

Geophysical Research Letters



RESEARCH LETTER

10.1029/2020GL089283

Key Points:

- Whether forecasts in the Lorenz Model are reliable or not depends on the ratio of initial ensemble spread to observational uncertainty
- Until predictability is lost in the Lorenz Model the level of over- or underconfidence increases with increasing lead-time

Correspondence to:

B. Mayer,
bjoern.mayer@mpimet.mpg.de

Citation:

Mayer, B., Düsterhus, A., & Baehr, J. (2021). When does the Lorenz 1963 Model exhibit the signal-to-noise paradox? *Geophysical Research Letters*, 48, e2020GL089283. <https://doi.org/10.1029/2020GL089283>

Received 10 JUN 2020

Accepted 3 DEC 2020

When Does the Lorenz 1963 Model Exhibit the Signal-To-Noise Paradox?

Björn Mayer^{1,2,3} , André Düsterhus⁴ , and Johanna Baehr¹ 

¹Center for Earth System Sustainability, Institute of Oceanography, Universität Hamburg, Hamburg, Germany,

²International Max Planck Research School on Earth System Modelling, Hamburg, Germany, ³Max Planck Institute for Meteorology, Hamburg, Germany, ⁴Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland

Abstract Seasonal prediction systems based on Earth System Models exhibit a lower proportion of predictable signal to unpredictable noise than the actual world. This puzzling phenomena has been widely referred to as the signal-to-noise paradox (SNP). Here, we investigate the SNP in a conceptual framework of a seasonal prediction system based on the Lorenz, 1963 Model (L63). We show that the SNP is not apparent in L63, if the uncertainty assumed for the initialization of the ensemble is equal to the uncertainty in the starting conditions. However, if the uncertainty in the initialization overestimates the uncertainty in the starting conditions, the SNP is apparent. In these experiments the metric used to quantify the SNP also shows a clear lead-time dependency on subseasonal timescales. We therefore, formulate the alternative hypothesis to previous studies that the SNP could also be related to the magnitude of the initial ensemble spread.

Plain Language Summary Comprehensive Earth System Models seem to be better at predicting the real observed climate system than expected based on their ability to predict their own modelled climate system. This puzzling phenomena is known as the signal-to-noise paradox (SNP) and its origin is still under intensive scientific debate with some studies pointing to deficiencies in the model formulation. In this study we investigate under which conditions the SNP can be obtained using a simple conceptual framework for a climate prediction system based on a simple dynamical model. Our results show that the SNP can be reproduced in the absence of model deficiencies if the model overestimates the observational uncertainty. We also investigate the development of the SNP on subseasonal timescales and find a clear dependency on the lead-time of the prediction. Our results lead us to formulate an alternative hypothesis to previous studies on the origin of the SNP.

1. Introduction

Recent studies identified a potential inconsistency in seasonal dynamical predictions of different climatic variables in the North Atlantic (Eade et al., 2014): The reported skill of these predictions as measured by the correlation coefficient between simulations and observations (actual predictability) exceeds expectations based on model simulations (model predictability) and the models are therefore underconfident. Since the model predictability is determined by the signal-to-noise ratio (Kumar, 2009), this inconsistency is also known as “signal-to-noise paradox” (SNP, Dunstone et al., 2016). The SNP is puzzling, since seasonal prediction systems usually tend to be overconfident (Weisheimer et al., 2019). Subsequent studies identified the SNP in seasonal predictions of the winter North Atlantic Oscillation (NAO; e.g., Baker et al., 2018; Dunstone et al., 2016; Scaife & Smith, 2018). However, the SNP is not limited to seasonal predictions in the North Atlantic, but also appears in a variety of different processes and timescales, such as in decadal predictions of temperature, precipitation and mean sea level pressure (Smith et al., 2019) as well as in seasonal predictions of European summer rainfall (Dunstone et al., 2018). The origin of the SNP is still unknown and this study aims to contribute to its identification.

For forecasts subject to the SNP, a variety of consequences emerge: More ensemble members than necessary are needed to extract the predictable signal (Strommen, 2020) and single ensemble members cannot be treated as viable representations of the actual world (Scaife & Smith, 2018). Therefore, explaining the origin of the SNP is of great importance for the climate prediction community.

© 2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Previous attempts of identifying the origin of the SNP have focused on either deficiency in the model formulation or statistical uncertainties. Studies focusing on model deficiencies proposed hypotheses such as too weak stratosphere-troposphere interactions (O'Reilly et al., 2018; Stockdale et al., 2015) and too weak ocean-atmosphere coupling (Ossó et al., 2020; Smith et al., 2014). The latter is substantiated by the apparently too weak persistence of the NAO (Zhang & Kirtman, 2019) and surface air temperature (Sévellec & Drijfhout, 2019) identified in uninitialized simulations. Since ocean-atmosphere feedback improves with increasing resolution, Sévellec and Drijfhout (2019) hypothesized that higher resolution prediction systems may suffer less from the SNP, however, Scaife et al. (2019) finds no discernible impact of atmospheric resolution. In contrast, Weisheimer et al. (2019) focus on statistical uncertainties in the quantities used to assess the level of underconfidence. Furthermore, Strommen and Palmer (2018) and Strommen (2020) propose that the SNP might be related to deficiencies of the models in representing the regime behavior of the NAO. Despite these previous efforts, there is up to now no scientific consensus on the origin of the SNP.

