https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

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

Jonathan Wider[1,2,3] and Jakob Zscheischler[1,2,3]

[1]Department of Compound Environmental Risks, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany
[3]Department of Hydro Sciences, TUD Dresden University of Technology, Dresden, Germany

Correspondence: Jonathan Wider (jonathan.wider@ufz.de)

Abstract. Weather generators allow anticipating unseen weather and help prepare for possible weather-related hazards by providing long continuous time series representative of a given climate. Accurately representing dependencies between variables and locations within weather generators is challenging yet important – ignoring them can result in biased risk estimates. Daily analog weather generators trivially capture spatial and multivariate dependencies within each single time step. These generators resample a historical dataset while ensuring that successive sampled days have consistent large-scale atmospheric fields, thereby also ensuring temporally consistent local weather to some extent. Nevertheless, analog weather generators so far underestimate temporal correlations and are limited by the length of the available dataset they sample from, usually observations or reanalysis data. We propose unseen-awg, an analog weather generator based on data from weather forecasts initialized with historical conditions (reforecasts) and apply it to Europe in a case study. Combined with a novel tuning strategy and block sampling, this large, high-resolution dataset representative of present-day climate allows unseen-awg to simulate weather for the full annual cycle, improve on the temporal continuity of the generated time series, and generate unseen extremes at a daily timescale. We demonstrate that unseen-awg captures both the distributional properties of the individual variables and the dependence between summer temperature and precipitation at the grid-cell scale. We further highlight its ability to simulate droughts and heatwaves of unprecedented spatial extent. Combined with climate impact models, unseen-awg holds great potential for assessing weather-related risks across sectors such as water, agriculture, and forestry, domains that require simulating multiple variables and spatial dependencies across a large number of locations.

1 Introduction

Weather generators are an important tool for exploring the full range of plausible weather states in a given climate. They allow anticipating risks and preparing for unprecedented extreme events (Kelder et al., 2025). Coupled with impact models, weather generators can, for example, help understand variability in ecosystem processes (Fatichi and Ivanov, 2014; Anand et al., 2024) or assess flood risks (Qin and Lu, 2014; Winter et al., 2020; Chen et al., 2025). Most approaches for simulating unobserved weather are based on multi-step parametric models (Wilks and Wilby, 1999). Yet a consistent major challenge for these models is to accurately simulate complex dependencies across large sets of locations and variables, especially in the tails of the distribution. Simulating weather fields with realistic spatial dependencies, however, is important for assessing risks associated with spatially compounding hazards such as concurrent floods in multiple catchments or large-scale droughts (Zscheischler et al., 2020; Li et al., 2026). Multivariate extreme value theory has been used to address this challenge to some extent, but the resulting models are typically restricted to temporally aggregated event footprints (Peard et al., 2026) or do not allow generating consistent daily weather (Boulaguiem et al., 2022). Other studies simulate spatial dependencies in variables of interest, but highlight the difficulty of simultaneously approximating temporal, spatial, and multivariate dependencies (Brunner et al., 2021). Recently, emulators and weather prediction models based on machine learning (ML) have risen to prominence (Kochkov et al., 2024; Watt-Meyer et al., 2025; Cresswell-Clay et al., 2025), but they are often trained on coarse-resolution output of general circulation models, simulate few variables, and recent studies raise questions about their ability to represent record-breaking events (Sun et al., 2025; Pasche et al., 2025; Zhang et al., accepted).

Analog weather generators (Yiou, 2014) offer an alternative with past applications including, for example, the construction of plausible worst-case heatwaves and cold spells (Yiou et al., 2023; Sippel et al., 2024). This type of weather generator simulates new time series by resampling a historical dataset while enforcing temporally consistent large-scale atmospheric circulation. As the time series is constructed step by step, consistency is achieved by ensuring that each newly appended day is either the actual subsequent day from the historical dataset or an analog of it, that is, a day on which the large-scale atmospheric state is similar. As a consequence, also local weather is indirectly constrained (Cariou et al., 2025). Analog weather generators simply construct a new temporal sequence from the existing dataset; therefore, they automatically represent spatial correlations within each individual time step with the same fidelity as the underlying dataset.

The main limitation of analog weather generators is typically the size of the underlying historical dataset, which restricts both the range of possible generated weather and the quality of the analogs. These generators simulate weather by temporally resampling historical observations; consequently, generated time series are only capable of exceeding past records in aggregate metrics (e.g., multi-day or seasonal means Yiou and Jézéquel, 2020), not for individual time steps. Additionally, even the closest analogs are often fairly dissimilar when the sample size is small, potentially creating discontinuities in the generated weather time series. One way to increase the sample size could be using data from climate models or numerical weather forecasts rather than direct observations or reanalysis data (Yiou et al., 2023). Climate models have rather coarse spatial resolutions and strong biases in comparison to observations, which limits their applicability for generating input data for climate impact models. An alternative are reforecasts (hindcasts), which are weather forecasts that are run together with modern operational forecasting systems; they use the same methodology but are started from historical initial conditions (Hamill et al., 2006). Therefore, reforecasts constitute a large dataset of high-resolution simulations that are continuously validated in an operational setting and closely tied to their observational initial conditions (ECMWF, 2023). The UNSEEN approach (UNprecedented Simulated Extremes using ENsembles; Thompson et al., 2017; Kelder et al., 2020) makes use of these data to study rare and extreme weather events for risk assessments (e.g., Van Den Brink et al., 2004; Klehmet et al., 2024; Benito et al., 2025), but its applicability is limited by the fact that available reforecasts tend to be either high in resolution but short in lead time, or long-running but coarsely resolved.

Here we introduce *unseen-awg*, a weather generator that extends the analog methodology to reforecast data. We demonstrate its usefulness by generating more than 10,000 years of stationary weather with realistic multi-variable and spatial dependencies

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

over long time scales for Europe. Grounded in an extensive dataset, unseen-awg samples from a larger pool of plausible weather states and can rely on closer analogs. Unlike earlier work that focused on single seasons, our method generates the full annual cycle and allows the generation of arbitrarily long weather time series from a present-day climate. In this way, we are able to sample many different types of unprecedented extremes that extend in space and time. To increase the fidelity of the generated weather, we develop a systematic approach for estimating unseen-awg's parameters by optimizing forecast skill. To the best of our knowledge, comparable optimization strategies have only been applied in analog forecast settings (Krouma et al., 2024), where optimization was performed over small, hand-picked sets of candidate values. Furthermore, we propose sampling states in multi-day blocks to improve temporal consistency. We present a flexible Python code base that allows user-specific modifications to many parts of the pipeline.

2 Data and methods

2.1 Data and pre-processing

Reforecasts, that is, weather forecasts with a modern forecasting system but initialized from historical conditions, exist mainly to measure and improve the calibration of prediction systems (Hamill et al., 2006, 2008; Hagedorn et al., 2008). In this study, we use a subset of the "Extended ensemble forecast hindcast" dataset by the European Centre for Medium-Range Weather Forecasts (ECMWF). The data consist of simulations with version 48r1 of ECMWF's Integrated Forecasting System (IFS; ECMWF, 2023). Reforecasts were produced twice a week at 00:00 UTC. On each selected day, a set of distinct ensemble reforecasts initialized with ERA5 reanalysis data (Hersbach et al., 2020) was generated, one for the same calendar date in each of the past 20 years (e.g., on 1 January 2024, reforecast start dates were 1 January 2004, 1 January 2005, $\ldots$, 1 January 2023).

The dataset contains one control and ten perturbed ensemble members; reforecasts run up to a lead time of 46 days at $6\,\mathrm{h}$ temporal and $0.4°$ horizontal resolution. We aggregate the data to daily scale for the 0th to 45th forecast day by taking the daily minimum, maximum, mean, or sum depending on the variable of interest. The full dataset spans June 2003–September 2023 and consists of reforecasts initialized on 128 different calendar dates in each of 20 years. Overall, the dataset contains $(1\,\text{control run} + 10\,\text{perturbed members}) \times (128 \times 20)\,\text{initializations} \times 46\,\text{days per forecast} \approx 3500$ years of daily weather states. While dependencies, especially for small forecast lead times and between consecutive initializations, clearly decrease the effective sample size, the dataset remains extensive.

Each daily weather state in the reforecast dataset can be identified with a coordinate vector $\boldsymbol{s} := (t_{\mathrm{init}}, t_{\mathrm{lead}}, m)$, where t_{init} is the initialization date, t_{lead} indicates by how many days the forecast predicts into the future, and m indexes the ensemble members. For each state, we downloaded variables characteristic of the large-scale atmospheric circulation $\boldsymbol{x}(\boldsymbol{s})$ ($Z_{500\mathrm{hPa}}$, geopotential height at $500\,\mathrm{hPa}$), and variables that are relevant for impacts: daily minimum, maximum, and mean $2\,\mathrm{m}$ temperature, and daily total precipitation sums ($T_{2\mathrm{m,min}}$, $T_{2\mathrm{m,max}}$, $T_{2\mathrm{m,mean}}$, P), summarized as $\boldsymbol{y}(\boldsymbol{s})$ (Fig. 1). We use $Z_{500\mathrm{hPa}}$ instead of alternative choices like sea-level pressure because it approximates the dynamical evolution of the atmosphere more directly (Gutzler and Shukla, 1984; van den Dool, 1989). While climate change induces trends in $Z_{500\mathrm{hPa}}$ (Christidis and Stott,

Table 1. Spatial extents and horizontal resolutions of the regions for computing analogs and simulating impact-relevant variables.

	Analogs	Impact-relevant
Latitude	30.0° N–72.5° N	30.2° N–71.8° N
Longitude	80.0° W–40.0° E	9.8° W–39.8° E
Resolution	2.5°	0.4°

2015), we work with data representative of the climate of a comparatively short period. As a ground-truth dataset, we use the years 2003–2023 of the ERA5 reanalysis dataset, which is produced using version 41r2 of IFS (Hersbach et al., 2020).

We define two areas for which we retrieve data (Table 1 and Fig. 1): $Z_{500\mathrm{hPa}}$ analogs are computed over a large but coarsely resolved region that influences weather in Europe, using an extent roughly similar to that in Yiou (2014). For impact-related weather variables, we use a smaller, more finely resolved region that roughly corresponds to the MED, WCE, and NEU regions used in the IPCC's Sixth Assessment Report (Iturbide et al., 2020). ERA5 data are interpolated to the chosen $0.4°$ horizontal resolution; bilinear (for $Z_{500\mathrm{hPa}}$ and $T_{2\mathrm{m}}$ aggregates) and conservative (for P) interpolations are used for regridding.

In both datasets, we set daily total precipitation values $P < 1\,\mathrm{mm\,d^{-1}}$ to zero to ensure non-negative precipitation and avoid drizzle effects, as is commonly done when working with $\mathrm{ERA5}_{2003-2023}$ data. Simple additive mean bias correction per grid point, day of year, and lead time is applied to temperature variables in the reforecast data using ERA5 as ground truth. Sensitivity tests (splitting the dataset in half by initialization time, then comparing the biases computed on each half) raised doubts on the robustness of the bias estimation for precipitation; therefore, bias correction is only applied to temperature variables (more details in Appendix A1). More elaborate and statistically efficient bias estimation methods are conceivable, but bias correction is not a focus of the study and applying a custom bias correction method to the provided uncorrected reforecast dataset is possible. To avoid limiting possible bias correction strategies, we do not set $P < 1\,\mathrm{mm\,d^{-1}}$ to zero in the dataset we provide; instead, this pre-processing step is executed in later stages in the evaluation pipeline.

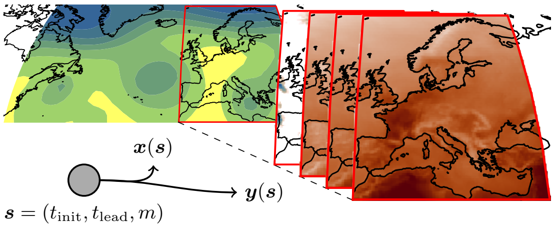

Figure 1. Illustration of the data used by the analog weather generator; $\boldsymbol{s}$ refers to a coordinate vector in the reforecast dataset with the corresponding large-scale atmospheric circulation $\boldsymbol{x}(\boldsymbol{s})$ (geopotential height at 500 hPa) and the impact-relevant variables $\boldsymbol{y}(\boldsymbol{s})$ (daily aggregated total precipitation, daily minimum, mean, and maximum $2\,\mathrm{m}$ temperatures). We visualize each variable with the same extent as used by unseen-awg (Table 1).

2.2 Methods

Our method, unseen-awg, generates weather data by reusing contiguous blocks of days from the reforecast dataset. The weather generator simply constructs a new sequence from the reforecast days while the full spatial fields and all variables within each day are copied verbatim. Therefore, we can conceptualize each generated time series S as a simple lookup table into the reforecast dataset

$$S = \left[\left(t_{\text{out}}^{(1)}, \boldsymbol{s}^{(1)}\right), \left(t_{\text{out}}^{(2)}, \boldsymbol{s}^{(2)}\right), \dots, \left(t_{\text{out}}^{(T)}, \boldsymbol{s}^{(T)}\right)\right]. \tag{1}$$

In particular, the generated spatiotemporal multivariate dataset is fully determined by a list of tuples $(t_{\text{out}}, \boldsymbol{s})$ consisting of a desired output time t_{out} and the reforecast weather state $\boldsymbol{s} = (t_{\text{init}}, t_{\text{lead}}, m)$ which the weather generator assigns to it. Time series are generated by alternating between two steps until a desired length is reached (Alg. 1). In the first step, a block of reforecast days is appended to the generated time series: starting from a state $\boldsymbol{s}_{\text{blockstart}}$, the coordinate vectors of all following states within a time τ are appended to the time series, i.e., $(t_{\text{init}}, t_{\text{lead}}, m)$, $(t_{\text{init}}, t_{\text{lead}} + 1, m), \dots, (t_{\text{init}}, t_{\text{lead}} + \tau - 1, m)$. The block-size parameter τ can be chosen between 1 and 44 days, one day less than the maximum t_{lead} in the dataset. The initial state of the time series, i.e., the first $\boldsymbol{s}_{\text{blockstart}}$ can be prescribed or randomly selected.

