Referee 3 (Report 1)

We are grateful to the referee for devoting their time to our manuscript. The valuable comments and suggestions will help us to improve the paper.

We will here respond to comments made:

The designed system is based on systems created by Lorenz (2005). The first and simplest of this type is the low-dimensional atmospheric system (L96) presented by Lorenz (1996). It is a nonlinear model, with N variables connected by governing equations

$$dX_n / dt = -X_{n-2}X_{n-1} + X_{n+1}X_{n-1} - X_n + F, \tag{1}$$

$n = 1, \ldots, N$. X_{n-2}, X_{n-1}, X_n, X_{n+1} are unspecified (i.e., unrelated to actual physical variables) scalar meteorological quantities (units), F is a constant representing external forcing, and t is time. The index is cyclic so that $X_{n-N} = X_{n+N} = X_n$ and variables can be viewed as existing around a latitude circle. Nonlinear terms of Eq. (1) simulate advection. Linear terms represent mechanical and thermal dissipation. The model quantitatively, to a certain extent, describes weather systems, but, unlike the well-known Lorenz model of atmospheric convection (Lorenz, 1963), it cannot be derived from any atmospheric dynamic equations. **The motivation was to formulate the simplest possible set of dissipative chaotically behaving differential equations that share some properties with the "real" atmosphere. One of the model's properties is to have 5 to 7 main highs and lows that correspond to planetary waves (Rossby waves) and several smaller waves corresponding to synoptic-scale waves. For Eq. (1), this is only valid for $N = 30$.** Lorenz (2005), therefore, introduced spatial continuity modification (L05). Eq. (1) is then rewritten to the form:

$$\frac{dX_n}{dt} = [X, X]_{L,n} - X_n + F, \tag{2}$$

where

$$[X, X]_{L,n} = \sum_{j=-J}^{J}{}' \sum_{i=-J}^{J}{}' \left(-X_{n-2L-i}X_{n-L-j} + X_{n-L+j-i}X_{n+L+j} \right) / L^2$$

If L is even, $\sum'$ denotes a modified summation, in which the first and last terms are to be divided by 2. If L is odd, $\sum'$ denotes an ordinary summation. Generally, L is much smaller than N and $J = L/2$ if K is even and $J = (L-1)/2$ if L is odd. **To keep a desirable number of main highs and lows, Lorenz (2005) suggested a ratio $N/L = 30$ and $F = 15$. The choice of parameters F, and the setting of time unit = 5 days, is also made to obtain a similar value of the largest Lyapunov exponent as the ECMWF forecasting system** (Lorenz, 2005).

A two-level (scales) system (L96-2) was introduced by Lorenz (1996) by coupling two such systems, each of which, aside from the coupling, obeys a suitably scaled variant of Eq. (1). There are N variables X_n plus $J \cdot N$ variables $Y_{j,n}$ defined for $n = 1, \ldots, N$ and $j = 1, \ldots, J$. Governing equations are:

$$dX_n / dt = -X_{n-2}X_{n-1} + X_{n+1}X_{n-1} - X_n + F - (c/b)\sum_{j=1}^{J} Y_{j,n}, \tag{3}$$

$$dY_{j,n} / dt = -cbY_{j-2,n}Y_{j-1,n} + cbY_{j+1,n}Y_{j-1,n} - cY_{j,n} + (c/b)X_n, \tag{4}$$

where c sets the rapidness of small scale compared to large scale, b sets the small scale amplitude size compared to large scale. $Y_{j,n-N} = Y_{j,n+N} = Y_{j,n}$ while $Y_{j+J,n} = Y_{j,n+1}$ and $Y_{j-J,n} = Y_{j,n-1}$. X_n represent the values of some quantity in N sectors of latitude circle, while the variables $Y_{j,n}$ ($Y_{1,1}, Y_{2,1}, \ldots, Y_{J,1}, Y_{1,2}, Y_{2,2}, \ldots, Y_{J,2}, Y_{3,1}, \ldots$) can represent some other quantity in JN sectors.

A two-level (scales) system introduced by Lorenz (2005) is:

$$dX_n / dt = [X, X]_{L,n} - X_n - cY_n + F, \tag{5}$$

$$dY_n / dt = b^2 [Y, Y]_{1,n} - bY_n + cX_n. \tag{6}$$

Eq. (6) is analogue to Eq. (1) (if we substitute *F* for X_n), and Eq. (5) is analogue to Eq. (2) (aside from the coupling where c is the coupling coefficient, and that Y_n fluctuates b times as rapidly, and their amplitude is reduced by the factor b).

(A) Different two-scale models in Lorenz (1996) and Lorenz (2005). Will it be feasible for providing a diagram for illustrating the grid system within the 2005 two-scale model?

Figure RR1 shows the similarity of the 1996 (Eqs. (3) and (4)) and 2005 (Eqs. (5) and (6)) two-scale systems in the attempt to maintain 5 to 7 main highs and lows and several smaller waves for large scales X_n. While for the 1996 two-scale system, this is ensured by a number of *N* large scale variables X_n close to 30 (and a number of *JN* variables for the small scales), for the 2005 system, it is ensured by linking the X_n variables as described in Eq. (2) (with the same number of small scale variables, however, determined from Eq. (1), Figure RR2). The 2005 two-scale system thus produces a smoother and more realistic evolution of the large-scale variable while maintaining properties similar to the 1996 system.

The systems used in this manuscript, which are described in Appendix A (of the manuscript), address one more condition that brings them closer to real systems. This condition is the fact that the large scale and small scale features in Eqs. (3) – (6) are represented by separate sets of variables instead of appearing as superimposed features of a single set. To satisfy this condition, the coupling of one small-scale variable and one large-scale variable is more realistic than the coupling that is present in the 1996 system (Eqs. (3) and (4)).

