JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000902

Stochastic Weather Generation for Scenario-Neutral Impact Assessments Using Simulation-Based Inference

Brian Groenke[1,2,3,4], **Jakob Wessel**[1,5], **Peter Miersch**[1,6], **Nadja Klein**[7], and **Jakob Zscheischler**[1,4,6]

[1]Department of Compound Environmental Risks, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany, [2]Permafrost Research Section, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany, [3]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [4]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig/Dresden, Germany, [5]Department of Mathematics and Statistics, University of Exeter, Exeter, UK, [6]Department of Hydro Sciences, TUD Dresden University of Technology, Dresden, Germany, [7]Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract Scenario-neutral and robust adaptation methods assess the vulnerability of climate-sensitive systems against a range of plausible climate conditions, independent of the socioeconomic scenarios typically used in climate modeling. Stochastic weather generators facilitate such analyses by enabling fast and flexible simulation of meteorological time series based on historical observations. Long-term changes in climate conditions are often described via corresponding changes in summary statistics or climate indices. However, adjusting stochastic weather generators to produce simulations consistent with perturbed summary statistics is challenging, especially for more complex statistics and weather generator models. We refer to this problem as climatology matching. In this work, we make two key contributions: First, we develop a flexible framework for stochastic weather generation based on Generalized Additive Models for Location, Scale, and Shape (GAMLSS). The proposed weather generator is capable of efficiently and accurately simulating daily temperature (mean, minimum, and maximum) and precipitation time series over multi-decadal time scales after being calibrated on historical data. Second, we propose a fully probabilistic formulation of the climatology matching problem, to which we apply techniques from the field of simulation-based inference (SBI). We evaluate our approach using weather station data from the German Weather Service and demonstrate its potential for scenario-neutral impact assessment by simulating realistic daily meteorological time series under various climate change conditions. Our method provides an efficient and flexible framework for stress-testing climate impact models with the potential to enhance the robustness of scenario-neutral impact assessments.

Plain Language Summary Understanding how to adapt to future changes in the climate requires modeling the impacts of those changes on key natural and manmade systems. Traditional methods for such impact assessments rely on projections generated by global climate models which are driven by socioeconomic emissions scenarios. Although valuable, these methods are only able to consider a relatively limited range of potential climate changes simulated by these models. Robust and systematic stress-testing of climate impacts requires the simulation of weather patterns under a broad range of conditions. In this work, we use state-of-the-art techniques in probabilistic modeling to address this problem. Our proposed method allows for fast and efficient simulation of realistic daily air temperature and precipitation time series under arbitrary climate conditions as defined by any relevant set of summary statistics, such as long-term averages, percentiles, or common climate indices. We apply our approach to weather station data from the German Weather Service and show how it can be used for climate change impact assessments by allowing researchers and practitioners to systematically simulate meteorological time series under a variety of localized scenarios.

Key Points:

- We propose a new method for adapting stochastic weather generators to arbitrary summary statistics using simulation-based inference
- We introduce a stochastic weather generator for air temperature and precipitation time series based on autoregressive GAMLSS
- We demonstrate how our methods can be used for scenario-neutral impact assessments using weather station data from Potsdam, Germany

Correspondence to:
B. Groenke and J. Wessel,
brian.groenke@pik-potsdam.de;
j.wessel@exeter.ac.uk

Citation:
Groenke, B., Wessel, J., Miersch, P., Klein, N., & Zscheischler, J. (2026). Stochastic weather generation for scenario-neutral impact assessments using simulation-based inference. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH000902. https://doi.org/10.1029/2025JH000902

Received 17 JUL 2025
Accepted 31 MAR 2026

Author Contributions:

Conceptualization: Brian Groenke, Jakob Wessel, Peter Miersch, Jakob Zscheischler
Data curation: Jakob Wessel, Peter Miersch
Formal analysis: Brian Groenke, Jakob Wessel, Nadja Klein
Funding acquisition: Jakob Wessel, Jakob Zscheischler
Investigation: Brian Groenke, Jakob Wessel, Peter Miersch
Methodology: Brian Groenke, Jakob Wessel, Peter Miersch

1. Introduction

Evaluating how natural and human systems respond to changes in weather patterns, such as those due to decadal climate variability or long-term climate change, requires high-resolution data on local weather conditions. For example, local time series of meteorological variables, such as air temperature and precipitation, are necessary inputs to many impact models, including hydrological or crop models, which are then used to estimate key quantities of interest like flood risk or crop yield. Climate model simulations based on socioeconomic scenarios represent our best knowledge about potential future weather patterns, and are routinely used as input for impact

Project administration: Jakob Zscheischler
Resources: Jakob Zscheischler
Software: Brian Groenke, Jakob Wessel
Supervision: Jakob Zscheischler
Validation: Brian Groenke, Jakob Wessel
Visualization: Brian Groenke, Jakob Wessel
Writing – original draft: Brian Groenke, Jakob Wessel
Writing – review & editing: Brian Groenke, Jakob Wessel, Peter Miersch, Nadja Klein, Jakob Zscheischler

models to conduct climate impact analysis (Warszawski et al., 2014). However, this traditional "top-down" approach of climate impact assessment can be difficult for decision makers to interpret and apply (Wilby & Dessai, 2010). Climate forcings derived from climate models are only available at coarse spatial scales and are typically biased when compared to observations, requiring downscaling and bias adjustment to be useful for impact assessment. Furthermore, climate model projections are subject to substantial uncertainties that are difficult to sample comprehensively.