In the second step, unseen-awg samples an analog of $\boldsymbol{s}_{\text{next}} := (t_{\text{init}}, t_{\text{lead}} + \tau, m)$, the true next state obtained by continuing along the reforecast. This sampled analog is subsequently used as the next $\boldsymbol{s}_{\text{blockstart}}$ (Fig. 2). For the sampling, we construct a set of candidate analogs and require their lead time to be so small that we can extract a τ-long block of days from the corresponding reforecast when starting the block with the candidate. Additionally, we demand that the analog's calendar date is within a window of half-width of Δ days (default $\Delta = 10$) of both the targeted output date t_{out} and the valid time of $\boldsymbol{s}_{\text{next}}$, $t_{\text{valid}}^{(\text{next})} = t_{\text{init}} + t_{\text{lead}} + \tau$. The windowed selections are introduced for computational efficiency and because differences in solar irradiation would otherwise necessarily lead to unrealistic dynamics (Lorenz, 1969). Constraining the set of candidates further is possible (Appendix A2).

Distances are computed between the large-scale circulation of each candidate state and $\boldsymbol{x}(\boldsymbol{s}_{\text{next}})$. Following common practice, we use the Euclidean distance $d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$ (e.g., Lorenz, 1969; Gutzler and Shukla, 1984; van den Dool, 1989; Krouma et al., 2022). For simplicity, the distance computations do not include area weights. Using the results, we define probabilities such that closer (more similar) analogs get sampled more frequently. We define

$$p(\boldsymbol{x}_{\text{c}}|\boldsymbol{x}_{\text{next}}) \propto \exp\left[\frac{-d\left(\boldsymbol{x}_{\text{c}}, \boldsymbol{x}_{\text{next}}\right)^2}{2\sigma^2}\right]. \tag{2}$$

Here, $\boldsymbol{x}_{\text{c}}$ is a shorthand notation for $\boldsymbol{x}(\boldsymbol{s}_{\text{c}})$, the circulation of a candidate state; $\boldsymbol{x}_{\text{next}}$ is defined similarly. The parameter σ determines how spread out the probability mass is; while it could vary with the day of year, we keep it constant for simplicity. Plugging the Euclidean distance into Eq. (2) leads to a Gaussian distribution $p(\boldsymbol{x}_{\text{c}}|\boldsymbol{x}_{\text{next}}) \propto \mathcal{N}(\boldsymbol{x}_{\text{c}}; \boldsymbol{\mu} = \boldsymbol{x}_{\text{next}}, \boldsymbol{\Sigma} = \sigma^2\mathbf{I})$.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

Instead of normalizing the probabilities of Eq. (2), we use the Gumbel-max trick (Maddison et al., 2014) to sample directly from the unnormalized log-probabilities, i.e., from $-d(\boldsymbol{x}_{\mathrm{c}},\boldsymbol{x}_{\mathrm{next}})^2/(2\sigma^2)$. Squared distances $d\left(\boldsymbol{x}_{\mathrm{c}},\boldsymbol{x}_{\mathrm{next}}\right)^2$ can be precomputed to speed up simulations with unseen-awg.

The presented distance measure and probability distribution are default choices, but the weather generator framework allows these components to be varied, the distance computation can be extended to include multiple days. Tunable parameters like σ provide additional flexibility for a given choice of model.

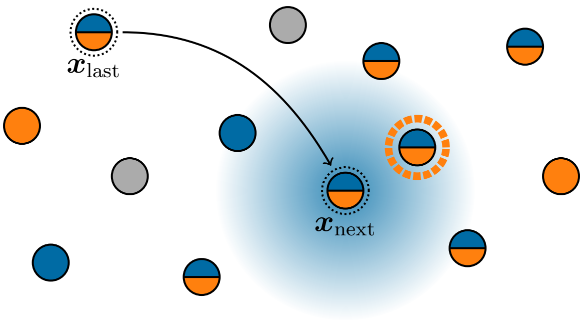

Figure 2. Illustration of the sampling procedure in unseen-awg. $\boldsymbol{x}_{\mathrm{last}}$ denotes the large-scale atmospheric circulation of the state $\boldsymbol{s}_{\mathrm{last}}$ at the end of the last block of days currently included in the generated time series. The weather generator assigns probabilities (blue shading) to each state in the reforecast dataset based on the distance between its large-scale atmospheric circulation and $\boldsymbol{x}_{\mathrm{next}}$, the large-scale circulation of the true next state that followed $\boldsymbol{s}_{\mathrm{last}}$ in the reforecast. Possible analogs are restricted to states with a calendar date that is similar enough to that of $\boldsymbol{x}_{\mathrm{next}}$ (blue filling) and to the calendar date of the targeted output time (orange filling). A new state is then randomly chosen among the samples that fulfil both conditions (orange and blue filling) based on the assigned probabilities and a new block of consecutive days starting from this state (orange dashed circle) is appended to the generated time series, see Alg. 1.

2.3 Optimizing weather generator parameters

By restricting the set of allowed analogs to those with $t_{\mathrm{valid}} := t_{\mathrm{init}} + t_{\mathrm{lead}}$ smaller than a fixed date, the weather generator can be used as an ensemble prediction system. We use the heuristic that a weather generator that produces skillful ensemble forecasts of $Z_{500\mathrm{hPa}}$ should generate more realistic weather than a generator with lower $Z_{500\mathrm{hPa}}$ forecast skill. Therefore, we choose parameter values that optimize the generator's predictive skill. We run t_{forecast}-long analog ensemble forecasts with block size $\tau = 1$ and 80 ensemble members and optimize the forecast skill of unseen-awg quantified through the Continuous Ranked Probability Score (CRPS, Gneiting and Raftery, 2007). For a 3-day forecast, we find an optimum at $\sigma \approx 2.5$ decameters (dam, Fig. 3a) and use this value as the default choice for unseen-awg. Results are robust to using 4-day forecasts instead (see Fig. A3). Identifying optima in forecast skill is easier for relatively short t_{forecast} and τ; for large t_{forecast}, the weather generator shows no skill over climatological forecasts. We expect parameters optimized for $\tau = 1$ to perform well with larger block sizes

Algorithm 1 Sampling a time series S with the weather generator.

```
Require: Initial coordinate vector $\boldsymbol{s}^{(0)} = \left(t_{\mathrm{init}}^{(0)}, t_{\mathrm{lead}}^{(0)}, m^{(0)}\right)$, set of all coordinate vectors in the dataset $\mathcal{C}_{\mathrm{all}}$, targeted output date $t_{\mathrm{current}}^{(0)}$ for the initial coordinate vector, block size $\tau$, length of the time series to be generated $n_{\mathrm{blocks}}$ given as a number of blocks of length $\tau$, probability model $p_\sigma$, distance function $d(\cdot,\cdot)$ between atmospheric fields, distance function $d_{\mathrm{date}}(\cdot,\cdot)$ between calendar dates, half-width $\Delta$ of sliding windows.
$S \leftarrow [\,], t_{\mathrm{current}} \leftarrow t_{\mathrm{current}}^{(0)}, \boldsymbol{s}_{\mathrm{blockstart}} \leftarrow \boldsymbol{s}^{(0)}$    // Initialize variables
for $i = 1$ to $n_{\mathrm{blocks}}$ do
    $(t_{\mathrm{init}}, t_{\mathrm{lead}}, m) \leftarrow \boldsymbol{s}_{\mathrm{blockstart}}$    // Unpack coordinate vector
    $S \leftarrow S \oplus [(t_{\mathrm{out}} = t_{\mathrm{current}} + j,\ (t_{\mathrm{init}}, t_{\mathrm{lead}} + j, m))]_{j=0}^{\tau-1}$    // Follow reforecast and concatenate with generated time series
    $t_{\mathrm{current}} \leftarrow t_{\mathrm{current}} + \tau$    // Update output time counter
    $\boldsymbol{s}_{\mathrm{next}} \leftarrow (t_{\mathrm{init}}, t_{\mathrm{lead}} + \tau, m)$    // Determine "true" next state in reforecast dataset
    $\mathcal{C} \leftarrow \{\boldsymbol{s} \in \mathcal{C}_{\mathrm{all}} \mid d_{\mathrm{date}}(t_{\mathrm{valid}}(\boldsymbol{s}), t_{\mathrm{valid}}(\boldsymbol{s}_{\mathrm{next}})) \leq \Delta\}$    // Form set of "candidate" analogs for $\boldsymbol{s}_{\mathrm{next}}$ based on date similarity
    $\mathcal{C} \leftarrow \{\boldsymbol{s} \in \mathcal{C} \mid d_{\mathrm{date}}(t_{\mathrm{valid}}(\boldsymbol{s}), t_{\mathrm{current}}) \leq \Delta\}$    // Restrict set of candidates to have date similar to assigned output date
    $\mathcal{D} \leftarrow \{d(\boldsymbol{x}(\boldsymbol{s}), \boldsymbol{x}(\boldsymbol{s}_{\mathrm{next}})) \mid \boldsymbol{s} \in \mathcal{C}\}$    // Compute (or load precomputed) distances
    Use $\mathcal{D}$ to sample $\boldsymbol{s}_{\mathrm{blockstart}}$ from $p_\sigma$, e.g., according to Eq. (2)
end for
return Time series $S = \left[\left(t_{\mathrm{out}}^{(1)}, \boldsymbol{s}^{(1)}\right), \left(t_{\mathrm{out}}^{(2)}, \boldsymbol{s}^{(2)}\right), \ldots, \left(t_{\mathrm{out}}^{(n_{\mathrm{blocks}}\tau)}, \boldsymbol{s}^{(n_{\mathrm{blocks}}\tau)}\right)\right]$
```

too, provided that the parameters only influence transitions between reforecast segments. Appendix A3 further details the forecast-based parameter optimization workflow and its results.

Time series generated with small block sizes τ lack temporal autocorrelation due to frequent transitions between reforecast blocks (Fig. 3b). While large-scale circulation variables like $Z_{500\mathrm{hPa}}$ constrain local surface variables to some extent, there is remaining variability (Cariou et al., 2025). Additionally, memory is lost during transitions between imperfect analogs. Decreasing the number of transitions by increasing the block size τ partially resolves this problem (Fig. 3b). However, setting the block size close to the number of days in each reforecast (i.e., close to 45) restricts the weather generator's flexibility, because larger blocks reduce the number of possible sequences and tie generated time series more closely to the weather evolution in the reforecast dataset. As a compromise between enhancing flexibility and preserving autocorrelation, we choose $\tau = 30$ as a default.

Unless stated otherwise, all following analyses are based on 500 21-year time series generated with the default setup of unseen-awg ($\sigma = 2.5\,\mathrm{dam}$, $\tau = 30$, and $\Delta = 10$).

3 Results

Combining the analog weather generator method with extensive reforecast data rather than a comparatively short historical dataset has notable advantages. Weather simulated with unseen-awg can exceed observed records on a daily scale. For example, generated time series can include unprecedented high and low temperature values (Fig. 4a).

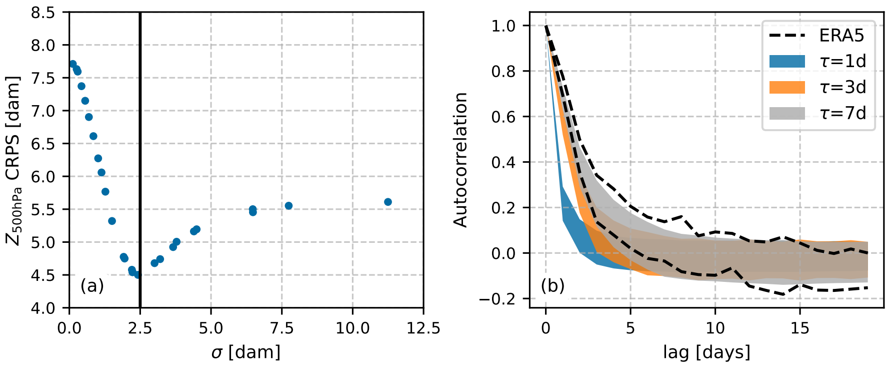

Figure 3. (a) Using the weather generator as an ensemble-forecast model for varying σ reveals an optimum in forecast skill (minimum of area-weighted average CRPS) around $\sigma = 2.5\,\mathrm{dam}$, which is subsequently used as the default value in unseen-awg. (b) Interquartile ranges for the temporal autocorrelation of $T_{\mathrm{2m,mean}}$ in the grid cell closest to Leipzig, computed over summer (JJA). For the ground truth ($\mathrm{ERA5}_{2003-2023}$, dashed), the autocorrelation is estimated within each summer and interquartile ranges are computed over years; for data simulated with unseen-awg (filled regions), quartiles are computed over all years in 500 generated 21-year time series.

An additional effect of the larger dataset size is that one can typically find better (closer) analogs among the reforecasts than in reanalysis datasets (Fig. 4b). Smaller distances to analogs imply transitioning to more similar states when sampling analogs (Fig. 4c-g), minimizing jumps between blocks of reforecast days. In the analyses presented in Fig. 4, we excluded candidate analogs with t_{valid} closer than 180 days to t_{valid} of the reference state – this is an optional configuration in unseen-awg that is not active by default (Appendix A2). Representing dependencies between variables and locations is important for capturing multivariate and spatially compounding events (Zscheischler et al., 2018), but is a challenge for many weather generators. Therefore, after validating its annual cycle and grid-cell-wise statistics in Sect. 3.1, we test unseen-awg's ability to represent the dependence between summer temperature and precipitation in Sect. 3.2, and explore whether it is able to simulate unprecedented spatially-extensive droughts in Sect. 3.3.

3.1 Annual cycle and marginal distributions

With the default parameter settings, unseen-awg well reproduces the annual cycle of $\mathrm{ERA5}_{2003-2023}$ (Appendix B1, Fig. B1). In particular, generated temperature time series are visually similar to their $\mathrm{ERA5}_{2003-2023}$ counterparts and there is no "lag" visible between the climatology of $\mathrm{ERA5}_{2003-2023}$ and the generated time series for all but unrealistically extreme settings of the weather generator parameters.