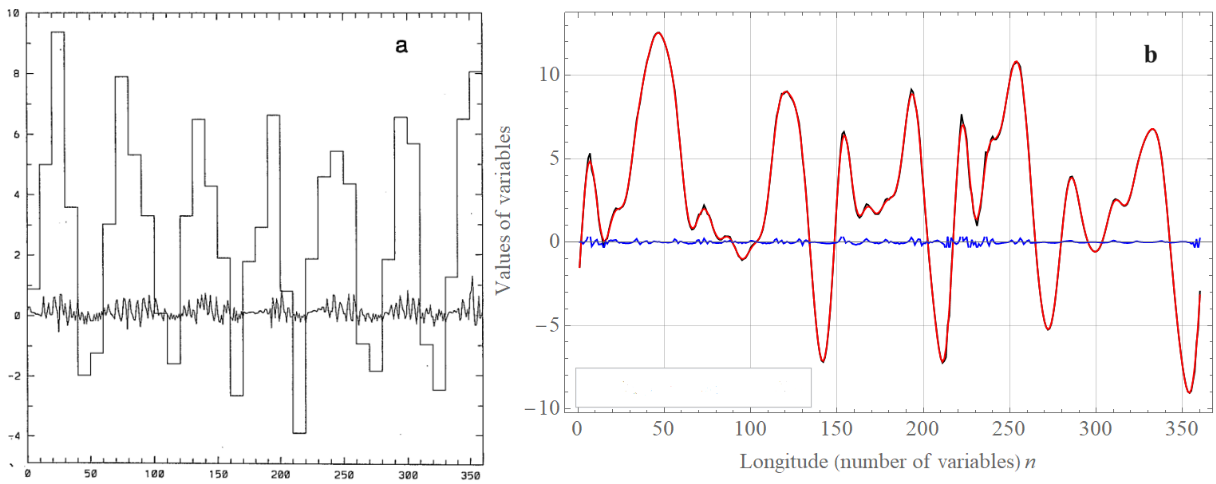

Figure RR1: Comparison of longitudinal profiles at one time of two-scale Lorenz systems (a) from 1996 (Eqs. (3) and (4)) and (b) from 2005 (Eqs. (5) and (6)).

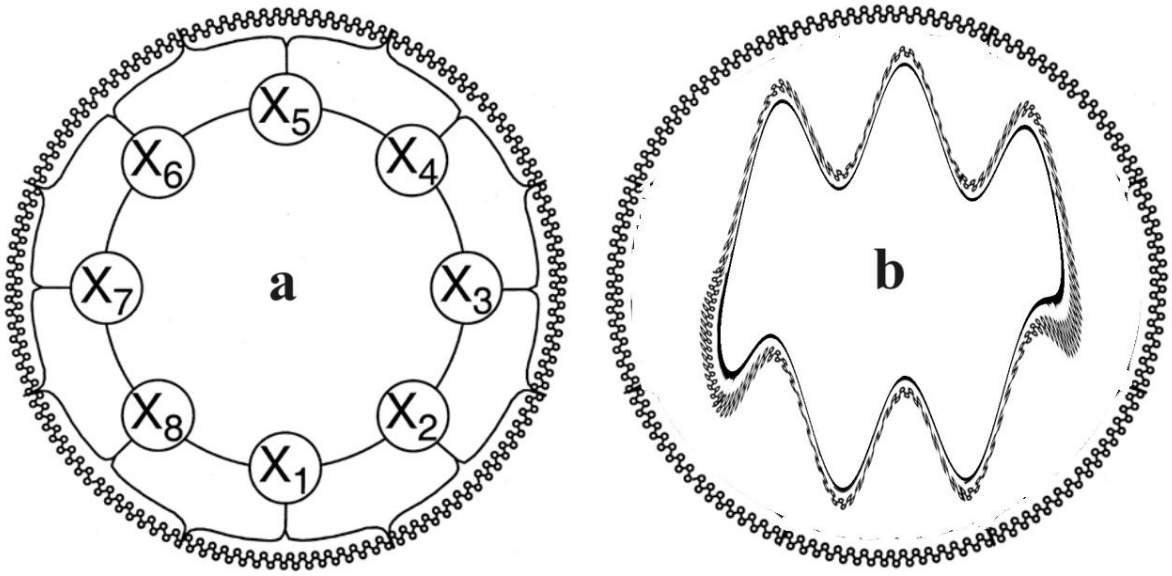

Figure RR2: Comparison of schematic illustrations of two-scale Lorenz systems (a) from 1996 (Eqs. (3) and (4)) (taken from Figure R2 of the referee report) and (b) from 2005, where the inner wave curve represents the large-scale variables described by Eq. (5), which produce 5-7 main waves, and where the outer curve represents the small-scale variables described by Eq. (6), which are not limited by the number of waves. In contrast to (a), one large scale variable is coupled to one small scale variable.

(B) Dependence of findings on temporal spacing (i.e., Δt) and "spatial" spacing (e.g., the number of sectors, N). It would be ideal for additional tests with a smaller $\Delta t = 10^{-5}$ (or $\Delta t = 10^{-4}$). Additionally, the choice of N and L should be explored since N = 960 and L = 32 were used in Lorenz (2005).

The choice of the variable $N = 360$ was made because the value of the largest Lyapunov exponent λ^{L05} of the system described by Eq. (2) ($F = 15$, time unit = 5 days) does not change for $N = 360$ and $N = 960$ (Table RR1) and therefore we chose the lower of the two values for computational efficiency.

N	λ^{L05}
30	0.70
60	0.29
90	0.35
120	0.32
150	0.33
360	0.33
960	0.33

Table RR1: Values of the largest Lyapunov exponent λ^{L05} for selected numbers of variables N in the 2005 Lorenz system (Eq. (2), $F = 15$, time unit = 5 days).

Figure RR3 compares the time evolution of the average value of the variables for the 2005 Lorenz system (Eq. (2)) with time step Δt=1/240 and Δt=1/2400. It can be seen that the values are similar. Given this, we use the larger time step dt=1/240 for faster computations.