As an alternative approach to the traditional scenario-led climate impact analysis, so-called *scenario-neutral* approaches have been increasingly used to explore the complex relationships and sensitivities within systems affected by climate variability and change. These "bottom-up" methods rely on detailed vulnerability analyses — essentially stress-tests — that examine how a system responds to a wide range of changes in key hydroclimatic variables (Culley et al., 2016; Prudhomme et al., 2010). This also allows for evaluating the robustness of different management or design strategies. So far, most of the development and application of scenario-neutral approaches has taken place within the water and environmental sciences, with particular focus on areas like water resource systems, sediment transport, and changes in streamflow and flood risk (Bennett et al., 2021).

In contrast to the use of climate models in scenario-led impact analysis, scenario-neutral approaches often rely on simulations of local weather conditions using stochastic weather generators (Wilks & Wilby, 1999). Stochastic weather generators are statistical models that can be used to simulate time series of meteorological variables. They range in complexity from simple Markov-chain models, such as the widely known Richardson weather generator (Richardson, 1981), to autoregressive models (Racsko et al., 1991) and complex multi-site and multivariable models which may include spatio-temporal dependencies between different variables (Ambrosino et al., 2014; Chandler, 2020; Kilsby et al., 2007; Yang et al., 2005). Hidden Markov Models have also been applied to infer transitions in latent climate/weather states (Ailliot et al., 2009; Hughes et al., 1999), as well as models based on transformed Gaussian random fields (Bárdossy & Plate, 1992), Poisson cluster processes (Onof et al., 2000), and the family of generalized linear (e.g., Chandler & Wheater, 2002) and additive models (Hyndman & Grunwald, 2000; Stoner & Economou, 2020; Underwood, 2009). Reviews and various perspectives are given by Maraun et al. (2010); Ailliot et al. (2015); Chandler (2020); Northrop (2024).

Weather generators are typically calibrated on meteorological data from weather stations and can then be used to create long synthetic time series to drive downstream models, such as in a sensitivity analysis investigating how the simulated system would respond to a wide range of weather conditions (Fatichi & Ivanov, 2014). This setting can also be used to test new data-driven approaches in a controlled environment (Anand et al., 2024; Fatichi & Ivanov, 2014; Marcolongo et al., 2022). Longer-term weather patterns are often described using *summary statistics*, such as averages, quantiles, frequencies, extreme values, or measures of variability, which characterize key aspects of the climate, often referred to as the *climatology* of a particular locality or region. Strategically, stress-testing a system on specific aspects of possible future change requires simulation of hydroclimatic time series data with altered long-term behavior, typically expressed by perturbing a predefined set of summary statistics. For relatively small sets of summary statistics such as annual or decadal averages, it is often possible to represent desired changes by applying scaling factors to observed or simulated time series data (Jones et al., 2010; Keller et al., 2019). However, this approach does not generalize to more complex summary statistics with potentially strong interdependencies. Furthermore, scaling factors alone are unable to represent changes in summary statistics that have a nonlinear relationship with the underlying data, such as frequencies, quantiles, or threshold-based climate indices. We hereafter refer to the problem of generating time series based on arbitrary summary statistics as *climatology matching*.

In this work, we propose a novel method for solving the climatology matching problem using recently developed computational tools from the field of simulation-based inference (SBI; Cranmer et al., 2020). We follow Guo et al. (2018), Culley et al. (2019), and McInerney et al. (2023) in casting climatology matching as an inverse problem, where the objective is to infer a set of parameters for a weather generator that is likely to produce time series with the desired climatic properties. Unlike previous work, however, we adopt a fully Bayesian treatment of the inverse problem and use SBI to obtain a probabilistic solution. One key advantage of our approach is that it fully accounts for both the epistemic uncertainty in the estimated parameters of the weather generator, as well as the aleatoric uncertainty arising from its internal variability.

The proposed method can, in principle, be applied to any (parametric) stochastic weather generator. However, the ability of the method to simulate the desired changes in climate conditions is inextricably linked to the efficacy

and efficiency of the underlying weather generator. Although many methods for stochastic weather generation have been proposed in recent years (Asong et al., 2016; Bennett et al., 2021; Chandler, 2020; Fatichi et al., 2011; Sommer & Kaplan, 2017), most lack the necessary computational efficiency and flexibility needed to fully realize our methodology for climatology matching. To address this, we first develop a novel approach for stochastic weather generation based on Generalized Additive Models for Location, Scale, and Shape (GAMLSS; Klein, 2024; Stasinopoulos et al., 2024). Our approach jointly models the predictive distributions of daily mean, minimum, and maximum air temperatures, as well as daily precipitation, while accounting for dependencies between these variables over time. We combine our proposed SBI approach with the newly developed weather generator and demonstrate the applicability of this method for scenario-neutral analyses by simulating realistic meteorological time series with perturbed summary statistics representing climate change trends for a weather station in Potsdam, Germany. To demonstrate the broad applicability of the method, applications of our method to additional stations in Germany are presented in Appendix E.

2. Autoregressive Weather Generation With GAMLSS

We first present a novel stochastic weather generator that builds upon the GAMLSS approach (Rigby & Stasinopoulos, 2005; Stasinopoulos et al., 2024) for weather generation introduced in Wessel and Chandler (2025). While Generalized Linear Models (GLMs) have gained popularity for weather generation in recent years (see e.g. Chandler & Wheater, 2002; Chandler, 2020; Keller et al., 2015; Stern & Coe, 1984; Yang et al., 2005), these models are limited to modeling the conditional mean. In contrast, GAMLSS enable the modeling of the entire distribution of quantities of interest as a function of a set of predictors. This approach offers the potential to create more realistic and better-calibrated models for complex phenomena.