Here, we investigate the occurrence of the SNP in a conceptual framework for an ensemble-based prediction system based on the dynamical Lorenz, 1963 Model (Lorenz, 1963). We conduct a variety of different experiments using different parameters in the initialization of the ensemble. We then analyze these experiments with the purpose of examining under which conditions the SNP occurs for seasonal predictions in our framework. We also analyze how the SNP evolves on subseasonal timescales in this framework, aiming at identifying characteristics of the SNP, which might be in future studies related to the SNP in comprehensive ESMs.

2. Methods

2.1. General Model Description

The Lorenz 1963 model (Lorenz, 1963) describes atmospheric convection in an idealized setup and comprises the following three equations:

$$\begin{aligned}\dot{x} &= -\sigma * x + \sigma * y, \\ \dot{y} &= -x * z + \rho * x - y, \\ \dot{z} &= +x * y - b * z,\end{aligned}$$

where the dot denotes the derivative of a quantity with respect to time. The three dimensionless parameters ρ , σ and b describe the physical properties of the fluid and the idealized setup (for details see Lorenz, 1963). We set them as $\sigma = 10$, $\rho = 28$, and $b = 8/3$. All numerical integrations are done using a fourth order Runge-Kutta method and a timestep of $\Delta t = 0.01$.

2.2. Creation of Idealized Hindcast Experiments

In the following the general setup of seasonal hindcast experiments in the L63 system is introduced, partly following Saetra et al. (2004):

1. Spinup

The model is integrated starting with the initial conditions [1,1,1] for 100.000 timesteps allowing the trajectory to evolve to the model attractor.

2. Generating initial conditions

The model is integrated for further 100.000 timesteps starting from the last timestep of the model-spinup. The set of all states that occur during this phase serves as a pool of initial conditions for the following hindcast experiments (Figure 1a, blue line).

3. Forecast

- (a) A random state from the pool of initial conditions is chosen as the initial state of the real world (Figure 1a, black dots).
- (b) The *reference run* is acquired by integrating the model for 2.000 time units starting from the initial state of the real world.
- (c) The observation of the initial state of the real world is acquired by adding normally distributed noise with zero mean to each dimension on the initial state. The standard deviation (*observational spread*) is denoted σ_o and the same for each dimension.

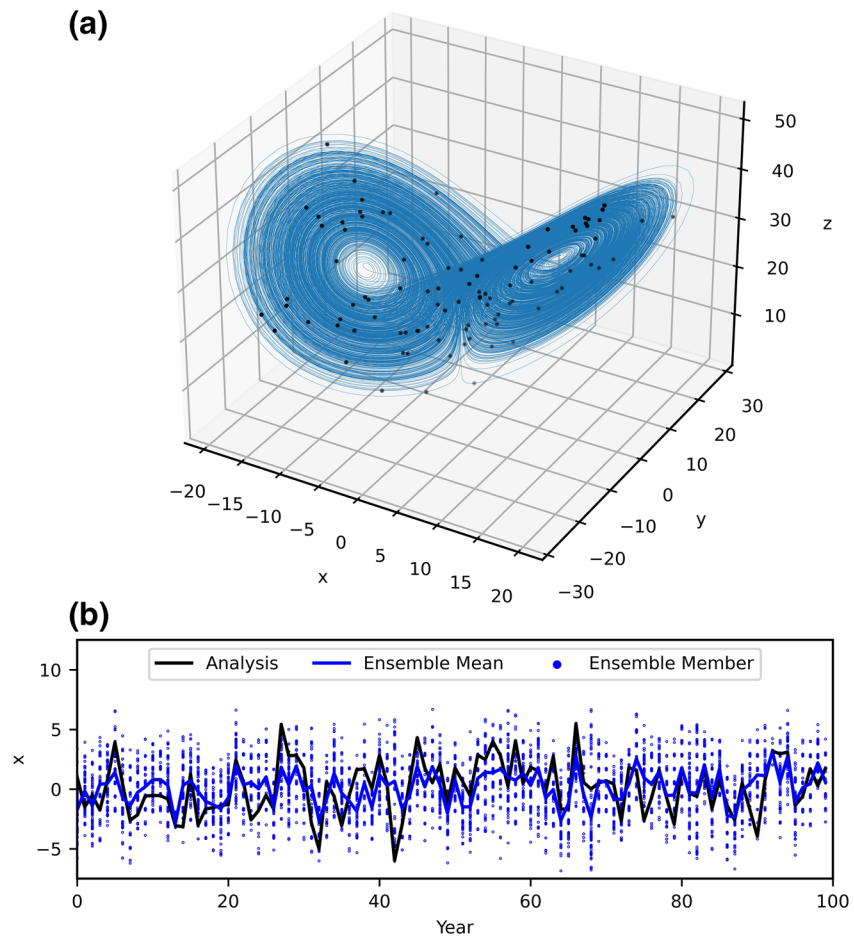


Figure 1. (a) Trajectory of the Lorenz-System used as a pool of initial conditions (blue line) for the initialization of hindcast experiments and 100 randomly chosen initial conditions (black dots). (b) Seasonal 100 years long hindcast experiment with 100 ensemble members in the Lorenz, 1963 Model, showing the analysis (black line), 100 individual ensemble forecasts for each year (blue dots) and the ensemble mean (blue line).