We further compare unseen-awg's ability to reproduce different quantiles of temperature and precipitation at the grid-cell scale. To this end, we compute quantiles of daily mean temperatures and wet-day precipitation amounts for unseen-awg simulations and the $\mathrm{ERA5}_{2003-2023}$ ground-truth data (the latter are shown in Fig. 5a–e, p–t). Deviations from the ground-truth data vary regionally and with quantiles (Fig. 5, f–j, u–y). For daily mean temperature, deviations are strongly reduced if the

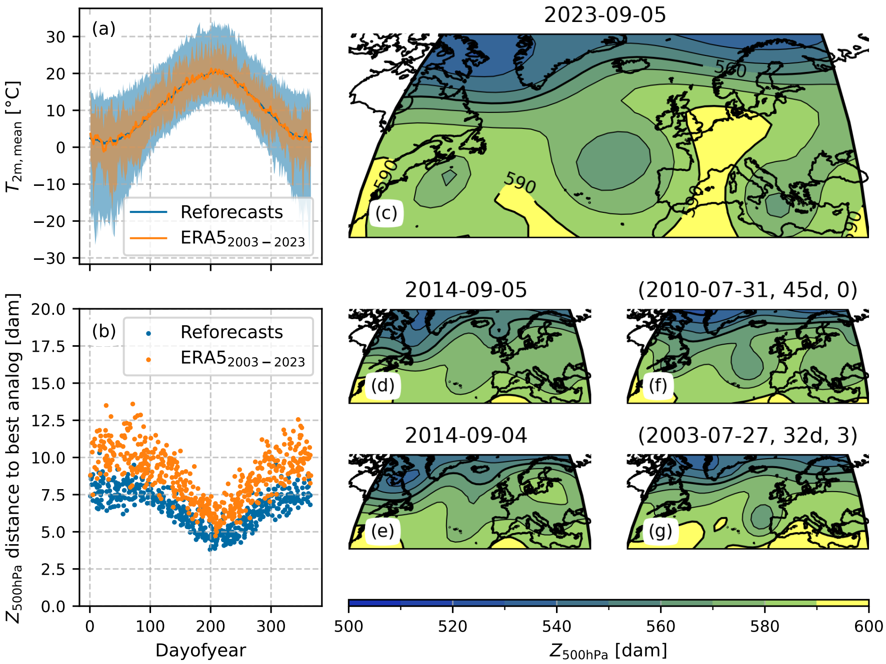

Figure 4. (a) Maximum and minimum daily mean 2 m temperature for each day of the year in the $0.4^\circ \times 0.4^\circ$ grid cell that includes Leipzig, Germany, for $\mathrm{ERA5}_{2003-2023}$ (orange) and reforecast data (blue). (b) Distance to the best analog for $\mathrm{ERA5}_{2003-2023}$ (orange) and reforecast data (blue). (c) Strong omega block in the $Z_{500\mathrm{hPa}}$ field on September 5, 2023. (d) Best and (e) second best analogs in $\mathrm{ERA5}_{2003-2023}$ to the event shown in (c). (f) Best and (g) second best analogs in the reforecast dataset to the event shown in (c). The tuples in the titles of (f) and (g) indicate forecast initialization time, lead time, and ensemble member.

reforecast data are bias corrected (Fig. 5k–o). Bias correction substantially reduces quantile deviations for daily minimum and maximum temperatures as well (not shown). For wet-day precipitation, relative deviations increase with increasing quantiles (Fig. 5u–y). Small-scale spatial structures are visible in the deviations and there is an overall tendency towards overestimated wet-day precipitation in generated data. We found bias estimation for precipitation to be unreliable (Sect. 2.1, Appendix A1); therefore, we do not apply bias correction for precipitation. We ran an unseen-awg weather generator with $\mathrm{ERA5}_{2003-2023}$ data instead of reforecast data and found strongly reduced quantile deviations for wet-day precipitation (Fig. C1). This suggests that deviations in the quantiles are primarily due to differences in the underlying dataset, not the weather generator methodology, observed discrepancies might be attributed to either systematic biases or the limited sample size of the $\mathrm{ERA5}_{2003-2023}$ dataset. Results are qualitatively similar when analyzing each season separately (not shown). As an additional diagnostic, we look at the distributions of daily mean 2 m temperature and total precipitation for selected grid cells by using kernel density estimation and quantile–quantile plots. For a sample grid cell containing Leipzig, Germany, the properties of the generated time series closely match those of the ground-truth dataset across all seasons (Fig. C2).

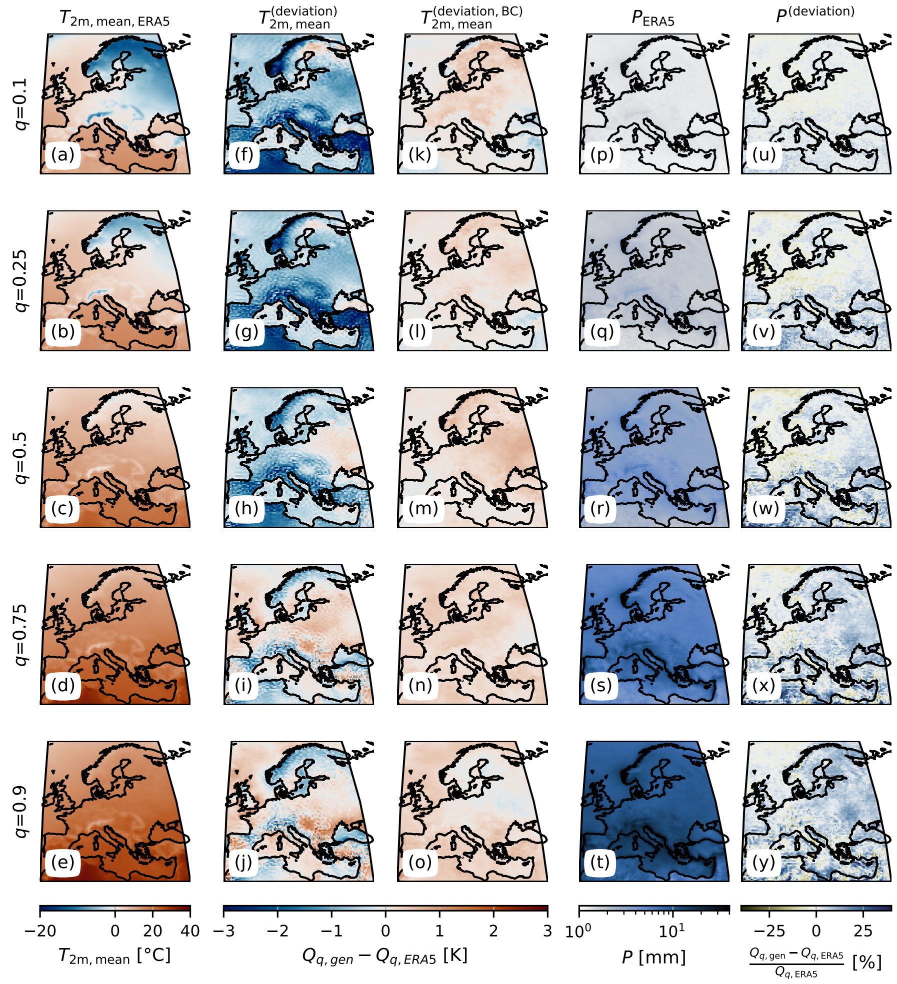

Figure 5. Empirical quantiles Q_q for $T_{\mathrm{2m,mean}}$ and total precipitation P (computed over days with $P \geq 1\,\mathrm{mm}$) in the $\mathrm{ERA5}_{2003-2023}$ dataset (panels a–e, p–t) and average deviation of the same quantiles in 500 generated 21-year time series (f–j, k–o, u–y). Deviations are computed additively for $T_{\mathrm{2m,mean}}$ (f–j, k–o) and as relative deviations for P (u–y). Panels (k)–(o) show the quantile deviations after simple additive bias correction has been applied to reforecast $T_{\mathrm{2m,mean}}$.

3.2 Correlation between summer temperature and precipitation

Capturing dependencies between seasonally aggregated variables with an analog weather generator is challenging: analogs – and consequently the transitions between blocks of reforecast data – are imperfect, and Z_{500hPa} does not fully determine local weather. At the same time, overlooking the dependencies between aggregate variables can distort estimated likelihoods of multivariate events. Due to the relevance of the dependence between summer temperature and precipitation for the likelihood of compound hot and dry summers (Zscheischler and Seneviratne, 2017), we compare the joint distributions of June-July-August (JJA) precipitation sums and JJA mean temperatures between $\mathrm{ERA5}_{2003-2023}$ and unseen-awg simulations.

When using a small block size ($\tau = 1$), the spread in the marginal distributions and the negative correlation between the seasonally aggregated variables appear underestimated; see Fig. 6a for an example in the grid cell containing Leipzig, Germany. These deficiencies are mitigated when using a larger block size (Fig. 6b). In particular, the average strength of the negative correlation over Europe tends to increase with increasing τ (Fig. 6c–f). A direct comparison with the correlation coefficients estimated from $\mathrm{ERA5}_{2003-2023}$ (Fig. 6g) is not advised due to ERA5's substantially smaller sample size which leads to very uncertain estimates (Bevacqua et al., 2023). We therefore use the rank $r \in [1, N+1]$ of the $\mathrm{ERA5}_{2003-2023}$ correlation coefficient among the coefficients computed from the N generated time series in each grid cell to compute p-values under the null hypothesis that the $\mathrm{ERA5}_{2003-2023}$ correlation coefficient is from the same distribution as the generated coefficients:

$$p = \min\left(1, \frac{2\min(r, N-r+2)}{N+1}\right). \tag{3}$$

The proportion of grid cells for which the null hypothesis is rejected at the $\alpha = 0.05$ significance level decreases with increasing τ (Fig. 6h–k): from 18.0 % of valid grid cells when using $\tau = 1$, to 10.0 % for $\tau = 30$ and further down to 9.1 % for $\tau = 44$.

3.3 Spatially extensive droughts

Europe experienced several large-scale high-impact drought events in the last decade, including events in 2018–2020 and 2022 (Bakke et al., 2020; Rakovec et al., 2022; Bevacqua et al., 2024). Droughts often last several months or even years, which means long temporal windows should be used for drought analysis. We construct a large dataset of spatially connected drought events generated with unseen-awg and test whether the distribution of the spatial extents of the events matches that obtained from $\mathrm{ERA5}_{2003-2023}$. We further explore generated events that exceed the observed maximal drought extent.

The following steps are performed per grid cell and per 21-year time series for both the 500 generated time series and the $\mathrm{ERA5}_{2003-2023}$ ground truth. We compute monthly precipitation sums and aggregate them into 12-month running totals. Given the running totals, we compute the empirical 20th percentile (per 21-year time series). For each time step in each time series, we then aggregate spatially connected grid cells for which total precipitation is below the 20th percentile and record the area A of the largest connected component (more details in Appendix B2).

We compare the distribution of A, the area of the largest connected drought event in each time step, between generated and ground-truth data. The largest connected 12-month drought event for $\mathrm{ERA5}_{2003-2023}$ occurs between January 2018 and January 2019 (Fig. 7a), which is a well-known historical event (Bakke et al., 2020). In the generated data, we find a roughly 50 % larger event, covering most of France and large parts of Scandinavia, Central, Eastern, and Southeastern Europe (Fig. 7b). Quantiles of the empirical distributions of A in $\mathrm{ERA5}_{2003-2023}$ fall inside the range of the area quantiles derived from the 500 generated time series, suggesting that there is no systematic under- or overestimation of drought extents in the generated data (Fig. 7c). Altered definitions of drought events (varying the threshold for defining droughts) yield qualitatively similar

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

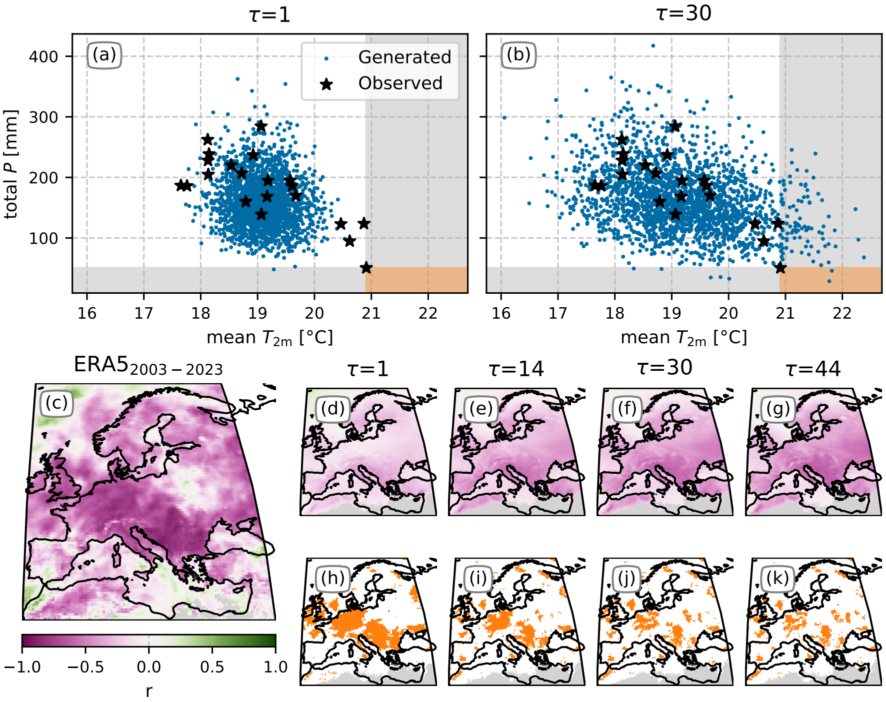

Figure 6. (a) Averaged daily mean 2 m temperature and precipitation totals of the 21 summers (JJA) in ERA5$_{2003-2023}$ in the grid cell closest to Leipzig, Germany (black stars) and in the 500 generated 21-year time series with block size $\tau = 1$ (blue markers). (b) Same as (a) but for $\tau = 30$. Gray and orange regions indicate unprecedented hot, dry, or hot-dry summers. (c) Pearson correlation between detrended JJA averaged daily mean 2 m temperature and detrended JJA total precipitation, estimated over the 21 years in ERA5$_{2003-2023}$. (d)–(g) Same for generated time series; shown is the mean over the 500 estimated correlation coefficients. (h)–(k) For each grid cell, p-values are estimated from the rank of the ERA5$_{2003-2023}$ correlation coefficient among the 500 generated coefficients. The grid cell is colored orange if $p < 0.05$. A cell is colored gray if estimating a correlation coefficient is impossible for at least one time series.

results (Appendix B2). We also found comparable results performing a related analysis using single-day spatially coniguous heat extremes (Appendix B2). This indicates that unseen-awg is able to simulate droughts and heatwaves with unprecedented spatial extent, while the analysis on the quantiles supports the realism of the generated event magnitudes.

