lead to improved predictive skill. The reported explanations for this non-monotonicity in predictive skill tend to be feature-specific, citing inconsistent interactions between the atmospheric processes above and below the melting level,<sup>44</sup> spurious local sensitivities to terrain height,<sup>45</sup> and an inadequate spectral transition from the synoptic scale to mesoscale.<sup>19</sup> Furthermore, there is some evidence suggesting that properly distributed vertical resolutions, more so than across-the-board refinements in the model vertical resolution, are important for simulating hurricane development.<sup>46,47</sup> In the past, problems caused by having inconsistent horizontal and vertical grid aspect ratios were also often highlighted,<sup>48–50</sup> but it appears that modern NWP models have more or less resolved this issue through proper tuning of energy cascades. Checking the model's compliance with the prescribed slopes of kinetic energy spectra has now become a standard practice in NWP modeling,<sup>51</sup> although there still remain concerns whether such a compliance is merely an artifact of having still-inadequate vertical grid resolutions.<sup>41</sup>

Just as chaos in the sense of L63 leads to the unevenness of the valley of uncertainty for the generalized Lorenz systems,<sup>11</sup> these feature-specific explanations for the non-monotonicity of predictability on model vertical resolution may turn out to be some of the ways through which the inherent chaos and nonlinearity are realized in the model atmosphere. Nevertheless, the relationship between the apparent non-monotonicity in predictive skill and that in predictability of the real atmosphere is still at large, and it is suspected that the solution to this problem will require a unifying explanation from the nonlinear dynamical perspective. Bridging the gap between the underlying chaotic dynamics and its manifestations in realistic model atmospheres through specific features necessitates finding a proper measure of atmospheric predictability in the sense of L63, that is, the measure of predictability strictly based on sensitive dependence on initial conditions. Now, in dynamical systems theory, we do have a tool for discerning between stability and chaos, namely, Lyapunov exponents. Using medium-complexity climate models, De Cruz *et al.*<sup>52</sup> demonstrated that the intrinsic stability properties measured using Lyapunov exponents can depend non-monotonically on model resolution; however, while Lyapunov exponents are a useful tool for characterizing instability in the dynamics,<sup>53</sup> what they measure is in some sense nontrivially different from the predictability envisioned by Lorenz in L63.<sup>54,55</sup> Moreover, reliably estimating Lyapunov exponents based on noisy data is known to be quite challenging, making it impractical for mass-scale deployment for measuring predictability in the context of real observational data or large ensemble model outputs.<sup>56</sup> In search of a measure of atmospheric predictability that may be more in tune with L63, we take a simpler, a kind of back-to-basics, approach to exploring what a reasonable definition of deviation time can look like for NWP model outputs. This is not to say that our approach represents the measure of predictability in the sense of L63; rather, by employing this basic measure of predictability, we attempt to provide a

perspective that may be in better alignment with what Lorenz envisioned in L63, so as to provide a bridge between the two perspectives of L63 and L69.

As an NWP model, we use the Weather Research and Forecasting (WRF) model, a comprehensive NWP system with emphasis on mesoscale meteorology.<sup>57</sup> Motivated by the initial results from studying the generalized Lorenz systems,<sup>11</sup> we use the estimated deviation times based on WRF model simulations of a real precipitation case to demonstrate non-monotonic dependence of atmospheric predictability on model vertical resolution in NWP. Note that, unlike in the Lorenz systems, a complete agreement among the deviation times computed from different output variables is not to be expected from the WRF model results. It is, therefore, important to assess to what extent the relationships between deviation time and model vertical resolution can be identified among different output variables and, in case there are large discrepancies, what kinds of connections can be made between the different variables and deviation time. Such an analysis can lead to the development of a more robust measure of predictability grounded in the chaotic dynamics of model atmospheres as well as lending further insight into the atmospheric predictability problem at large.

### III. RESULTS

#### A. Deviation time in the Lorenz systems revisited

In this subsection, we expand the twin-experimental results of Moon *et al.*<sup>11</sup> by performing ensemble experiments on their  $(3N)$ - and  $(3N + 2)$ -dimensional generalization of the Lorenz system. The  $(3N)$ -dimensional generalized Lorenz systems<sup>11</sup> consist of ordinary differential equations in variables  $X_1, \dots, X_N, Y_1, \dots, Y_N$ , and  $Z_1, \dots, Z_N$ . The  $(3N + 2)$ -dimensional systems have two additional variables,  $Y_{N+1}$  and  $Z_{N+1}$  (see Moon *et al.*<sup>11</sup> for the equations and their derivations). It is reiterated that raising the model dimension according to the generalization by Moon *et al.*<sup>11</sup> corresponds to raising the vertical resolution of the associated fluid convection model. This can be interpreted as refining the model vertical resolution in a more realistic atmospheric model, such as WRF, as will be discussed in Subsection III B.

For integration of the control members, the same initial conditions, parameter values, and numerical integration schemes are used. We set all variables to 0 initially except  $X_1 = 1$ . The parameter values used throughout are  $r = 500$ ,  $\sigma = 50$ , and  $b = 8/3$ , where  $r$  is the Rayleigh parameter,  $\sigma$  is the Prandtl number, and  $b$  is a geometric parameter. For numerical integration, the fourth-order Runge–Kutta method with time step size  $\Delta\tau = 1.0 \times 10^{-4}$  is used. Corresponding to each dimension, four perturbed ensemble members are generated by perturbing the initial condition for  $X_1$  by  $\varepsilon$ , which is drawn from the uniform distribution  $\mathcal{U}(-1.0 \times 10^{-4}, 1.0 \times 10^{-4})$ .

The results are summarized in Fig. 1. Here, two types of deviation time are defined. Type I deviation time marks the time of significant deviation among the ensemble members, that is, when the mean absolute error between the perturbed and control member solutions exceeds the threshold value. Type II deviation time marks the time when there is a significant deviation among the absolute errors.

Suppose  $m$  solutions with perturbed initial conditions are generated in an ensemble experiment using a generalized Lorenz system. Consider the perturbed solutions  $Z_1^{(1)}(\tau), \dots, Z_1^{(m)}(\tau)$  for variable  $Z_1$  in time  $\tau$  and denote the control (unperturbed) solution by  $Z_1^{(c)}(\tau)$ . Following Moon *et al.*,<sup>11</sup> type I deviation time is defined to be the time  $\tau_d$  when the first instance of threshold exceedance occurs in the mean absolute difference between the perturbed and control solutions, that is, the smallest  $\tau_d$  such that

$$\frac{1}{m} \sum_{i=1}^m |Z_1^{(i)}(\tau_d) - Z_1^{(c)}(\tau_d)| > \delta, \quad (1)$$

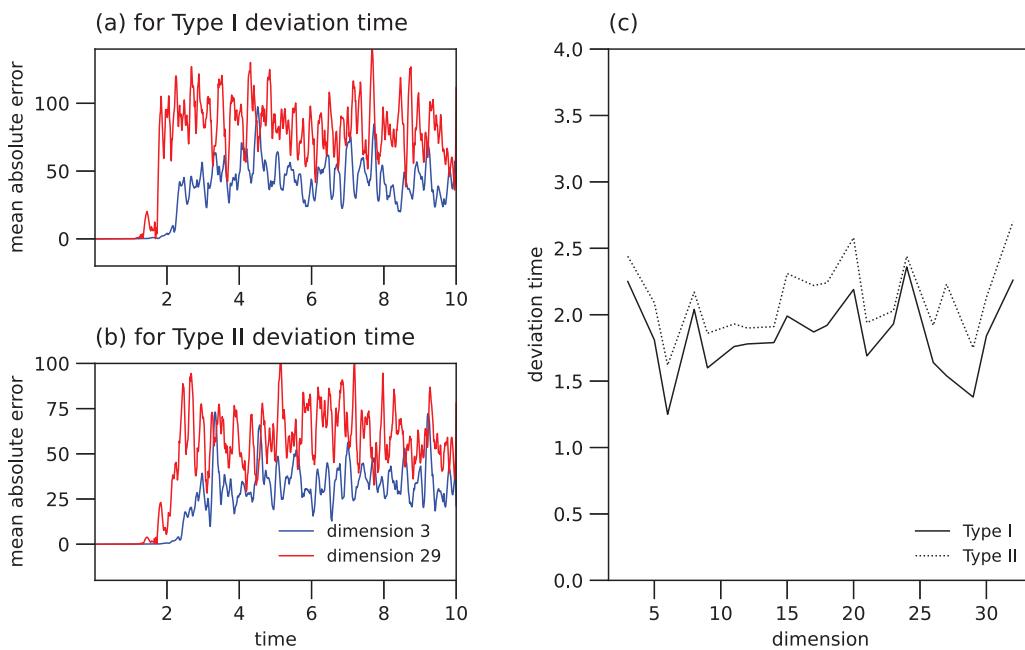
where  $\delta$  is the threshold magnitude chosen for type I deviation time. Type II deviation time is newly defined in this study as the time  $\tau_d$  when the first instance of threshold exceedance

$$\frac{1}{\binom{m}{2}} \sum_{i,j=1}^m \left| |Z_1^{(i)}(\tau_d) - Z_1^{(c)}(\tau_d)| - |Z_1^{(j)}(\tau_d) - Z_1^{(c)}(\tau_d)| \right| > \delta \quad (2)$$

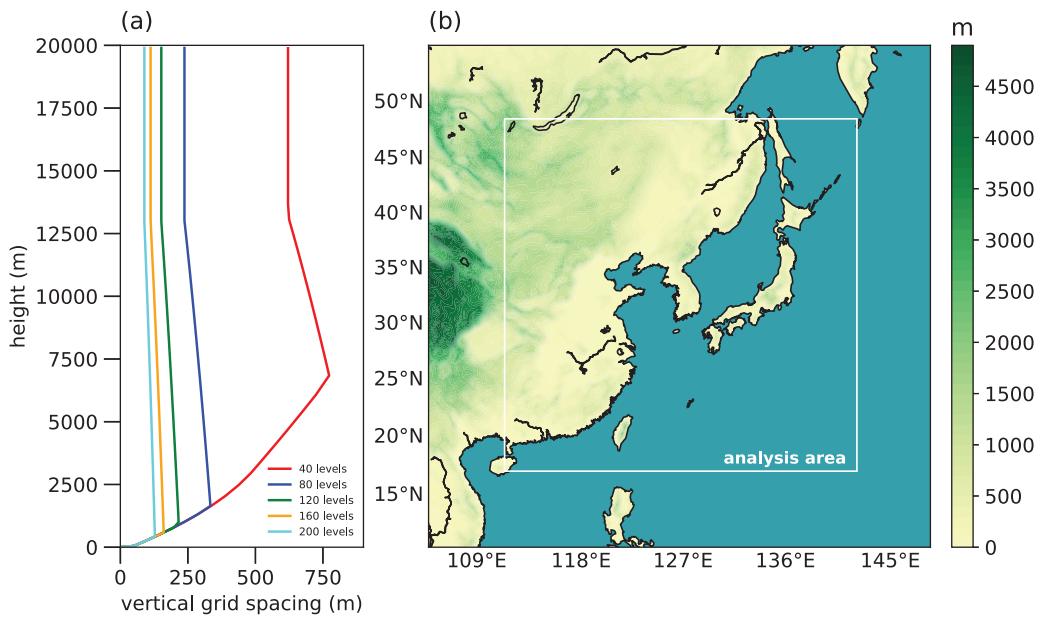
occurs, where  $\binom{m}{2}$  is the binomial coefficient and  $\delta$  is the threshold magnitude chosen for type II deviation time. In this study, the threshold magnitude of  $\delta = 10$  is used for calculating both type I and type II deviation times. Moon *et al.*<sup>11</sup> found the threshold deviation magnitude of 10 to be suitable for variable  $Z_1$  given the initial condition perturbation magnitude of  $1.0 \times 10^{-4}$  in  $X_1$ . Here,  $\delta = 10$  was chosen as a result of extensive testing on extended Lorenz and Lorenz-like systems. It would be interesting in a future study to conduct a systematic investigation or statistical testing to determine the sensitivity of the deviation time with respect to the threshold value.

Type I deviation time, therefore, simply tracks the sudden growth of error as the Lorenz system is known to exhibit. In this sense, type I deviation time is compatible with the deviation time defined in the context of twin experiments in Moon *et al.*<sup>11</sup> On the other hand, type II deviation time follows the differences among the error trajectories; therefore, it is only defined for an ensemble of size of at least two.