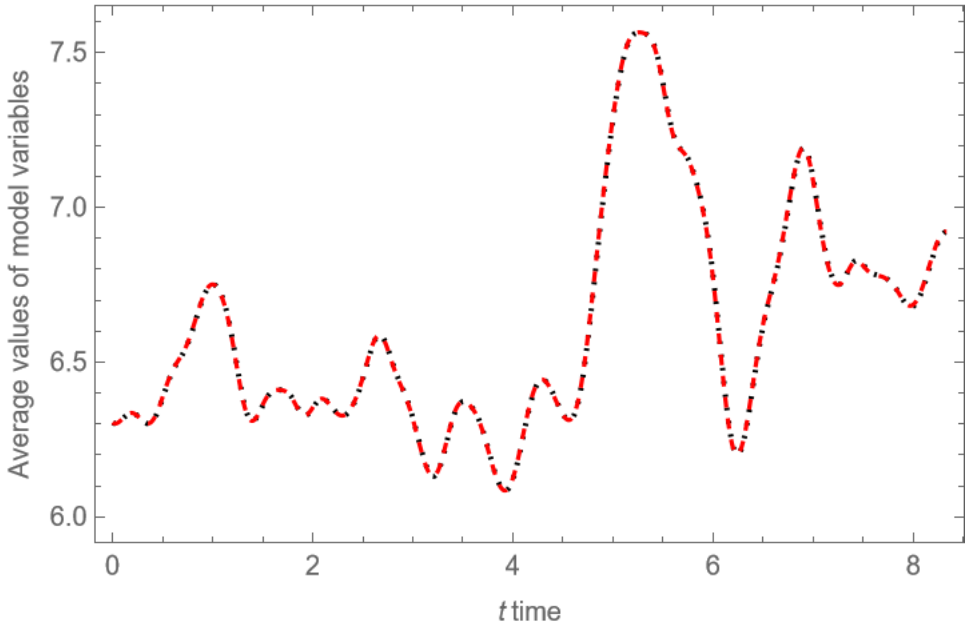

Figure RR3: Comparison of the time evolution of the mean value of the variables (N = 360) for the 2005 Lorenz system (Eq. (2)) based on the same initial conditions with time step Δt=1/240 (red dashed curve) and Δt=1/2400 (black dotted curve).

(C) Impact of model's configuration and complexity on critical points (equilibrium points).

Based on the linearization theorem, critical points of the Lorenz systems could roughly indicate the local behavior of the solutions. As a result, initial error growth should display a dependence on the equilibrium state. Please consider identifying the appearance of the critical points and perform stability analysis using the Jacobian matrix of the linearized system at each of the critical points.

While for analytical studies the instability of fixed points (critical points) is certainly of high interest, we are interested in the typical error growth and therefore focus on the Lyapunov exponent on the chaotic attractor. Since the phase space is so high dimensional, we are not even sure that unstable fixed points are embedded in the chaotic attractor or whether they are outside, as they are in the Lorenz 1963 low dimensional model. We therefore calculate the maximal LE numerically in the following way: a reference trajectory (considered the "truth" or verification) and a trajectory which is the numerical solution of the systems with a given error, are repeatedly generated. For this scheme to be meaningful, we have to ensure that the reference trajectory is on the system's attractor and that the repetition of this scheme samples the whole attractor with correct weights (the invariant measure). We solve this issue in the following way: We first integrate the system over ten years (175200 steps), starting from arbitrary initial conditions, and assume that after discarding this transient, the trajectory is on the attractor. We continue to integrate this single trajectory and consider segments of it as reference trajectories for error growth, i.e., the many reference trajectories are simply segments of one very long trajectory, which ensures not only that all these segments are located on the attractor but that in addition, they sample the attractor according to the invariant measure.

Larger F may produce a larger eigenvalue (a larger real part of the eigenvalue), suggesting a larger growth rate. Based on the following preliminary analysis of the one- and two-scale models with the same value of the forcing parameter F, the effective forcing parameter for the two-scale model is smaller, yielding a smaller leading eigenvalue (i.e., a smaller real part of the eigenvalue). This is consistent with the finding that Figures 5 and 6 display larger growth rates (λ) within the one-scale system (e.g., L05-1) than the two-scale system (e.g., L05-2).

Figure RR4 compares the error growth rates of the L05-1 (Eq. (A1) in manuscript), L05-2 (Eq. (A8) in manuscript), and L05-3 (Eq. (A9) in manuscript) systems. In contrast to the reviewer's findings, the figure shows the smallest growth rate for the L05-1 system and the largest for the L05-3 system. We confirm that the effective forcing for slow variables is weaker, indicating a smaller growth rate within the two-scale model, as compared to the one-scale model. However, it should be noted that in Figure RR4 the values of the single-scale system (L05-1) are not compared with the large-scale values of the multi-scale systems (L05-2 and L05-3), but are compared with the total values of the L05-2 and L05-3 systems, where the large-scale and small-scale features are appearing as superimposed features of a single set.