4 Discussion

The outlined analog weather generator is able to generate data for a present-day climate and incorporates mechanisms for preserving autocorrelation as well as multivariate and spatial dependencies over Europe. By resampling 30-day stretches and using reforecast data, we are able to substantially improve the sampling pool and temporal consistency of unseen-awg. The weather generator can be used to assess present-day risks of complex compound events that extend in time, space, or across

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

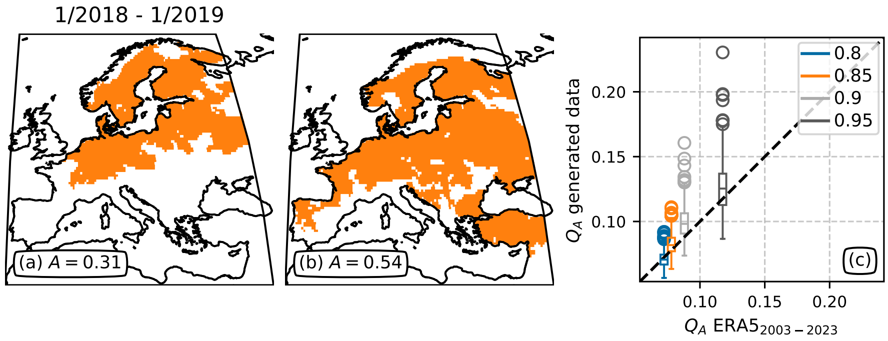

Figure 7. (a) Largest connected component in which every land grid cell has an annual precipitation total below its 20th percentile in the $\mathrm{ERA5}_{2003-2023}$ time series of monthly rolling 12-month precipitation sums. (b) Same for time series generated with unseen-awg using quantiles computed on the generated time series. (c) Distribution of connected areas under drought: The x-axis indicates the quantiles Q_A of A, the area covered by the largest connected drought event in each time step, computed over the $\mathrm{ERA5}_{2003-2023}$ dataset. The box plots show the respective quantiles of A in the 500 time series generated with unseen-awg.

variables. Furthermore, the output of the weather generator can be used to force hydrological, vegetation or agricultural models to study weather-related impacts such as floods, forest mortality and crop failure risks. In this way, the weather generator can also be used for stress testing climate-sensitive system, for instance by generating plausible storylines of worst-case events (Sillmann et al., 2021).

Preserving autocorrelation (Sect. 2.3) and simulating realistic aggregated variables such as seasonal averages (Sect. 3.2) are key challenges for dynamic analog weather generators. The large-scale atmospheric circulation poses a constraint on local weather whose strength depends on scale, variable, region, and process (Cariou et al., 2025; Thompson, 2025). Sampling data in blocks of days can improve the representation of processes that are only weakly constrained by the circulation. Identifying high-quality analogs in large-scale atmospheric fields has long been known to be challenging (Lorenz, 1969). The distance to the closest analog scales exponentially with dimensionality of the system (van den Dool, 1994); therefore increasing the dataset size has limited effects ("curse of dimensionality"). Indeed, the distances to the closest analogs in the reforecast dataset are smaller than those in the $\mathrm{ERA5}_{2003-2023}$ dataset, but of the same order of magnitude (Fig. 4b). Although the reforecast dataset is much larger, typical distances are still substantial: when excluding data from within 180 days of the reference state, the distance to the best analog is comparable to the distance between two ensemble forecast members at a lead time of roughly 6 days (Fig. B4, further analyses in Appendix B3). Computing analogs over a smaller area reduces the number of degrees of freedom, but this improves analog quality at the cost of weaker constraints on the dynamical evolution of the system (van den Dool, 1994). Analog approaches in attribution or forecasting usually operate with sets of analogs (van den Dool, 1994; Lguensat et al., 2017; Platzer et al., 2021), but these methods do not transfer easily to the setting of generating a single realistic weather time series. Despite these known limitations, we illustrate that using reforecast data and longer stretches before switching to the next analog substantially improves key properties of the generated weather.

The fidelity of the weather generated by unseen-awg is tied to the fidelity of the underlying dataset, which has known limitations compared to real observations, in particular for precipitation (Lavers et al., 2022). However, by their nature, reforecasts are continuously validated in an operational setting, and they are more closely tied to real-world climate variability than free-running simulations with climate models (Kelder et al., 2020). Caution is advised when interpreting return levels of generated data: reforecast data do not necessarily constitute a representative sample of possible weather states. For small lead times, members of the same ensemble reforecasts are similar to each other and to the ground truth, causing a denser sampling of this subregion of possible weather than expected. On the other hand, some plausible weather states may be never be generated by unseen-awg if their occurrence depends on values of long-term oscillations in the climate system that were not observed between 2003 and 2023.

Generating weather data with unseen-awg is fast: sampling a 21-year time series with default parameter choices and using precomputed similarities takes less than 5 minutes on a single CPU core and uses less than $1\,\mathrm{GB}$ of memory. Each generated time series can be stored as a look-up table (Sect. 2.2) and requires only $300\,\mathrm{KB}$ of memory. However, storing the precomputed distances requires roughly $500\,\mathrm{GB}$ of disk space when using the full reforecast dataset. Distances can also be computed on the fly, but this slows down simulations by orders of magnitude. The size of the reforecast dataset poses challenges. Computational and memory footprints of the weather generator scale linearly with dataset size when lazy computations are used and quadratically if distances are precomputed. While precomputing is already demanding given the size of reforecast dataset, it may become prohibitively expensive for even larger datasets.

There are rapid advances in the application of machine learning to forecasting and emulating the climate system. While ML weather prediction models reach impressive forecast accuracy, some tend to produce blurry atmospheric fields for long time series (Lam et al., 2023) or become unstable (Karlbauer et al., 2024). The performance of ML methods on record-breaking extreme events remains uncertain (Zhang et al., accepted), which are of particular interest when studying impacts. ML-based emulators of the atmosphere or coupled atmosphere-ocean models exist (Kochkov et al., 2024; Watt-Meyer et al., 2025; Cresswell-Clay et al., 2025). However, they are often coarsely resolved, may lack variables of interest, and in some cases they exhibit limited long-term stability (Kochkov et al., 2024). More work is needed to assess their ability to reproduce extreme weather.

Future work may try to improve unseen-awg through more elaborate measures of distance. A low-dimensional representation of the atmospheric circulation might be obtained through dimensional reduction methods like principal component analysis/empirical orthogonal functions (e.g., Zorita and von Storch, 1999) or using machine learning. Ideally, distance measures should depend on the day of year or even the last state included in a time series. However, learning such a measure from data is challenging. Future research could focus analog computation on impact-related variables by learning a tailored representation from $Z_{500\mathrm{hPa}}$ fields or by including more variables in the similarity computation (e.g., Horton et al., 2017). Including covariates to better capture slow modes of variability is also conceivable and computing analogs between sequences of days instead of single days is already possible within our framework. Alternative reforecast datasets are available with longer lead times and lower spatial resolution, or shorter lead times and higher resolution (Vitart and Robertson, 2018; ECMWF, 2023). Depending on the problem at hand, different choices might be desirable. Coupling unseen-awg with impact models can be used to guide the selection of evaluation metrics for the weather generator.

5 Conclusion

We present an approach for generating long stretches of spatiotemporal weather data over Europe representative of present-day climate. Our method, unseen-awg, combines an analog weather generator with unseen data from a large reforecast dataset. It thus bridges recent approaches for anticipating unprecedented weather (Kelder et al., 2025) with the analog approach. By enabling the creation of large forcing datasets for climate impacts models, one can use it to assess present-day climate risks across different climate-sensitive sectors. To ease user uptake, we develop a flexible python framework that implements unseen-awg, present suggestions for optimizing parameters, and evaluate it from multiple points of view and across a set of analyses directly tied to weather-related impacts.

Unseen-awg inherits challenges from the approaches it builds on: it uses weather forecast data instead of observations, and generated time series include imperfect transitions between blocks of analog states. At the same time, it also has clear advantages over the individual approaches: arbitrarily long time series can be generated, we can find better analogs than in observational datasets, and unlike traditional analog weather generators, the generated time series can contain unprecedented events at daily scale.

Given this trade-off and the necessary careful validation of the weather generator's output for each system of interest, unseen-awg holds promise for studying many types of impacts, especially various types of compound events for which capturing spatiotemporal aspects in the analysis matters. To this end, we provide more than 10,000 years of synthetic daily European weather data, including total precipitation and minimum, maximum, and mean temperature. Further variables can be downloaded from ECMWF's operational archive. We also release a weather generator instance with precomputed distances for rapid generation of additional data, as well as the preprocessed reforecast dataset needed to develop new weather generators.

Appendix A: Methodology

A1 Bias estimation

We estimate biases for each impact-relevant variable, location, day of year, and lead time separately. First, climatologies are computed for $\mathrm{ERA5}_{2003-2023}$ and the reforecast dataset per variable, location, day of year and lead time. Forecast initializations in the reforecast dataset are not uniformly distributed. Therefore, to compute daily climatologies we group data for a given location and lead time by the day of year of its valid time ($t_{\mathrm{valid}} = t_{\mathrm{init}} + t_{\mathrm{lead}}$) and then compute the average within each group. We impute values for days of year with no data using linear interpolation and periodic boundary conditions to bridge the year-end discontinuity. Finally, we smooth the resulting climatology by applying a temporal rolling mean with a default window size of 61 days (half-width of 30 days). To prevent inconsistent sample sizes from biasing the estimates, we only use a subset of $\mathrm{ERA5}_{2003-2023}$ by restricting its temporal domain to those time steps available in the reforecast dataset for a given lead time. Biases are then computed additively for temperature variables and as relative biases for precipitation:

$$\begin{aligned}
\mathrm{Bias}_{T_{2\mathrm{m}}}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})} &= \overline{T_{2\mathrm{m},\mathrm{reforecasts}}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})}} - \overline{T_{2\mathrm{m},\mathrm{ERA5}_{2003-2023}}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})}} \\
\mathrm{Bias}_{P}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})} &= \overline{P_{\mathrm{reforecasts}}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})}} \,/\, \overline{P_{\mathrm{ERA5}_{2003-2023}}^{(\mathrm{doy},\mathrm{loc},t_{\mathrm{lead}})}} - 1
\end{aligned}$$

Overlines indicate climatological means, doy and loc are shorthands for day of year and location, and the $\mathrm{ERA5}_{2003-2023}$ quantities include a t_{lead} index due to subsetting to available reforecast dates. The bias computation for $2\,\mathrm{m}$ temperatures is the same for $T_{2\mathrm{m},\mathrm{min}}$, $T_{2\mathrm{m},\mathrm{max}}$, and $T_{2\mathrm{m},\mathrm{mean}}$. For precipitation, we include all days (i.e., also dry days) in the computation of the climatology and the bias.

To test the reliability of the bias estimation, we split the reforecast dataset into subsets and assess how similar the estimates from different subsets are. This is done by randomly splitting it into two halves by its initialization time steps t_{init}. Qualitatively, patterns seem consistent between the subsets for $2\,\mathrm{m}$ temperature biases (Fig. A1 for $T_{2\mathrm{m},\mathrm{mean}}$), but for precipitation, the bias estimation yields patterns that vary strongly between the two random subsets (Fig. A2), with especially large variability in dry regions. Results for $T_{2\mathrm{m},\mathrm{min}}$ and $T_{2\mathrm{m},\mathrm{max}}$ are comparable to those for $T_{2\mathrm{m},\mathrm{mean}}$ (not shown). The main findings do not change when t_{lead} is held constant instead of the day of year or when chronological splitting is used instead of random splitting (not shown).

A2 Restrictions on the set of candidate states

In addition to the constraints presented in Sect. 2.2 (an analog must be a valid start of a new block; its calendar has to be within a window of half-width Δ around both the targeted output date and the valid time of the true next state), users can further restrict the set of candidates. These additional restrictions can be set for each simulated time series separately. By default, only the constraints presented in Sect. 2.2 are employed.

Useful additional implemented restrictions include avoiding direct repeats of the same coordinate vector $\boldsymbol{s}_{\mathrm{last}}$, assuring a minimal calendar distance from $t_{\mathrm{valid}}(\boldsymbol{s}_{\mathrm{next}})$, or disallowing the selection of analogs past a certain date. The last option is particularly useful when employing the weather generator as a forecast model; the second-to-last could be used to avoid sampling analogs from within the same year.

In practice, the constraints are implemented within the module that defines probability models, e.g., together with the one defined in Eq. (2), by setting the probability of sampling a state for which the condition is not fulfilled to zero (or equivalently setting the log-probability to $-\infty$).

