

Responses to Reviewers

Submission ID: EGUSPHERE-2024-2574

Title: Quantifying the local predictability of the 2021 sudden stratospheric warming event using a novel nonlinear method

Author(s): Guiping Zhang, Xuan Li, Yang Li, Quanliang Chen,
and Xin Zhou

January 17, 2025

Responses to Reviewer 2

II-A: Summary

The manuscript investigates the predictability limit of the 2021 Northern Hemisphere Sudden Stratospheric Warming (SSW) event using a fairly novel technique known as BaSIC (BAckward Searching for the Initial Condition). Using ERA5 as verification, the authors choose two subseasonal-to-seasonal (S2S) model forecasts in which to show how far in advance the dynamical models could predict the SSW event. As expected, the ensemble-mean forecasts have the best skill (space and time) under 2 weeks before the start date of the SSW event, agreeing with previous works. However, using the BaSIC method on zonal wind conditions, the authors find that predictability may be extended to 14-17 days before the event. Then, the authors examine error growth in stratospheric fields to show that the largest errors in the stratosphere for this event exist over the middle to high latitudes of Eurasia and North America, suggesting that these regions may be the biggest contributors to the errors in the forecast models' performance for the 2021 SSW.

Response: Thank the reviewer for sparing time to go through the manuscript and pointing out many important comments.

II-B: Overall Opinion

The manuscript presents an interesting way of trying to understand how to examine predictability limits for extreme events like major SSWs. However, I have several concerns about the paper that make it not ready for publication at this time. These concerns are mainly twofold: (1) the BaSIC methodology is inadequately described and, with all of the acronyms used, is very confusing to follow, and (2) the paper lacks actual scientific advancement in identifying the source of errors, particularly when it only focuses on stratospheric fields. The writing is generally good but could also use some more proofreading and refinement. At this time, I think that the paper requires significantly more work before it can be published. Therefore, I am unfortunately recommending that it be rejected with the opportunity for resubmission later.

Response: Thank the reviewer for sparing time to go through the manuscript, highlighting very important issues and providing helpful comments and valuable suggestions to improve the manuscript. According to the reviewer's suggestions, we

have revised the manuscript seriously and carefully. More details and point-to-point responses to the reviewer's comments are listed as follows.

II-C: Major Concern

1. Poor Description of BaSIC. While the authors have referenced other papers that use the BaSIC technique, there is a lack of clarity on several terms, equations, and their use throughout the paper. For example, I am very uncertain what CIS and ES refer to (Lines 161 and 162 and elsewhere). I am also unclear what it means that a random state vector x_0 "loses its predictability" (Line 167) - e.g., what is the measure for "predictability?" Much more clarity in the methods section needs to be applied in order for just basic understanding of the method, which can then be used for further critiquing its use in the paper.

Response: Thanks for the reviewer pointing out this issue, and apology for poor description of BaSIC method. The BaSIC technique used in our study aims to quantify the predictability limit of specific events from the perspective of forecast error growths. The criterion on how to estimate predictability limit includes three steps. Firstly, we should determine a specific or target state, whose predictability limit is to be quantified. Because the predictability limit of target state estimated by the BaSIC method is based on the perspective of nonlinear growth of forecast errors. Secondly, we need to investigate the nonlinear growth of forecast errors. Specifically, the growth of initial forecast errors from different initialization date should be analyzed. Thirdly, after the analysis of initial forecast error growths, corresponding initial state (CIS) should be determined. Determining the corresponding initial state should follow a criterion. If the forecast error of an initial state prior to this target state grows with time, and exceeds the threshold value at the time of the target state, then this initial state is the CIS which we need to find. After the determination of the CIS, the timespan between the corresponding initial state and the target state is defined as the predictability of the target state. It should be noted that the threshold value employed is the attractor radius, which is the standard deviation of a variable in a long time series. The above mentioned steps are how to estimate predictability limit of a specific state.

To make readers more understandings on the BaSIC technique, we have rewritten the

introduction of this method. More details can be found in the revised manuscript (Lines 169-197 and 614-710).

In addition, the ES and CIS refer to extreme state and corresponding initial state, respectively. The extreme state and ES have been replaced with target state and TS in the revised manuscript, respectively.

Apology for the poor description of predictability. Clarifications on "loses its predictability" have been added from lines 687 to 701. Predictability is closely related to the growth of the initial error. Slight difference of two states (\mathbf{x}_0 and \mathbf{x}'_0) on initially nearby trajectories in phase space grows over time. Owing to the chaotic nature of nonlinear dynamical systems, the slight difference will evolve to exceed a certain threshold value, then predictability of the state \mathbf{x}_0 can be considered to be lost.

Detailed procedure on how to measure the practical predictability of the 2021 SSW event can be found in the revised manuscript from lines 169 to 197.