The time series of mean absolute error needed for computation of type I and type II deviation times are shown in Figs. 1(a) and 1(b), respectively, for the generalized Lorenz systems of dimensions 3 and 29, that is, dimensions  $3N$  with  $N = 1$  and  $(3N + 2)$  with  $N = 9$  (see Appendix A for the equations). As expected, the deviations between the perturbed and control solutions clearly experience sudden increases, exceeding the threshold magnitude afterward [Fig. 1(a)]. It is notable that the absolute differences among the deviations also rise suddenly as shown in Fig. 1(b). Type I and type II deviation times can therefore be reliably estimated with the choice of a threshold magnitude  $\delta = 10$ . Figure 1(c) reveals the non-monotonic dependence of both types of deviation time on dimension. The non-monotonicity can be explained by changes in the relative position of the given parameter combinations with respect to the borders against non-chaotic regions in the parameter space.<sup>11</sup> In other words, changes in dimension trigger shifts in the underlying bifurcation structure of the generalized Lorenz system. This is an interesting revelation particularly in connection with potentially important implications for operational weather forecasting. Recall that changes in the dimensionality of a generalized Lorenz system correspond to changes in the model vertical resolution. If this idea



**FIG. 1.** Ensemble-experimental results using the 3- and 29-dimensional generalized Lorenz systems of Moon *et al.*<sup>11</sup> Time series of (a) the mean absolute errors (mae) indicating differences in  $Z_1$  between the perturbed and control members and (b) the mae of the differences over all possible perturbed member pairs. (c) Type I and II deviation times as a function of dimension.



**FIG. 2.** (a) Vertical grid spacing distributions corresponding to using 40, 80, 120, 160, and 200 levels in the WRF model. (b) The single domain configuration with terrain height. Inside of the box with white edges corresponds to the analysis area.

can be applied to operational weather models, it leads to the conclusion that refining the vertical resolution of such a model may not necessarily lead to improved predictability because the very action of resolution change would trigger unpredictable shifts in the model's underlying bifurcation structure.

Note that for dimensions 3 and 29, the two types of deviation time do not seem too far apart from each other. In fact, Fig. 1(c) shows that across all dimensions from 3 to 32, type II deviation time closely follows type I deviation time but with a slight delay, and this delay appears to be controlled by the ensemble configuration based on repeated numerical experiments showing that the gap between the two deviation times shrinks under certain circumstances, such as having greater ensemble sizes or smaller initial perturbation magnitudes. The unreliability of ensemble error magnitudes seems to occur mainly due to sudden increases in the ensemble errors themselves, stemming from the solutions' sensitive dependence on initial conditions. For this reason, here, it does not matter which one of the two types of deviation time is used for measuring relative changes in predictability depending on the model vertical resolution. In other systems, however, such as NWP models whose unreliability of ensemble error has sources other than in their solutions' sensitive dependence on initial conditions, such an agreement between the two types of deviation time may not always hold.

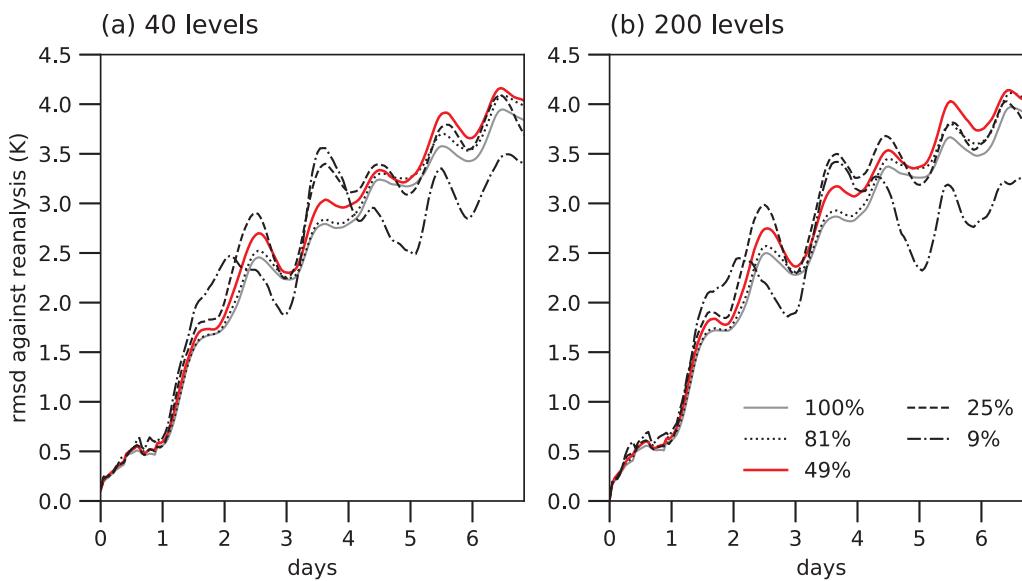
## B. WRF model control simulations

For WRF model simulations, we consider a typical summer precipitation case over East Asia, characterized by the presence of a quasi-stationary front.<sup>58</sup> Version 4.1.2 of the WRF model<sup>57</sup> is used for the single domain simulations with varying number of model levels ranging from 40 to 200 levels. Using 40 model levels corresponds to the vertical resolution commonly used in NWP models

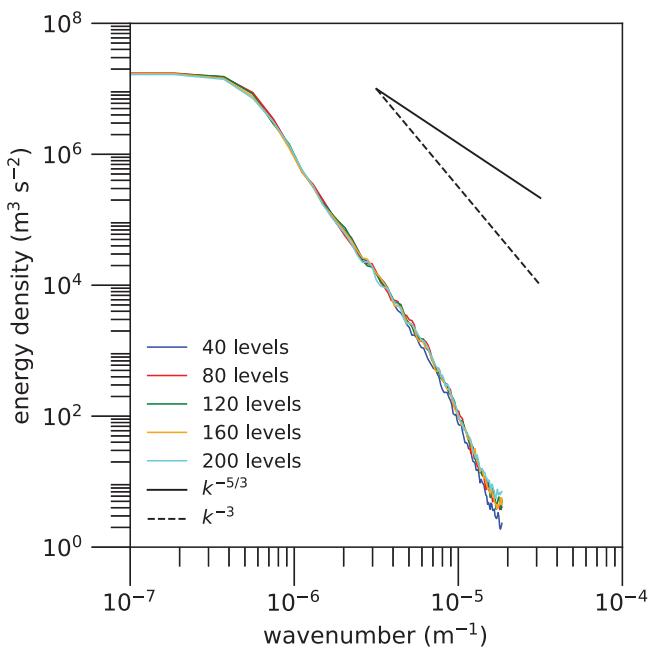
**TABLE I.** WRF model domain and physics configurations.

Time domain	00:00 UTC 30 June 2016–00:00 UTC 7 July 2016 time step: 20 s
Spatial resolution	Horizontal: 200 × 200 with a grid size of 27 km Vertical: 40, 80, 120, 160, 200 levels
Physics options	YSU PBL scheme <sup>78</sup> Dudhia shortwave radiation scheme <sup>79</sup> RRTM longwave radiation scheme <sup>80</sup> Unified Noah land surface model <sup>81</sup> Kain–Fritsch convection scheme <sup>82</sup> WSM6 microphysics scheme <sup>83</sup>
Initial and boundary conditions	ECMWF ERA5 (hourly, 31 × 31 km <sup>2</sup> ) <sup>61</sup>

around 20 years ago.<sup>59</sup> With the increment of 40 model levels at a time, we refine the model vertical resolution up to 200 model levels, which corresponds to a somewhat finer vertical resolution than the 137 model levels currently in use for high-resolution forecasts at the European Centre for Medium-Range Weather Forecasts;<sup>60</sup> hence, 200 model levels represent the expected improvements in model vertical resolution for operational NWP in the near future. The vertical grid spacing corresponding to each number of model levels follows WRF's default configuration as shown in Fig. 2(a). The model configuration, including the horizontal resolution and physics parameterization schemes used, is summarized in Table I. For initial and boundary conditions, the ECMWF Reanalysis 5 (ERA5) data<sup>61</sup> are used.



**FIG. 3.** Time series of the rmse in the 500-hPa air temperature between the ERA5 data (reanalysis) and control simulation results using (a) 40 model levels and (b) 200 model levels inside the analysis areas given by the concentric rectangles taking up 100%, 81%, 49%, 25%, and 9% of the model domain.



**FIG. 4.** Time-averaged kinetic energy spectra for simulations with a different number of model levels. Straight lines with spectral slopes corresponding to  $k^{-5/3}$  and  $k^{-3}$  are given in the upper right corner for reference.

As shown in Fig. 2(b), a rectangular subdomain is chosen as the analysis area in order to mitigate the unwanted influence of the hourly input of lateral boundary conditions. To see the extent of the effects the lateral boundary conditions have, the root-mean-square differences (rmsd) between the 500-hPa air temperature from the reanalysis data and that from the control simulations inside the chosen analysis areas are plotted in Fig. 3. In all cases, the same lateral boundary conditions are applied. It appears that having buffer zones around the analysis area increases the rmsd compared to using no buffer zones (100%) especially at local rmsd peaks. On the other hand, having too much of the buffer zones alters the overall rmsd growth patterns (e.g., the 25% and 9% cases in Fig. 3), indicating that certain important features are left out from the analysis area. The rectangular analysis area taking up 49% of the model domain is chosen for the present study. In both simulations with 40 and 200 model levels, its rmsd in air temperature against the reanalysis data exhibits a healthy growth of error (Fig. 3).

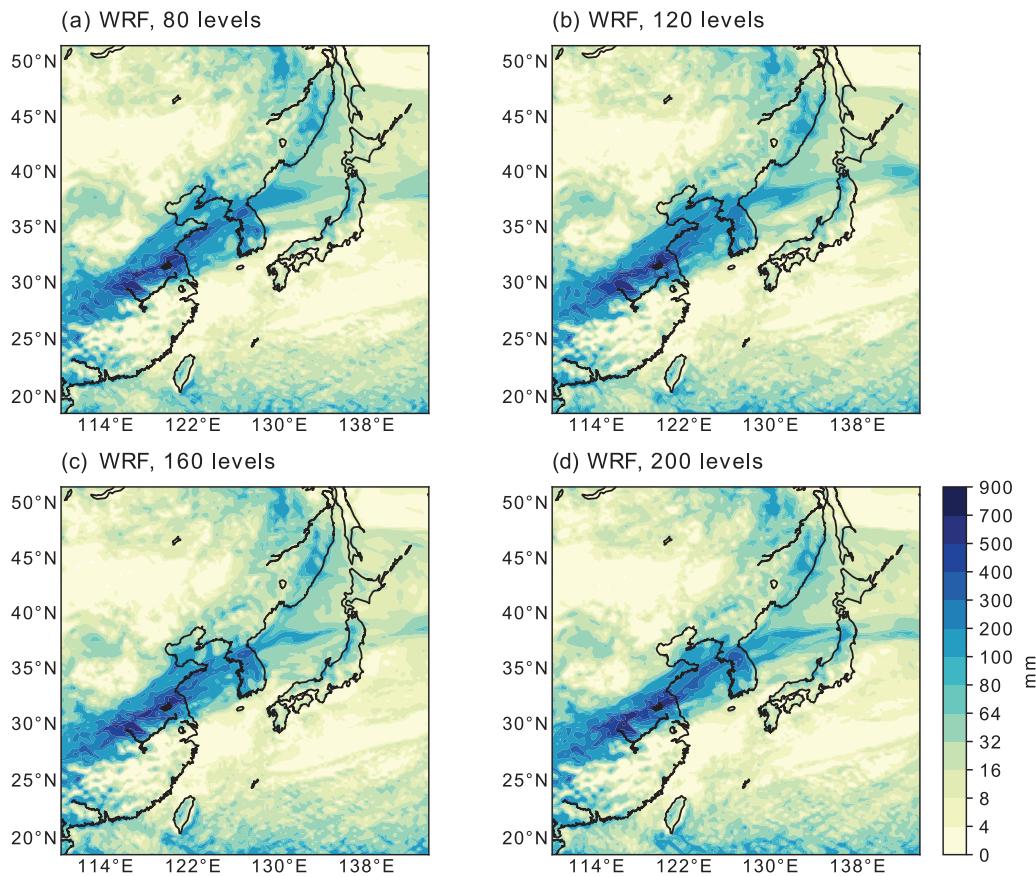
Increasing or decreasing the number of model levels in this way does not seem to result in any strange breakdown of the kinetic energy spectra (Fig. 4). Furthermore, insofar as the control simulations are concerned, the simulated accumulated precipitation amount patterns, while varying in details, do not appear to harbor any kind of biases in terms of location depending on the number of model levels used, adequately capturing the key features of the simulation results relevant to the predictability analysis, including the elongated east–west front characterizing the summer precipitation case over East Asia [Fig. 5; see also Figs. 11(c) and 11(d)].

### C. WRF model ensemble experiments and deviation time

For ensemble experiments using the WRF model, column-wise perturbations are applied to the initial air temperature by magnitudes up to 0.01 K. More precisely, for each grid point on the horizontal plane, the perturbation is given by  $\varepsilon \in \{\varepsilon^+, \varepsilon^-\}$  pulled from a uniform distribution of random numbers, and the same  $\varepsilon$  is applied for all grid points belonging to the same model column. Corresponding to each batch of simulations differing in the model vertical resolution, two members, p1 and p2, with positive initial perturbations  $\varepsilon^+ \sim \mathcal{U}(0, 0.01)$  and two members, n1 and n2, with negative initial perturbations  $\varepsilon^- \sim \mathcal{U}(-0.01, 0)$  are generated. With four members in each ensemble, there are six possible pairs to be compared against one another.