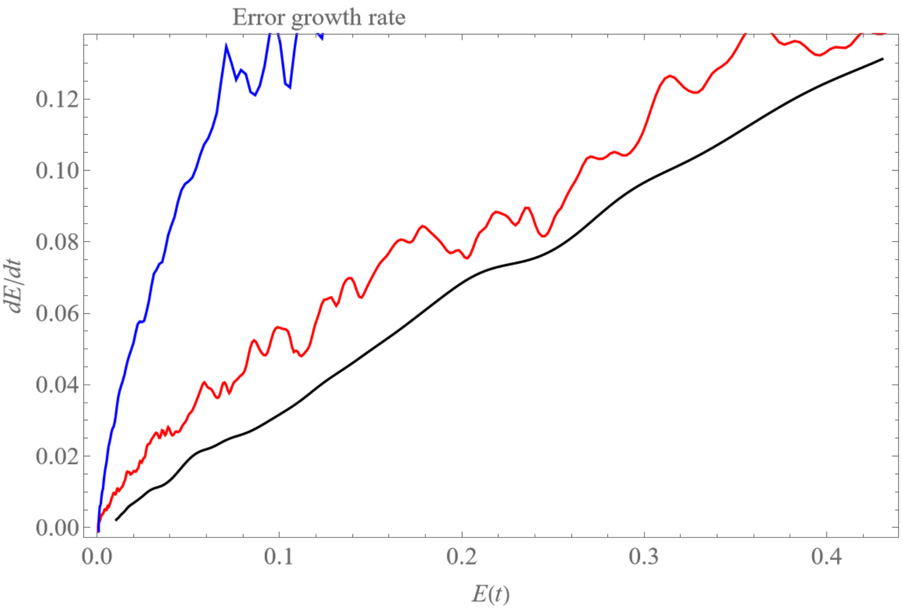

Figure RR4: Initial error growth tendency (rate) dE/dt as a function of the error magnitude E for L05-1 system (Black, Eq. (A1) in manuscript), for L05-2 system (Red, Eq. (A8) in manuscript), and for L05-3 system (Blue, Eq. (A9) in manuscript).

A justification for the use of the L05-2 (Eq. (A8) in manuscript) and L05-3 (Eq. (A9) in manuscript) systems as the "reality" and the L05-1 system as the "model." is presented in the manuscript (Lines 210-220):

“This approach is justified by the fact that the L05-2 and L05-3 systems can be viewed as a variant of the L05-1 system:

$$dX_{tot,n} / dt = \left[X_1, X_1 \right]_{L,n} - X_{1,n} + \tilde{F}_n(t), \quad (12)$$

where $\tilde{F}_n(t) = b^2[X_2, X_2]_{1,n} + c[X_2, X_1]_{1,n} - bX_{2,n} + F$ for the L05-2 system and $\tilde{F}_n(t) = b_1^2[X_2, X_2]_{1,n} + b_2^2[X_3, X_3]_{1,n} + c_1[X_2, X_1]_{1,n} + c_2[X_3, X_2]_{1,n} - b_1 X_{2,n} - b_2 X_{3,n} + F$ for the L05-3 system are treated as a forcing, which varies in a complicated manner with time. We parameterize these small-scale phenomena contained in $\tilde{F}_n(t)$ by the average value of these phenomena, which is close to zero, and therefore we can write:

$$\langle \tilde{F}_n(t) \rangle \approx F = 15, \quad (13)$$

where $\langle \ldots \rangle$ represents the mean calculated over a long orbit on the L05-2 and L05-3 systems attractors."

Please provide justifications for the choice of b1 = 10 for the two-scale system but b1 = 1 for the three-scale system. Additionally, within the three-scale system, are nonlinear terms (e.g., c1 and c2 in Eq. A9) applied for coupling the "sub-systems" for the small- and medium-scale variables with the large-scale system? Please comment on the impact of c1 and c2 on system's stability.

The parameters of any multi-level Lorenz's system (L96-2, L05-2, L05-3) should be set so that all levels behave chaotically (the largest Lyapunov exponent of each level is positive) and that all levels have a significant difference in amplitudes and fluctuation rates. For the L-96 system (Eq. (1)), the chaotic behavior is determined by the value of F, and the number of variables N. Lorenz (2005) states that as long as $N \geq 12$ chaos is found when $F > 5$ (for $N = 4$ it is when $F > 12$ and for $N > 6$ when $F > 8$). In cases such as the L96-2 system (Eqs. (3) and (4)), where the forcing F acts only on the largest scale, the chaotic behavior of smaller scales is created by coupling. The size of the coupling is cascaded from the largest scale to the smaller ones. Because the values of the largest scale variables are determined by the forcing F, the F value indirectly affects the smaller scales' chaotic behavior and must be chosen large enough to ensure chaotic behavior through coupling for all scales (levels). For the L05-2 system (Eq. (A8)), variables are superposed features of a single set calculated by Eqs (A4) and (A5). In addition to those mentioned above, this procedure affects the chaotic behavior, amplitude, and fluctuation rate of the levels, and the choice of I between 10 and 20 may be optimal (Lorenz, 2005). In order to maintain the required properties of the two scales L05-2 system, Lorenz (2005) chose N = 960, L = 32, I = 12, F = 15, b = 10, and c = 2.5 (**note that for L05-2 and L05-3 systems it is not possible to directly determine the amplitude and fluctuation rate of smaller scales using spatiotemporal scaling factors *b*, because these values are mainly determined by the procedure for expressing variables and the length of the intervals** $[-I, I]$).

For the L05-3 system (Eqs. (A9) – (A12)), it is necessary to specify eight parameters. We tested that the values of coupling coefficients c_1 and c_2 do not affect the L05-3 system compared to the values of other parameters, and therefore for simplification $c_1 = 1$ and $c_2 = 1$. The parameter $F = 15$ is set the same as for other L05 systems. For the medium scale amplitude to be approximately ten times smaller than the large scale amplitude and the small scale amplitude to be approximately ten times smaller than the medium scale amplitude and for the scales to have different oscillation rates, the spatiotemporal scale factors are chosen b_1 = 1 and b_2 = 10 and interval lengths $I_1 = 20$, and $I_2 = 10$. $N = 360$ turned out to be most suitable for the chaotic behavior of all three levels (found experimentally).

(D) Separations of initial and model errors

We fully agree with the comment. We simulate the initial error growth in the same systems (perfect model assumption), and the model error growth with zero initial error (perfect initial conditions assumption). Combination of both is studied in section 3.3 of the manuscript.

(E) Validity of error saturation for periodic attractors and coexisting attractors. Have you observed periodic solutions? Can you comment on the validity of error saturation for periodic solutions? Have you observed multistability in your ensemble runs?

In our research, we focused only on the average value of error growth (over variables and number of runs). We set all the scales through the parameters of the Lorenz systems to behave chaotically (details can be found in Bednar and Kantz (2022)) and the evolution of the average error growth did not show signs of periodic solution or multistability.

