

Response to Reviewer 1

Manuscript ID: egusphere-2026-2325

“unseen-awg v1.0: spatio-temporal weather generation using analogs and unseen data”

Jonathan Wider, Jakob Zscheischler

October 5, 2026

Below, we respond to the reviewer’s assessment of our manuscript. Their comments are shown in *gray italics*, our replies in regular text.

General comments

The paper presents a novel tool, unseen-awg, which is an analog weather generator that re-samples ECMWF reforecast data rather than reanalysis. The combination to the best of my knowledge is novel with a large reforecast pool as the sampling basis, with a forecast-skill criterion for tuning σ , and multi-day block sampling to preserve autocorrelation. The methodology, as well as results and discussions are well put forth. The limitations of the study are also openly acknowledged, and the code and data release allow for ease of replication. I view this as a valuable contribution and recommend publication following minor to moderate revision once the following comments have been addressed

Response: We thank the reviewer for their positive assessment of our manuscript and their critical, yet constructive feedback. In the following, we reply to each point individually.

Major comments

1. Lines 82-84: The manuscript notes that the dependencies between the ensemble members and between consecutive initializations reduces the effective sample size. Since the generation of unprecedented daily extremes is a central claim, a quantitative estimate of the unique/independent states that exist (particularly in the tail) may strengthen the argument further.

Response: Effective sample size depends on predictability: it is smaller for more predictable variables and types of events (e.g. [Kelder et al., 2020](#)). If events are predictable, members from the same ensemble forecast will be more dependent; the same is true for forecasts from nearby initialization times. Therefore, the effective sample size depends on the type of event under consideration, and we cannot give a general answer about the effective sample size. We will make this clearer in the revised manuscript. For unprecedented spatially connected daily heat extremes (Appendix B2), we will look into the distribution of events with a record-breaking spatial extent (see response to major comment 2).

2. In appendix B2, the authors point out that the largest connected heat event is drawn from the same reforecast state at different threshold levels, while this may indeed be a interesting result, it could also point to limited no of distinct candidate states present in the extreme tail.

Response: Given that in Appendix B2 we deal with events on a daily timescale, we can assess the extremeness of the selected state directly within the bias-corrected reforecast dataset. We will extract JJA days from this dataset, compute the 90th, 95th, and 98th percentiles from the bias-corrected dataset, and run the connected-component algorithm to the resulting threshold-exceedance masks. We will then study how the event in Appendix B2 ranks among the set of spatially compounding heat extremes. This will address the extremeness of the event in

Appendix B2 within the reforecast dataset. We will also test how many distinct large events from this dataset are actually included in simulations with our weather generator.

3. Line 130-131: Unweighted Euclidean distance may give more importance or weightage to the high latitude cells. This may be addressed/justified using area weighting.

Response: A multitude of distance measures is used in the literature. We originally chose the unweighted Euclidean distance for simplicity. Strictly speaking, we computed the root mean squared error (RMSE), which differs from the Euclidean distance only by a scaling factor that does not affect analog selection if σ is adjusted accordingly but leads to different numerical values for the distances. This was described incorrectly in the original manuscript and will be corrected in the revised version.

We will use weighted RMSE in the revised manuscript.

Additionally, we have noticed an issue with the weather generator that caused it to use only a subset of all available reforecast samples in autumn (see our response to Reviewer 2, major comment 5). We will fix this in the revised version and rerun all analyses with the corrected weather generator.

4. The extreme events highlighted in Fig. B3 and Appendix B2 seem to come from fairly long lead times (26–28 days), where reforecasts have drifted towards the model’s own climate. The additive bias correction applied per lead time takes care of the mean drift, but not the variance? It may be useful to show that the t_{lead} distribution of the generated extremes aren’t preferentially drawn from long leads with inflated spread?

Response: Two effects make the lead-time distribution of the generated time series non-uniform:

1. Through sampling in blocks with $\tau = 30$ days, more states with medium lead times are included than states with very short or very long lead times, because block starts can vary but each block must be $\tau = 30$ days long.
2. For small t_{lead} , the forecast ensemble members are still statistically dependent, and their effective sample size is lower than for long lead times. Therefore, even in a situation without any biases, the chance that the nearest neighbor for a state at the end of a block (with large t_{lead}) will have small t_{lead} is reduced.

We do not consider either of these effects to be problematic per se, but acknowledge that, in combination with lead-time-dependent biases, they may skew the distribution of generated weather data. To make this aspect more transparent, we will include a histogram of lead times in the simulated time series in the Appendix of the revised manuscript and reference it in the main text.

A separate, somewhat independent question is whether extremes follow a different lead-time distribution from the bulk of the generated time series. We will test whether the distribution of lead times during extreme events differs from that of non-extremes included in the dataset. We will also try to assess whether there is bias in the variance with increasing lead times (“inflated spread”). We interpret “inflated spread” as excess climatological variance. We will test for biases in climatological variance for example by selecting a single ensemble member and computing the standard deviation for each day of the year across different years.

Computing quantities relative to thresholds defined on the generated data itself should provide some robustness to biases in the standard deviation.

Minor comments

- | 1. L231: you mention roughly 50% increase; however 0.31 to 0.54 is more than 70% ?
- | 2. L236: contiguous instead of coniguous?
- | 3. Fig. B1 caption refers to panel (d); but the panels are labelled as (c1), (c2). Please correct.
- | 4. In Appendix B2, you report that the state selected at all three thresholds as ($t_{\text{init}} = 29$ July 2012, $t_{\text{lead}} = 28, m = 8$), whereas the corresponding panel titles in Fig. B3 read as (2022-06-24, 26d, 6)?

Response: We thank the reviewer for pointing out these errors and inconsistencies. We will fix them in the revised manuscript.

References

Kelder, T., Müller, M., Slater, L. J., Marjoribanks, T. I., Wilby, R. L., Prudhomme, C., Bohlinger, P., Ferranti, L., and Nipen, T.: Using UNSEEN Trends to Detect Decadal Changes in 100-Year Precipitation Extremes, *npj Climate and Atmospheric Science*, 3, 47, <https://doi.org/10.1038/s41612-020-00149-4>, 2020.