A number of different model output variables are available for computing deviation time in WRF model simulations. Here, we consider air temperature, geopotential height, and horizontal wind speed all at the 500-hPa level as well as accumulated precipitation amount. With the Lorenz systems, the absolute differences between the perturbed and control member solutions are tracked through time. In the context of the output variables from the WRF model where the comparisons are made between two-dimensional objects (e.g., values on an isobaric surface), the rmsd over the analysis area are used in place of the absolute differences between the perturbed and control members as done in the Lorenz system experiments. Note that the variables in the Lorenz system experiments are one-dimensional, and therefore, their L2 norms (rmsd) and their L1 norms (absolute difference) are equivalent. We obtain four rmsd time series corresponding to each batch as shown in Figs. 6–9 for the 500-hPa air temperature, 500-hPa geopotential height, 500-hPa horizontal wind speed, and accumulated precipitation amount, respectively. In such a time series, corresponding to type I deviation time are the rising rmsd time series above a certain threshold. Corresponding to type II deviation time are the times marking the sudden widening of the gaps among the individual rmsd time series. Figures 6–9 reveal that, unlike in the Lorenz systems, type I and type II deviation times in the realistic model atmosphere simulated by the WRF model have distinct characteristics. Note that whether to compute the rmsd among vertically interpolated variables corresponding to a specific pressure level or to consider the variables at all model levels for the rmsd can be a matter of debate. Our preliminary results indicate that the two methods lead to similar conclusions, however.

The rmsd time series for the 500-hPa air temperature in the WRF model ensemble experiments do not possess clear-cut times of a significant increase; rather, they initially experience rapid growths, followed by a gentler growth almost to the point of saturation in some cases (Fig. 6). This is largely in agreement with previously reported ensemble error growth patterns.<sup>28,29,33,62–64</sup> Lorenz<sup>54</sup> supposed that this rapid error growth in the short term is not related to chaos but dominated by small-scale turbulent dynamics perhaps involving “enormous latent heat energy released by convection.”<sup>28</sup> For this reason, type I deviation time based on the 500-hPa air temperature in the WRF model simulations is not as clearly pronounced and varies depending on the choice of threshold.



**FIG. 5.** Spatial distribution of accumulated precipitation amount for the duration of the entire time domain based on the control simulation with (a) 80, (b) 120, (c) 160, and (d) 200 model levels.

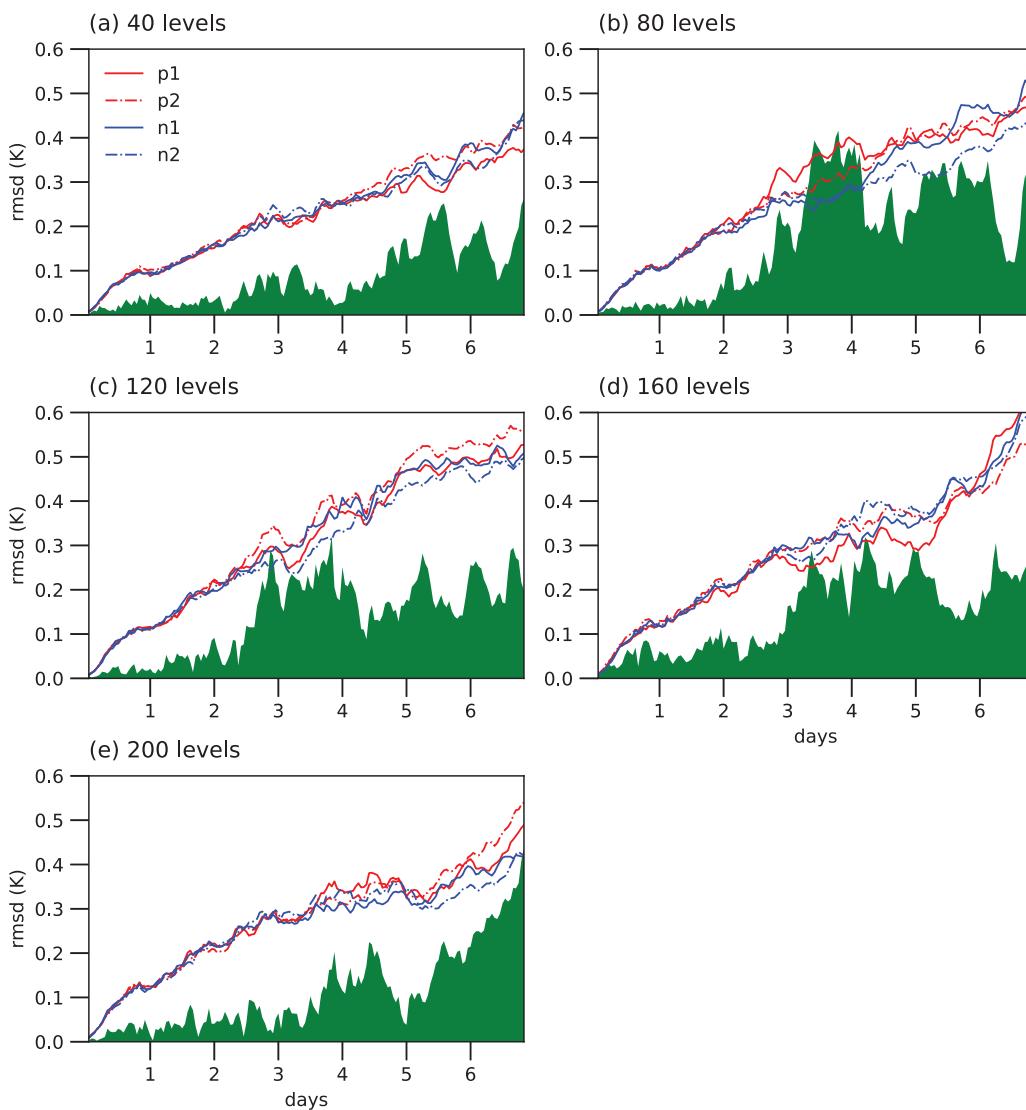
The rmsd time series for other variables likewise feature a rapid rising pattern for up to  $\sim 1$  day immediately followed by a brief slowdown (Figs. 7–9). For this, notice the concave-down shape of the time series in Figs. 7–9 as they start to plateau following an inflection point, prior to which there is observed initial acceleration of rmsd in a concave-up fashion. Note that this initial rapid-rise and slowdown are distinct from the longer-term rise and fall of the rmsd, which in some cases will accelerate at an even faster rate than the initial rate. The rmsd time series for the 500-hPa geopotential height additionally experience the return of a rapid rising pattern around day 4 (Fig. 7).

Given the suddenly visible onset of rapidly rising rmsd, it is possible to detect a relatively clear-cut type I deviation time for the 500-hPa geopotential height with an appropriate choice of threshold deviation. In other words, it might be that type I deviation time is clearly identifiable only for certain variables, which can, of course, depend also on the particular case under consideration among various other factors. In contrast, type II deviation time overall behaves similarly to what we see in the Lorenz system experiments; in all cases, there exists a relatively clear breaking point when the

individual rmsd curves start to split from one another. Therefore, using a threshold to estimate type II deviation time should yield relatively consistent results.

In Fig. 10, the estimated type I and type II deviation times for different output variables are plotted as a function of model vertical resolution. For reference, we additionally compute deviation times based on the moist exergy distance<sup>65</sup> between the perturbed and control members. The moist exergy distance is a squared norm combining the kinetic energy and the moist-air available enthalpy used to facilitate fair and consistent comparisons between the perturbed and reference thermodynamic states.<sup>65,66</sup> Since it is inevitable in realistic NWP outputs that estimated deviation times would depend on the choice of a variable, deviation time based on moist exergy distance can serve as a unifying metric with regard to how atmospheric predictability responds to having more number of model levels. Detailed information about the calculation of moist exergy distance can be found in Appendix B.

For type I deviation times in Fig. 10(a), we use as threshold the magnitude amounting to 50% of the maximum rmsd averaged over the four perturbed members. As a function of the number of model

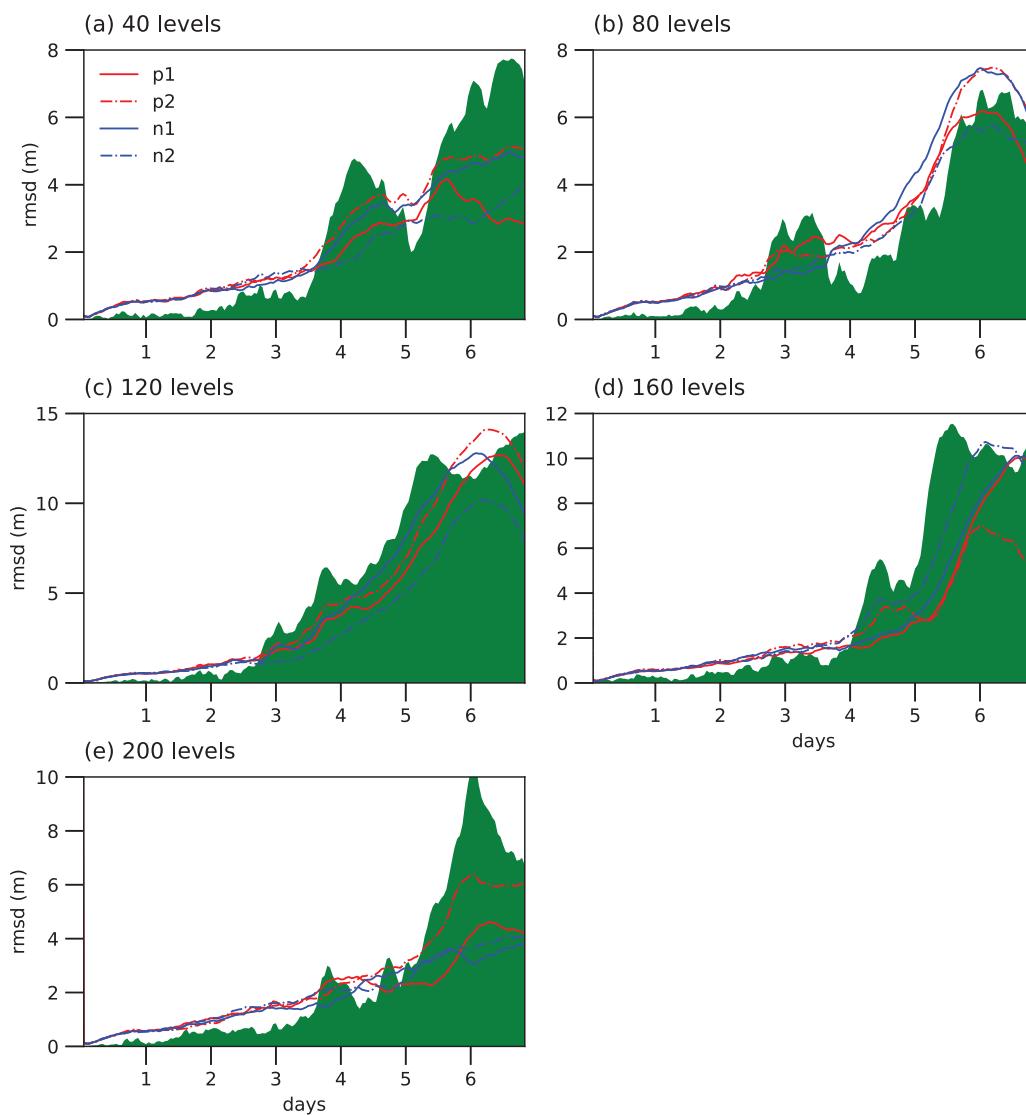


**FIG. 6.** Time series of the rmsd in the 500-hPa air temperature between the perturbed ( $p_1, p_2, n_1, n_2$ ) and control runs with (a) 40, (b) 80, (c) 120, (d) 160, and (e) 200 model levels. The shaded area indicates the sum of the absolute differences between all six possible pairs of rmsd.

levels, type I deviation times for most output variables, including moist exergy distance, conform to a sort of an inverted-tilde shape with a peak deviation time at the vertical resolution corresponding to 160 model levels. Type I deviation time computed based on the 500-hPa geopotential height is an exception, which monotonically increases with the number of model levels. These overall patterns are robustly maintained even when a different threshold percentage is used, indicating that even though type I deviation time absolutely depends on the choice of threshold, the relative rise and fall of deviation time with model vertical resolution do not and thus can still be used to measure relative changes in atmospheric predictability with respect to model vertical resolution. Table II summarizes the

estimated type I deviation times for different variables based on different threshold percentages.