Before describing our weather generation model in more detail, we briefly outline the general idea of GAMLSS. Let Y be a univariate quantity of interest and $\boldsymbol{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^d$ be some covariates which are to be used as predictors for Y. In GAMLSS, one assumes that $Y|\boldsymbol{x}$ has an arbitrary parametric distribution with density $p(Y|\boldsymbol{\eta}, \boldsymbol{x})$, where $\boldsymbol{\eta} \equiv \boldsymbol{\eta}(\boldsymbol{x}) = (\eta_1, \ldots, \eta_d)^\top$ is a $d-$dimensional vector of distribution parameters. A simple example is a Gaussian regression $\mathcal{N}(Y|\boldsymbol{\eta})$ with parameters $\boldsymbol{\eta} = (\mu, \sigma)^\top$. In a traditional mean regression model, the mean μ is modeled as a function of the predictors but the standard deviation σ is treated as a constant. In contrast, GAMLSS allows each distribution parameter to be modeled via linear predictors of the form,

$$g_l(\eta_l) = \boldsymbol{x}_l^\top \boldsymbol{\beta}_l, \tag{1}$$

where $\boldsymbol{x}_l \subset \boldsymbol{x}$, $\boldsymbol{\beta}_l$ are the distribution parameter-specific regression coefficients, and g_l are so-called *link functions* that map the distribution parameters onto the real line. For example, if η_l is strictly positive (as is the case with scale parameters), a log-link, $g_l(\eta_l) = \log(\eta_l)$, is the canonical choice. The GAMLSS framework includes many types of weather generators as special cases, such as the Richardson-type weather generators (Richardson, 1981) and other Markov-chain based models (Furrer & Katz, 2008; Grunwald & Jones, 2000; Katz & Parlange, 1998; Katz & Zheng, 1999; Wilks, 2010; Wilks & Wilby, 1999), as well as autoregressive models (Nguyen et al., 2024; Racsko et al., 1991).

2.1. Model Specification

We are primarily interested in stochastic modeling of daily average temperature, T_{avg}, and precipitation, P over time. We additionally consider daily minimum and maximum air temperatures, T_{min}, and T_{max}, since these variables are required by many hydrological and agricultural models in order to parameterize processes such as evapotranspiration which are affected by the diurnal cycle. Unlike the mean, daily minimum and maximum air temperatures are inherently constrained to follow the ordering, $T_{\text{min}} < T_{\text{avg}} < T_{\text{max}}$. To address this issue, we reparameterize T_{min} and T_{max} in terms of the daily temperature range and skew,

$$T_{\text{range}} = T_{\text{max}} - T_{\text{min}}, \quad T_{\text{skew}} = (T_{\text{avg}} - T_{\text{min}})/T_{\text{range}}. \tag{2}$$

Here, T_{range} corresponds to the diurnal temperature range for any given day and T_{skew} models the degree of asymmetry in the diurnal temperature distribution. Note that T_{skew} is here a measure of skewness in the daily

temperature distribution, but does not correspond to the statistical definition of skewness as the third standardized moment. Our approach can be seen as a generalization of the parameterization proposed by Kilsby et al. (2007), which corresponds to a special case where $T_{\text{skew}} = 1/2$. Modeling T_{skew} explicitly allows us to model variations in the asymmetry of the diurnal temperature distribution, relaxing the common assumption that $T_{\min}$ and $T_{\max}$ are distributed symmetrically around the mean. This is well-justified by physical understanding: for example, cloud cover and humidity are known to vary diurnally in many regions which can result in the daily distribution of temperatures being skewed rather than symmetric.

We model each meteorological variable $T_{\text{avg}}, T_{\text{range}}, T_{\text{skew}}, P$ autoregressively as functions of their k previous days' values, with the exception of T_{skew} which showed minimal autocorrelation. We choose $k = 2$ since this appeared sufficient to capture most of the autocorrelation in our calibration data sets. In order to capture seasonality, we include so-called "Fourier-features" of the form $\mathbf{f}^{(t)} = (\sin(2\pi\omega t), \cos(2\pi\omega t))$ with periods of 182.25 and 365 days for each of the variables, corresponding to the leading frequencies based on a (truncated) fast Fourier transform. Since we also aim to model dependencies between variables, we formulate the joint distribution over all four variables such that it can be factorized into a series of conditional distributions which capture these dependencies.