Specific Comments:

(1) Please check consistency in the capitalization of the initial letters of words within a title.

We checked and fixed it. Thank you for pointing this out. (Lines 1-2).

(2) Lines 45-50, the application of the Lyapunov exponent (LE) is not accurate. A global LE represents a long-term average of "local" growth rates (determined by the separations of two nearby trajectories). Initial separations should remain small. Local growth rates may vary with time. As a result, Eq. (1) with a constant growth rate is valid only for a finite time interval. During different time internals, different growth rates may appear. Note that in addition to one positive LE, solution's boundedness is another important feature that defines a chaotic system.

We have added information about boundedness and validity for a finite time interval. (Lines 47-48)

(3) Lines 45-55, please consider referring to the growth rates in Eqs. (1) and (2) as the exponential growth rate (with a J-shaped curve) and logistic growth rate (with a S-shaped curve), respectively.

We changed the description of Eqs. (1) and (2). (Lines 816-819, 828-830, 841-843, 855-856, 870-871).

(4) Line 80, the term "error growth laws" should be rephrased since they are not necessarily physical laws.

We replaced the term law with the term hypothesis. (Lines 81, 307)

(5) Lines 122, statements are not accurate. Unless additional forcing terms are introduced, improving model's spatial or temporal resolution does not necessarily enhance instability. (Please think of a convergent Taylor series.)

We added to the introduction: "Buizza (2010), Magnusson and Kallen (2013) or Jacobson (2001) show that improving the model's spatial and temporal resolution will improve the ability to predict, especially for short forecast range (Buizza, 2010). However, the cited studies work with models that do not model small spatiotemporal phenomena (they are parameterized) and whose initial condition error magnitude is larger than the magnitude of these phenomena. We have verified the fact that the high resolution model (that models small scales) is less stable than the low resolution model (that doesn't model small scales) against initial condition errors (Bednar and Kantz, 2022; Budanur and Kantz, 2022), and that therefore the issue of omitting small scales has another facet. Our new approach models and omits small spatiotemporal scales using…" (Lines 129-135)

(6) Lines 128-130: it is wired that the two-scale system contains large- and small-scale systems while the three-scale system adds a medium scale, in addition to large- and small-scale flows. Any justifications?

It would be more natural to take the L05-2 and L05-1 systems as the model and the L05-3 system as the reality.). A variant where the L05-2 system was used as the model and the L05-3 system as the "reality" was also tested. The resulting model error growth is approximately identical to the previous variant (L05-1 system as the model and L05-3 system as the "reality"). That's why we chose the settings we present. Further, it would be more natural for the L05-2 system to have a small scale comparable to the medium scale of the L05-3 system. However, our intention was to be close to the L05-2 system presented by Lorenz (2005), whose small scale is equivalent to the small scale of our L05-3 system.

(7) Lines 160-165, have you observed coexisting attractors (e.g., more than one attractors) in your ensemble runs? (e.g., see multistability in Van Kekem and Sterk 2018a,b, 2019; Pelzer et al., 2020).

In our research, we focused only on the average value of error growth (over variables and number of runs) and we did not observe signs of multistability.

(8) Line 170, does the statement "errors might even shrink in short times" indicates the existence of a stable manifold?

Yes, the Lorenz L05-systems possess rather high dimensional stable manifolds, along which trajectories are attracted towards the attractor. Calculation of the Lyapunov-dimension done by us for L05-2 show this very clearly, the attractor dimension is much smaller than the phase space dimension, where the attractor is the unstable manifold. But the statement on line 170 does not indicate the existence of a stable manifold but the fact that initial perturbations might not point into the locally most unstable direction.

(9) Lines 194, while N=360 was used in this study, N=960 was appied in Lorenz (2005).

Thank you for pointing this out. The problem is already discussed in comment (B).

(10) line 186, how many time steps for the transfer of error to the small-scale variables?

The error would immediately (one time step) propagate into the small-scale variables.

(11) Section 3.1, please confirm whether the leading LE in the L05-1 system is larger (smaller) than that in the L05-2 (L05-03) system.

Figure RR4 compares the error growth rates of the L05-1 (Eq. (A1) in manuscript), L05-2 (Eq. (A8) in manuscript) and L05-3 (Eq. (A9) in manuscript) systems. The figure shows the smallest growth rate for the L05-1 system and the largest for the L05-3 system (therefore also for LE). It should be noted that for the L05-2 and L05-3 systems, the error growth rate is scale dependent.

(12) Line 382-394: The key point that higher resolution model produces better predictability is acceptable. However, it is not clear whether Figure 10 is sufficient to support this point. Please see details in the last specific comment below.

Please see the discussion at comment (17)

(13) Line 656: The statement "Based on the fact that scale-dependent error growth implies an intrinsic predictability limit" is not accurate. A finite growth rate may indicate a limit for practical predictability. By comparison, a finite intrinsic predictability is established by the feature of chaos (e.g., sensitive dependence on initial condition, SDIC; e.g., Shen, Pielke Sr., and Zeng, 2023)

Our statement really refers to the finite intrinsic predictability that is established by the features of chaos. The statement is based on Brisch and Kantz (2019), Bednar and Kantz (2022), and Budanur and Kantz (2022).

(14) Lines 612 - 623, discussions are duplicated; they are the same as those in Lines 600-611.

We deleted the duplicated part. Thank you for pointing this out.

(15) Line 715, the parameter "K" should be replaced by "L".

We replaced K by L. Thank you for pointing this out.

(16) Line 716, Lorenz (2005) did not explicitly suggest the ratio of N/L = 30 nor provide justification for the choice of N = 960 and L = 32.

We assume the requirement for a model to have 5 to 7 main highs and lows that correspond to planetary waves (Rossby waves) and several smaller waves corresponding to synoptic-scale waves, and we follow the text of Lorenz (2005) on the pages 1579 (Fig. RR5) and 1580 (Fig RR6).

