Responses to Reviewer 1's Comments

Note to the Editor and all reviewers: We already posted six responses online and these responses will be summarized here. Following the comments and suggestions from the Editor and reviewers, we have revised the manuscript by

- moving the original Section 3.6 regarding the Lilly's formula for two discretization methods into Appendix B to avoid repeated discussions of the scale factor Jacobian,
- adding a few paragraphs, and
- making editorial changes to improve readability (see the manuscript with tracked changes).

> I recommend a major revision of the paper before it can be considered for publication. That is so because it contains only a (rather elementary) mathematical analysis of the works by Lorenz and Lilly, without consideration of the underlying physical processes, and also because it is rather poorly written.

We appreciate your feedback. To avoid repetition, we've reorganized the manuscript. To address other concerns, we've provided three concise responses, each addressing a specific topic as listed below, and summarized them in this final report. We trust our responses meet your expectations.

> The paper is meant to 'revisit' the works by Lorenz and Lilly, which have led to the well-known conclusion that the range of deterministic weather forecasts has an ultimate limit of something around two or three weeks. That conclusion has been amply confirmed by numerical experiments performed with models of increasing spatial resolution and physical realism (see, *e.g.*, Zhang *et al.*, 2019).

To address the comments, our responses (R1A) offer details showing that (1) Zhang et al.'s findings cannot be directly applied to validate Lorenz's and Lilly's formulas because of differing evaluation criteria; (2) Zhang et al. failed to provide compelling reasons for choosing a new tunable parameter in the modified Logistic equation; and (3) Zhang et al. (2019) suggested the potential for increased predictability for certain variables and certain low-frequency weather systems, such as MJOs.

> The authors' main point is that the works by Lorenz and Lilly, although they lead to a similar conclusion, are actually very different from their very starting point, and that further study is necessary as to the predictability of the atmospheric flow.
>
> As just said, the paper is poorly written, with lengthy developments of secondary interest, useless repetitions and inclusion of elementary mathematical material that should not be necessary. This confuses the reader and I had actually some difficulty in even following the logical thread of the paper.

We've significantly revised the manuscript to eliminate repetitive discussions and enhance readability.

From what I understand, the significant part of the paper begins with the introduction of the eddy turnover time $\tau(k)$ (Eq. 5). That quantity is introduced with a reference to Vallis (2006), without appropriate explanation as to its physical significance nor on how it has been determined. The only indication in the paper is that $\tau(k)$ is *the time for a parcel with velocity v_k to move a distance of 1/k, with v_k being the velocity associated with wavenumber k.* (ll. 298-299). More information would be necessary, be that only to refresh the reader's memory. I simply note that, since v_k is defined as the velocity associated with wavenumber k, the variations of v_k with k contain the same basic information as the spectrum of kinetic energy, which is considered later in the paper. That should be mentioned explicitly.

To address the comments above, our responses R1B provide detailed discussions that delve into the physical relationship between kinetic energy, velocity, and turnover time. While it's still uncertain whether the physical significance of Lilly's integral of turnover times has been widely accepted within the community, we propose a possible interpretation for Lilly's formulas that could serve as an alternative measure for predicting predictability horizons. Nevertheless, it requires further effort to determine whether such turbulence-based findings can be applied to estimate the predictability of weather patterns.

Additionally, since the concept of turnover times cannot be directly applied to analyze the data obtained from Lorenz's 1969 model, which was based on a conservative PDE, we simplified our discussions to focus on the sum and integral of the turnover times in Eq. (6) and their dependence on two discretization methods in the revised manuscript.

The authors then proceed to estimate predictability times by integrating the turnover time over two different grids in spectral space. They find (Eqs 9a-b) that the integral on the exponential grid (Eq. 1a) is finite while the integral on the linear grid (Eq. 1b) is infinite. The former being sparser for large values of the wavenumber k, it is obvious that the corresponding integral will be smaller. From a physical point of view, what should be considered there is how fast an uncertainty at wavenumber k propagates to larger scales, and how the propagation relates to the turnover time. That should determine on which kind of discretized grid an integral of the turnover time can be physically significant. Although I presume that has been done by other authors, that basic question is not even mentioned, nor is any reference given about it. The authors totally miss here a critical point.

The above concerns have also been addressed in R1B response file. The Lilly's integral with respect to ln(k) (Lilly 1990) is consistent with Lorenz's hypothetical assumption that is not supported by data in Lorenz (1969). While such an integral was documented in Vallis (2006), it is very challenging to have any studies that discussed the detailed meaning of the integral with respect to ln(k). In fact, in the previous round of review, no reviewers can share additional references that provide reasons for such an integral: an integral with respect to ln(k).