Let $\boldsymbol{Y} = [\mathbf{T}_{\text{avg}}|\mathbf{T}_{\text{range}}|\mathbf{T}_{\text{skew}}|\mathbf{P}]$ be a matrix of dimension $N_t \times 4$, where N_t is the length of the time series for each variable. In order to capture the dependencies between variables, we factorize the joint density into a series of conditionals as

$$\begin{aligned} p\left(T_{\text{avg}}^{(t)}, T_{\text{range}}^{(t)}, T_{\text{skew}}^{(t)}, P^{(t)}\right) = {} & p\left(T_{\text{avg}}^{(t)} | T_{\text{avg}}^{(t-1)}, T_{\text{avg}}^{(t-2)}\right) \\ & \cdot p\left(T_{\text{range}}^{(t)} | T_{\text{avg}}^{(t)}, T_{\text{range}}^{(t-1)}, T_{\text{range}}^{(t-2)}\right) \\ & \cdot p\left(T_{\text{skew}}^{(t)} | T_{\text{avg}}^{(t)}\right) \\ & \cdot p\left(P^{(t)} | T_{\text{avg}}^{(t)}, P^{(t-1)}, P^{(t-2)}\right), \end{aligned} \tag{3}$$

where dependence on the exogenous covariates $\boldsymbol{x}^{(t)}$ is suppressed for brevity. Decompositions into conditional densities like the above are often used for multivariable weather generators and we refer to Chandler (2020) for a discussion of this. Here we condition on T_{avg} as "primary" variable as it is usually modeled with less error than the other variables. We assume conditional independence of $T_{\text{range}}, T_{\text{skew}}$ and P, given T_{avg} and find little evidence in the data of a dependence beyond the one induced by T_{avg}. Figure 1 contains a schematic representation of how the model and joint distribution in Equation 3 are built. We next describe the specific parametric choices for the conditional distributions in Equation 3. Appendix A also contains a list of all covariates used for each distribution parameter and each variables.

Mean air temperature: We model the distribution of $T_{\text{avg}}^{(t)}$ via a Gaussian distribution, where both the mean $\mu_T^{(t)}$ and standard deviation $\sigma_T^{(t)}$ are modeled as functions of the predictors. Specifically, we assume

$$T_{\text{avg}}^{(t)} | \boldsymbol{x}_{T,\mu}^{(t)}, \boldsymbol{x}_{T,\sigma}^{(t)} \sim \mathcal{N}\left(\mu_T^{(t)}, \sigma_T^{2\,(t)}\right), \quad \mu_T^{(t)} = \boldsymbol{\beta}_{T,\mu}^{\top} \boldsymbol{x}_{T,\mu}^{(t)}, \quad \log\sigma_T^{(t)} = \boldsymbol{\beta}_{T,\sigma}^{\top} \boldsymbol{x}_{T,\sigma}^{(t)}, \tag{4}$$

where the covariate vectors $\boldsymbol{x}_{T,\mu}^{(t)}$ and $\boldsymbol{x}_{T,\sigma}^{(t)}$ include constant terms for the intercept, as well as a set of time-varying Fourier features to capture seasonal variations in both mean temperature and variability. To account for temporal dependence, the covariate sets for the location-parameter $\boldsymbol{x}_{T,\mu}^{(t)}$ additionally includes lagged values $T_{\text{avg}}^{(t-1)}$ and $T_{\text{avg}}^{(t-2)}$, as well as their interactions with the seasonal Fourier terms. These interactions allow the model to capture potential seasonal changes in autocorrelation, which are frequently observed for temperature.

Daily temperature range: T_{range} is a strictly positive variable, with a marginal distribution that is light right tailed and has essentially zero probability mass at values close to zero. We explored various distributions to model T_{range}, including truncated-normal, log-normal, and Gamma distributions. One issue that was persistent across all of these distributions was the presence of right tails that were too heavy, resulting in simulated diurnal

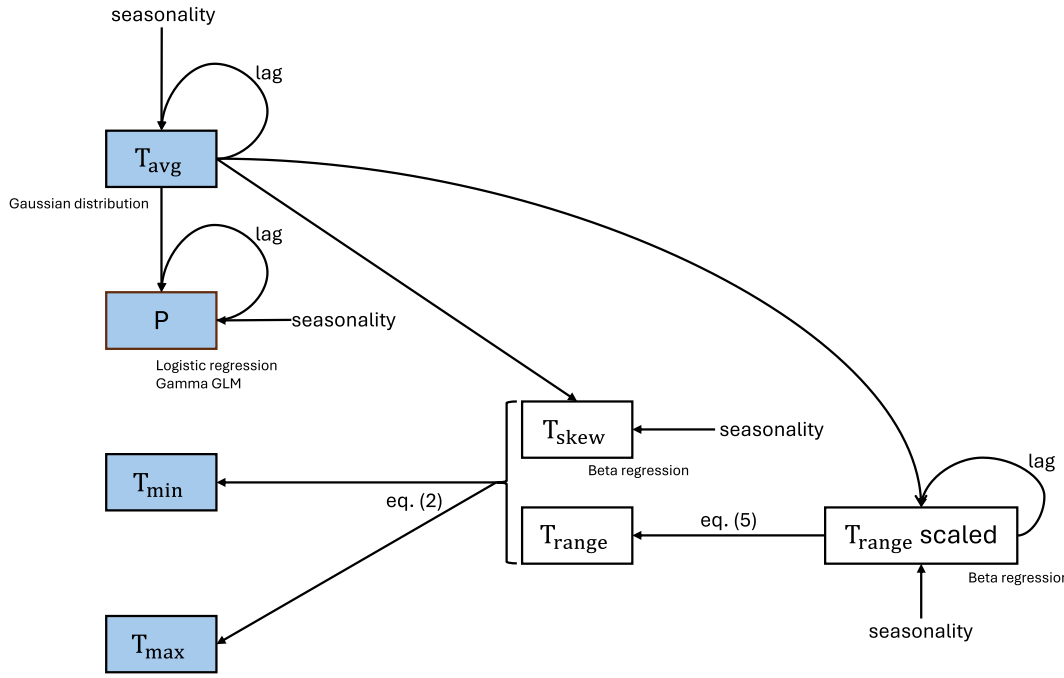

Figure 1. Modeling approach to daily precipitation (P) and average (T_{avg}), minimum (T_{min}) and maximum (T_{max}) temperature.