For type II deviation times plotted in Fig. 10(b), 5% of the mean maximum deviation among the individual rmsd trajectories is chosen as the threshold deviation magnitude. Table III summarizes the estimated type II deviation times for different variables based on different threshold percentages. Note that due to type II deviation time being relatively clearly identifiable, the outcome is expected to be fairly insensitive to the choice of threshold within reasonable bounds. The overarching pattern for type II deviation time with respect to the number of model levels is a U-shaped dipping, which suggests that the reliability of ensemble



**FIG. 7.** The same as Fig. 6 except 500-hPa geopotential height.

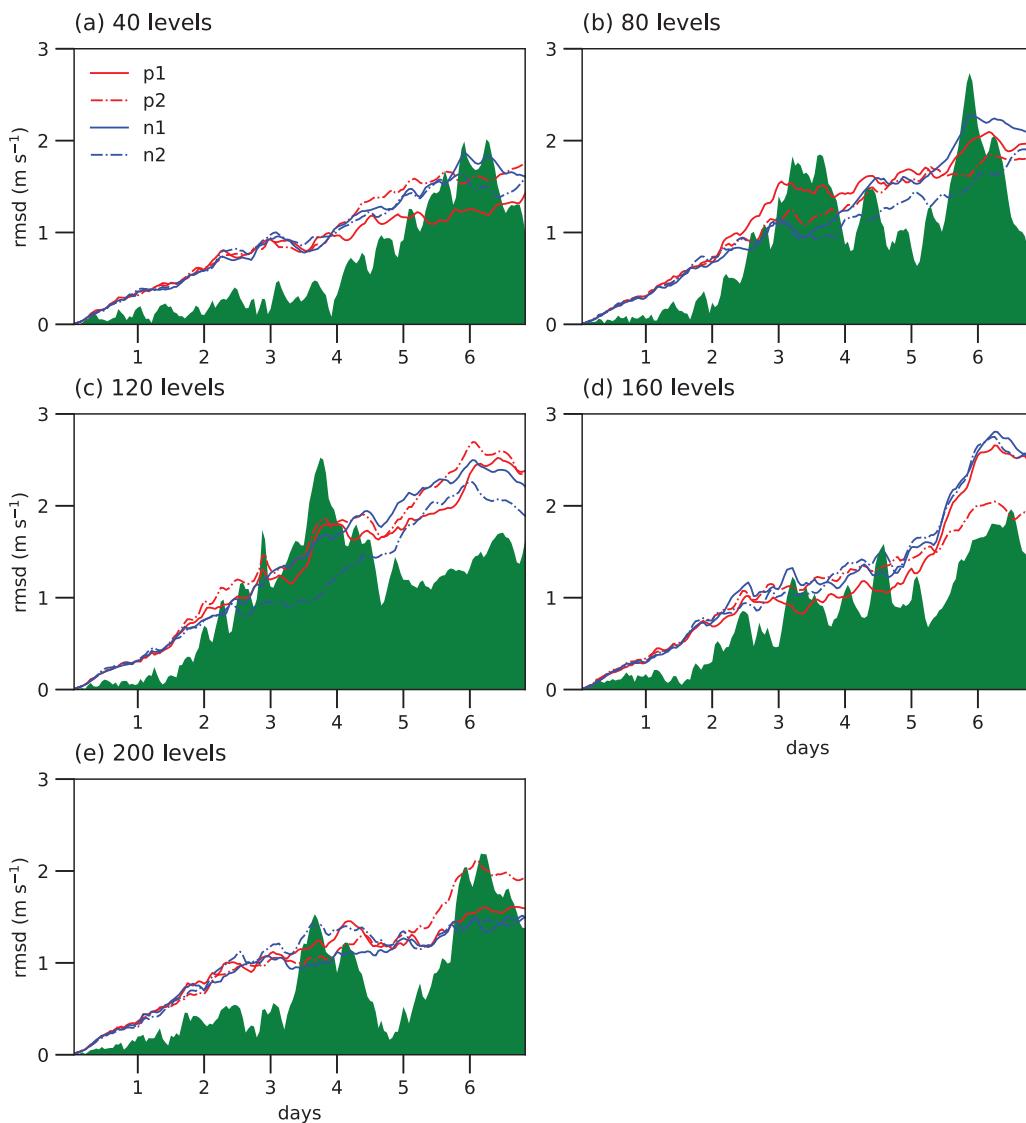
error, at least initially, decreases with refinements in the vertical resolution.

Note that estimating type II deviation time comes down to detecting the time when there appears to be an abrupt increase in deviations among the rmsd trajectories. In order to check for the robustness of our estimations of type II deviation times, we have adopted an alternative method using the slopes of the sums of all possible deviations between any two rmsd trajectories (shaded areas in Fig. 6). Since even the sums tend to be fairly noisy, for this method to work, we must set the running window wide enough so that we only capture sudden large increases in slope. For the rmsd trajectories based on the 500-hPa air temperature, abrupt rises in deviations occur as the slope exceeds  $0.1 \text{ K day}^{-1}$ . Table IV shows

that type II deviation times computed using this *slope method* on the 500-hPa air temperature rmsd trajectories are consistent with those computed using the threshold method. The downside of using the slope method is the lack of consistent threshold slope across different output variables, but for reasons explained in Subsection III D, the slope method can be useful for back-computing the type II deviation times at individual grid points of the model domain.

#### D. Spatial distribution of deviation time

We have so far established deviation time as a measure of atmospheric predictability or at least the aspects of it that are associated

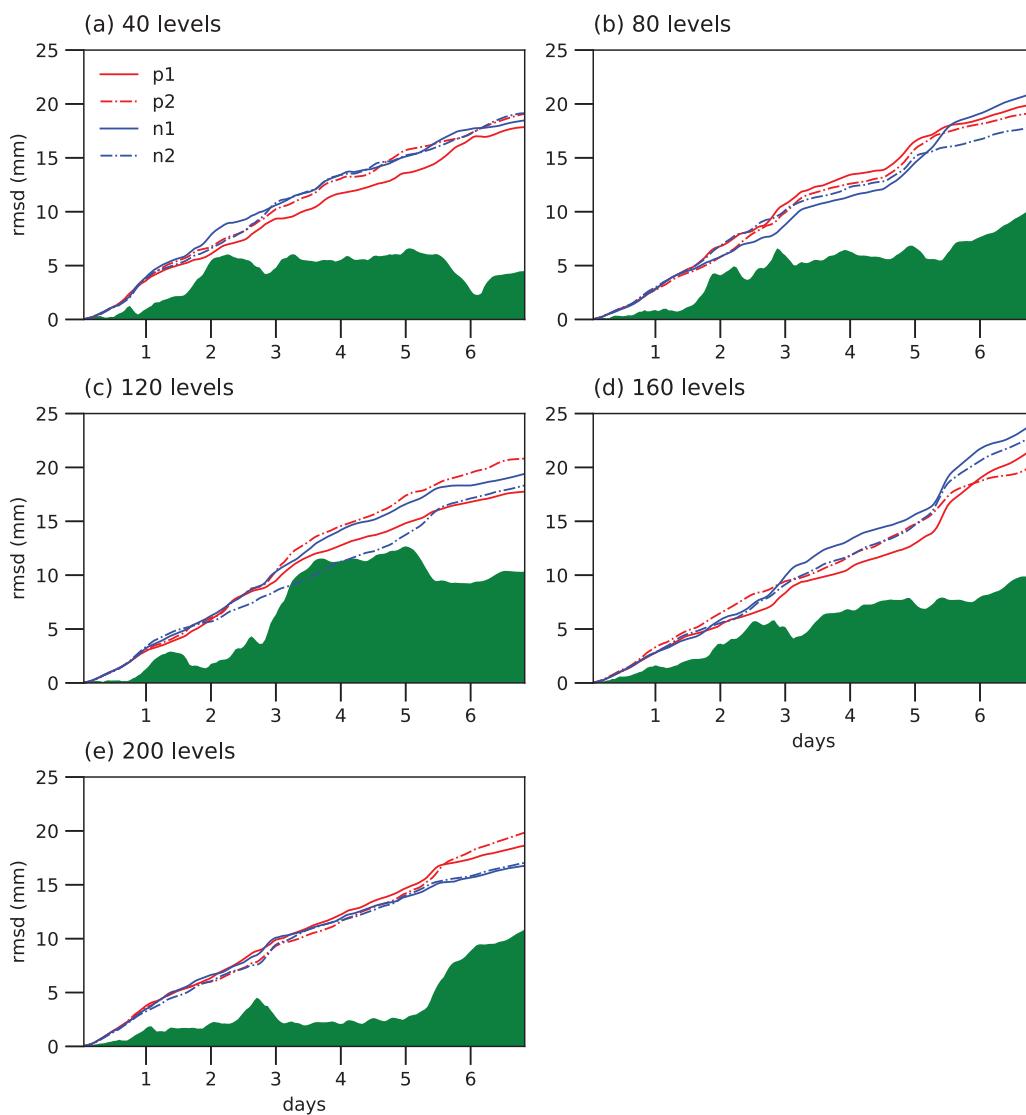


**FIG. 8.** The same as Fig. 6 except 500-hPa horizontal wind speed.

with sensitive dependence on initial conditions as expounded in L63. We have also seen that, as in the generalized Lorenz systems, the relationship between deviation time and model vertical resolution can be highly non-monotonic. The question remains as to whether deviation time also reflects aspects about atmospheric predictability that are influenced by specific features of the model atmosphere.

Figure 11 shows horizontal distributions of type I and type II deviation times computed at individual grid points based on the 500-hPa air temperature in simulations with 40 model levels together with the spatial patterns of accumulated precipitation amount according to both the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (IMERG) data<sup>67</sup> and

the control simulation with 40 model levels. The threshold magnitude for type I deviation time is determined using the rmsd time series following the 50% rule, and this threshold magnitude is then used to estimate deviation times for the individual grid points. For computing the type II deviation times at individual grid points, we use the slope method. This is partly because rmsd is an excellent smoothing tool, hiding the noisy fluctuations that are detectable in the time series of ensemble error dispersion at individual grid points. Any threshold deviation magnitude obtained based on the rmsd results will be met at individual grid points in a matter of a few time steps. The slope method, on the other hand, detects more enduring changes with the time scale of  $\sim 1$  day.



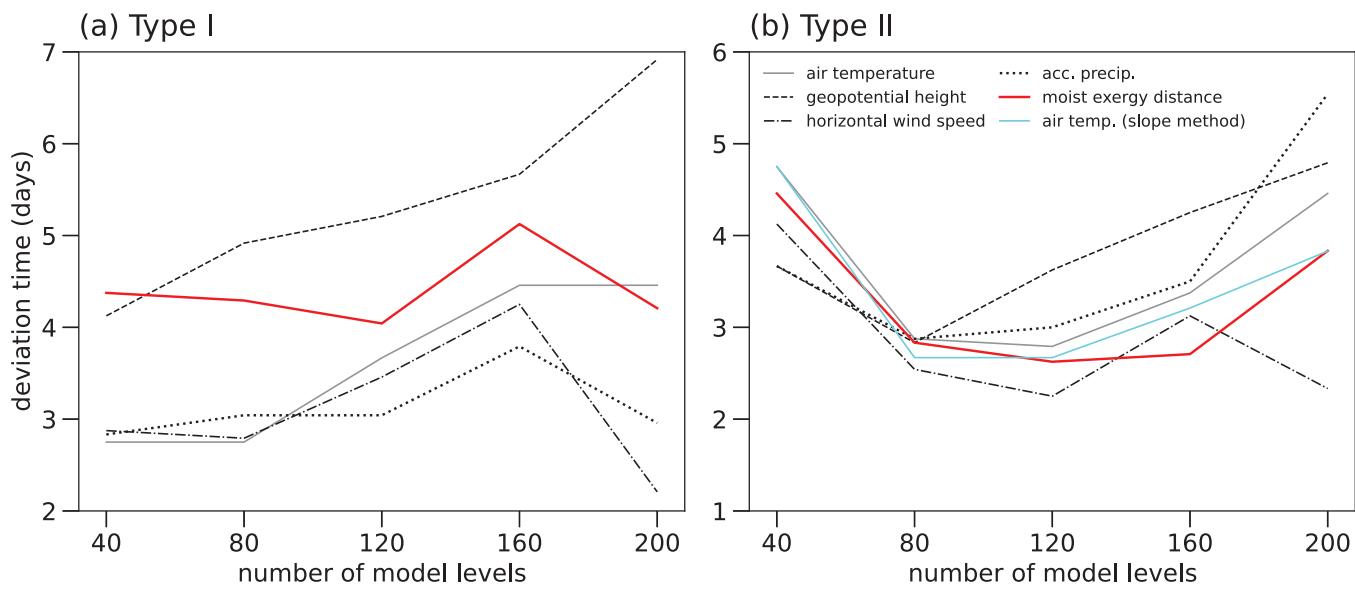
**FIG. 9.** The same as Fig. 6 except accumulated precipitation amount.

A quick glance at Fig. 11 reveals that there exists a weakly negative relationship between accumulated precipitation amount and the deviation times at individual grid points based on the 500-hPa air temperature. For example, a large amount of accumulated precipitation is recorded over the Korean Peninsula and along the Yangtze River in mainland China, which coincides with the area over which some of the shortest deviation times are found. Overall, this relationship seems to be more pronounced over land than over the sea.

The weakly negative relationship is also visible in the scatterplots of deviation times against accumulated precipitation amount in Fig. 12. In fact, the scatterplots reveal that accumulated precipitation amount acts as a limiting factor for deviation time.