Via email discussions with Prof. Vallis, we learned that a non-uniform grid discretization might be compatible with self-similarity through the specific energy cascade: $2^{n-1}k_L \ldots \rightarrow 4k_L \rightarrow 2k_L \rightarrow k_L$ (for the inverse cascade) or "$k_L \rightarrow 2k_L \rightarrow 4k_L \rightarrow \cdots 2^{n-1}k_L \ldots$." (for the direct cascade). However, we contend that it hasn't been proven that all weather systems exhibit self-similarity. For instance, baroclinic waves at wavenumber 10 aren't included in the aforementioned scenario. Therefore, we propose that the integral with respect to ln(k) cannot fully capture the scale interactions in weather and climate. Furthermore, in the revised manuscript text, we highlighted that the Lorenz 1969 linear, multiscale model was constructed based on mode-mode interactions, resulting in each model interacting with all modes in the system. This mode-mode interaction doesn't align with the above cascade.

> Another point (subsection 3.4.2) is relative to the coefficient $2^{-2/3}$, which is present in both Lorenz's and Lilly's approaches. The authors show (Table 1 and subsection 3.4.2 together with the associated Figure 4) that this coefficient is not defined by Lorenz with any real accuracy. But they do not really mention how it comes into Lilly's approach through Eq. (5) and the hypothesis of a -5/3 power law for the KE spectrum. Again, additional explanations may, be necessary there.

The detailed responses to the above comments are provided in responses R1B. In that section, the mathematical expression for the turnover time of the KE -5/3 power is presented, resulting in the turnover time $\tau(k) = C_0 k^{-\frac{2}{3}}$. In Lilly's formula, the use of the non-uniform grid, $k_j = 2^{j-1} k_L$, leads to the presence of the common factor of $2^{-\frac{2}{3}}$ in the turnover times. In the original Lorenz's idea, the common factor is based on the fixed ratio of two consecutive "saturation time differences. However, our reexamination of Lorenz's Table (i.e., Table 1 in the manuscript) does not support this hypothetical ratio.

> Actually, the point I have found of most interest in the paper is the fact that Lorenz, although he used in Lorenz (1969d) a linear nonturbulent model, found a predictability time of about the same magnitude as Lilly, who used a nonlinear turbulent model. That fact, which is certainly of great interest, is not further discussed in the paper, but I accept it could be considered as going beyond its scope.

In our responses R1C, we highlighted that in simple models or formulas, the model time may not accurately represent real-world time. Without verifying the time evolution of a specific model against observations, it becomes difficult to determine whether such a model accurately simulates the true nature of weather. Consequently, qualitative predictability estimates should be the primary focus when applying Lorenz's and Lilly's formulas.

We acknowledged the effectiveness of using theoretical models and simple models to qualitatively estimate predictability. For instance, the Lorenz 1963 model is widely accepted to illustrate finite predictability within chaotic systems. However, theoretical models and simple models relied on the pre-assumption of time scales to provide quantitative estimates, which can be challenged by falsifying the assumption of time scales as well as the assumptions for the models (e.g., the absence of significant forcing or dynamics).

In Responses R1C, we provide a concise overview of key studies in atmospheric predictability, including Charney et al. (1966), Lorenz (1969d), Lilly (1972, 1973, 1990), and Vallis (2006). Three different time scales were used in these studies. They may suggest similar conclusions but with different assumptions. (Note that different conclusions were reported in Vallis 2006).

For example, Lorenz (1969d) assumed that one model time unit represents six real-world days. Under this assumption, while Lorenz (1969d) suggested a predictability limit of 16.8 days, Lorenz (1972) reported a limit of 20.6 days. By comparison, as shown in Figure 3 in the main text, Lilly (1973) reported that the sum of turnover times is 2.7 $\tau(k_L)$, including an "adjustable" time scale $\tau(k_L)$ for quantitative predictability estimates. As discussed in Responses R1C and in Shen et al. (2024), extrapolating a doubling time of 5 days in a GCM to a two-week predictability limit also implies an assumption of time scale (i.e., the ratio between 5 days and 2 weeks).

> On a different aspect, the paper is full of elementary developments in basic calculus, with which most readers can be expected to be already fully familiar (for instance the development from Eq. (18) to Eq. (21b)). And there is certainly no need to *review the concept of the Jacobian* (l. 393) (the Jacobian turns out to be no more in the present case than a scalar derivative, and not a full determinant as is usually meant by the word).
>
> By eliminating those useless developments as well as many equally useless repetitions, the length of the paper as it stands could be substantially reduced.