Figure 4a shows typical profiles produced by Eq. (8) when $N = 240$ and $F = 10$ for selected values of K. When $K = 2$, there are nearly as many waves as if Eq. (1) had been retained. Increasing K to 4 decreases the number, but there are still too many compared with Fig. 1a. When $K = 8$, one evidently succeeds in producing an acceptable number of major waves, although weaker smaller-amplitude waves are superposed. In drawing the curve I have, as usual, connected the successive values of X_n with straight-line segments, but these are hard to detect. Any other reasonable interpolation procedure would have produced an indistinguishable curve. Increasing K to 16, 32, or 64 lengthens the waves still more, and, evidently, one can produce any wavenumber desired by choosing K judiciously.

Since the ratio N/K is 30 in the third profile, whose dominant wavenumber agrees most closely with Fig. 1a, where $N = 30$, there is a suggestion that the appearance of a profile may depend largely upon N/K. Figure 4b is constructed with $F = 10$ and $N/K = 30$ in each profile, and with N successively doubling from 30 in the leading profile to 960 in the final one. The conjecture seems to be well supported; the profiles in Fig. 4b show little resemblance to any profile in Fig. 4a except the third one.

With $N = 960$ and again with $K = 32 = N/30$ and $F = 10$, Fig. 5a has been constructed in the manner of Fig. 1a; it shows profiles produced by Eq. (8) at 6-h intervals for five days. Again, at least for the five days, the major crests and troughs retain their identities, while minor ones come and go. One can conclude that Model II is ready for some applications for which Model I would have been inadequate.

For Model I the doubling time for small errors, as seen in Fig. 2a, depends strongly upon F, but is nearly independent of N if N is not too small. For Model II, with $K > 1$, it also proves to depend strongly upon F while being nearly independent of N and K if N/K is not too small, but, for a given value of F, it is much smaller when $K > 1$ than when $K = 1$. Thus, for the values used in Fig. 5a, the doubling time is about four days—considerably longer than expected in the atmosphere. It can be restored to a more nearly atmospheric value by increasing F.

Figure 5b is constructed like Fig. 5a, again with $N =$

Figure RR5: Page 1579 in Lorenz (2005).

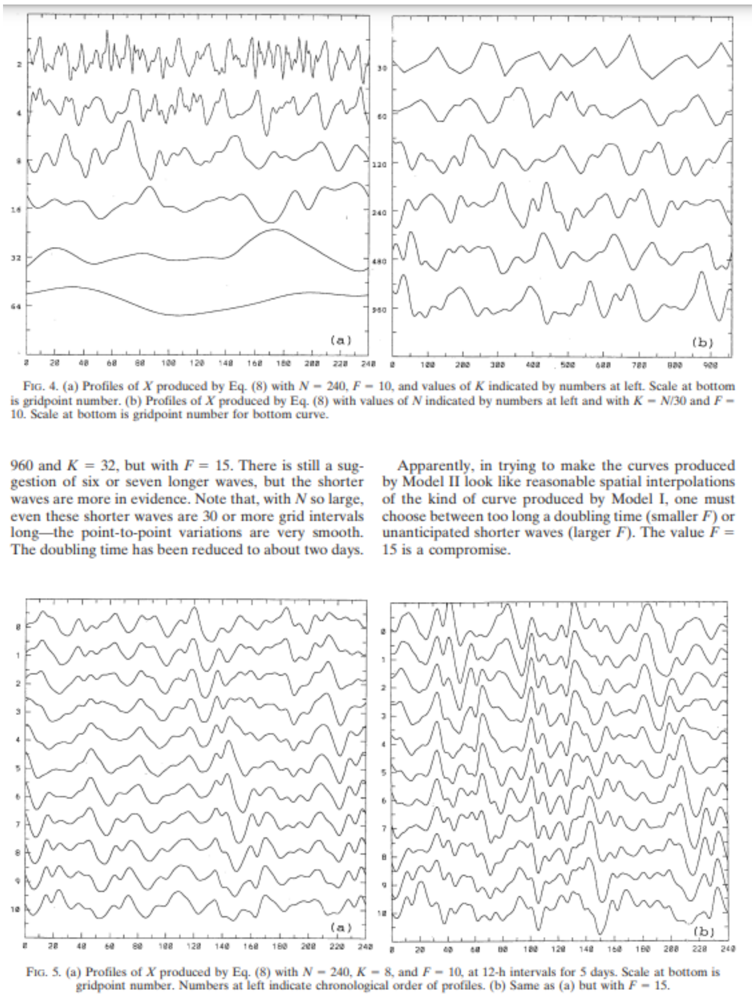

FIG. 4. (a) Profiles of X produced by Eq. (8) with $N = 240$, $F = 10$, and values of K indicated by numbers at left. Scale at bottom is gridpoint number. (b) Profiles of X produced by Eq. (8) with values of N indicated by numbers at left and with $K = N/30$ and $F = 10$. Scale at bottom is gridpoint number for bottom curve.

960 and $K = 32$, but with $F = 15$. There is still a suggestion of six or seven longer waves, but the shorter waves are more in evidence. Note that, with N so large, even these shorter waves are 30 or more grid intervals long—the point-to-point variations are very smooth. The doubling time has been reduced to about two days.

Apparently, in trying to make the curves produced by Model II look like reasonable spatial interpolations of the kind of curve produced by Model I, one must choose between too long a doubling time (smaller F) or unanticipated shorter waves (larger F). The value $F = 15$ is a compromise.

FIG. 5. (a) Profiles of X produced by Eq. (8) with $N = 240$, $K = 8$, and $F = 10$, at 12-h intervals for 5 days. Scale at bottom is gridpoint number. Numbers at left indicate chronological order of profiles. (b) Same as (a) but with $F = 15$.

Figure RR6: Page 1580 in Lorenz (2005).