A small amount of accumulated precipitation does not always lead to a long deviation time, but a larger amount of accumulated precipitation shortens the longest estimated deviation time. In other words, predictability tends to be poorer where there occurs heavy precipitation, and yet lighter precipitation does not always mean improved predictability because there are other factors that can still limit atmospheric predictability without precipitation getting involved.

Interestingly, having a different vertical resolution tends not to drastically change the total accumulated precipitation amount or the maximum accumulated precipitation amount, which determines the horizontal axis-intercept of the line of suppression. Moreover, an inverse relationship is found between the total and maximum



**FIG. 10.** (a) Type I deviation times for each number of model levels estimated with the 50% threshold magnitude for member-averaged moist exergy distance or rmsd in the 500-hPa air temperature, 500-hPa geopotential height, 500-hPa horizontal wind speed, and accumulated precipitation amount. (b) The same as (a) except type II deviation times estimated with the 5% threshold magnitude for mean absolute errors among the rmsd or moist exergy distances between the perturbed and control members. The additional curve (cyan) represents type II deviation times for the 500-hPa air temperature estimated based on the “slope method” with the threshold slope of  $0.1 \text{ K day}^{-1}$ .

accumulated precipitation amounts. In other words, comparing among the different vertical resolution cases (Table V), the greater the maximum accumulated precipitation is (among all grid points), the smaller the total accumulated precipitation tends to be, implying that the changes largely involve horizontal redistributions of the accumulated precipitation amounts. This is despite that the overall deviation time, regardless of the type, clearly changes with respect to the number of model levels (Fig. 10).

Now, consider the collection of points each corresponding to the maximum accumulated precipitation amount for a discretization of deviation time in one of the scatterplots in Fig. 12. We can imagine the line of best fit for these points dashing through the scatterplot near the diagonal from the upper left corner to the lower right corner. Since this line suppresses the deviation times at individual grid points, we refer to this line as the *line of suppression* (see Appendix C for the computational details). The line of suppression with respect to precipitation amounts is an interesting feature of the spatial distribution of deviation time, and information it contains, such as its slope and axis-intercepts, may turn out to play an important role in unveiling the relationship between the inherent and feature-specific factors (i.e., precipitation patterns) affecting atmospheric predictability. Even as the partial control of the simulated precipitation amounts over atmospheric predictability is largely maintained under different model vertical resolutions, the underlying chaotic dynamics of the model atmosphere leads to changes in the spatial distribution of atmospheric predictability. The spatial distribution of atmospheric predictability, in turn, must be determined by how model vertical resolution interacts with localized features or particular meteorological phenomena characterizing

the case. Analyses making use of ideas such as the line of suppression can lead to a reconciliation between the nonlinear dynamical explanation and feature-specific explanations for non-monotonic dependence of atmospheric predictability on model vertical resolution.

Note that there may also be other approaches with great potential for yielding meaningful results, such as searching for a scaling law in a nonlinear fit for the *boundary points*, that is, the scatter points located around the “boundary” of concentrated clusters to the left vs the relatively sparsely populated area to the right in Fig. 13. In Fig. 13, an alternative nonlinear fit is considered in the form of

$$\tau_d^{\text{boundary}} = A(\text{acc})^\alpha, \quad (3)$$

where  $\tau_d^{\text{boundary}}$  is the deviation time at the “boundary” (i.e., the maximum for a given *acc* excluding outliers),  $A$  is a real coefficient, *acc* stands for accumulated precipitation amount, and  $\alpha$  is a scaling exponent. In addition to the information gained from the lines of suppression, the estimated scaling exponents may provide some additional insight. It is therefore suggested that further research be pursued along these ideas.

## E. Connections to predictive skill

Deviation time is thought to measure a predictability notion that is inherent to a given system. In an operational context, however, predictive skill can be more relevant. How predictability in the sense of deviation time directly translates to predictive skill is unknown, but if the factors holding back the systems from having

**TABLE II.** Type I deviation times (days) when each of the average rmsd in the 500-hPa air temperature, 500-hPa geopotential height, 500-hPa horizontal wind speed, accumulated precipitation amount, and moist exergy distance between the perturbed and control runs exceeds the threshold magnitude given by 30%, 50%, and 70% of the maximum average rmsd corresponding to each model vertical resolution setup. .... marks estimated deviation time exceeding 6 days.

Threshold	Variable	Deviation time				
		40 levels	80 levels	120 levels	160 levels	200 levels
30%	Air temperature	1.63	1.63	1.96	2.29	1.83
	Geopotential height	3.46	3.75	4.42	5.29	5.54
	Horizontal wind speed	1.79	1.75	1.96	2.13	1.46
	Accumulated precipitation	1.67	1.92	2.00	2.42	1.71
	Moist exergy distance	3.63	3.08	3.21	3.79	3.00
50%	Air temperature	2.75	2.75	3.67	4.46	4.46
	Geopotential height	4.13	4.92	5.21	5.67	6.92
	Horizontal wind speed	2.88	2.79	3.46	4.25	2.21
	Accumulated precipitation	2.83	3.04	3.04	3.79	2.96
	Moist exergy distance	4.38	3.08	3.21	3.79	3.00
70%	Air temperature	4.83	4.63	5.04	...	...
	Geopotential height	5.29	5.38	5.71	...	...
	Horizontal wind speed	4.42	4.42	4.46	5.54	3.79
	Accumulated precipitation	4.17	4.79	4.17	5.38	4.46
	Moist exergy distance	5.38	5.71	5.21	5.63	5.83

**TABLE III.** Type II deviation times (days) when each of the mean absolute differences among the rmsd in the 500-hPa air temperature, 500-hPa geopotential height, 500-hPa horizontal wind speed, accumulated precipitation amount, and moist exergy distance between the perturbed and control runs exceeds the threshold percentages of 3%, 5%, 7%, and 10% of the maximum average rmsd. .... marks estimated deviation time exceeding 6 days.

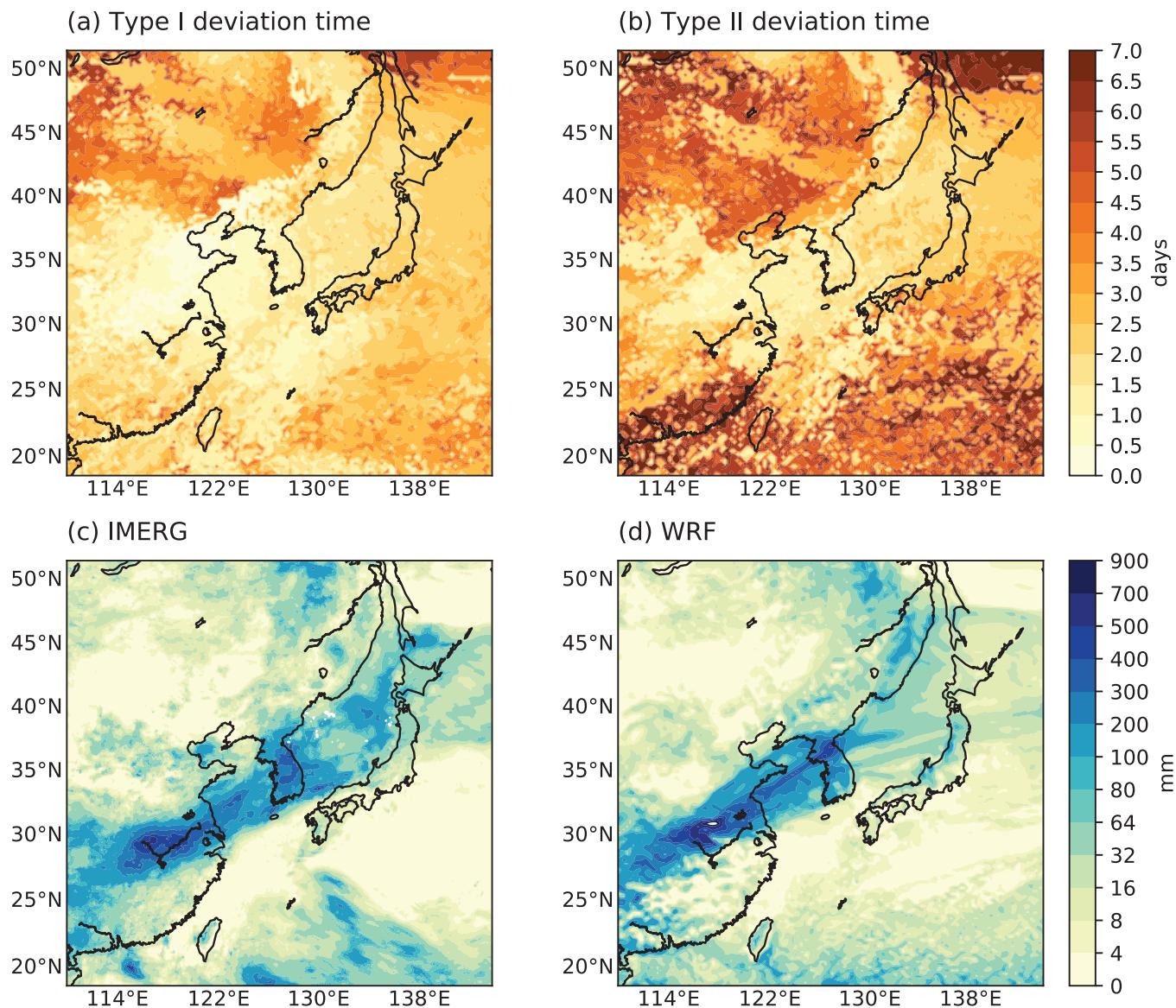
Threshold	Variable	Deviation time				
		40 levels	80 levels	120 levels	160 levels	200 levels
3%	Air temperature	2.71	2.08	2.71	3.17	3.75
	Geopotential height	2.79	2.75	3.00	4.08	3.71
	Horizontal wind speed	2.46	2.21	1.96	2.29	1.71
	Accumulated precipitation	2.79	1.88	2.58	2.21	2.58
	Moist exergy distance	4.13	2.46	2.42	2.50	3.33
5%	Air temperature	4.75	2.88	2.79	3.38	4.46
	Geopotential height	3.67	2.83	3.63	4.25	4.79
	Horizontal wind speed	4.13	2.54	2.25	3.13	2.33
	Accumulated precipitation	3.67	2.88	3.00	3.50	5.54
	Moist exergy distance	4.46	2.83	2.63	2.71	3.83
7%	Air temperature	5.33	2.88	2.92	...	...
	Geopotential height	3.75	3.00	3.75	4.33	5.42
	Horizontal wind speed	4.25	2.63	2.58	3.17	3.46
	Accumulated precipitation	3.75	...	3.21	5.75	5.75
	Moist exergy distance	4.63	2.96	2.83	3.04	5.38
10%	Air temperature	...	3.42	...	...	...
	Geopotential height	3.88	5.54	4.96	5.25	5.75
	Horizontal wind speed	5.08	3.04	2.92	5.75	3.50
	Accumulated precipitation	3.88	...	3.71	...	...
	Moist exergy distance	5.21	3.21	3.29	3.38	5.63

infinite predictability and infinite predictive skill overlap to some degrees, it may be reasonable to expect some correlation between these two measurements. In particular, the non-monotonic dependence of predictability on model resolution observed through deviation time can influence the overall model conditions, leading to a non-monotonic dependence of predictive skill on model resolution.

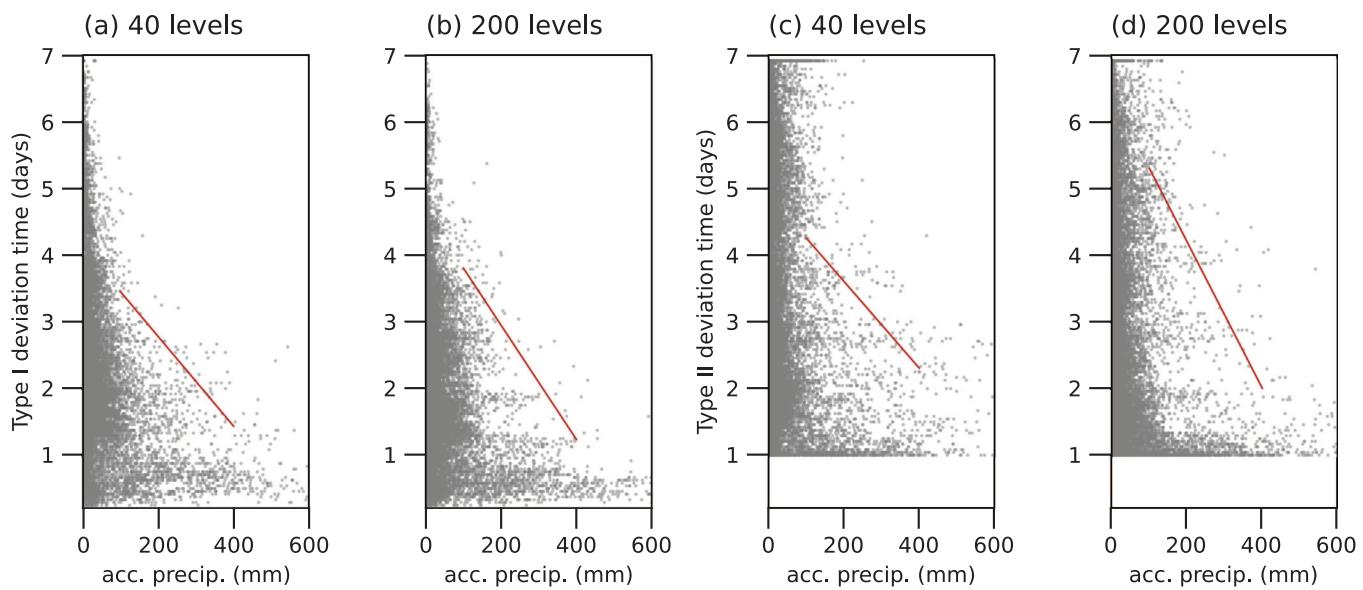
Traditionally, a model's predictive skill is gauged through lead time experiments,<sup>32,68</sup> employing a measure of differences appropriate for the target variable, such as the equitable threat score (ETS) for precipitation.<sup>44,69,70</sup> Underlying these experiments is the

assumption that the target simulations with the shortest lead time possess the best prediction skills, which may or may not hold in practice. See Appendix D for details on the calculations of ETS.