- (d) A 100-member ensemble is initialized by adding normally distributed noise with zero mean to each dimension of the observation. The standard deviation (*initial ensemble spread*) is denoted σ_e and the same for each dimension.
- (e) The time series for verification of the hindcast (*verifying analysis*), is obtained by adding normally distributed noise with zero mean on each dimension of the reference run with standard deviations corresponding to the observational spread for each dimension.

4. Hindcast Experiment

The forecast (Step 3) is repeated 100 times each time with a different randomly chosen initial state. The set of these 100 forecasts combined with the verifying analysis is denoted a *hindcast experiment*.

2.3. Experiment Descriptions

We conduct in total four sets of experiments, each comprising multiple hindcast experiments, which are all calculated over 100 years and with 100 ensemble members with different and randomly chosen initial states. The first set consists of 100 hindcast experiments where initial and observational spread are equal ($\sigma_e = \sigma_o = 0.01$) for all experiments and which is therefore further referred to as the *equal initial spread* set. The second and third set also comprise 100 hindcasts, however the magnitude of the initial ensemble spread underestimates or overestimates the observational spread ($\sigma_o = 0.01$) by one order of magnitude and which is therefore further referred to as the *low initial spread* and *high initial spread* set, respectively. The fourth set

(*incremental initial spread*) consists of 501 experiments where the initial ensemble spread is incremented over two orders of magnitude from $\sigma_e = 0.001$ to $\sigma_e = 0.1$, spaced evenly on a log scale with constant observational spread $\sigma_o = 0.01$.

2.4. Postprocessing and Statistical Analysis

2.4.1. Postprocessing

The raw hindcast experiments (forecasts and verifying analysis) are averaged over 0.1 time units (10 time-steps) using a running mean. According to Palmer (1993), this period amounts to approximately 1 day in the real atmosphere. A month then corresponds to the average over 30 consecutive days (300 timesteps) and accordingly we define the season as the average over 90 consecutive days (900 timesteps). For a seasonal prediction (Figure 1b) we leave a gap of 1 month (300 timesteps) between the initialization and the beginning of the season as general practice in some operational seasonal forecasting systems (e.g., Frhlich et al., 2020; Johnson et al., 2019).

We normalize the timeseries by subtracting their mean and dividing by the standard deviation. Mean and standard deviation are calculated over time and ensemble members. In this study all results are obtained for predictions of the x-component.

2.4.2. Calculation of the RPC

To test whether the amount of predictable signal relative to total variability is comparable in model and actual world, Eade et al. (2014) introduced the *ratio of predictable components* (RPC) as the ratio between the *predictable component in the actual world* (PC_{act}) and the *predictable component in the model* (PC_{mod}):

$$RPC = \frac{PC_{act}}{PC_{mod}}. \quad (1)$$

The predictable component in the actual world is estimated by the *actual predictability* (ACP) calculated by the correlation coefficient between the mean of the full ensemble and the verifying analysis. In this study we estimate the predictable component in the model by the *model predictability* (MOP) using the so called perfect model approach (e.g., Weisheimer et al., 2019). We first choose a single ensemble member as a substitute for the verifying analysis and calculate the ensemble mean over the remaining reduced 99-member ensemble. The correlation between the reduced ensemble mean and the substitute for the verifying analysis yields one sample of the MOP. This process is repeated for every ensemble member. The mean over these samples is then used to estimate the mean MOP (e.g., Ehsan et al., 2013; Kumar et al., 2014).

2.5. Calculation of the RPC for Smaller Ensemble Sizes

In order to investigate the dependency of the SNP on the number of ensemble members we also select reduced ensembles from the full 100-member ensemble, according to the following procedure: First, a random permutation of the full ensemble is generated. The first ensemble member in this permutation is considered to be the substitute for the verifying analysis. Subsequently, ensemble mean predictions of different ensemble sizes are generated by iterative averaging over the remaining ensemble members: The second ensemble member in this specific permutation is gives one possibility of a 1-member ensemble, while the average over the second and third ensemble member gives one possibility of a 2-member ensemble and so forth. The correlation of the ensemble mean prediction for every ensemble size with the verifying analysis (substitute of the verifying analysis) then yields one sample for the ACP (MOP). This process is repeated 100 times starting each time with a different permutation and the mean over all samples is then used to calculate the mean ACP and MOP.

2.6. Selection of Representative Experiments

When we conduct multiple hindcast experiments with different randomly chosen initial conditions, we select one experiment whose ACP as well as RPC are close to their corresponding mean values over all hindcast experiments. This selection is done by first normalizing the ACP as well as the RPC over all

hindcast-experiments and then calculating the root mean square (RMS) of the normalized ACP and RPC index. The hindcast experiment with the smallest RMS-value is considered to be representative of the average behavior.