To avoid repetitive discussions, we've relocated Section 3.6 to Appendix B. Based on interactions with reviewers in the previous round, we believe it's essential to maintain related discussions in the main text and Appendix B for future verification. This is because it was difficult for some reviewers to recognize that the application of the non-uniform grid yields to the scale factor of 1/k in the Lilly's formula, which is a Jacobian. However, we welcome further suggestions to simplify or eliminate these discussions.

> I make a final remark. There is some truth in the existence of at least a practical limit to deterministic weather forecasts. Is the fact that Lorenz and Lilly have reached the same conclusion, with the same approximate value for the limiting value, purely accidental, or is there a common basic truth in the two approaches ? That question should, if not discussed, at be least explicitly mentioned.
>
> I suggest a major revision of the paper, with inclusion of a physical discussion in the approaches of both Lorenz and Lilly, particularly on the concept of eddy turnover time, and elimination of lengthy and useless developments and repetitions.

Historically, the concept of the specific common factor first emerged from the sequence of saturation time differences observed in Lorenz's (1969d) work. Later, Lilly (1972, 1973, 1990) formulated the idea using turnover times. Since Lorenz and Lilly were friends, it's plausible that some ideas were shared between them. However, it's noteworthy that none of Lilly's papers were cited in Lorenz's book titled "The Essence of Chaos" published in 1993 or in Lorenz's significant predictability study published in 1996 (Lorenz 1996, 2006). Given these circumstances, it's reasonable to question the validity of applying a geometric series (i.e., Lorenz's or Lilly's formula) to predictability estimates. Furthermore, Reeves' interview with Lorenz in 2007 (Reeves, 2014) confirmed that a robust predictability limit was not established using Lorenz models. Instead, both Lorenz's book and the interview suggested that the two-week predictability limit was determined based on a doubling time of 5 days and reported in Charney et al. (1966) (refer to a review by Shen et al. 2024).

Additionally, we identified physical inconsistencies based on the physical definitions of saturation time scales and turnover times. Furthermore, mathematical analysis revealed discrepancies between Lorenz's and Lilly's formulas. Consequently, our study challenges the validity of applying the integral of turnover times for the quantitative estimation of the predictability limit in weather.

As discussed in our chaos studies (e.g., Shen 2014), Lorenz (1963b) proposed two types of predictability: intrinsic and practical. Intrinsic predictability depends on the nature of the flow, while practical predictability is determined by mathematical formulas and data. Ideally, a perfect determinism of intrinsic predictability could provide an upper bound for practical predictability. However, despite over six decades since the 1960s, while theoretical models and formulas effectively provide qualitative estimates of predictability (e.g., finite predictability within Lorenz chaotic systems), no robust upper limit has been established. In contrast, real-world models continuously yield improved predictions, leading to increased practical predictability. Moreover, recent advancements in AI-powered models have outperformed traditional PDE-based prediction models. Therefore, the absence of a robust predictability limit reiterated in our recent studies motivates further research to explore the predictability limit using various approaches.

Links for the Posted Responses:

- Shen, Pielke Sr., and Zeng, 2024: Responses Part 1A (R1A): "A reevaluation of Figure 3 in Zhang et al. (2019)". https://doi.org/10.5194/egusphere-2024-2228-AC1
- Shen, Pielke Sr., and Zeng, 2024: Responses Part 1B (R1B): "A Brief Note on Turbulence-based Turnover Time." https://doi.org/10.5194/egusphere-2024-2228-AC2
- Shen, Pielke Sr., and Zeng, 2024: Responses Part 1C (R1C): "Qualitative Predictability Estimates Using Lilly's Formula and Comparative Insights" https://doi.org/10.5194/egusphere-2024-2228-AC3

Relevant Responses:

- Shen, Pielke Sr., and Zeng, 2024: Responses to Editor: Additional discussions of Zhang et al. and the validity of the revised Logistic equation. https://doi.org/10.5194/egusphere-2024-2228-AC5
- Shen, Pielke Sr., and Zeng, 2024: Responses Part 2A (R2A): "A Brief Note on Turbulence-based Turnover Time" (this is different from R1B). https://doi.org/10.5194/egusphere-2024-2228-AC4
- Shen, Pielke Sr., and Zeng, 2024: Responses Part 2B (R2B): "A Brief Note on Bistability, Duality, and Dimensional Transitions in Recent Turbulence Studies" https://doi.org/10.5194/egusphere-2024-2228-AC6