Figure 14 shows the ETS for lead time experiments targeting the spatial distribution of accumulated precipitation amount in the last 24 h of the time domain over the analysis area simulated by the model using the same model vertical resolution configuration but with the lead time of 1 day. The lead time ensembles are produced by increasing the lead time by one day at a time up to 7 days covering the entire time domain. The differences in the 24-h



**FIG. 11.** Spatial distribution of (a) type I deviation times and (b) type II deviation times at individual grid points based on the 500-hPa air temperature in simulations with 40 model levels. Spatial distribution of accumulated precipitation amount for the duration of the entire time domain (c) according to the IMERG data and (d) based on the control simulation with 40 model levels.



**FIG. 12.** Scatterplots of type I deviation times at individual grid points based on the 500-hPa air temperature and the 50% threshold magnitude against accumulated precipitation amount in simulations with (a) 40 and (b) 200 model levels. (c)–(d) The same as (a)–(b) except type II deviation times at individual grid points estimated using the slope method with a  $0.1 \text{ K day}^{-1}$  threshold slope. Note that since the slope method requires a full 1-day-long interval, the estimated type II deviation times start at 1. The straight lines (red) are the approximate lines of suppression.

accumulated precipitation amount patterns between the target and other model runs with longer lead times are measured by the ETS.

The lead time experiments confirm that, at least in this particular precipitation case, the WRF model simulations with at least 40 model levels exhibit non-monotonic dependence of atmospheric predictability on model vertical resolution. For instance, given the threshold ETS of 0.3, the simulations with 200 levels exceed this threshold with the longest lead time as expected, but this is followed by the simulations with 40 levels, the coarsest vertical resolution configuration considered in this study. In all simulations, the chosen threshold ETS is met some time between 4 and 5 days of lead time. Evidently, this time window containing the threshold-exceedance lead times and also the ranking among the simulations with different numbers of model levels change with the choice of threshold ETS.

**TABLE IV.** Type II deviation times (days) based on the 500-hPa air temperature in simulations with varying number of model levels, estimated using the slope method with different threshold slopes.

Threshold slope ( $0.1 \text{ K day}^{-1}$ )	Deviation time				
	40 levels	80 levels	120 levels	160 levels	200 levels
0.8	3.25	2.13	2.67	3.21	3.83
1.0	4.75	2.67	2.67	3.21	3.83
1.2	4.83	2.88	2.75	3.25	3.83
1.4	5.00	2.88	2.79	3.25	3.92

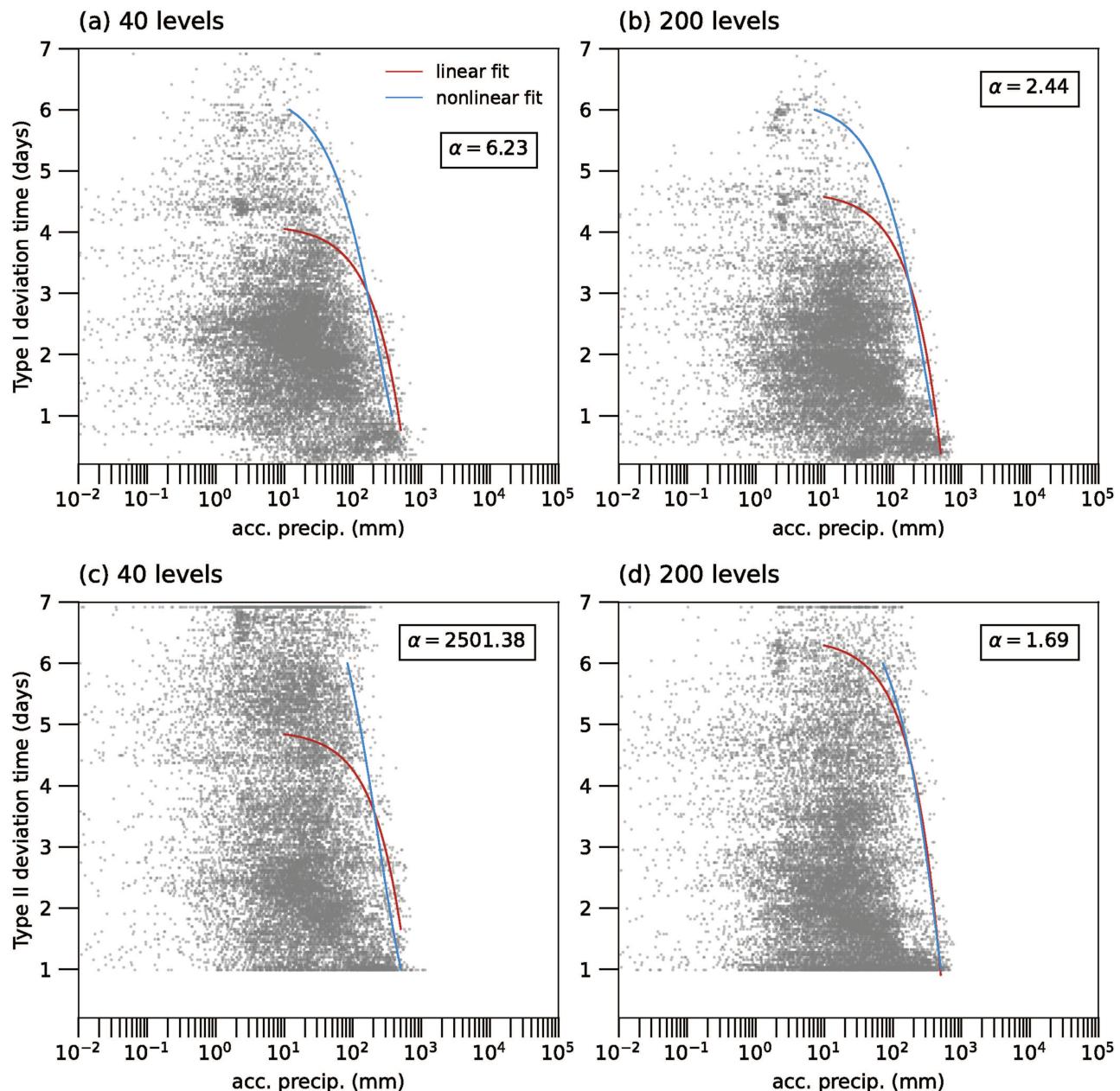
#### IV. DISCUSSION

A few years after the publication<sup>7</sup> of his groundbreaking research on chaotic dynamics and why it might suggest the existence of an inherent limit of atmospheric predictability, Lorenz<sup>8</sup> showed that the predictability limit can also be a consequence of the rapid upscale propagation of small errors overshadowing any benefit from the enhanced accuracy of the model resolving smaller-scale motions. These two ideas are indeed related to some extent, but they were often misleadingly fused into a single concept summarily referred to as the butterfly effect, causing a great deal of confusion among posterity,<sup>9</sup> however, with the advent of ensemble forecasting as well as time-averaging of forecast outputs, the theoretical limit suggested by L69 has been overcome.<sup>71</sup>

In this study, we reiterate the argument that sensitive dependence on initial conditions, a defining characteristic of chaos as well as the theoretical foundation for the practice of ensemble weather forecasting, captures something different and perhaps more

**TABLE V.** Total accumulated precipitation amount (mm) as well as the maximum over the analysis area in the control simulations each with 40, 80, 120, 160, and 200 model levels.

	Accumulated precipitation				
	40 levels	80 levels	120 levels	160 levels	200 levels
Total	38.5	40.0	40.8	40.4	40.6
Max	1123.7	796.5	679.9	695.1	752.4

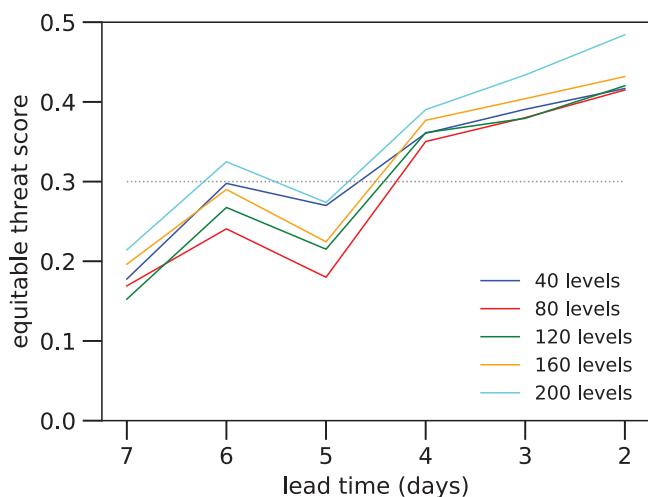


**FIG. 13.** The same as Fig. 12 except the horizontal axis is in log-scale. In addition, the lines of suppression (red) from Fig. 12 are compared against the curves obtained using a nonlinear fit (blue). The  $\alpha$  value in each subplot indicates the scaling exponent for the nonlinear fit. The corresponding values of coefficient  $A$  [from Eq. (3)] are (a)  $-9.26 \times 10^{-4}$ , (b)  $-2.90 \times 10^{-3}$ , (c)  $-1.71 \times 10^{-6}$ , and (d)  $-4.36 \times 10^{-3}$ .

fundamental about weather than what are afforded by the scale-interaction explanations alone.<sup>72,73</sup> Accordingly, deviation time has been put forward as a measure of predictability more strictly based on sensitive dependence on initial conditions.

We were able to identify a non-monotonic dependence of atmospheric predictability on model vertical resolution in the

ensemble experiments using both a simplified model of Lorenz-type and the more realistic WRF model. For the Lorenz system experiments, the dimensionality of the system was associated with its vertical resolution as a model for thermal convection.<sup>11</sup> In the WRF model experiments, the model vertical resolution was controlled by the number of model levels used in each simulation. Such



**FIG. 14.** Equitable threat scores (ETs) targeting accumulated precipitation amount in the last 24 h of the time domain in simulations with varying lead time durations and model levels. An example threshold ETS of 0.3 is indicated by the dotted line.

an observation can hold grave implications for the future of NWP and ensemble forecasting. Buizza<sup>74</sup> notes that future developments in ensemble forecasting will entail a move toward higher resolution as well as an integrated approach to medium-range, sub-seasonal, and seasonal ensembles aimed at improving the medium-range predictability. It is widely acknowledged that extending the medium-range predictability even by a few days of lead time can bring about tremendous societal benefits.<sup>75</sup> The non-monotonic dependence of predictability on model resolution demonstrated using deviation time in this study and the patterns seen in predictive skill in agreement with the deviation time results may suggest that future developments in ensemble forecasting can be more challenging than how it is currently envisioned.

A robust confirmation of the aforementioned ideas requires further in-depth study. It should be noted that typical efforts to improve model resolutions entail refining the horizontal resolution or simultaneous refinements of both the horizontal and vertical resolutions. Since part of the motivation here has been to expand upon and present a real-case analog of the findings from the twin experiments using the generalized Lorenz systems in Moon *et al.*,<sup>11</sup> the analysis and discussions in this study have largely focused on the dependence of predictability on the vertical resolution only. Incorporating higher wavenumber modes in the horizontal direction in the derivation of higher-dimensional Lorenz systems is doable, but a simultaneous incorporation of both horizontal and vertical wavenumber modes may prove challenging, if not impossible, if the aim is to form a Lorenz-like system of ordinary differential equations.

Another concern is how to approach the disconnect between the results in the generalized Lorenz systems and those in the WRF model outputs. Numerical experiments using the generalized Lorenz systems had been the motivation for applying the deviation time

idea to a realistic model atmosphere, but it should also be acknowledged that the Lorenz systems, even at the level of complexity of the high-dimensional generalizations, are far too removed from being a realistic model of the atmosphere. It is therefore suggested that a future study of atmospheric predictability based on deviation times consider an additional “middle-of-the-road” model, such as a model for direct numerical simulation (DNS)<sup>76</sup> or the WRF model in an idealized setting.