3. Results and Discussion

3.1. Dependency of Reliability in Seasonal Predictions on Initial Ensemble Spread and Observational Uncertainty

We first analyze the SNP for seasonal predictions of the x-component in the Lorenz 1963 Model with a lead-time of 1 month for the *equal initial spread experiments*. Since the predictability in the Lorenz 1963 System depends on the position of the initial state on the Lorenz Attractor (e.g., Huai et al., 2017), estimates of ACP and MOP are also subject to uncertainty related to the position of the initial states chosen. Therefore, both ACP, as well as RPC, exhibit variability across the 100 different hindcast experiments with the same general model setup but different initial conditions (Figure 2a). Our analysis shows that the sample mean ACP over all 100 hindcast experiments in the equal initial ensemble spread set amounts to 0.515 ± 0.008 , which is comparable to the hindcast skill in a typical seasonal forecast system for the winter NAO (e.g., Dobrynin et al., 2018; Baker et al., 2018; Scaife et al., 2014). The distribution of the RPC over all 100 equal initial spread hindcast experiments exhibits a peak around the value of one (Figure 2a) and the sample mean RPC amounts to 0.985 ± 0.013 , which is close to the perfect reliable case for which the RPC is expected to be equal to one. We then select the one out of the 100 hindcast experiments (Figure 2a), which is representative of the average behavior (Section 2.6). This representative experiment is further analyzed for the ACP as well as MOP at different sizes of the ensemble, allowing us to investigate the uncertainty associated with the selection of ensemble members independent of the uncertainty associated with the position of the initial states. Our results show that independent of the ensemble size the mean ACP is within the interquartile range of the MOP for different permutations of the ensemble and therefore ACP and MOP are found to be almost equal (Figure 2b).

These results obtained in our equal initial spread experiment set indicate that, while individual hindcast experiments might be overconfident ($RPC < 1$) or underconfident ($RPC > 1$) simply by chance, the average of the RPC over all hindcast experiment appears to be close to reliable ($RPC = 1$). This finding is also independent of the ensemble size and robust with respect to random permutations of the ensemble for a representative hindcast experiment. In comparison, the results of a similar analysis for seasonal predictions of the NAO in a comprehensive ESM (Scaife & Smith, 2018) reports values for the RPC larger than two.

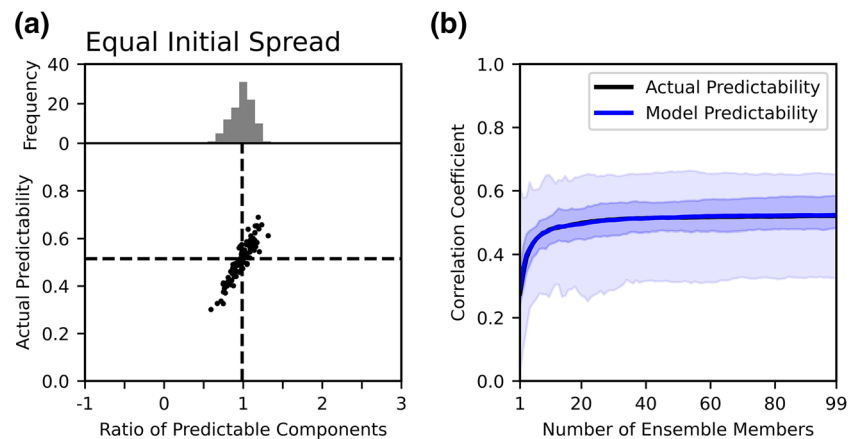


Figure 2. (a) ACP against RPC for 100 seasonal hindcast experiments at 1 month lead-time with the same model setup and equal initial ensemble spread and observational spread but different initial conditions. The horizontal (vertical) dashed line indicates the mean of the RPC (ACP) over all hindcast experiments. (b) Mean ACP (black line) and Mean MOP (blue line) in dependence of the number of ensemble members used for the prediction. Means are taken over 100 random permutations of the ensemble. The dark shading indicates the interquartile range, while the light shading indicates the minimum and maximum over all permutations.