A3 Parameter tuning

To select values of the parameter σ in Eq. (2), we use the heuristic that a weather generator that produces good ensemble forecasts should also generate more realistic weather time series than a weather generator with bad forecast skill. As described in Appendix A2, it is possible to disallow analogs with a valid time later than a fixed date from being included in a generated

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

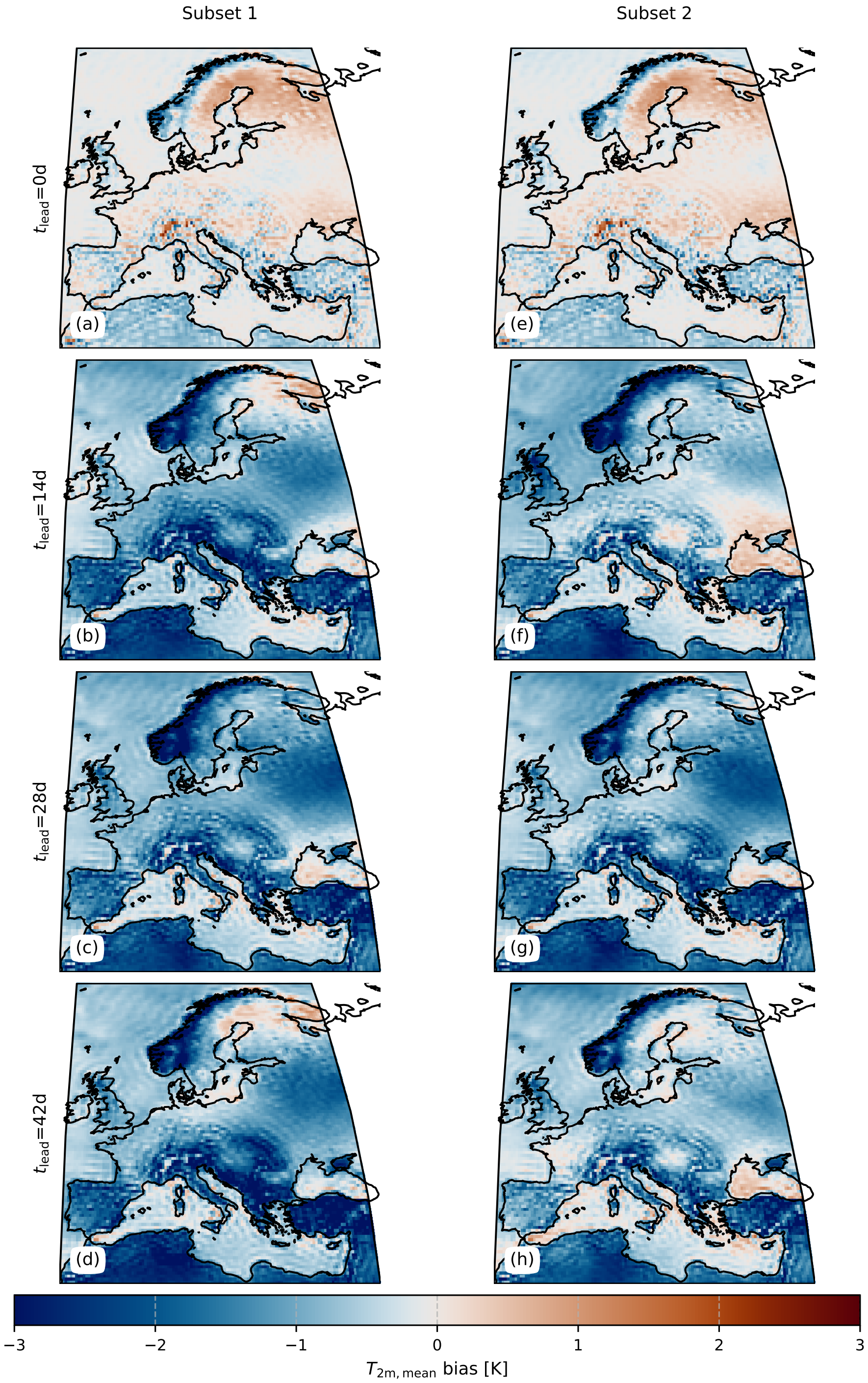

Figure A1. Additive $T_{\mathrm{2m,mean}}$ bias (reforecast climatological mean minus $\mathrm{ERA5}_{2003-2023}$ climatological mean) for 1 January and varying lead-time t_{lead}. The reforecast dataset was split into subsets 1 and 2 randomly along the t_{init} dimension and the bias was computed for each subset separately.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

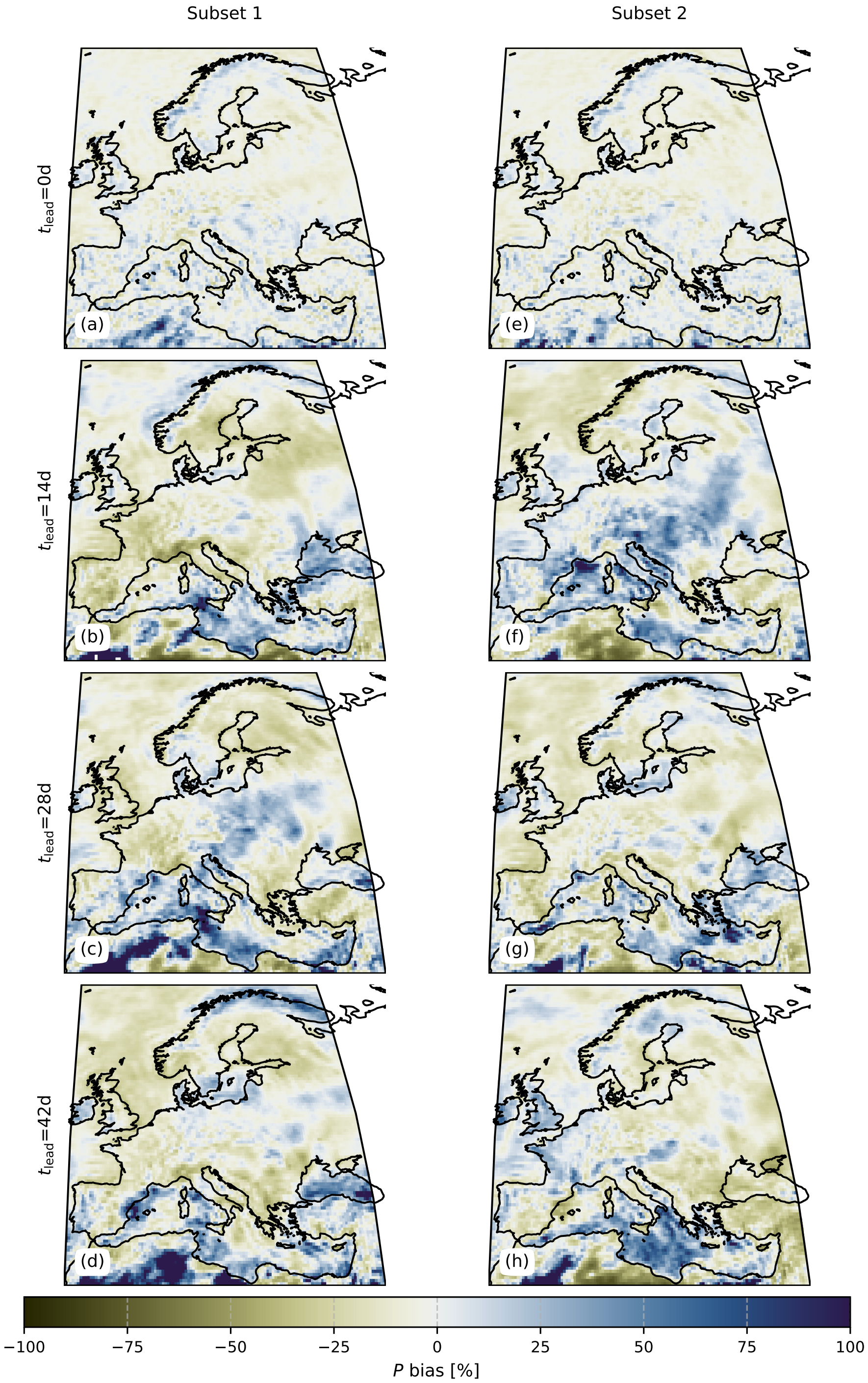

Figure A2. Relative mean precipitation bias (reforecast climatological mean divided by $\mathrm{ERA5}_{2003-2023}$ climatological mean minus 1) for 1 January and varying lead-time t_{lead}, computed including both wet and dry days. The reforecast dataset was split into subsets 1 and 2 randomly along the t_{init} dimension and the bias was computed for each subset separately.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

time series. We set this fixed date as the start dates of the analog ensemble weather forecasts, thereby preventing the weather generator from using "future weather" in its generated time series.

We use all weather states with $t_{\text{lead}} = 0$ and unperturbed ensemble member $m = 0$ to construct a ground-truth dataset. We initialize the analog forecasts with a state in this ground-truth dataset, generate a forecast time series using the block size $\tau = 1$, and compare results against the corresponding sample in this ground-truth dataset. This results in a restriction on the possible analog forecast lengths, because ground-truth states are typically separated by intervals of 3 or 4 days (for most weeks, our reforecast dataset contains two reforecast starting dates, see Sect. 2.1).

We choose 500 initial conditions randomly among all points in the ground-truth dataset for which we also have a valid ground-truth data point against which we can compare predictions (e.g., for a possible initial time t_0 and 3-day forecasts, we make sure that also the sample 3 days after t_0 is in the ground-truth dataset defined above). We also ensure that all months appear roughly equally often in the set of initial conditions. Ensemble forecasts are created using 80 ensemble members and compared to two baselines: persistence and (probabilistic) climatological forecasts. The climatological forecast uses the empirical distribution over all past data with the same day of year as the target t_{valid} as its prediction, while the persistence forecast simply returns its initial condition. Forecast performance is quantified by the area-weighted mean Continuous Ranked Probability Score (CRPS; Gneiting and Raftery, 2007).

For small σ, the analog forecast behaves like a persistence forecast. This is because in this case the weather generator chooses the most similar state to be the true next state with exceedingly high probability and, with the restriction of not sampling from the future, the most similar state to $(t_{\text{init}}, t_{\text{lead}} = 1, m = 0)$ will most likely be $(t_{\text{init}}, t_{\text{lead}} = 0, m = 0)$ or potentially some other ensemble member m'. When σ is large, the weather generator will assign roughly equal probabilities to each candidate state; therefore, the behavior gets similar to climatology forecasts (Fig. A3). At intermediate scales, we find an optimum σ^*, for which the analog CRPS is smaller (better) than that of the persistence and climatology forecasts. There are subtle differences between the two edge case behaviors of analog forecasts, and persistence and climatology forecasts respectively: for very small σ, we tend to find a slightly better forecast skill than for persistence, because the analog forecast is effectively the closest analog of $(t_{\text{init}}, t_{\text{lead}} + 1, m)$ in the dataset before the forecast initialization. Therefore, the analog weather generator has a small advantage through this one-day "look-ahead". For very large σ, differences in skill compared to climatology arise from the limited size of the analog ensemble forecast (which leads to an incomplete estimation of the full climatological distribution).

For a forecast lead time of 3 days, $\sigma^* \approx 2.5\,\text{dam}$ (Fig. A3a), while for a weather generator using the same methodology but $\text{ERA5}_{2003-2023}$ data, we find $\sigma^* \approx 4.0\,\text{dam}$ (Fig. A3c). The optima stay at similar values when evaluating 4-day forecasts instead (Fig. A3b, d). One would expect a smaller optimal sigma for a dataset with better analogs (i.e., closer nearest neighbors); indeed, we observe a smaller optimal σ for the reforecast dataset compared to when $\text{ERA5}_{2003-2023}$ is used. Distances to the best analogs in these datasets are studied in Appendix B3.

Other parameters could be tuned similarly, e.g., one could also compare alternatives to the default choices for computing distances between analogs and for assigning probabilities based on the distances in Eq. (2).

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

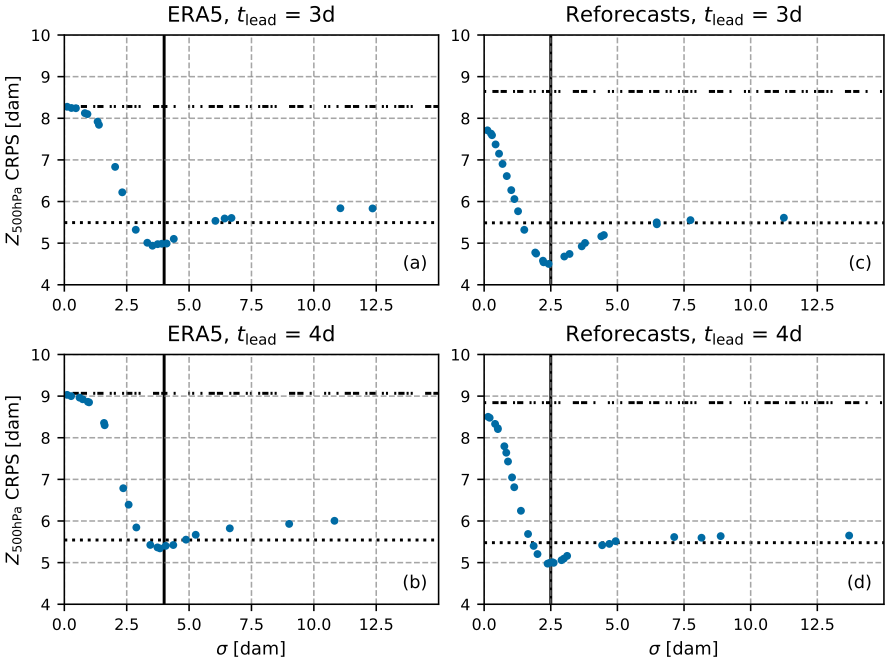

Figure A3. Tuning the σ parameter in Eq. (2) of unseen-awg with the goal of optimizing 3-day or 4-day forecast accuracy. (a, b) Weather generator using $\mathrm{ERA5}_{2003-2023}$. (c, d) Weather generator using reforecast data; (c) shows the same data as Fig. 3a. Each data point is the area-weighted CRPS-average over the entire spatial domain and over forecasts started from 500 randomly selected initializations. Dash-dotted line: average CRPS of persistence forecasts for the same set of initial conditions. Dotted line: same for probabilistic climatological forecasts. Optuna (Akiba et al., 2019) was used to efficiently explore plausible values of σ, vertical lines indicate default choices of $\sigma = 2.5\,\mathrm{dam}$ for the reforecast-based weather generator and $\sigma = 4.0\,\mathrm{dam}$ for the ERA5-based weather generator.

Appendix B: Additional analyses

B1 Annual cycle

The generated time series do not show a major lag in their annual cycle when compared to $\mathrm{ERA5}_{2003-2023}$ (Fig. B1a). For an arbitrary year and the grid cell closest to Leipzig, Germany, true and generated time series are visually similar (Fig. B1b). Analyzing the differences between the day of year of t_{valid}, the true date of the sample within the reforecast dataset, and day of year of t_{out}, the date that gets assigned to the sample in the generated time series, we obtain a symmetric histogram with most cases showing small lags (Fig. B1c). Here, we exclude days with t_{out} in DJF from the time series to simplify the analysis.