temperature ranges that were unrealistically high compared to the observations. As a solution, we consider the scaled temperature range, τ, defined as

$$\tau^{(t)} = \frac{T^{(t)}_{\text{range}}}{T^{\text{max}}_{\text{range}}} \in (0,1), \tag{5}$$

where $T^{\text{max}}_{\text{range}}$ is the maximum possible temperature range. Since $\tau^{(t)} \in (0,1)$, it is natural to model its conditional distribution with a Beta distribution (Ferrari & Cribari-Neto, 2004, see Appendix A for the density), where both the logit-transformed mean μ_τ and logarithmic dispersion ϕ_τ are functions of the predictors, that is,

$$\tau^{(t)} | \boldsymbol{x}^{(t)}_{\tau,\mu}, \boldsymbol{x}^{(t)}_{\tau,\phi} \sim \mathcal{B}\left(\mu^{(t)}_\tau, \phi^{(t)}_\tau\right), \quad \log \frac{\mu^{(t)}_\tau}{1-\mu^{(t)}_\tau} = \boldsymbol{\beta}^\top_{\tau,\mu} \boldsymbol{x}^{(t)}_{\tau,\mu}, \quad \log \phi^{(t)}_\tau = \boldsymbol{\beta}^\top_{\tau,\phi} \boldsymbol{x}^{(t)}_{\tau,\phi}, \tag{6}$$

Both $\boldsymbol{x}^{(t)}_{\tau,\mu}$ and $\boldsymbol{x}^{(t)}_{\tau,\phi}$ include an intercept as well as Fourier features accounting for seasonality and two time lags $\tau^{(t-1)}, \tau^{(t-2)}$ as predictors, both as main effects and as interactions with seasonality to account for seasonal variations in the autocorrelation of the diurnal temperature range. Finally, $\boldsymbol{x}^{(t)}_{\tau,\mu}$ also includes the simulated value $T^{(t)}_{\text{avg}}$ at the current time step, as linear effect and in interaction with seasonality, in order to account for correlations between average temperature and the diurnal temperature range.

Daily temperature skew: Similar to τ, the temperature skew, T_{skew}, is double-bounded on the unit interval and thus can be appropriately modeled as a Beta distribution,

$$T^{(t)}_{\text{skew}} | \boldsymbol{x}^{(t)}_{s,\mu}, \boldsymbol{x}^{(t)}_{s,\phi} \sim \mathcal{B}\left(\mu^{(t)}_s, \phi^{(t)}_s\right), \quad \log \frac{\mu^{(t)}_s}{1-\mu^{(t)}_s} = \boldsymbol{\beta}^\top_{T_{\text{skew}},\mu} \boldsymbol{x}^{(t)}_{s,\mu}, \quad \log \phi^{(t)}_s = \boldsymbol{\beta}^\top_{T_{\text{skew}},\phi} \boldsymbol{x}^{(t)}_{s,\phi}. \tag{7}$$

Both, $\boldsymbol{x}^{(t)}_{s,\mu}$ and $\boldsymbol{x}^{(t)}_{s,\phi}$ include Fourier features and the simulated value $T^{(t)}_{\text{avg}}$ at the current time step as predictors, as well as an interaction between them, in order to promote consistency between the skew and average temperature.

Observed time series of temperature skew appear to have very little autocorrelation, thus we do not include lagged variables as predictors.

Precipitation: We model the occurrence and intensity of precipitation separately, as is common in weather generators for precipitation (Chandler, 2020; Richardson, 1981; Wilks & Wilby, 1999). This is justified by the fact that the cumulative distribution function of precipitation can be written as

$$\begin{aligned}\Pr(P \le x) &= \Pr(P \le x|P=0)\Pr(P=0) + \Pr(P \le x|P > 0)\Pr(P > 0)\\ &= \Pr(P=0) + \Pr(P \le x|P > 0)(1 - \Pr(P=0)).\end{aligned} \tag{8}$$

Here $\Pr(P = 0) = 1 - \Pr(P > 0)$ corresponds to the probability of no precipitation (non-occurrence) and $F_R(x) \coloneqq \Pr(P \le x|P > 0)$ can be derived from a model for the precipitation intensity R, given a positive precipitation. We model precipitation occurrence $O^{(t)}$ at time t as a Bernoulli-distributed indicator variable:

$$O^{(t)}|\boldsymbol{x}_{P_O}^{(t)} \sim \mathrm{Ber}\left(p_O^{(t)}\right), \quad \log\frac{p_O^{(t)}}{1-p_O^{(t)}} = \boldsymbol{\beta}_{P_O}^{\top}\boldsymbol{x}_{P_O}^{(t)}. \tag{9}$$

For the precipitation intensity R, we found the Gamma distribution to provide a reasonable fit (see Appendix A for the employed density that results in the location parameter $\mu_R^{(t)}$ being equal to the expectation),

$$R^{(t)}|\boldsymbol{x}_{R,\mu}^{(t)}, \boldsymbol{x}_{R,\lambda}^{(t)} \sim \mathrm{Ga}\left(\mu_R^{(t)}, \sigma_R^{(t)}\right), \quad \log\mu_R^{(t)} = \boldsymbol{\beta}_{R,\mu}^{\top}\boldsymbol{x}_{R,\mu}^{(t)}, \quad \log\lambda_R^{(t)} = \boldsymbol{\beta}_{R,\lambda}^{\top}\boldsymbol{x}_{R,\lambda}^{(t)}. \tag{10}$$