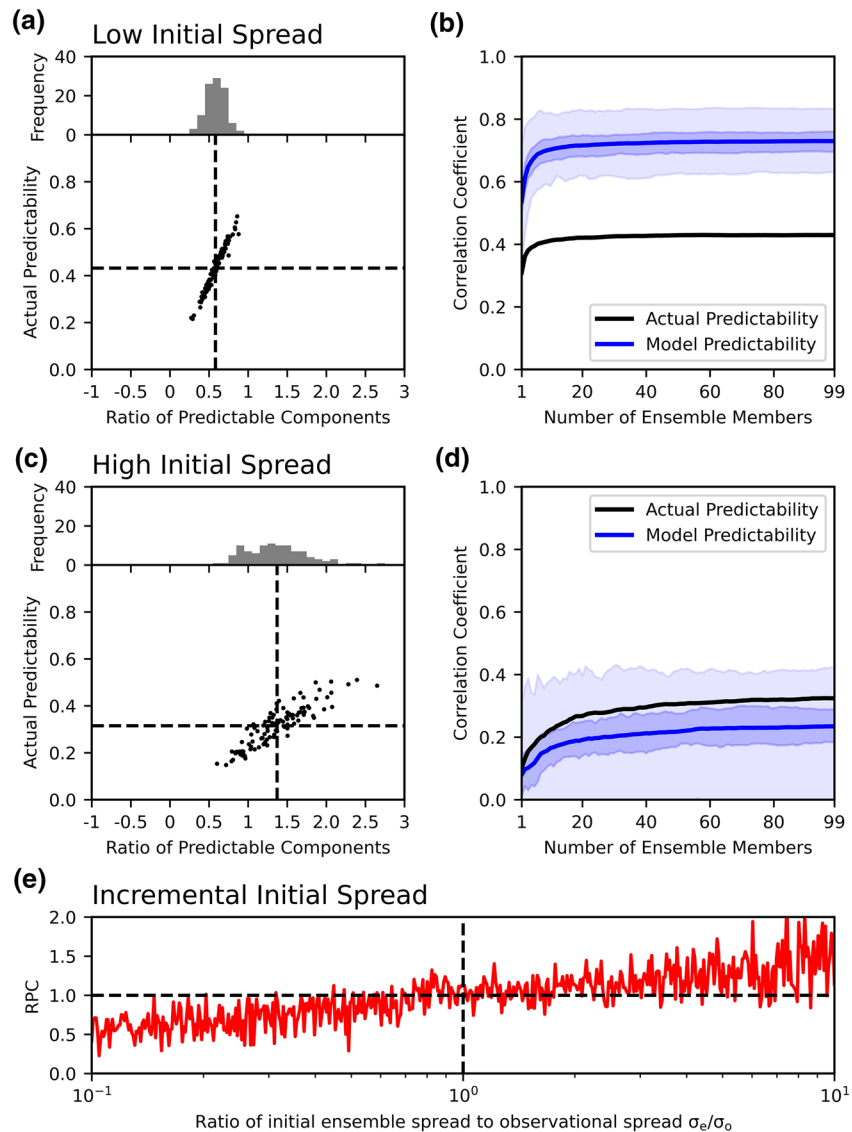


Figure 3. (a and c) ACP against RPC for 100 seasonal hindcast experiments in the low (high) initial spread case, with the same model setup but different initial conditions (lower panel). The horizontal (vertical) dashed line indicates the mean of the RPC (ACP). Histogram of the RPC (upper panel). (b and d) Mean ACP (black line) and Mean MOP (blue line) in dependence of the number of ensemble members for different sizes of the ensemble for one representative hindcast experiment in the Low (High) Ensemble Spread Set. Means are taken over 100 random permutations of the ensemble as described in Section 2.4. The dark shadings in (b and d) indicate the interquartile range, while the light shading indicates the minimum and maximum over 100 random permutations of the ensemble. (e) RPC in dependence of the ratio of initial ensemble spread to observational spread for 501 different hindcast experiments.

Therefore, if the model is perfect and the initial ensemble spread represents the observational spread, the conceptual framework does not show the SNP.

The previous result raises the question whether this overall statement about the occurrence of the SNP in our conceptual framework changes if the initial ensemble spread and the observational spread are not equal. To test whether this is the case, we analyze the seasonal forecasts of the x-component for the *high initial spread* experiment set and *low initial spread* experiment set.

In the case of *low initial spread* (Figure 3a), the sample mean ACP (RPC) over all 100 hindcast experiments amounts to 0.431 ± 0.009 (0.584 ± 0.012). The case for a reliable forecast ($RPC = 1$) is not included in the

sample mean RPC, not even in the extremes of the distribution (Figure 3a). For the representative experiment (Section 2.6), the mean MOP exceeds the mean ACP independent of the ensemble size (Figure 3b). Furthermore, the mean ACP lies outside the range of the MOP acquired by random permutations of the ensemble. The ensemble appears to be overconfident.

In contrast, seasonal forecasts in the *high initial spread* case exhibit a sample mean ACP (RPC) of 1.369 ± 0.008 (0.315 ± 0.038) and $RPC = 1$ is only just included in the extremes of the distribution. It is worth noting that the ACP, as well as the mean RPC found in this study for the *high initial spread* case, are lower than the ones found in seasonal predictions of the winter NAO in Scaife and Smith (2018). However, the investigated GloSea5 system (MacLachlan et al., 2014) belongs to the systems with the highest reported skill and the results of Baker et al. (2018) suggest that systems with higher skill exhibit higher values of the RPC.

The analysis of the representative experiment (Section 2.6) in the *high initial spread* set shows that the mean ACP exceeds the mean MOP independent of the ensemble size (Figure 3d). For most sizes of the ensemble the mean ACP also exceeds the interquartile range of the MOP for different permutations of the ensemble, however it still lies within the range of the extremes. Compared to the *equal* and *low initial spread* sets the uncertainty associated with the permutation of the ensemble is highest in the case of *high initial spread*.

These results show that increasing (decreasing) the ratio of initial ensemble spread to observational spread in the initialization by one order of magnitude leads to—on average—underconfident (overconfident) hindcasts (Figures 3a–3f), with the former resembling the situation empirically found in seasonal predictions of the NAO in comprehensive ESMs (e.g., Baker et al., 2018; Scaife & Smith, 2018).

Similar to the previous set of experiments, we first analyze the RPC in the *incremental initial spread case* for seasonal predictions in the Lorenz 1963 Model for each individual hindcast experiment. The analysis of seasonal forecasts in the *incremental initial spread* set shows a systematic dependency between the RPC and the magnitude of initial ensemble spread to observational spread (Figure 3e). A linear ordinary least square regression between the RPC and the decimal logarithm of the initial ensemble spread to observational spread reveals a statistically significant slope ($p < 0.01$) of 0.450 ± 0.014 . This means that depending on the relative size of the initial ensemble spread to the observational spread the seasonal forecasts based on the Lorenz 1963 Model can appear to be overconfident or underconfident and in general the level of over- or underconfidence increases with increasing miss match between initial ensemble spread to observational spread. Since these results are all obtained in a perfect model framework, they suggest that probably even if comprehensive ESMs at the core of seasonal prediction systems were to be perfect, the same paradoxical result found in these systems could also be obtained simply by initializing the ensemble with a too large initial spread when compared to the observational spread.