Under extreme conditions (σ very small, $\tau = 1$), we obtain a skewed distribution (Fig. B1d, blue histogram). In this case, because of the small σ, the generated time series follow a single reforecast without doing any transitions until there are no more forecast days left that could be included next (i.e., reaching the sample with $t_{\mathrm{lead}} = 44$). In some cases, this last state in the reforecast will itself be the best analog of its successor and will therefore be repeated in the time series with high probability.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

This leads to the days of year of t_{valid} lagging behind those of t_{out}. The window of half-width Δ centered on t_{out} restricts how large these lags can get. It is possible to choose a configuration of the weather generator that precludes repeats by excluding them from the list of valid candidates in the sampling process (see Sect. A2), but as this is not a concern for default parameter settings, the setting is not active by default.

Having a windowed selection of analogs centered on t_{out} partially enforces that generated time series follow an annual cycle. Having a small window is therefore beneficial, but it reduces the set of valid analogs. This is less of a problem in the much larger reforecast dataset than when working with $\text{ERA5}_{2003-2023}$ or observational data.

B2 Spatially compounding extreme events

After binarized datasets are obtained as described in Sect. 3.3, we aim to identify spatially connected components within each 12-month aggregate time step. To do this, we use the scikit-image python library (van der Walt et al., 2014) and search for connected components using 8-connectivity. Only land area is included in the dataset while oceans are masked out. We compute the area-weighted size of each component and only keep the largest event for each time step. We compute the area of the largest component relative to the total land area within our study domain. We repeat the analysis in Sect. 3.3 using different quantile thresholds: above, we defined drought conditions as having a 12-monthly precipitation sum below the 20th percentile in the respective grid cell. We repeat the analysis for the 10th and 5th percentile, obtaining comparable results (Fig. B2).

We also analyze spatially connected heat events using a similar methodology. To study single hot days, the following operations are performed per grid cell and time series (for each generated and $\text{ERA5}_{2003-2023}$ ground-truth time series). We filter the time series to only include days in JJA and compute the empirical 90th, 95th and 98th percentile over the selected subset. We then binarize the time series as above depending on whether the corresponding percentile is exceeded and extract the largest spatially connected components within each day. Again, the Q–Q plots computed from the distribution of true and generated event area A over time do not indicate clear trends towards overestimation or underestimation. For the $\text{ERA5}_{2003-2023}$ reanalysis data, the largest connected event occurs during the 2010 heat wave for all tested thresholds (Fig. B3). We observe substantially larger extents in the generated dataset. Interestingly, the same state in the reforecast dataset ($t_{\text{init}} = $ 29 July 2012, $t_{\text{lead}} = 28, m = 8$) gets selected for all three thresholds.

To study short events like daily heat extremes it is not strictly necessary to use unseen-awg as this application is not limited by the maximum reforecast lead time in the reforecast dataset. Potentially, the weather generator helps produce a more realistic distribution by avoiding oversampling dependent states at small lead times but one would have to study this more systematically.

B3 Distances to best analogs

While the distance between a base state and its best analog (nearest neighbor) is smaller than the distance to a random analog (a random state from a similar calendar date), both distances are still on a similar order of magnitude: the average distance to the best analog is roughly $40\,\%$ of the distance to a random analog in the reforecast dataset (Table B1, roughly $60\,\%$ in the $\text{ERA5}_{2003-2023}$ dataset). We exclude potential analogs within 180 d of each reference state to avoid sampling strongly dependent atmospheric conditions.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

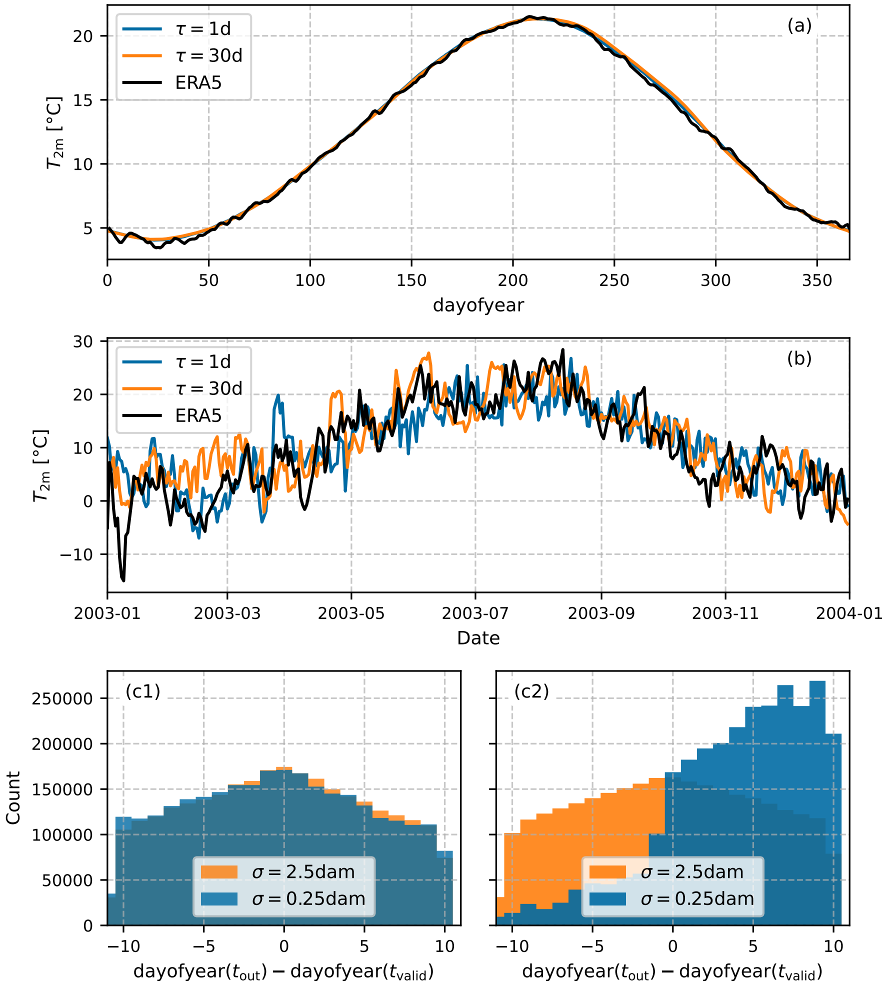

Figure B1. Comparing the annual cycle of generated time series to ERA5 (2003–2023). For all panels and each configuration, we generated 500 21-year time series using unseen-awg with default parameter values unless mentioned otherwise. (a) True and generated $T_{\mathrm{2m,mean}}$ for each day of the year, averaged over time, space (area-weighted), and generated time series. (b) True $T_{\mathrm{2m,mean}}$ time series for the grid cell closest to Leipzig in 2003 and one randomly chosen generated time series. The initializations of generated time series are random, so no one-to-one comparison is possible with the ERA5 year. (c) Pooled histogram of the lag between the day of year of t_{out}, the "output time" the weather generator assigns to a sample, and t_{valid}, its actual date in the reforecast dataset. (d) Same as (c) but using $\tau = 1$.

Although the reforecast dataset is much larger than $\mathrm{ERA5}_{2003-2023}$, the nearest neighbors are not much closer, a consequence of the "curse of dimensionality". Another consequence is that the distance to the kth best analog varies only weakly with k (Fig. B4). Consequently, for each base state, there are typically many analogs in the reforecast dataset closer than the best analog in $\mathrm{ERA5}_{2003-2023}$: In roughly half the cases, distances to 300 analogs in the reforecast dataset are smaller than

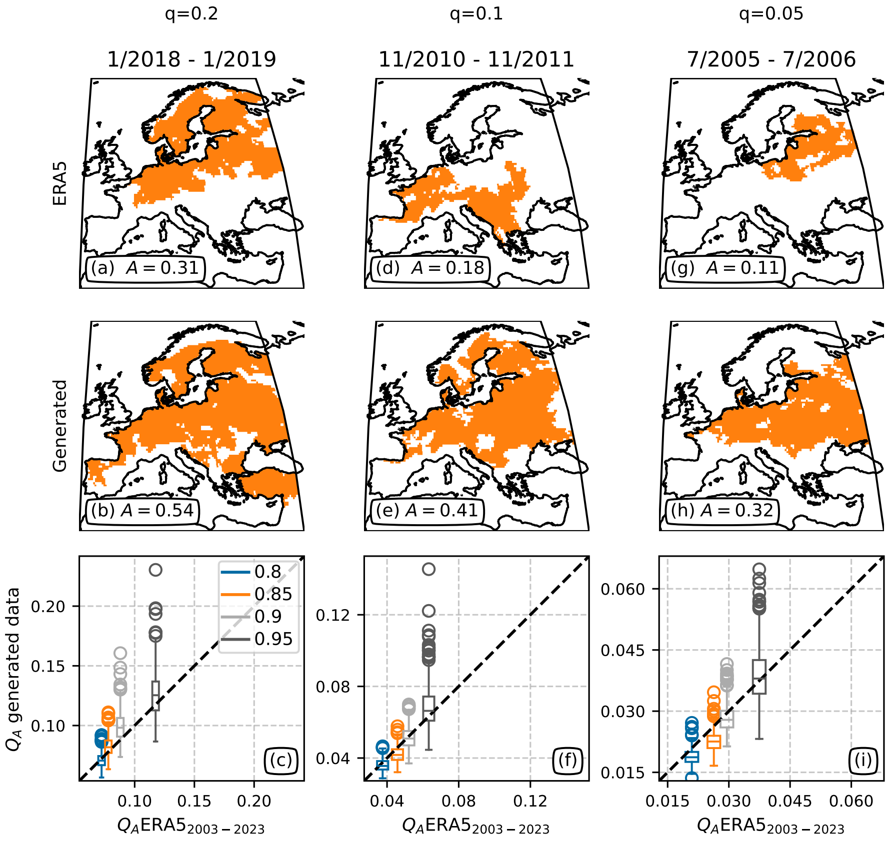

Figure B2. Same as Fig. 7, but in addition to the analysis for annual precipitation totals below the 20th percentile only (a)–(c), it also include analyses for $q = 0.1$ (d)–(f) and and $q = 0.05$ (g)–(i).

the distance to the best analog in $\mathrm{ERA5}_{2003-2023}$. This may seem surprising given that the distance to the best analog in the reforecast dataset is not much smaller than that to the best analogs in $\mathrm{ERA5}_{2003-2023}$ (Table B1).

We compare typical distances between a base state and its closest analog to the typical distance between two members of the same ensemble forecast at a given t_{lead}. To do so, we compute typical distances between ensemble members for various lead times in the reforecast dataset. Typically, a base state and its best analog in the reforecast dataset are roughly as far apart as two ensemble forecast members at a lead time of 6 days (Fig. B4, roughly equal for 7 days or 8 days for the $\mathrm{ERA5}_{2003-2023}$ dataset).

Appendix C: Figures

This section collects additional figures not associated with any of the appendix sections.

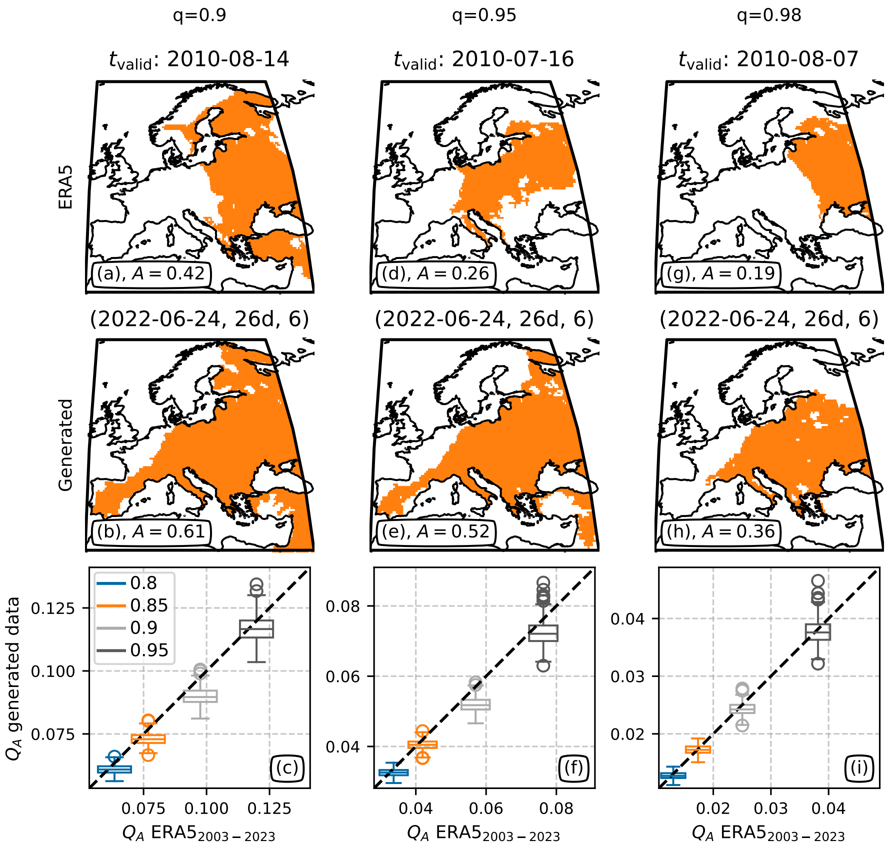

Figure B3. (a), (d), (g) Largest connected component in which every grid cell has a higher $T_{\mathrm{2m,mean}}$ than the q quantile in the $\mathrm{ERA5}_{2003-2023}$ JJA time series. (b), (e), (h) Same for 500 time series generated with unseen-awg using quantiles of the respective generated time series. (c), (f), (i) Studying the distribution of connected areas under heat. The x-axis indicates quantiles Q_A of A, the area covered by the largest connected heat event, computed over the the $\mathrm{ERA5}_{2003-2023}$ dataset. The box plots show quantiles of A in the 500 time series generated with unseen-awg.