The predictor vectors $\boldsymbol{x}_{P_o}^{(t)}, \boldsymbol{x}_{R,\mu}^{(t)}$ and $\boldsymbol{x}_{R,\lambda}^{(t)}$ each include an intercept and a set of Fourier features to capture seasonal variation. To account for temporal dependence $\boldsymbol{x}_{P_o}^{(t)}$ and $\boldsymbol{x}_{R,\mu}^{(t)}$, include $\log(1 + x)$-transformed precipitation amounts from the two previous days ($P^{(t-1)}$ and $P^{(t-2)}$); and the predictors for the occurrence model $\boldsymbol{x}_{P_o}^{(t)}$ additionally include binary indicators of precipitation occurrence on these days. All autocorrelation effects enter the model both as main effects, as well as in interaction with the Fourier features, as the strength of temporal dependence of rainfall is known to vary seasonally in the mid-latitudes (Chandler & Wheater, 2002). Finally, simulated mean air temperature $T_{\mathrm{avg}}^{(t)}$ is also included as a predictor in $\boldsymbol{x}_{P_o}^{(t)}, \boldsymbol{x}_{R,\mu}^{(t)}$, both again as main effect and as interactions with the Fourier features, since the relationship between temperature and precipitation is found to vary throughout the year.

2.2. Calibration on Historical Data

GAMLSS are commonly fit using (penalized) maximum likelihood estimation, and the corresponding confidence intervals are based on asymptotic theory. However, it has been shown empirically (Klein et al., 2015), that these confidence intervals tend to be too narrow and this issue is even more pronounced when computing transformations of the original regression coefficients. Klein et al. (2015) develop a Bayesian formulation of GAMLSS and show that the credible intervals derived from posterior samples are more reliable. We therefore adopt a Bayesian approach to weather generation by placing priors on the 124 free regression coefficients of the weather generator, collected in

$$\boldsymbol{\theta} = \left(\boldsymbol{\beta}_{T,\mu}, \boldsymbol{\beta}_{T,\sigma}\boldsymbol{\beta}_{\tau,\mu}, \boldsymbol{\beta}_{\tau,\phi}, \boldsymbol{\beta}_{T_{\mathrm{skew}},\mu}, \boldsymbol{\beta}_{T_{\mathrm{skew}},\phi}, \boldsymbol{\beta}_{P_O}, \boldsymbol{\beta}_{R,\mu}, \boldsymbol{\beta}_{R,\lambda}\right)^{\top}.$$

We use zero-centered, independent normal distributions, $\mathcal{N}\left(0, \sigma_\beta^2 I\right)$, as calibration priors for all of the coefficients, $\boldsymbol{\beta}$. In most cases, we set the prior variance to $\sigma_\beta = 1$; for autoregressive lag coefficients, however, we use a reduced variance of $\sigma_\beta^2 = 0.5$ in order to discourage unrealistically large autocorrelations and large values of β which might result in divergent dynamics.

Due to the complex structure of the likelihood, the posterior density is not available in closed form. Markov Chain Monte Carlo (MCMC) algorithms are generally considered the gold-standard for approximate posterior inference (Gelman et al., 2013). However, MCMC does not scale well to large data sets such as multi-decadal daily time

series. Furthermore, it provides only samples from the posterior rather than a closed-form approximation. In light of this, we instead opt for stochastic variational inference (SVI; Hoffman et al., 2013) as a more efficient method for approximating the posterior of $\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{x}$. We choose a multivariate Gaussian as the variational approximation, which we denote as $q_{\mathrm{SVI}}(\boldsymbol{\theta}) = \mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{SVI}}, \Sigma_{\mathrm{SVI}}\right)$, with the covariance parameterized in terms of its Cholesky decomposition. Once the posterior approximation is obtained, realizations of simulated time series can be easily generated. A pseudocode algorithm for this is given in Algorithm 1 in the Appendix.

Our models are implemented in Python using the `numpyro` probabilistic programming language (Bingham et al., 2019; Phan et al., 2019), which is based on the `jax` framework (Bradbury et al., 2018) for high performance differentiable programming. Calibration on 50 years of daily data using SVI takes ∼2 min on commercial hardware (Apple M4 CPU, 16 Gb RAM). Once calibrated, the model is capable of simulating 100 synthetic realizations of the 50-year time series of daily data (a total of 50,000 years) in about 10 s.

3. Climatology Matching With Simulation-Based Inference

The weather generator described in Section 2, once calibrated for a particular site, enables highly efficient simulation of multiple realizations of meteorological time series based on the weather patterns observed in the calibration data. Large-scale climatological patterns in the observed data are implicitly encoded in the posterior distribution over model parameters, $p(\boldsymbol{\theta}|\boldsymbol{Y})$. However, this distribution represents only a relatively narrow range of climatic conditions similar to the climate of the calibration period. For some applications, such as scenario-neutral climate impact analysis, we are interested in simulating conditions that require parameters outside of this range. Climate conditions are often summarized using a set of summary statistics, $\boldsymbol{s} \coloneqq h(\boldsymbol{Y})$, which effectively summarize the observed climatology over annual or multi-annual time periods. The precise choice of h is arbitrary and context-dependent; however, it is typically a function that deterministically aggregates the time series into a vector summarizing the long-term climate. In order to "adjust" a weather generator to produce plausible time series under an altered climate, we must find one or more parameter vectors, $\boldsymbol{\theta}^{\star}$, which produce time series that follow a predefined set of target summary statistics, $\boldsymbol{s}^{\star} = \boldsymbol{s}_{\mathrm{obs}} + \Delta\boldsymbol{s}$. Here, $\boldsymbol{s}_{\mathrm{obs}}$ refers to the observed summary statistics computed from the calibration data, $\boldsymbol{Y}$, and $\Delta\boldsymbol{s}$ refers to a perturbation reflecting the climatic change of interest.