3.2. Lead-Time Dependency of Statements About Over-or Underconfidence on Subseasonal Timescales

Using the same representative experiments chosen from the three different sets of hindcast experiments (*low initial spread*, *equal initial spread*, *high initial spread*), we further analyze the evolution of the RPC on monthly timescales for different sizes of the ensemble (Figures 4a–4f). In the 1st month, after initialization all three representative hindcast experiments exhibit values of the ACP and MOP close to one for all sizes of the ensemble (Figures 4a, 4c, and 4e).

In the 2nd month after initialization the difference between the experiments becomes clearly apparent (Figures 4b, 4d, and 4f). The representative *low initial spread* experiment is overconfident with the MOP exceeding the ACP over all sizes of the ensemble (Figure 4b). The representative *equal initial spread* experiment is close to reliable with ACP and MOP being almost equal over all sizes of the ensemble (Figure 4d). The representative *high initial spread* experiment is underconfident with the ACP exceeding the MOP over all sizes of the ensemble (Figure 4f).

We investigate this time-dependency further in the *incremental spread* set for predictions of monthly means at different lead-times with 1 day increments (Figure 4g). Here, we take the absolute value of the RPC since for long lead-times ACP and MOP can take on opposing signs by chance leading to negative values for the

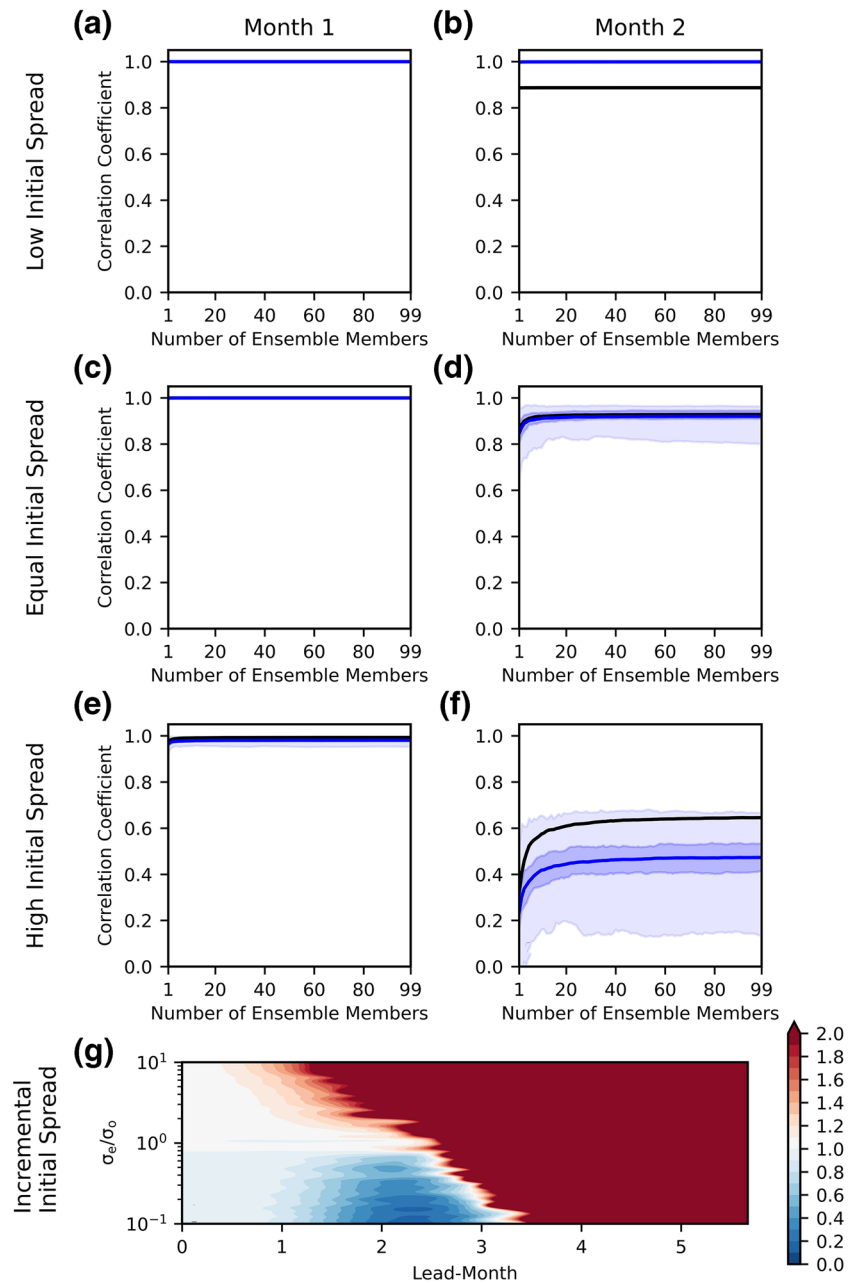


Figure 4. Mean ACP (black line) and Mean MOP (blue line) for monthly predictions for the representative experiment in the low ([a]: 1st month, [b] 2nd month), equal ([c]: 1st month, [d] 2nd month) and high ([e]: 1st month, [f] 2nd month) initial spread set for different ensemble-sizes. Means are taken over 100 random permutations of the ensemble. The dark shadings in (a–f) indicate the interquartile range, while the light shading indicates the minimum and maximum over all permutations. (g) Absolute RPC for monthly predictions in dependence of the ratio of initial ensemble spread to observational spread and lead time for 501 different hindcast experiments. Absolute RPC values are smoothed using a 2-dimensional gaussian filter with a standard deviation of 5 in both dimensions.