Table B1. Mean Euclidean distances between large-scale atmospheric circulation states and their best and random analogs (i.e., randomly selected from set of all valid candidates). Estimates are computed over 1000 randomly selected base states. 2.5%–97.5% confidence intervals for the estimated mean are computed through bootstrapping with 9999 bootstrap resamples.

Dataset	$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{n}_1(\boldsymbol{x})\|$ [dam]	$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{n}_{random}(\boldsymbol{x})\|$ [dam]
Reforecasts	6.50 (6.42, 6.59)	15.39 (15.30, 15.48)
$\mathrm{ERA5}_{2003-2023}$	8.94 (8.83, 9.05)	15.56 (15.47, 15.65)

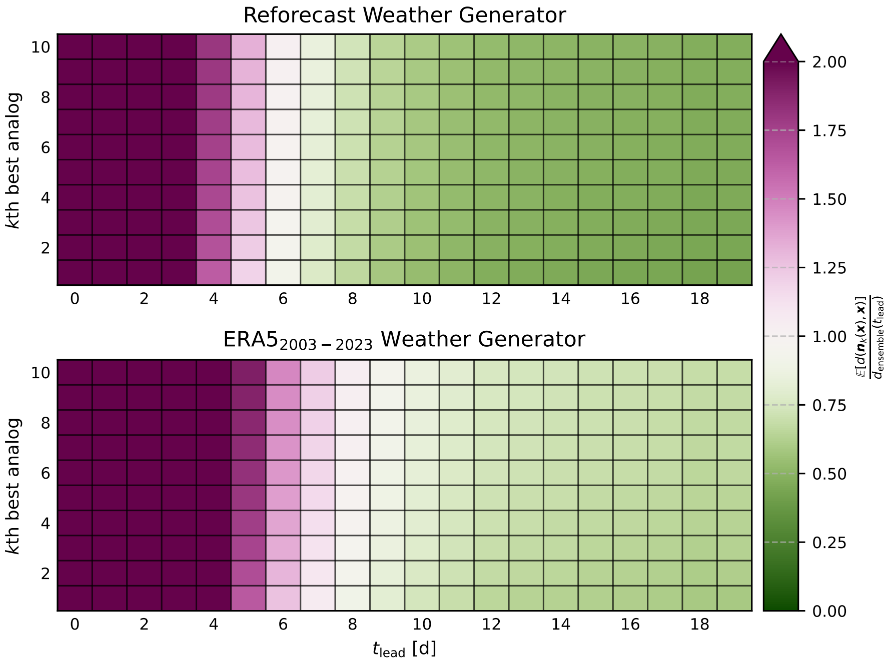

Figure B4. Ratio of the average Euclidean distance between a random large-scale atmospheric state and its kth best analog (i.e., kth nearest neighbor) to the average Euclidean distance between two random members of a single reforecast ensemble forecast at lead time t_{lead}.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

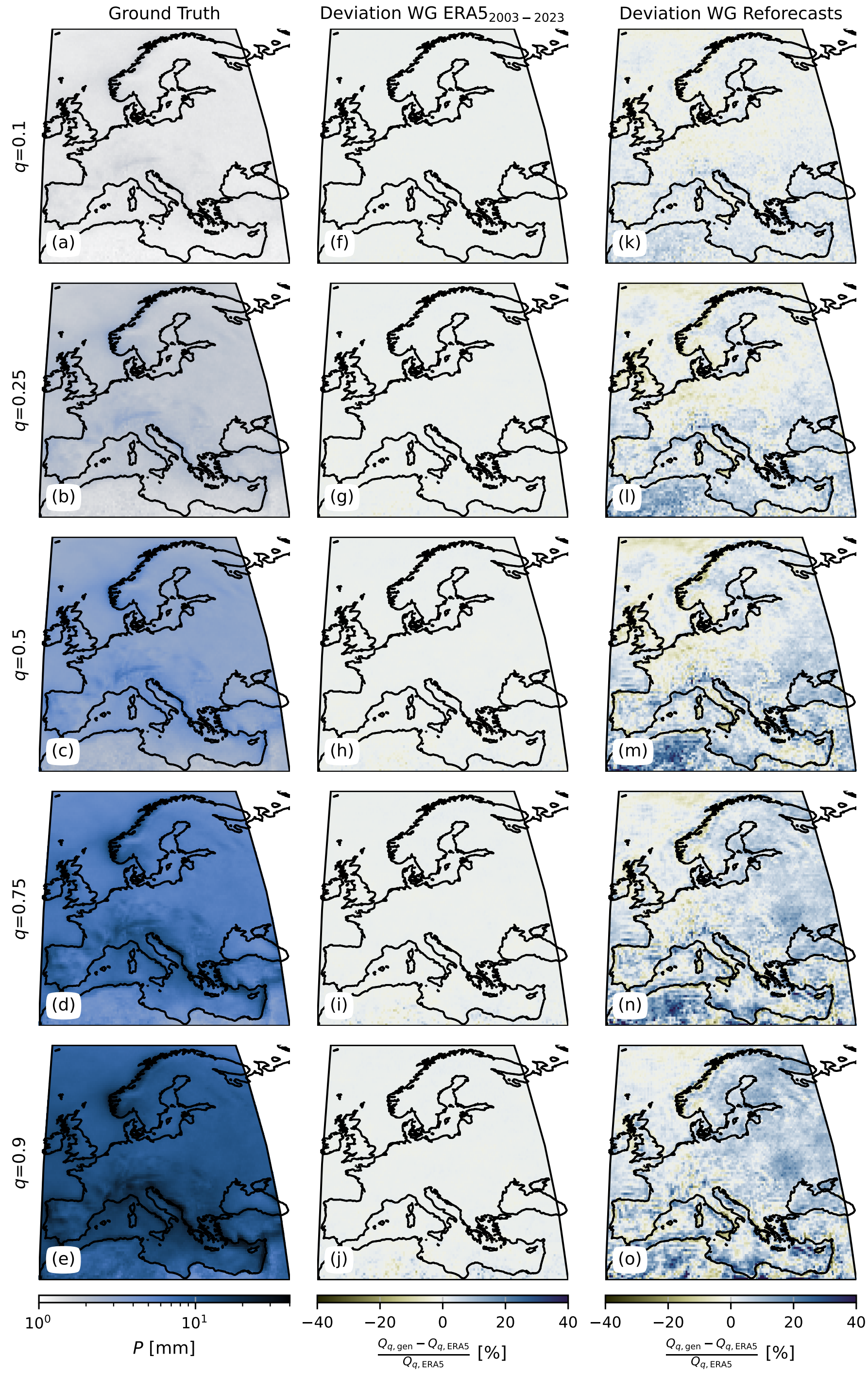

Figure C1. (a)–(e) Quantiles Q_q for daily total precipitation P in the $\mathrm{ERA5}_{2003-2023}$ dataset (computed on days with $P \geq 1\,\mathrm{mm}$). Average deviation of the same quantile in 500 21-year time series generated with unseen-awg weather generators based on $\mathrm{ERA5}_{2003-2023}$ (f)–(j) and reforecast data (k)–(o). Deviations are computed relatively as indicated in the colorbar labels.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

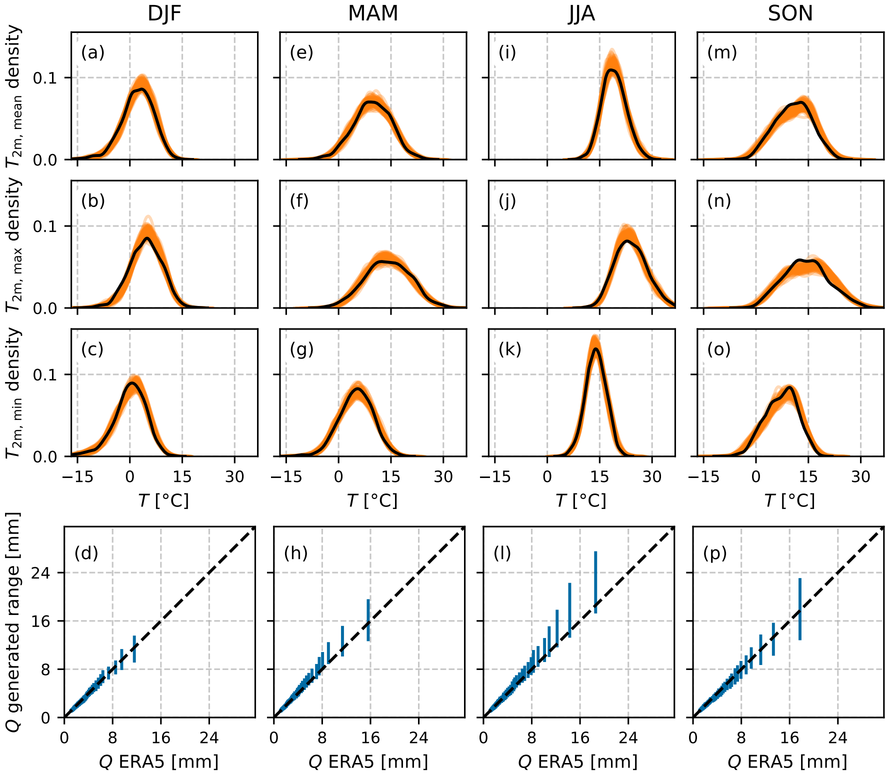

Figure C2. Marginal statistics of true and generated time series in grid cell closest to Leipzig, Germany, split by season. (a–c), (e–g), (i–k), (m–o) Kernel density estimate plots for $\mathrm{ERA5}_{2003-2023}$ (black) and 500 21-year time series generated with the standard setup of unseen-awg (orange). (d), (h), (l), (p) Quantiles $q = 0.025, 0.050, \ldots, 0.975$ of total precipitation on days with $P \geq 1\,\mathrm{mm}$ for $\mathrm{ERA5}_{2003-2023}$ (x-axis) and the generated time series (y-axis, min–max ranges computed over the values obtained from the 500 time series).

Code and data availability. We release an instance of unseen-awg with precomputed similarities and 500 21-year simulations with the weather generator are available (https://doi.org/10.26050/WDCC/unsawg_wg; Wider and Zscheischler, 2026b). Datasets of geopotential height at 500hPa and impact-relevant variables stored too (https://doi.org/21.14106/43da459a79e9f91e817e0b8690d494e8e91a00a5; Wider and Zscheischler, 2026c). Here, we archived preprocessed data from ECMWFs “Extended ensemble forecast hindcast” dataset (ECMWF, 2023) and the ERA5 reanalysis (Hersbach et al., 2020). All data are available under a CC-BY 4.0 license. The model code uses an MIT license and is split into a main repository for the unseen-awg weather generator itself (https://doi.org/10.5281/zenodo.19698709; Wider and Zscheischler, 2026d), and a secondary repository for reproducing our evaluation results (https://doi.org/10.5281/zenodo.19698739; Wider and Zscheischler, 2026a). We also created a user guide for unseen-awg (https://jonathanwider.github.io/unseen-awg; Wider and Zscheischler, 2026e). The user guide contains instructions on how to utilize the archived data within the unseen-awg framework.

Author contributions. Conceptualization: JW, JZ. Data curation: JW. Formal analysis: JW. Funding acquisition: JZ. Investigation: JW. Methodology: JW, JZ. Project administration: JW, JZ. Resources: JZ. Software: JW. Supervision: JZ. Validation: JW, JZ. Visualization: JW, JZ. Writing – original draft: JW. Writing – review & editing: JW, JZ

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank Daniel Klotz for fruitful discussions in early stages of the project. The authors acknowledge the financial support by the Federal Ministry of Research, Technology and Space of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI. The scientific results have been computed at the High-Performance Computing Cluster EVE, a joint effort of both the Helmholtz Centre for Environmental Research - UFZ and the German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig. We also acknowledge constructive feedback by Helmholtz Software Consulting. The text was revised and proofread using artificial intelligence methods. We also used AI assisted coding tools when implementing unseen-awg.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19, pp. 2623–2631, Association for Computing Machinery, New York, NY, USA, ISBN 978-1-4503-6201-6, https://doi.org/10.1145/3292500.3330701, 2019.

Anand, M., Bohn, F. J., Camps-Valls, G., Fischer, R., Huth, A., Sweet, L.-b., and Zscheischler, J.: Identifying Compound Weather Drivers of Forest Biomass Loss with Generative Deep Learning, Environmental Data Science, 3, e4, https://doi.org/10.1017/eds.2024.2, 2024.

Bakke, S. J., Ionita, M., and Tallaksen, L. M.: The 2018 Northern European Hydrological Drought and Its Drivers in a Historical Perspective, Hydrology and Earth System Sciences, 24, 5621–5653, https://doi.org/10.5194/hess-24-5621-2020, 2020.

Benito, I., Ward, P. J., Aerts, J. C. J. H., Eilander, D., and Muis, S.: Uncovering Unprecedented Storm Surges in European and Mediterranean Coastlines, Environmental Research Letters, 20, 124 066, https://doi.org/10.1088/1748-9326/ae2529, 2025.

Bevacqua, E., Suarez-Gutierrez, L., Jézéquel, A., Lehner, F., Vrac, M., Yiou, P., and Zscheischler, J.: Advancing Research on Compound Weather and Climate Events via Large Ensemble Model Simulations, Nature Communications, 14, 2145, https://doi.org/10.1038/s41467-023-37847-5, 2023.

Bevacqua, E., Rakovec, O., Schumacher, D. L., Kumar, R., Thober, S., Samaniego, L., Seneviratne, S. I., and Zscheischler, J.: Direct and Lagged Climate Change Effects Intensified the 2022 European Drought, Nature Geoscience, 17, 1100–1107, https://doi.org/10.1038/s41561-024-01559-2, 2024.

Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K., and Engelke, S.: Modeling and Simulating Spatial Extremes by Combining Extreme Value Theory with Generative Adversarial Networks, Environmental Data Science, 1, e5, https://doi.org/10.1017/eds.2022.4, 2022.

Brunner, M. I., Gilleland, E., and Wood, A. W.: Space–Time Dependence of Compound Hot–Dry Events in the United States: Assessment Using a Multi-Site Multi-Variable Weather Generator, Earth System Dynamics, 12, 621–634, https://doi.org/10.5194/esd-12-621-2021, 2021.