As described by Guo et al. (2018) and Culley et al. (2019), this implies an inverse problem of mapping from the target climatology, described by $\boldsymbol{s}^{\star}$, to weather generator parameters which would, in expectation, produce time series matching that climatology. Guo et al. (2018) proposed to solve the inverse problem by applying numerical optimization techniques to a Richardson-type weather generator. However, their approach necessitates fixing the random seed of the weather generator in order to make the optimization problem tractable; this neglects the uncertainty implied by the internal variability of the weather generator, which may lead to biased parameter estimates and underestimation of uncertainty for any calibration period of finite length. We argue, therefore, that the inherent stochasticity of the weather generator necessitates a fully probabilistic treatment of the inverse problem.

3.1. Probabilistic Formulation of the Inverse Problem

The inverse problem of mapping from a particular choice of summary statistics to parameters is fundamentally ill-posed; that is, for any fixed choice of summary statistics, there necessarily exists an infinite number of parameter settings which could hypothetically produce a time series matching those statistics. A key advantage to adopting a Bayesian approach, however, is that such ill-posed problems are transformed into well-posed problems of probabilistic inference. Under the Bayesian interpretation, we seek to obtain a posterior distribution over parameters, $\boldsymbol{\theta}$, given the target summary statistics $\boldsymbol{s}^{\star}$,

$$p\left(\boldsymbol{\theta}|\boldsymbol{s}^{\star}\right) \propto p(\boldsymbol{\theta}) p\left(\boldsymbol{s}^{\star}|\boldsymbol{\theta}\right), \tag{11}$$

$$p\left(\boldsymbol{s}^{\star}|\boldsymbol{\theta}\right) = \int_{y} p\left(\boldsymbol{s}^{\star}|\tilde{\boldsymbol{Y}}\right) p\left(\tilde{\boldsymbol{Y}}|\boldsymbol{\theta}\right) d\tilde{\boldsymbol{Y}}. \tag{12}$$

Here, $p(\boldsymbol{\theta})$ corresponds to a prior distribution over weather generator parameters which we refer to as the *climatology matching prior*, and $p(\boldsymbol{s}^{\star}|\boldsymbol{\theta})$ is the corresponding likelihood which describes how likely the target

summary statistics would be for any given choice of parameters. The distribution $p(\tilde{\boldsymbol{Y}}|\boldsymbol{\theta})$ characterizes the predictive distribution over simulated time series for a particular choice of parameters $\boldsymbol{\theta}$, and $p(\boldsymbol{s}^\star|\tilde{\boldsymbol{Y}})$ specifies the likelihood of a particular choice of target summary statistics, $\boldsymbol{s}^\star$, for simulated time series $\tilde{\boldsymbol{Y}} \in \mathcal{Y}$. In the typical case where $\boldsymbol{s} = h(\tilde{\boldsymbol{Y}})$ is assumed strictly deterministic, this distribution reduces to a Dirac-delta with nonzero density only on simulated time series where the summary statistics exactly match the target. Note, however, that we could alternatively choose any arbitrary density for $p(\boldsymbol{s}^\star|\tilde{\boldsymbol{Y}})$ that assigns non-zero weight to simulated statistics based on their distance from the target values.

The posterior distribution on the left-hand side of Equation 11 describes the most likely values of the weather generator parameters, given the target summary statistics, $\boldsymbol{s}^\star$. Predictive uncertainty in the simulated summary statistics $\tilde{\boldsymbol{s}}$ is then represented by the *posterior predictive* distribution,

$$p(\tilde{\boldsymbol{s}}|\boldsymbol{s}^\star) = \int_\Theta p(\tilde{\boldsymbol{s}}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\boldsymbol{s}^\star)\, d\boldsymbol{\theta}, \tag{13}$$

which marginalizes the predicted summary statistics over the climatology matching posterior in Equation 11.

Unfortunately, unlike the calibration procedure discussed in Section 2.2, the use of low-dimensional summary statistics requires integration over all possible time series simulated by the weather generator, as seen on the right-hand side of Equation 12. Such an integral is clearly intractable given the high-dimensionality of the simulated time series, thereby precluding the use of standard inference algorithms such as SVI that require computation of the likelihood. As such, a different class of algorithms is necessary in order to approximate the climatology matching posterior distribution.