(17) page 40, line 870-875, Figure 10. Figure's title and captions are confusing. Since L05-02 and L05-03 systems were used to provide the "ground true" (or reference) for computing errors, these errors do not represent the errors of the L05-02 and L05-03 systems, respectively, the growth of initial errors within the L05-02 or L05-03 system does contribute to the growth of differences of the solutions between the L05-1 and L05-02 (or L05-03) systems.

For a comparison in Figures 5-7, let's simply choose λ+, = 0.33, 0.29, and 0.46 for the L05-1, L05-2, and L05-3 systems, respectively. The comparison of the above selected growth rates produces a consistent finding that larger differences (in error growths) are reported in Figure 10b than in Figure 10a. However, on the other hand, considering differences between the L05-02 and L05-03 systems, the differences may produce the largest growth rates as compared to those in Figure 10a and Figure 10b.

The question under investigation in this paper is whether omitting small scale atmospheric phenomena, which contribute little to the final value, will improve the predictability of the resulting value. In other words, how does the average forecast error growth change in a model where small-scale phenomena are omitted but where model errors are therefore introduced, compared to a model where all phenomena are present but the average forecast error growth is scale-dependent. So if we use L05-02 and L05-03 systems to provide the "ground true" (or reference) then, when searching for an answer to the research question, it is reasonable to use the results presented in Figure 10.

Figures 5-7 show that, the L05-1 system is a classical chaotic system with the largest Lyapunov exponent of about $\lambda \approx 0.33$ 1/ day. The data of the L05-2 and L05-3 are best approximated by the power law . For a power law: $\lambda_p(E) := \frac{d\ln(E)}{dt} = \frac{\dot{E}}{E} = aE^{-\sigma}$, with an exponent σ and a coefficient $a > 0$, the error growth rate $\lambda(E) \approx \frac{1}{\Delta t}\ln(E(t+\Delta t)/E(t))$ is expected to be a function of the error magnitude E, and is not constant as for classical chaotic systems. For exponential growth (classical chaos) $E_{exp}(t) = E_0 e^{\lambda_{\exp} t}$ and for an initial error E_0 going to zero, the time t_{lim} at which the error reaches a limiting value E_{lim}, goes to infinity: $t_{lim} = \frac{\ln E_{lim} - \ln E_0}{\lambda_{\exp}} \to \infty \;\; for \;\; E_0 \to 0.$ However, a strict predictability limit t_{lim} exists for scale-dependent error growth even when the initial error E_0 vanishes. For a description by a power law dE_p, the predictability limit t_{lim} is: $t = \left(E^b(t) - E_0^b\right)/\left(a \cdot b\right) \;\; \to \;\; t_{lim} = E_{lim}^b/(a \cdot b) < \infty \;\; for \;\; E_0 \to 0.$

It is true that if we show the growth of the model and initial error in Figure 10, this is the initial error of the L05-1 system, but this is consistent with the question under investigation. At the same time, Figure 10 compares the strictly model error growth (no initial error) with the strictly initial error growth (L05-2, L05-3 systems), where the initial error is limiting towards zero and is then a strict predictability limit.

References:

Bednář, H., and Kantz, H.: Prediction error growth in a more realistic atmospheric toy model with three spatiotemporal scales, Geosci. Model Dev., 15, 4147–4161, https://doi.org/10.5194/gmd-15-4147-2022, 2022.

Brisch, J., and Kantz, H.: Power law error growth in multi-hierarchical chaotic system-a dynamical mechanism for finite prediction horizon, New J. Phys., 21, 1–7, https://doi.org/10.1088/1367-2630/ab3b4c, 2019.

Budanur, N. R., Kantz, H.: Scale-dependent error growth in Navier-Stokes simulations, Phys. Rev. E, 106, 1–7, https://doi.org/10.1103/PhysRevE.106.045102, 2022.

Buizza, R.: Horizontal resolution impact on short- and long-range forecast error, Quarterly Journal of the Royal Meteorological Society, 136, 1020–1035, https://doi.org/10.1002/qj.613, 2010.

Jacobson, M. Z.: GATOR-GCMM: 2. A study of day- and nighttime ozone layers aloft, ozone in national parks, and weather during the SARMAP field campaign, J. Geophys. Res., 106, 5403-5420, https://doi.org/10.1029/2000JD900559, 2001.

Lorenz, E. N.: Predictability: a problem partly solved, in: Predictability of Weather and Climate, edited by: Palmer, T., and Hagedorn, R., Cambridge University Press, Cambridge, UK, 1–18, https://doi.org/10.1017/CBO9780511617652.004, 1996.

Lorenz, E. N.: Designing chaotic models, J. Atmos. Sci., 62, 1574–1587, https://doi.org/10.1175/JAS3430.1, 2005.

Magnusson, L., and Kallen, E.: Factors Influencing Skill Improvements in the ECMWF Forecasting System, Mon. Wea. Rev., 141, 3142–3153, https://doi.org/10.1175/MWR-D-12-00318.1, 2013.

Referee 4 (Report 2)

We are grateful to the referee for devoting their time to our manuscript. The valuable comments and suggestions will help us to improve the paper.

We will here respond to comments made:

The abstract states, "This system shows that omitting small spatiotemporal scales will reduce predictability more than modeling it. In other words, a system with model error (omitting phenomena) will not improve predictability." However, this conclusion is not new. The abstract of Jacobson (2001), for example, states, "Statistics from outer nested domains indicated that the coarser the grid spacing, the greater the underprediction of ozone." Table 2 of the same paper quantifies the impact of grid spacing on model accuracy against data for 25 parameters, including meteorological (wind speed/direction, temperature, pressure, RH), and air quality parameters, in each of four nested domains. The paper concludes (Section 6), "For many parameters…accuracy improved from the coarsest to finest regional domains." Please include a discussion of Jacobson (2001) in your Introduction and indicate whether any other reference you are aware of have also shown the conclusion you are making (that omitting spatiotemporal scales reduces model predictability against data) through a comparison of model results at different scales with data.