In this study, we have only analyzed the simulations of a particular summer precipitation case over East Asia; however, there are many different cases that can be considered (e.g., different types of precipitation events found in this region, such as those involving typhoons). It would be interesting to see if we can find any pattern to deviation time across the numerical simulations of other cases featuring different types of precipitation events. Such an investigation can further elucidate connections between the atmospheric predictability measured through deviation time and that influenced by specific features of the model atmosphere. While the general principles laid out in the work should still be applicable in the broad sense, we further note that the ensemble experiments were carried out under a particular configuration and a fixed set of choices for the physics parameterizations in the WRF model, wherefore using a different model or model setup can lead to different model behavior. For example, a few studies, including On *et al.*,<sup>77</sup> report that the selection of cumulus parameterization scheme can influence performance outcomes as much as the model resolution does. A broadly configured meta-analysis employing a diverse set of models and model configurations for testing the deviation time concept may lead to a better understanding of the vertical resolution problem in the context of atmospheric predictability.

## ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the two anonymous reviewers who provided highly detailed comments and feedback on this manuscript. This work was supported by the Small Grant for Exploratory Research (SGER) program through the National Research Foundation of Korea (No. NRF-2018R1D1A1A02086007) and the Basic Science Research Program through the National Research Foundation of Korea (No. NRF-2021R1A6A3A01086440).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Sungju Moon:** Conceptualization (equal); Data curation (lead); Formal analysis (lead); Funding acquisition (supporting); Investigation (lead); Methodology (equal); Project administration (supporting); Visualization (lead); Writing – original draft (lead); Writing – review and editing (supporting). **Jong-Jin Baik:** Conceptualization (equal); Funding acquisition (lead); Investigation (supporting); Methodology (equal); Project administration (lead); Resources

(lead); Supervision (lead); Writing – review and editing (supporting). **Hyo-Jong Song:** Conceptualization (equal); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Visualization (supporting); Writing – review and editing (supporting). **Ji-Young Han:** Conceptualization (equal); Methodology (equal); Visualization (supporting); Writing – review and editing (supporting).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX A: THE LORENZ SYSTEMS OF DIMENSIONS 3 AND 29

The Lorenz system<sup>7</sup> consists of the following three ordinary differential equations in variables  $X_1, Y_1, Z_1$ :

$$\dot{X}_1 = -d_1\sigma X_1 + \frac{\sigma}{d_1} Y_1, \quad (\text{A1})$$

$$\dot{Y}_1 = rX_1 - d_1 Y_1 - X_1 Z_1, \quad (\text{A2})$$

$$\dot{Z}_1 = -bZ_1 + X_1 Y_1, \quad (\text{A3})$$

where the dots indicate the first derivative with respect to time  $\tau$ ,  $b$  is a geometric parameter, and  $d_k$  is defined for all positive integer  $k$  such that

$$d_k = \frac{(2k-1)^2 + a^2}{1 + a^2}, \quad (\text{A4})$$

where  $a$  is usually set as  $1/\sqrt{2}$ . The equations for the 29-dimensional extension of the Lorenz system following the generalization by Moon *et al.*<sup>11</sup> read

$$\dot{X}_k = -d_k\sigma X_k + \frac{\sigma}{d_k} Y_k, \quad (\text{A5})$$

$$\dot{Y}_k = rX_k - d_k Y_k + \sum_{(i,j) \in P_k} (jX_i Z_j \text{sgn}_Y), \quad (\text{A6})$$

$$\dot{Z}_k = -k^2 b Z_k + \sum_{(i,j) \in Q_k} (kX_i Y_j \text{sgn}_Z) \quad (\text{A7})$$

for  $k = 1, 2, \dots, 9$  and

$$\dot{Y}_{10} = -d_{10} Y_{10} + \sum_{(i,j) \in P_{10}} (jX_i Z_j \text{sgn}_Y), \quad (\text{A8})$$

$$\dot{Z}_{10} = -10^2 b Z_{10} + \sum_{(i,j) \in Q_{10}} (10X_i Y_j \text{sgn}_Z), \quad (\text{A9})$$

where

$$P_k = \{(i,j) : i + j = k\} \cup \{(i,j) : j - i = k - 1\} \cup \{(i,j) : i - j = k\}, \quad (\text{A10})$$

$$Q_k = \{(i,j) : |j - i| = k\} \cup \{(i,j) : j + i = k + 1\}, \quad i \text{ and } j \text{ integers}, \quad (\text{A11})$$

and

$$\text{sgn}_Y = \begin{cases} -1 & \text{if } j - i = k - 1, \\ 1 & \text{otherwise,} \end{cases} \quad (\text{A12})$$

$$\text{sgn}_Z = \begin{cases} -1 & \text{if } |j - i| = k, \\ 1 & \text{otherwise.} \end{cases} \quad (\text{A13})$$

## APPENDIX B: COMPUTATION OF MOIST EXERGY DISTANCE

Following Marquet *et al.*,<sup>65</sup> moist exergy distance is made up of four quadratic components integrated over the whole domain (or analysis area),

$$N_{\text{KE}} = \iiint \frac{(u')^2 + (v')^2}{2} \frac{dm}{\Sigma}, \quad (\text{B1})$$

$$N_T = \iiint \left[ \frac{C_{p_d} T_r}{\bar{T}^2} \right] \frac{(T')^2}{2} \frac{dm}{\Sigma}, \quad (\text{B2})$$

$$N_p = \iiint \left[ \frac{R_d T_r}{\bar{p}_s^2} \right] \frac{(p_s')^2}{2} \frac{dm}{\Sigma}, \quad (\text{B3})$$

$$N_{\text{vapor}} = \iiint \left[ \frac{R_v T_r}{\bar{r}_v} \right] \frac{(r_v')^2}{2} \frac{dm}{\Sigma}, \quad (\text{B4})$$

where the primed quantities  $u', v', T', p_s',$  and  $r_v'$  are departures from the control zonal wind velocity  $u$ , meridional wind velocity  $v$ , air temperature  $T$ , surface pressure  $p_s$ , and water vapor mixing ratio  $r_v$ , respectively.  $C_{p_d}$  is the specific heat of dry air,  $R_d$  is the dry-air gas constant,  $R_v$  is the water vapor gas constant, and  $T_r$  is the reference temperature taken to be 300 K following Borderies *et al.*<sup>66</sup> The horizontal averages that vary with height are denoted with an overline. The differential mass  $dm$  is divided by the horizontal surface area  $\Sigma$  so that volume integrals over  $dm/\Sigma$  represent energy per unit horizontal area.

## APPENDIX C: ESTIMATION OF THE LINES OF SUPPRESSION

In what follows, the *ad hoc* method used to estimate the lines of suppression in Fig. 12 is described. For each deviation time along the vertical axis of the scatterplot in Fig. 12 (ranging from 0 to 7 days for type I and from 1 to 7 days for type II), the *boundary point*—the point with the greatest accumulated precipitation amount excluding the *out-of-boundary points*—is identified. The out-of-boundary points are considered to be any point to the right of the diagonal line from the top left corner to the bottom right corner of a plot window. In each subplot in Fig. 12, a straight line connecting the boundary points defines the boundary between the area to the left with highly concentrated clusters of scatters and the area outside of this to the right. Once the boundary points are identified, a line fitting algorithm, such as `numpy.polyfit`, can be used to find the slope

and axis-intercepts of the line of suppression. It should be noted that there are other approaches that may also be helpful in understanding the spatial distribution of deviation times with respect to accumulated precipitation amounts (e.g., the utilization of a nonlinear fit; see also Fig. 13 and accompanying discussions).

## APPENDIX D: CALCULATION OF ETS

The equitable threat score (ETS) is computed following the updated calculation scheme proposed by Wang<sup>70</sup> as follows:

$$\text{ETS} = \frac{H - R}{O + F - H - R}, \quad (\text{D1})$$

where  $O$  is the number of grid points at which the 24-h accumulated precipitation amount exceeds the given threshold in the target simulation with 1-day-long lead time (target area),  $F$  is the number of grid points exceeding the threshold in ensemble member simulations with different lead times (forecast area),  $H$  is the number of grid points in the intersection of the area corresponding to  $O$  and the area corresponding to  $F$ , and  $R = FO/M$ , where  $M$  is the total number of grid points in the model domain.

## REFERENCES

- <sup>1</sup>S. Park, “A unified convection scheme (UNICON). Part I: Formulation,” *J. Atmos. Sci.* **71**, 3902–3930 (2014).
- <sup>2</sup>T. Ahmed, H.-G. Jin, and J.-J. Baik, “A physically based raindrop–cloud droplet accretion parametrization for use in bulk microphysics schemes,” *Q. J. R. Meteorol. Soc.* **146**, 3368–3383 (2020).
- <sup>3</sup>E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, 2003).
- <sup>4</sup>P. Bauer, A. Thorpe, and G. Brunet, “The quiet revolution of numerical weather prediction,” *Nature* **525**, 47–55 (2015).
- <sup>5</sup>S. G. Benjamin, J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter, “100 years of progress in forecasting and NWP applications,” *Meteorol. Monogr.* **59**, 13.1–13.67 (2019).
- <sup>6</sup>F. Zhang, Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, “What is the predictability limit of midlatitude weather?,” *J. Atmos. Sci.* **76**, 1077–1091 (2019).
- <sup>7</sup>E. N. Lorenz, “Deterministic nonperiodic flow,” *J. Atmos. Sci.* **20**, 130–141 (1963).
- <sup>8</sup>E. N. Lorenz, “The predictability of a flow which possesses many scales of motion,” *Tellus* **21**, 289–307 (1969).
- <sup>9</sup>T. N. Palmer, A. Döring, and G. Seregin, “The real butterfly effect,” *Nonlinearity* **27**, R123–R141 (2014).
- <sup>10</sup>E. N. Lorenz, “Atmospheric predictability as revealed by naturally occurring analogues,” *J. Atmos. Sci.* **26**, 636–646 (1969).
- <sup>11</sup>S. Moon, J. M. Seo, and J.-J. Baik, “High-dimensional generalizations of the Lorenz system and implications for predictability,” *Phys. Scr.* **95**, 085209 (2020).
- <sup>12</sup>A. N. Kolmogorov, “The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers,” *Dokl. Akad. Nauk SSSR* **30**, 299–303 (1941).
- <sup>13</sup>R. H. Kraichnan, “Inertial ranges in two-dimensional turbulence,” *Phys. Fluids* **10**, 1417 (1967).
- <sup>14</sup>J. J. Tribbia and D. P. Baumhefner, “Scale interactions and atmospheric predictability: An updated perspective,” *Mon. Weather Rev.* **132**, 703–713 (2004).
- <sup>15</sup>G. D. Nastrom and K. S. Gage, “A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft,” *J. Atmos. Sci.* **42**, 950–960 (1985).
- <sup>16</sup>J. A. Weyn and D. R. Durran, “Ensemble spread grows more rapidly in higher-resolution simulations of deep convection,” *J. Atmos. Sci.* **75**, 3331–3345 (2018).
- <sup>17</sup>T. Selz, “Estimating the intrinsic limit of predictability using a stochastic convection scheme,” *J. Atmos. Sci.* **76**, 757–765 (2019).
- <sup>18</sup>R. Rotunno and C. Snyder, “A generalization of Lorenz’s model for the predictability of flows with many scales of motion,” *J. Atmos. Sci.* **65**, 1063–1076 (2008).
- <sup>19</sup>S. Brune and E. Becker, “Indications of stratified turbulence in a mechanistic GCM,” *J. Atmos. Sci.* **70**, 231–247 (2013).
- <sup>20</sup>D. R. Durran and M. Gingrich, “Atmospheric predictability: Why butterflies are not of practical importance,” *J. Atmos. Sci.* **71**, 2476–2488 (2014).
- <sup>21</sup>M. J. P. Cullen, “The impact of high vertical resolution in the Met Office Unified Model,” *Q. J. R. Meteorol. Soc.* **143**, 278–287 (2017).
- <sup>22</sup>T. Y. Leung, M. Leutbecher, S. Reich, and T. G. Shepherd, “Atmospheric predictability: Revisiting the inherent finite-time barrier,” *J. Atmos. Sci.* **76**, 3883–3892 (2019).
- <sup>23</sup>Y. Q. Sun and F. Zhang, “A new theoretical framework for understanding multiscale atmospheric predictability,” *J. Atmos. Sci.* **77**, 2297–2309 (2020).
- <sup>24</sup>J. H. Curry, “A generalized Lorenz system,” *Commun. Math. Phys.* **60**, 193–204 (1978).
- <sup>25</sup>P. Reiterer, C. Lainscsek, F. Schürrer, C. Letellier, and J. Maquet, “A nine-dimensional Lorenz system to study high-dimensional chaos,” *J. Phys. A: Math. Gen.* **31**, 7121–7139 (1998).
- <sup>26</sup>B.-W. Shen, “Nonlinear feedback in a five-dimensional Lorenz model,” *J. Atmos. Sci.* **71**, 1701–1723 (2014).
- <sup>27</sup>D. Adamec, “Predictability of quasi-geostrophic ocean flow: Sensitivity to varying model vertical resolution,” *J. Phys. Oceanogr.* **19**, 1753–1764 (1989).
- <sup>28</sup>J. Harlim, M. Oczkowski, J. A. Yorke, E. Kalnay, and B. R. Hunt, “Convex error growth patterns in a global weather model,” *Phys. Rev. Lett.* **94**, 228501 (2005).
- <sup>29</sup>F. Judt, “Insights into atmospheric predictability through global convection-permitting model simulations,” *J. Atmos. Sci.* **75**, 1477–1497 (2018).
- <sup>30</sup>J. M. Lewis, “Roots of ensemble forecasting,” *Mon. Weather Rev.* **133**, 1865–1885 (2005).
- <sup>31</sup>S. Yoden, “Atmospheric predictability,” *J. Meteorol. Soc. Jpn.* **85B**, 77–102 (2007).
- <sup>32</sup>F. Zhang, C. Snyder, and R. Rotunno, “Effects of moist convection on mesoscale predictability,” *J. Atmos. Sci.* **60**, 1173–1185 (2003).
- <sup>33</sup>R. Buizza, “Horizontal resolution impact on short- and long-range forecast error,” *Q. J. R. Meteorol. Soc.* **136**, 1020–1035 (2010).
- <sup>34</sup>C. Hohenegger and C. Schär, “Atmospheric predictability at synoptic versus cloud-resolving scales,” *Bull. Am. Meteorol. Soc.* **88**, 1783–1794 (2007).
- <sup>35</sup>R. Mureau, F. Molteni, and T. N. Palmer, “Ensemble prediction using dynamically conditioned perturbations,” *Q. J. R. Meteorol. Soc.* **119**, 299–323 (1993).
- <sup>36</sup>Z. Toth and E. Kalnay, “Ensemble forecasting at NCEP and the breeding method,” *Mon. Weather Rev.* **125**, 3297–3319 (1997).
- <sup>37</sup>D. J. Stensrud, J.-W. Bao, and T. T. Warner, “Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems,” *Mon. Weather Rev.* **128**, 2077–2107 (2000).
- <sup>38</sup>A. Atencia and I. Zawadzki, “Analogs on the Lorenz attractor and ensemble spread,” *Mon. Weather Rev.* **145**, 1381–1400 (2017).
- <sup>39</sup>W. C. Skamarock and S.-H. Park, “Vertical resolution requirements in atmospheric simulation,” *Mon. Weather Rev.* **147**, 2641–2656 (2019).
- <sup>40</sup>A. A. Scaife, J. Camp, R. Comer, P. Davis, N. Dunstone, M. Gordon, C. MacLachlan, N. Martin, Y. Nie, H.-L. Ren, M. Roberts, W. Robinson, D. Smith, and P. L. Vidale, “Does increased atmospheric resolution improve seasonal climate predictions?,” *Atmos. Sci. Lett.* **20**, e922 (2019).
- <sup>41</sup>M. L. Waite, “Dependence of model energy spectra on vertical resolution,” *Mon. Weather Rev.* **144**, 1407–1421 (2016).
- <sup>42</sup>G. A. Meehl, L. Goddard, G. Boer, R. Burgman, G. Branstator, C. Cassou, S. Corti, G. Danabasoglu, F. Doublas-Reyes, E. Hawkins, A. Karspeck, M. Kimoto, A. Kumar, D. Matei, J. Mignot, R. Msadek, A. Navarra, H. Phlmann, M. Rienerker, T. Rosati, E. Schneider, D. Smith, R. Sutton, H. Teng, G. J. Van Oldenborgh, G. Vecchi, and S. Yeager, “Decadal climate prediction: An update from the trenches,” *Bull. Am. Meteorol. Soc.* **95**, 243–267 (2021).
- <sup>43</sup>Y. Q. Sun and F. Zhang, “Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect,” *J. Atmos. Sci.* **73**, 1419–1438 (2016).