2. Does this method really show longer predictability windows? The authors make a key point of highlighting that, while the dynamical models in the aggregate (key term) have poor prediction skill at leads longer than 2 weeks, the BaSIC technique highlights that the 2021 major SSW event had a predictability limit of 14-17 days. I have a couple of issues with this statement. First, the authors are comparing the skill for ONE event against the skill of models for MANY events. For any extreme event phenomena (heavy rainfall, heatwaves, etc.), one can find evidence that even a model can perform better than expected for MULTI-MODEL means for ONE event versus a collection of them. This is why we have aggregate statistics for model performance. Scientific interest certainly lies in outlier events (e.g., events in which the model did exceptionally well or poorly forecasting at long leads), but the comparison here of an aggregate of 14 days predictability limit vs. 14-17 days for the BaSIC method isn't the right one, in my opinion. Moreover, is a gain of maybe 3 days really useful or even significant for this event? Overall, I find this to be a significant weakness of the authors' main argument for this manuscript.

Response: Thanks for the good comments from the reviewer. The BaSIC method employed in the study reveals longer predictability windows than operational forecast skills of two S2S models. It should be noted that the BaSIC method focuses on estimating the practical predictability of specific events. And the practical predictability

is different from the forecast skills. Actually, the practical predictability is the upper limit of forecast skills with the presence of both the initial errors and model errors. In this study, we firstly evaluated the forecast skills and found the ECMWF and CMA models show relatively good performances within two weeks. However, the practical predictability limit is higher than the forecast skills. Then, how long is the practical predictability limit? We are more concerned with this issue. Therefore, the key point of this work is to quantify the practical predictability limit of the 2021 major SSW event, and we use the BaSIC method to address this issue. The result showed that the 2021 major SSW event had a predictability limit of 17 days using the two single models. We have rewritten the BaSIC method to make potential readers more understand it (Lines 169-197 and 614-710).

Apology for poor descriptions to lead misunderstandings. In practice, we haven't compared the skill for one event against the skill of models for many events. We only focus on the 2021 SSW event, and are not concerned with other SSW events. It should be noted that we have evaluated the average forecast skills of the zonal wind (north of 60°N, 10-hPa) within 15 days by ECMWF and CMA centers from December 2020 to January 2021 (Fig. 3). However, this evaluation of average forecast skills in winter is independent from that of the forecast skills of the 2021 SSW. In addition, we also have not carried out the comparison of forecast skills between one model and multi-model means. In this work, we only use two S2S models to investigate the practical predictability of the 2021 SSW event. During the investigation, we performed analysis using the reforecast data from the respective model. And we have not carried out the analysis of multi-model means. Since the multi-model means are not needed in evaluation of average forecast skills or quantifying the upper limit of practical predictability. How to obtain the practical predictability limit (17 days) of the 2021 SSW event, and corresponding analysis can be found from lines 412 to 455 of the revised manuscript.

For synoptic-scale events, the practical predictability limit is widely recognized within two weeks. However, the SSWs are subseasonal to seasonal events. After applying the BaSIC method to the 2021 SSW event, we obtained a practical predictability limit of

17 days, further demonstrating that the SSW event has a longer predictability window. Besides, we also have not performed the comparison of an aggregate of 14 days predictability limit vs. 14-17 days for the BaSIC method. In fact, the time lengths of two S2S reforecasts are 46 and 60 days for ECMWF and CMA models, respectively. We just showed the reforecasts of 14 days, since the reforecasts of longer time have little forecast skills. Anyway, the evaluation of forecast skills in 14 days is independent from the upper limit of practical predictability of 17 days calculated by the BaSIC method.

We have added more clarifications to make potential readers more understand our work in the revised manuscript (Lines 412-455).

3. Source of Errors is too limited and does not offer significant scientific advancement. In using their novel technique, the authors also discuss how they are able to find sources of errors for the forecasts which subsequently can result in poorer predictability. However, the authors only look at stratospheric fields, motivated by some literature cited in the Introduction. While certainly there can errors in the evolution of these fields (or maybe even initialization errors), another major source for changes in the polar stratospheric flow fields is wave driving from the troposphere. Models can have significant errors (and/or biases) in wave driving, which subsequently cascade into stratospheric circulation errors. For example, Schwartz et al. (2022) showed that biases in stationary waves within several subseasonal prediction models significantly impact the upward propagation of waves into the stratosphere. This paper does not even consider these errors for this major SSW. I also do not see suggestions for ways to improve even the errors shown in the paper within the Discussion section. These types of analyses and suggestions are where the scientific advancements could be made with this paper and thus make it a useful publication. As such, I recommend that the authors rework their "sources of error" section of the paper to account for tropospheric fields and wave driving and make suggestions for improvement.