Cariou, E., Cattiaux, J., Qasmi, S., Ribes, A., Cassou, C., and Doury, A.: Linking European Temperature Variations to Atmospheric Circulation With a Neural Network: A Pilot Study in a Climate Model, Geophysical Research Letters, 52, e2024GL113 540, https://doi.org/10.1029/2024GL113540, 2025.

Chen, M., Mahto, S. S., He, X., Jun, C., Paschalis, A., Peleg, N., Mascaro, G., and Fatichi, S.: Record-Breaking Rainfall: A Stochastic Approach for Its Prediction, npj Natural Hazards, 2, 98, https://doi.org/10.1038/s44304-025-00148-6, 2025.

Christidis, N. and Stott, P. A.: Changes in the Geopotential Height at 500 hPa under the Influence of External Climatic Forcings, Geophysical Research Letters, 42, https://doi.org/10.1002/2015gl066669, 2015.

Cresswell-Clay, N., Liu, B., Durran, D. R., Liu, Z., Espinosa, Z. I., Moreno, R. A., and Karlbauer, M.: A Deep Learning Earth System Model for Efficient Simulation of the Observed Climate, AGU Advances, 6, e2025AV001 706, https://doi.org/10.1029/2025AV001706, 2025.

ECMWF: IFS Documentation CY48R1 - Part V: Ensemble Prediction System, https://doi.org/10.21957/E529074162, 2023.

Fatichi, S. and Ivanov, V. Y.: Interannual Variability of Evapotranspiration and Vegetation Productivity, Water Resources Research, 50, 3275–3294, https://doi.org/10.1002/2013WR015044, 2014.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, Journal of the American Statistical Association, 102, 359–378, https://doi.org/10.1198/016214506000001437, 2007.

Gutzler, D. S. and Shukla, J.: Analogs in the Wintertime 500 Mb Height Field, Journal of the Atmospheric Sciences, 41, 177–189, https://doi.org/10.1175/1520-0469(1984)041<0177:AITWMH>2.0.CO;2, 1984.

Hagedorn, R., Hamill, T. M., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures, Monthly Weather Review, 136, 2608–2619, https://doi.org/10.1175/2007MWR2410.1, 2008.

Hamill, T. M., Whitaker, J. S., and Mullen, S. L.: Reforecasts: An Important Dataset for Improving Weather Predictions, Bulletin of the American Meteorological Society, 87, 33–46, https://doi.org/10.1175/BAMS-87-1-33, 2006.

Hamill, T. M., Hagedorn, R., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation, Monthly Weather Review, 136, 2620–2632, https://doi.org/10.1175/2007MWR2411.1, 2008.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Horton, P., Obled, C., and Jaboyedoff, M.: The Analogue Method for Precipitation Prediction: Finding Better Analogue Situations at a Sub-Daily Time Step, Hydrology and Earth System Sciences, 21, 3307–3323, https://doi.org/10.5194/hess-21-3307-2017, 2017.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., Van Den Hurk, B., and Vera, C. S.: An Update of IPCC Climate Reference Regions for Subcontinental Analysis of Climate Model Data: Definition and Aggregated Datasets, Earth System Science Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

Karlbauer, M., Cresswell-Clay, N., Durran, D. R., Moreno, R. A., Kurth, T., Bonev, B., Brenowitz, N., and Butz, M. V.: Advancing Parsimonious Deep Learning Weather Prediction Using the HEALPix Mesh, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 021, https://doi.org/10.1029/2023MS004021, 2024.

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.

Kelder, T., Heinrich, D., Klok, L., Thompson, V., Goulart, H. M. D., Hawkins, E., Slater, L. J., Suarez-Gutierrez, L., Wilby, R. L., Coughlan De Perez, E., Stephens, E. M., Burt, S., Van Den Hurk, B., De Vries, H., Van Der Wiel, K., Schipper, E. L. F., Carmona Baéz, A., Van Bueren, E., and Fischer, E. M.: How to Stop Being Surprised by Unprecedented Weather, Nature Communications, 16, 2382, https://doi.org/10.1038/s41467-025-57450-0, 2025.

Klehmet, K., Berg, P., Bozhinova, D., Crochemore, L., Du, Y., Pechlivanidis, I., Photiadou, C., and Yang, W.: Robustness of Hydrometeorological Extremes in Surrogated Seasonal Forecasts, International Journal of Climatology, p. joc.8407, https://doi.org/10.1002/joc.8407, 2024.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural General Circulation Models for Weather and Climate, Nature, 632, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y, 2024.

Krouma, M., Yiou, P., Déandreis, C., and Thao, S.: Assessment of Stochastic Weather Forecast of Precipitation near European Cities, Based on Analogs of Circulation, Geoscientific Model Development, 15, 4941–4958, https://doi.org/10.5194/gmd-15-4941-2022, 2022.

Krouma, M., Specq, D., Magnusson, L., Ardilouze, C., Batté, L., and Yiou, P.: Improving Subseasonal Forecast of Precipitation in Europe by Combining a Stochastic Weather Generator with Dynamical Models, Quarterly Journal of the Royal Meteorological Society, p. qj.4733, https://doi.org/10.1002/qj.4733, 2024.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: Learning Skillful Medium-Range Global Weather Forecasting, Science, p. eadi2336, https://doi.org/10.1126/science.adi2336, 2023.

Lavers, D. A., Simmons, A., Vamborg, F., and Rodwell, M. J.: An Evaluation of ERA5 Precipitation for Climate Monitoring, Quarterly Journal of the Royal Meteorological Society, 148, 3152–3165, https://doi.org/10.1002/qj.4351, 2022.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R.: The Analog Data Assimilation, Monthly Weather Review, 145, 4093–4107, https://doi.org/10.1175/MWR-D-16-0441.1, 2017.

Li, J., Zscheischler, J., and Bevacqua, E.: Global Record-Shattering Breadbasket Droughts Emerge from Moderately Extreme Regional Events, Nature Communications, 17, 2577, https://doi.org/10.1038/s41467-026-70700-z, 2026.

Lorenz, E. N.: Atmospheric Predictability as Revealed by Naturally Occurring Analogues, Journal of the Atmospheric Sciences, 26, 636–646, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2, 1969.

Maddison, C. J., Tarlow, D., and Minka, T.: A* Sampling, Advances in neural information processing systems, 27, 2014.

Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J., and Engelke, S.: Validating Deep Learning Weather Forecast Models on Recent High-Impact Extreme Events, Artificial Intelligence for the Earth Systems, 4, e240 033, https://doi.org/10.1175/AIES-D-24-0033.1, 2025.

Peard, A., Mo, Y., and Hall, J. W.: Simulating Spatial Multi-Hazards with Generative Deep Learning, Natural Hazards and Earth System Sciences, 26, 1663–1683, https://doi.org/10.5194/nhess-26-1663-2026, 2026.

Platzer, P., Yiou, P., Naveau, P., Tandeo, P., Filipot, J.-F., Ailliot, P., and Zhen, Y.: Using Local Dynamics to Explain Analog Forecasting of Chaotic Systems, Journal of the Atmospheric Sciences, 78, 2117–2133, https://doi.org/10.1175/JAS-D-20-0204.1, 2021.

Qin, X. S. and Lu, Y.: Study of Climate Change Impact on Flood Frequencies: A Combined Weather Generator and Hydrological Modeling Approach*, Journal of Hydrometeorology, 15, 1205–1219, https://doi.org/10.1175/JHM-D-13-0126.1, 2014.

Rakovec, O., Samaniego, L., Hari, V., Markonis, Y., Moravec, V., Thober, S., Hanel, M., and Kumar, R.: The 2018–2020 Multi-Year Drought Sets a New Benchmark in Europe, Earth's Future, 10, e2021EF002 394, https://doi.org/10.1029/2021EF002394, 2022.

Sillmann, J., Shepherd, T. G., Van Den Hurk, B., Hazeleger, W., Martius, O., Slingo, J., and Zscheischler, J.: Event-Based Storylines to Address Climate Risk, Earth's Future, 9, e2020EF001 783, https://doi.org/10.1029/2020EF001783, 2021.

Sippel, S., Barnes, C., Cadiou, C., Fischer, E., Kew, S., Kretschmer, M., Philip, S., Shepherd, T. G., Singh, J., Vautard, R., and Yiou, P.: Could an Extremely Cold Central European Winter Such as 1963 Happen Again despite Climate Change?, Weather and Climate Dynamics, 5, 943–957, https://doi.org/10.5194/wcd-5-943-2024, 2024.

Sun, Y. Q., Hassanzadeh, P., Zand, M., Chattopadhyay, A., Weare, J., and Abbot, D. S.: Can AI Weather Models Predict Out-of-Distribution Gray Swan Tropical Cyclones?, Proceedings of the National Academy of Sciences, 122, e2420914 122, https://doi.org/10.1073/pnas.2420914122, 2025.

Thompson, V.: Circulation Analogues Cannot Identify Changes in Rainfall Extremes, Environmental Research Letters, 20, 121 003, https://doi.org/10.1088/1748-9326/ae20a6, 2025.

https://doi.org/10.5194/egusphere-2026-2325
Preprint. Discussion started: 22 June 2026

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E.: High Risk of Unprecedented UK Rainfall in the Current Climate, Nature Communications, 8, 107, https://doi.org/10.1038/s41467-017-00275-3, 2017.

Van Den Brink, H. W., Können, G. P., Opsteegh, J. D., Van Oldenborgh, G. J., and Burgers, G.: Improving 10^4 -year Surge Level Estimates Using Data of the ECMWF Seasonal Prediction System, Geophysical Research Letters, 31, 2004GL020 610, https://doi.org/10.1029/2004GL020610, 2004.

van den Dool, H. M.: A New Look at Weather Forecasting through Analogues, Monthly Weather Review, 117, 2230–2247, https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2, 1989.

van den Dool, H. M.: Searching for Analogues, How Long Must We Wait?, Tellus A, Dynamic Meteorology and Oceanography, 46, 314, https://doi.org/10.3402/tellusa.v46i3.15481, 1994.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T.: Scikit-Image: Image Processing in Python, PeerJ, 2, e453, https://doi.org/10.7717/peerj.453, 2014.

Vitart, F. and Robertson, A. W.: The Sub-Seasonal to Seasonal Prediction Project (S2S) and the Prediction of Extreme Events, npj Climate and Atmospheric Science, 1, 3, https://doi.org/10.1038/s41612-018-0013-0, 2018.

Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., and Bretherton, C. S.: ACE2: Accurately Learning Subseasonal to Decadal Atmospheric Variability and Forced Responses, npj Climate and Atmospheric Science, 8, 205, https://doi.org/10.1038/s41612-025-01090-0, 2025.

Wider, J. and Zscheischler, J.: Evaluation Code for 'Unseen-Awg: Spatio-temporal Weather Generation Using Analogs and Unseen Data', Zenodo, https://doi.org/10.5281/zenodo.19698739, 2026a.

Wider, J. and Zscheischler, J.: An Instance of the Analog Weather Generator Unseen-Awg with Precomputed Similarities and a Large Set of Generated Weather Data, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.26050/WDCC/unsawg_wg, 2026b.

Wider, J. and Zscheischler, J.: Preprocessed Atmospheric Circulation and Impact-Relevant Variables from ERA5 and "Extended Ensemble Forecast Hindcast" (ECMWF) for Unseen-Awg Simulations, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/21.14106/43da459a79e9f91e817e0b8690d494e8e91a00a5, 2026c.

Wider, J. and Zscheischler, J.: Unseen-Awg: Spatio-temporal Weather Generation Using Analogs and Unseen Data, Zenodo, https://doi.org/10.5281/zenodo.19698709, 2026d.

Wider, J. and Zscheischler, J.: User Guide for "Unseen-Awg: Spatio-Temporal Weather Generation Using Analogs and Unseen Data", https://jonathanwider.github.io/unseen-awg/, 2026e.

Wilks, D. S. and Wilby, R. L.: The Weather Generation Game: A Review of Stochastic Weather Models, Progress in Physical Geography: Earth and Environment, 23, 329–357, https://doi.org/10.1177/030913339902300302, 1999.

Winter, B., Schneeberger, K., Förster, K., and Vorogushyn, S.: Event Generation for Probabilistic Flood Risk Modelling: Multi-Site Peak Flow Dependence Model vs. Weather-Generator-Based Approach, Natural Hazards and Earth System Sciences, 20, 1689–1703, https://doi.org/10.5194/nhess-20-1689-2020, 2020.

Yiou, P.: AnaWEGE: A Weather Generator Based on Analogues of Atmospheric Circulation, Geoscientific Model Development, 7, 531–543, https://doi.org/10.5194/gmd-7-531-2014, 2014.

Yiou, P. and Jézéquel, A.: Simulation of Extreme Heat Waves with Empirical Importance Sampling, Geoscientific Model Development, 13, 763–781, https://doi.org/10.5194/gmd-13-763-2020, 2020.

Yiou, P., Cadiou, C., Faranda, D., Jézéquel, A., Malhomme, N., Miloshevich, G., Noyelle, R., Pons, F., Robin, Y., and Vrac, M.: Ensembles of Climate Simulations to Anticipate Worst Case Heatwaves during the Paris 2024 Olympics, npj Climate and Atmospheric Science, 6, 188, https://doi.org/10.1038/s41612-023-00500-5, 2023.

Zhang, Z., Fischer, E., Zscheischler, J., and Engelke, S.: Numerical Models Outperform AI Weather Forecasts of Record-Breaking Extremes, Science Advances, accepted.

Zorita, E. and von Storch, H.: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, Journal of Climate, 12, 2474–2489, https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2, 1999.

Zscheischler, J. and Seneviratne, S. I.: Dependence of Drivers Affects Risks Associated with Compound Events, Science Advances, 3, e1700 263, https://doi.org/10.1126/sciadv.1700263, 2017.

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future Climate Risk from Compound Events, Nature Climate Change, 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., and Vignotto, E.: A Typology of Compound Weather and Climate Events, Nature Reviews Earth & Environment, 1, 333–347, https://doi.org/10.1038/s43017-020-0060-z, 2020.