- <sup>44</sup>E. A. Aligo and W. A. Gallus, Jr., "On the impact of WRF model vertical grid resolution on midwest summer rainfall forecasts," *Weather Forecast* **24**, 575–594 (2009).
- <sup>45</sup>B. C. Ancell and C. F. Mass, "Structure, growth rates, and tangent linear accuracy of adjoint sensitivities with respect to horizontal and vertical resolution," *Mon. Weather Rev.* **134**, 2971–2988 (2006).
- <sup>46</sup>D.-L. Zhang and X. Wang, "Dependence of hurricane intensity and structures on vertical resolution and time-step size," *Adv. Atmos. Sci.* **20**, 711–725 (2003).
- <sup>47</sup>S. K. Kimball and F. C. Dougherty, "The sensitivity of idealized hurricane structure and development to the distribution of vertical levels in MM5," *Mon. Weather Rev.* **134**, 1987–2008 (2006).
- <sup>48</sup>R. S. Lindzen and M. Fox-Rabinovitz, "Consistent vertical and horizontal resolution," *Mon. Weather Rev.* **117**, 2575–2583 (1989).
- <sup>49</sup>M. J. Pecnick and D. Keyser, "The effect of spatial resolution on the simulation of upper-tropospheric frontogenesis using a sigma-coordinate primitive equation model," *Meteorol. Atmos. Phys.* **40**, 137–149 (1989).
- <sup>50</sup>P. G. Persson and T. T. Warner, "Model generation of spurious gravity waves due to inconsistency of the vertical and horizontal resolution," *Mon. Weather Rev.* **119**, 917–935 (1991).
- <sup>51</sup>W. C. Skamarock, "Evaluating mesoscale NWP models using kinetic energy spectra," *Mon. Weather Rev.* **132**, 3019–3032 (2004).
- <sup>52</sup>L. De Cruz, S. Schubert, J. Demaeyer, V. Lucarini, and S. Vannitsem, "Exploring the Lyapunov instability properties of high-dimensional atmospheric and climatic models," *Nonlinear Process. Geophys.* **25**, 387–412 (2018).
- <sup>53</sup>C. Stegemann, H. A. Albuquerque, R. M. Rubinger, and P. C. Rech, "Lyapunov exponent diagrams of a 4-dimensional Chua system," *Chaos* **21**, 033105 (2011).
- <sup>54</sup>E. N. Lorenz, "Predictability—A problem partly solved," in *Predictability of Weather and Climate*, edited by T. Palmer and R. Hagedorn (Cambridge University Press, Cambridge, 2006), Chap. 3, pp. 40–58.
- <sup>55</sup>T. Bódai, "Predictability of threshold exceedances in dynamical systems," *Physica D* **313**, 37–50 (2015).
- <sup>56</sup>X. Zeng, R. Eykholt, and R. A. Pielke, "Estimating the Lyapunov-exponent spectrum from short time series of low precision," *Phys. Rev. Lett.* **66**, 3229–3232 (1991).
- <sup>57</sup>W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, Z. Liu, J. Berner, W. Wang, J. G. Powers, M. G. Duda, D. M. Barker, and X.-Y. Huang, "A description of the advanced research WRF version 4," NCAR Technical Notes NCAR/TN-556+STR (2019), p. 145.
- <sup>58</sup>N. Bei and F. Zhang, "Impacts of initial condition errors on mesoscale predictability of heavy precipitation along the Mei-Yu front of China," *Q. J. R. Meteorol. Soc.* **133**, 83–99 (2007).
- <sup>59</sup>E. Lee, E.-H. Lee, and I.-J. Choi, "Impact of increased vertical resolution on medium-range forecasts in a global atmospheric model," *Mon. Weather Rev.* **147**, 4091–4106 (2019).
- <sup>60</sup>See <https://www.ecmwf.int/en/forecasts/documentation-and-support> for ECMWF, "L137 Model Level Definitions" (2021); accessed 24 March 2021.
- <sup>61</sup>See <https://cds.climate.copernicus.eu/cdsapp#!/home> for Copernicus Climate Change Service (C3S), "ERA5: Fifth Generation of ECMWF Atmospheric Reanalysis of the Global Climate" (Copernicus Climate Change Service Climate Data Store (CDS), 2017); accessed 20 December 2019.
- <sup>62</sup>C. Hohenegger and C. Schär, "Predictability and error growth dynamics in cloud-resolving models," *J. Atmos. Sci.* **64**, 4467–4478 (2007).
- <sup>63</sup>J. Blázquez, N. L. Pessacq, and P. L. M. Gonzalez, "Simulation of a baroclinic wave with the WRF regional model: Sensitivity to the initial conditions in an ideal and a real experiment," *Meteorol. Appl.* **20**, 447–456 (2013).
- <sup>64</sup>L. S. R. Froude, L. Bengtsson, and K. I. Hodges, "Atmospheric predictability revisited," *Tellus* **65A**, 19022 (2013).
- <sup>65</sup>P. Marquet, J.-F. Mahfouf, and D. Holdaway, "Definition of the moist-air exergy norm: A comparison with existing "moist energy norms"," *Mon. Weather Rev.* **148**, 907–928 (2020).
- <sup>66</sup>M. Borderies, O. Caumont, J. Delanoë, V. Ducrocq, N. Fourrié, and P. Marquet, "Impact of airborne cloud radar reflectivity data assimilation on kilometre-scale numerical weather prediction analyses and forecasts of heavy precipitation events," *Nat. Hazards Earth Syst. Sci.* **19**, 907–926 (2019).
- <sup>67</sup>G. J. Huffman, E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and J. Tan, "GPM IMERG Final Precipitation L3 Half Hourly 0.1 Degree × 0.1 Degree V06" (Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, 2019); accessed 29 January 2021. 10.5067/GPM/IMERG/3B-HH/06
- <sup>68</sup>E. N. Lorenz, "Atmospheric predictability experiments with a large numerical model," *Tellus* **34**, 505–513 (1982).
- <sup>69</sup>J. T. Schaefer, "The critical success index as an indicator of warning skill," *Weather Forecast* **5**, 570–575 (1990).
- <sup>70</sup>C.-C. Wang, "On the calculation and correction of equitable threat score for model quantitative precipitation forecasts for small verification areas: The example of Taiwan," *Weather Forecast* **29**, 788–798 (2014).
- <sup>71</sup>C. M. Wheeler, H. Zu, A. H. Sobel, D. Hudson, and F. Vitart, "Seamless precipitation prediction skill comparison between two global models," *Q. J. R. Meteorol. Soc.* **143**, 374–383 (2017).
- <sup>72</sup>A. A. Tsonis and J. B. Elsner, "Chaos, strange attractors, and weather," *Bull. Am. Meteorol. Soc.* **70**, 14–23 (1989).
- <sup>73</sup>D. Schertzer and S. Lovejoy, "Uncertainty and predictability in geophysics: Chaos and multifractal insights," in *The State of the Planet: Frontiers and Challenges in Geophysics*, Geophysical Monograph Series Vol. 150, edited by R. S. J. Sparks and C. J. Hawkesworth (American Geophysical Union, Washington, DC, 2004), pp. 317–334.
- <sup>74</sup>R. Buizza, "Introduction to the special issue on "25 years of ensemble forecasting"," *Q. J. R. Meteorol. Soc.* **145**, 1–11 (2019).
- <sup>75</sup>D. A. Lavers, F. Pappenberger, and E. Zsoter, "Extending medium-range predictability of extreme hydrological events in Europe," *Nat. Commun.* **5**, 5382 (2019).
- <sup>76</sup>J. Park, S. Moon, J. M. Seo, and J.-J. Baik, "Systematic comparison between the generalized Lorenz equations and DNS in the two-dimensional Rayleigh–Bénard convection," *Chaos* **31**, 073119 (2021).
- <sup>77</sup>N. On, H. M. Kim, and S. Kim, "Effects of resolution, cumulus parameterization scheme, and probability forecasting on precipitation forecasts in a high-resolution limited-area ensemble prediction system," *Asia-Pac. J. Atmos. Sci.* **54**, 623–637 (2018).
- <sup>78</sup>S.-Y. Hong, Y. Noh, and J. Dudhia, "A new vertical diffusion package with an explicit treatment of entrainment processes," *Mon. Weather Rev.* **134**, 2318–2341 (2006).
- <sup>79</sup>J. Dudhia, "Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model," *J. Atmos. Sci.* **46**, 3077–3107 (1989).
- <sup>80</sup>E. Mlawer, S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave," *J. Geophys. Res.* **102**, 16663–16682, <https://doi.org/10.1029/97JD00237> (1997).
- <sup>81</sup>M. Tewari, F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, "Implementation and verification of the unified NOAH land surface model in the WRF model," in *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction* (American Meteorological Society, 2004), pp. 11–15.
- <sup>82</sup>J. S. Kain, "The Kain–Fritsch convective parameterization: An update," *J. Appl. Meteorol.* **43**, 170–181 (2004).
- <sup>83</sup>S.-Y. Hong and J.-O. Lim, "The WRF single-moment 6-class microphysics scheme (WSM6)," *J. Kor. Meteorol. Soc.* **42**, 129–151 (2006).