RPC. For lead-times up to about 1 month the different experiments with a different ratio of initial ensemble spread to observational show little deviations of the absolute RPC from one. It is only after this period, when the predictive skill has sufficiently dropped that the difference between the MOP and ACP becomes apparent. Individual hindcast experiments with a too large initial ensemble spread ($\sigma_e/\sigma_o > 1$) are predominantly underconfident, while individual hindcast experiments with a too low initial ensemble spread ($\sigma_e/\sigma_o < 1$) are predominantly overconfident. On average, the RPC grows (declines) with increasing lead-time when

the initial ensemble spread is larger (smaller) than the observational spread until it reaches a maximum (minimum) shortly before the prediction skill is lost. When the prediction skill is lost, the absolute RPC takes on large positive numbers by construction.

Our results for the SNP on subseasonal time-scales indicate that in the 1st month after the initialization hindcast experiments with different ratios of initial ensemble spread to observational spread show little deviations of the RPC from the expected value of one. However, the only difference between these experiments is their ratio of initial ensemble spread to observational spread that, even though not clearly apparent in the 1st month after initialization, determines when and to which degree the hindcast evolves to become over- or underconfident in the 2nd month.

4. Conclusion

Based on four sets of experiments using a conceptual framework for a seasonal prediction system based on the Lorenz, 1963 Model (Lorenz, 1963), we conclude that the SNP is not apparent if the initial ensemble spread represents the observational spread. However, the SNP can occur if in the process of initialization, the ensemble spread is overestimated compared to the observational spread. Since reference runs and forecast runs are time integrations of the same model using the same parameterization, we effectively create a perfect model framework where uncertainties in the model formulation can be ruled out when interpreting these results.

Zhang and Kirtman (2019) conclude that the persistence in uninitialized simulations is weaker in models than in reanalysis products. According to the authors their results point toward a fundamental model problem instead of a problem in the initialization procedures, which is contrary to the present study. However, their analysis investigates the SNP in uninitialized ESMs, while our study investigates the SNP in an initialized conceptual model and therefore both setups differ from the initialized ESMs, subject to the SNP. The degree to which persistence in uninitialized simulations is related to forecast skill in initialized simulations as well as the degree to which the results in the conceptual model are related to ESMs remains uncertain in both studies and we therefore find that our results are not mutually exclusive.

Our results suggest that the magnitude of the initial ensemble spread relative to the observational spread could be an alternative hypothesis on the origin of the SNP. Based on these results we suggest to extend the set of candidates being potentially responsible for the SNP and include in further investigations of the SNP also an analysis of how well the initial ensemble spread actually represents the observational uncertainty. While it would be clearly desirable to conduct such a study in comprehensive ESMs, such studies will inevitable face challenges such as how to quantify the observational uncertainty as well as how to separate the effect of initial ensemble spread from other effects occurring in comprehensive models. Nevertheless, our results indicate that despite these challenges further investigations in comprehensive ESMs might be worthwhile.

Acknowledgments

This work was funded by the Copernicus Climate Change Service (contract number C3S 330 DWD), and by Germany's Excellence Strategy – EXC 2037 'Climate, Climatic Change, and Society' (CliCCS) – Project Number: 390683824, as contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. Andre Dusterhus is supported by A4 (Aigéin, Aeráid, agus athrú Atlántaigh), funded by the Marine Institute and the European Regional Development fund (grant: PBA/CC/18/01). The authors thank Sebastian Brune for his suggestions and the Climate Modelling Group at Universität Hamburg. Data used in this study has been self-generated. Open access funding enabled and organized by Projekt DEAL.

Data Availability Statement

The code used to generate and analyze the data as well as to create all figures has been made available (BjoernMayer92, 2020). The original data generated for this study as well as the code has also been archived in the Climate and Environmental Retrieval and Archive (CERA) (https://cera-www.dkrz.de/WDCC/ui/ceraresearch/entry?acronym=DKRZ_LTA_1075_ds00003)

References

- Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018). An intercomparison of skill and overconfidence/underconfidence of the wintertime north Atlantic oscillation in multimodel seasonal forecasts. *Geophysical Research Letters*, *45*(15), 7808–7817. <https://doi.org/10.1029/2018gl078838>
- Bjoern Mayer92 (2020). *Bjoernmayer92/signal-to-noise-paradox-in-the-lorenz-model: First version*. Zenodo. <https://doi.org/10.5281/zenodo.4418958>
- Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., et al. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, *45*(8), 3605–3614. <https://doi.org/10.1002/2018gl077209>
- Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., et al. (2016). Skilful predictions of the winter North Atlantic oscillation one year ahead. *Nature Geoscience*, *9*(11), 809–814. <https://doi.org/10.1038/ngeo2824>

- Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Fereday, D., O'Reilly, C., et al. (2018). Skillful seasonal predictions of summer European rainfall. *Geophysical Research Letters*, *45*(7), 3246–3254. <https://doi.org/10.1002/2017gl076337>
- Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628. <https://doi.org/10.1002/2014gl061146>
- Ehsan, M. A., Kang, I.-S., Almazroui, M., Abid, M. A., & Kucharski, F. (2013). A quantitative assessment of changes in seasonal potential predictability for the twentieth century. *Climate Dynamics*, *41*(9–10), 2697–2709. <https://doi.org/10.1007/s00382-013-1874-x>
- Frhlich, K., Dobrynin, M., Isensee, K., Gessner, C., Paxian, A., Pohlmann, H., et al. (2020). The German Climate Forecast System: GCFS. *Earth and Space Science Open Archive*, *28*, 235–237. <https://doi.org/10.1002/essoar.10502582.1>
- Huai, X.-W., Li, J.-P., Ding, R.-Q., Feng, J., & Liu, D.-Q. (2017). Quantifying local predictability of the Lorenz system using the nonlinear local Lyapunov exponent. *Atmospheric and Oceanic Science Letters*, *10*(5), 372–378. <https://doi.org/10.1080/16742834.2017.1346457>
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). Seas5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*(3), 1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>
- Kumar, A. (2009). Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, *137*(8), 2622–2631. <https://doi.org/10.1175/2009mwr2814.1>
- Kumar, A., Peng, P., & Chen, M. (2014). Is there a relationship between potential and actual skill?. *Monthly Weather Review*, *142*(6), 2220–2227. <https://doi.org/10.1175/mwr-d-13-00287.1>
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020%3C0130:dnf%3E2.0.co;2](https://doi.org/10.1175/1520-0469(1963)020%3C0130:dnf%3E2.0.co;2)
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., et al. (2014). Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1072–1084. <https://doi.org/10.1002/qj.2396>
- Ossó, A., Sutton, R., Shaffrey, L., & Dong, B. (2020). Development, amplification, and decay of Atlantic/European summer weather patterns linked to spring north Atlantic sea surface temperatures. *Journal of Climate*, *33*(14), 5939–5951. <https://doi.org/10.1175/jcli-d-19-0613.1>
- O'Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L. J., & MacLeod, D. (2018). The importance of stratospheric initial conditions for winter north Atlantic oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 131–146. <https://doi.org/10.1002/qj.3413>
- Palmer, T. N. (1993). Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, *74*(1), 49–65. [https://doi.org/10.1175/1520-0477\(1993\)074\(0049:erapart\)2.0.co;2](https://doi.org/10.1175/1520-0477(1993)074(0049:erapart)2.0.co;2)
- Saetra, Ø., Hersbach, H., Bidlot, J.-R., & Richardson, D. S. (2004). Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, *132*(6), 1487–1501. [https://doi.org/10.1175/1520-0493\(2004\)132\(1487:eooot\)2.0.co;2](https://doi.org/10.1175/1520-0493(2004)132(1487:eooot)2.0.co;2)
- Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and north American winters. *Geophysical Research Letters*, *41*(7), 2514–2519. <https://doi.org/10.1002/2014gl059637>
- Scaife, A. A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., et al. (2019). Does increased atmospheric resolution improve seasonal climate predictions?. *Atmospheric Science Letters*, *20*(8), e922. <https://doi.org/10.1002/asl.922>
- Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *NPJ Climate and Atmospheric Science*, *1*(1), 28. <https://doi.org/10.1038/s41612-018-0038-4>
- Sévellec, F., & Drijfhout, S. S. (2019). The signal-to-noise paradox for interannual surface atmospheric temperature predictions. *Geophysical Research Letters*, *46*(15), 9031–9041. <https://doi.org/10.1029/2019gl083855>
- Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *2*(1), 13. <https://doi.org/10.1038/s41612-019-0071-y>
- Smith, D. M., Scaife, A. A., Eade, R., & Knight, J. R. (2014). Seasonal to decadal prediction of the winter north Atlantic oscillation: emerging capability and future prospects. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 611–617. <https://doi.org/10.1002/qj.2479>
- Stockdale, T. N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions and the predictability of the arctic oscillation. *Geophysical Research Letters*, *42*(4), 1173–1179. <https://doi.org/10.1002/2014gl062681>
- Strommen, K. (2020). Jet latitude regimes and the predictability of the north Atlantic oscillation. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2368–2391. <https://doi.org/10.1002/qj.3796>
- Strommen, K., & Palmer, T. N. (2018). Signal and noise in regime systems: A hypothesis on the predictability of the north Atlantic oscillation. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 147–163. <https://doi.org/10.1002/qj.3414>
- Weisheimer, A., Decremier, D., MacLeod, D., O'Reilly, C., Stockdale, T. N., Johnson, S., et al. (2019). How confident are predictability estimates of the winter north Atlantic oscillation?. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 140–159. <https://doi.org/10.1002/qj.3446>
- Zhang, W., & Kirtman, B. (2019). Understanding the signal-to-noise paradox with a simple Markov model. *Geophysical Research Letters*, *46*(22), 13308–13317. <https://doi.org/10.1029/2019gl085159>