Response: Thanks for the reviewer's suggestion. We are sorry for the error source section to lead some misunderstandings. In fact, we would like to reveal the sensitive areas of forecast error growths based on the ERA5 reanalysis and S2S reforecasts. It is mainly because forecast errors over sensitive areas grow more rapidly than other areas, limiting the overall forecast skills. Investigating the source of errors from external

conditions is not our intention in this work. We agree that it is of great scientific interest to study sources of errors, and we will work on the source of errors from external conditions in further work. To eliminate the potential misunderstandings, we have added subheadings in this section. And the subheading of this section is "Evolution of sensitive forecast error growths".

We agree that tropospheric wave activities have influenced the stratospheric flows, and errors in tropospheric wave driving will cascade into stratospheric circulation errors. However, our study is not involved with the external sources of the practical predictability of 2021 SSW event. The aim of this work is to directly estimate its practical predictability limit based on the reanalysis data and the reforecast data. The reason why we use stratospheric data is that the observational data contained all the dynamical information from the external factors. Researchers have successfully obtained predictability limits of climate and weather events based on the observational data (Ding et al., 2010; Li and Ding, 2013; Ding et al., 2016). Some references are listed below. We hold the same opinion as theirs. In this study, we think that the zonal wind data at 10-hPa contained the dynamical information both from stratospheric and tropospheric wave activities. Therefore, to obtain the practical predictability limit of the 2021 SSW event, we just directly need to analyze the reanalysis data and reforecast data.

Anyway, investigating the tropospheric error sources of practical predictability limit of the 2021 SSW event have important scientific significances. And we will carry out it in further work.

We have added more explanations in the revised manuscript from lines 354 to 368, and from lines 501 to 508.

Ding, R., Li, J. and Seo, K.-H. Predictability of the Madden–Julian oscillation estimated using observational data. *Monthly Weather Review* 138, 1004-1013, <https://doi.org/doi.org/10.1175/2009MWR3082.1>, 2010.

Li, J. and Ding, R. Temporal-spatial distribution of the predictability limit of monthly

sea surface temperature in the global oceans. *International Journal of Climatology* 33, <https://doi.org/doi.org/10.1002/joc.3562>, 2013.

Ding, R., Li, J., Zheng, F., Feng, J. and Liu, D. Estimating the limit of decadal-scale climate predictability using observational data. *Climate Dynamics* 46, 1563-1580, <https://doi.org/doi.org/10.1007/s00382-015-2662-6>, 2016.

II-D: More Minor/Specific Comments

1. Line 10. Change the start of the sentence to "Sudden stratospheric warmings (SSWs) are..." Same with Line 32.

Response: Revised (Lines 11 and 33).

2. Line 48. Change "tropospheric fluctuations" to "tropospheric waves."

Response: We have revised it (Line 49).

3. Figure 2 (and others). The pressure level at which these fields are plotted is not indicated in the caption or accompanying text. This is very important to indicate and should be labeled.

Response: Our analyses are all based on the pressure level of 10-hPa. We have labeled in the revised manuscript (Lines 266-268).

4. Figure 3. There is no indication the months/season to which these errors refer. Please be more specific in the figure caption and text.

Response: Figure 3 presents the average daily forecast errors from December 2020 to January 2021. We have added specific clarifications in the figure caption in the revised manuscript (Lines 281-283).

5. Figure 4 and accompanying text. There is no discussion of how different ensemble members perform (note, for example, there are some ECMWF ensemble members show a major SSW at 2 week leads). This is important, as the authors only compare their method to ensemble-mean statistics.

Response: We have redrawn this figure and added more discussions of how different ensemble members perform in the revised manuscript (Lines 301-329).

6. Lines 269-270. I don't know what "lost their forecast skills" means exactly. Is there a quantitative measure for loss of skill?

Response: Apology for unclear description. We intended to demonstrate the forecast skills decreased so much that the accuracy of forecasts was very low. To eliminate the misunderstandings, we have changed it to "showed lower forecast skills" in the revised manuscript (Line 342). Generally, we qualitatively describe the forecast skills from the forecast errors. If the forecast errors are large, the forecast skills are low, and vice versa.

7. Figure 5. There is no indication of what field is plotted or what the units are.

Response: At 10-hPa, and the unit is m/s. We have illustrated this in the revised manuscript (Lines 345-347).

8. Appendix. I do not know what the script "A" represents in Equations A1, A2, and A3.

Response: \mathcal{A} is the calligraphic style of script "A" in Equations A1, A2, and A3, and it represents the compact attractor (Lines 641).

9. Data Availability. There is no indication that the authors are making their code and/or datasets publicly accessible. Please consider adding a Zenodo and/or Github repository for this.

Response: We have created a Zenodo link and put it in Code and data Availability (Lines 717-719).