We added to the abstract: "that significantly affect the ability to predict" (Line 11)

We added to the introduction: "Buizza (2010), Magnusson and Kallen (2013) or Jacobson (2001) show that improving the model's spatial and temporal resolution will improve the ability to predict, especially for short forecast range (Buizza, 2010). However, the cited studies work with models that do not model small spatiotemporal phenomena (they are parameterized) and whose initial condition error magnitude is larger than the magnitude of these phenomena. We have verified the fact that the high resolution model (that models small scales) is less stable than the low resolution model (that doesn't model small scales) against initial condition errors (Bednar and Kantz, 2022; Budanur and Kantz, 2022), and that therefore the issue of omitting small scales has another facet. Our new approach models and omits small spatiotemporal scales using…" (Lines 129-135)

Abstract. Also, what is missing in the abstract is a summary of results relative to model resolution. How much does improving the resolution, say by a factor of 2 in each the north-south and east-west direction, reduce the error over a specified period of time?

A comparison of how much an improvement in resolution reduce the error over a specified period of time is made in Section 3.4 (lines 367-422). This full comparison is too extensive for the requirements of the abstract, so we have restricted information in the abstract to: "This system shows that omitting small spatiotemporal scales that significantly affect the ability to predict will reduce predictability more than modeling it. In other words, a system with model error (omitting phenomena) will not improve predictability." (lines 11-13).

We are also more interested in the general qualitative perspective (whether omitting small scale phenomena that contribute little to the forecasted product but significantly affect the ability to predict this product will improve the predictability of the resulting value) than in specific quantitative values, because these depend on the parameters of the particular system and its setting.

The authors use the ECMWF model. Please clarify what parameters this model conserves. Does it conserve mass, momentum, kinetic energy, vorticity, enstrophy, and/or potential enstrophy? Do you hypothesize that the non-conservation of some of these properties may affect the results. Can you hypothesize whether results using the ECMWF would give different results from those of a different model, such as the UCLA GCM, which conserves different properties (mass, kinetic energy, vorticity, and potential enstrophy in that case)?

We used 500 hPa geopotential height values of ECMWF systems calculated as 25 annual averages over the Northern Hemisphere (20–90) obtained daily from 1 January 1987 to 31 December 2011. Data was obtained from Magnusson (2013).

Magnusson and Kallen (2013) summarized the development of the ECMWF system during that period: "Since the operational start in 1979, the ECMWF forecast model and the data assimilation system have been continuously developed. Among the important upgrades is the introduction of four-dimensional variational data assimilation (4D-Var) at the end of 1997 and subsequent changes in the use of data in the assimilation were undertaken (Simmons and Hollingsworth 2002). One important change here was the upgrade of the usage of raw microwave radiances from the Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) and Advanced TIROS TOVS (ATOVS) satellite-borne instruments in the year 2000. A major change in the model physics took place in 2007 when changes to the convection scheme and the vertical diffusion were introduced (Bechtold et al. 2008). A comprehensive description of the changes between 2005 and 2008 is given in Jung et al. (2010)." Unfortunately, we did not find in the cited articles what parameters the systems conserve (we suppose that it is based on the primitive equations and hence conserves mass and momentum, but certainly there is some damping (modeling viscosity), so that energy might not be conserved).

Regarding the question of whether non-conservation of some of these properties may affect the results. Drift described in Section 2.4 is a general description of how to characterize a model error and is therefore universal. The extension described in Section 4.1 describes the time evolution of the drift generated at each time step using exponential growth. The universality of this hypothesis has to be confirmed.

Along those lines, in general, do you think the conclusions drawn with this model apply to other models?

We examined whether omitting atmospheric phenomena, which contribute little to the final value, will improve the predictability of the resulting value. For this, we used the L05 systems defined by Lorenz (2005) and Bednar and Kantz (2022) and the ECMWF systems with data from Magnusson (2013). We have shown that omitting atmospheric phenomena, which contribute little to the final value, **will not** improve the predictability of the resulting value. The average prediction error grows faster in a model where small-scale phenomena are omitted, but the model error is therefore created, compared to a model where all phenomena are present, but the average forecast error growth is scale-dependent. We think that our conclusions are general and can by applied to other models.

References:

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Quarterly Journal of the Royal Meteorological Society, 134, 1337–1351, https://doi.org/10.1002/qj.289, 2008.

Bednář, H., and Kantz, H.: Prediction error growth in a more realistic atmospheric toy model with three spatiotemporal scales, Geosci. Model Dev., 15, 4147–4161, https://doi.org/10.5194/gmd-15-4147-2022, 2022.

Budanur, N. R., Kantz, H.: Scale-dependent error growth in Navier-Stokes simulations, Phys. Rev. E, 106, 1–7, https://doi.org/10.1103/PhysRevE.106.045102, 2022.

Buizza, R.: Horizontal resolution impact on short- and long-range forecast error, Quarterly Journal of the Royal Meteorological Society, 136, 1020–1035, https://doi.org/10.1002/qj.613, 2010.

Jacobson, M. Z.: GATOR-GCMM: 2. A study of day- and nighttime ozone layers aloft, ozone in national parks, and weather during the SARMAP field campaign, J. Geophys. Res., 106, 5403-5420, https://doi.org/10.1029/2000JD900559, 2001.

Jung, T., and Coauthors: The ECMWF model climate: Recent progress through improved physical parametrizations, Quarterly Journal of the Royal Meteorological Society, 136, 1145–1160, https://doi.org/10.1002/qj.634, 2010.

Magnusson, L.: Factors Influencing Skill Improvements in the ECMWF Forecasting System, available from personal repository: linus.magnusson@ecmwf.int [data set], 2013.

Magnusson, L., and Kallen, E.: Factors Influencing Skill Improvements in the ECMWF Forecasting System, Mon. Wea. Rev., 141, 3142–3153, https://doi.org/10.1175/MWR-D-12-00318.1, 2013.

Simmons, A., and A. Hollingsworth: Some aspects of the improvement in skill of numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 128, 647–677, https://doi.org/10.1256/003590002321042135, 2002.