3.2. Simulation-Based Inference

Statistical inference for problems with intractable likelihoods is commonly referred to as simulation-based (or "likelihood-free") inference (SBI; Cranmer et al., 2020). Early work investigated the use of kernel density estimators in the one-dimensional case (Diggle & Gratton, 1984), while Tavaré et al. (1997) and Beaumont et al. (2002) later introduced a more flexible class of methods referred to as Approximate Bayesian Computation (ABC; Sisson et al., 2018). ABC aims to approximate the "true" likelihood using a distance function $d(\boldsymbol{s},\boldsymbol{s}')$ defined over a set of summary statistics. ABC methods, however, have three key limitations (Papamakarios & Murray, 2016): First, they do not provide direct estimates of the posterior density but rather samples which permit only (noisy) calculations of expectations, similar to Markov chain Monte Carlo simulations. Second, the samples are not actually distributed according to the "true" posterior but rather an approximation controlled by a chosen acceptance threshold, ϵ and the distance metric d. Third, as $\epsilon \to 0$ and the dimensionality of the parameter space increases, the number of simulations required becomes intractably large, making ABC inefficient for problems such as climatology matching.

The limitations of traditional ABC methods, along with the advent of so-called *neural density estimators* in the field of probabilistic deep learning, have inspired the use of modern techniques such as mixture-density networks (Bishop, 1994) and normalizing flows (Papamakarios et al., 2021; Rezende & Mohamed, 2016) to directly approximate the posterior distribution from simulated data (Papamakarios & Murray, 2016). This task is now commonly referred to as *neural posterior estimation* (NPE; Lueckmann et al., 2017). We apply the *automated posterior transformation* (APT) algorithm of Greenberg et al. (2019), also sometimes referred to as (S)NPE-C. APT approximates the posterior distribution Equation 11 by maximizing the likelihood of the simulated parameter-data pairs under a so-called *proposal posterior*,

$$\tilde{q}(\boldsymbol{\theta}|\boldsymbol{s}^\star) \propto \frac{\tilde{p}(\boldsymbol{\theta})}{p(\boldsymbol{\theta})} q_{\boldsymbol{\psi}}(\boldsymbol{\theta}|\boldsymbol{s}^\star) \tag{14}$$

where $q_\psi(\boldsymbol{\theta}|\boldsymbol{s}^\star)$ is any suitable neural density estimator parameterized by $\boldsymbol{\psi}$ and $\tilde{p}(\boldsymbol{\theta})$ is the "proposal" distribution from which parameter vectors are sampled, initially chosen to be the prior. Given a finite set of simulated

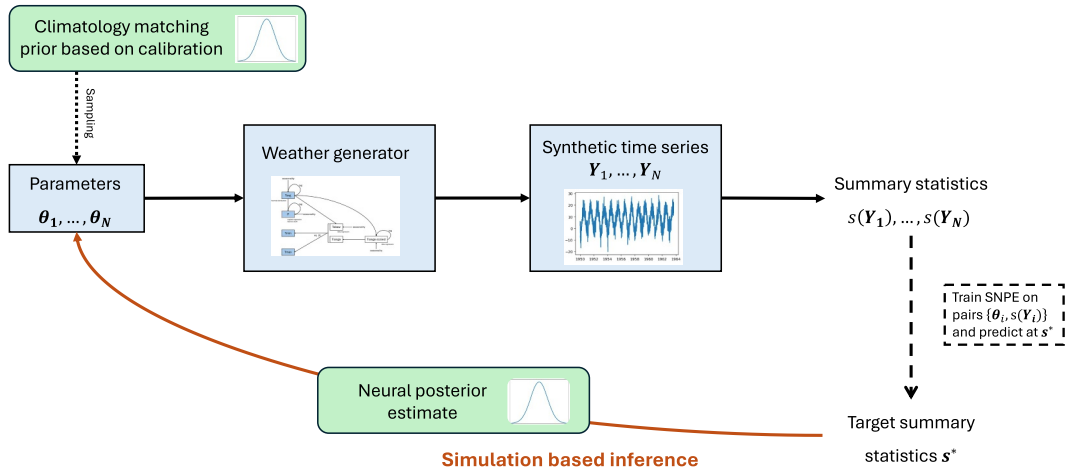

Figure 2. SBI approach to generate synthetic time series matching summary statistics of interest.

parameters and summary statistics $\{(\boldsymbol{\theta}_i, \boldsymbol{s}_i)\}_{i=1}^N$, $q_{\boldsymbol{\psi}}$ can be trained by minimizing $\mathcal{L}(\boldsymbol{\psi}) = -\sum_{i=1}^N \log q_{\boldsymbol{\psi}}(\boldsymbol{\theta}_i|\boldsymbol{s}_i)$. As $N \to \infty$, it can be shown that sufficiently expressive choices of $q_{\boldsymbol{\psi}}$ will converge to the true posterior distribution (Greenberg et al., 2019; Papamakarios & Murray, 2016).

Since the prior is rarely an efficient proposal distribution for sampling from the posterior, it is often advantageous to split inference into multiple, sequential "rounds", where the previous round's posterior is used as the next round's proposal. This is typically referred to as *sequential* NPE (SNPE). Sequential inference enables faster convergence by increasing sampling efficiency within each inference round. Another key benefit of the APT algorithm is that, once an estimate of $\boldsymbol{\psi}^\star$ is obtained, the resulting density estimator can be used to approximate the posterior distribution of $\boldsymbol{\theta}$ given any value of the summary statistics $\boldsymbol{s}^\star$ via $q_{\hat{\boldsymbol{\psi}}}(\boldsymbol{\theta}|\boldsymbol{s}^\star)$ in an "amortized" fashion without the need for re-training as required by, for example, ABC. See Appendix B and Appendix D for further discussion on diagnostics and amortized inference in climatology matching.

3.3. Application of SBI to Stochastic Weather Generators

We now outline a general procedure for generating realistic synthetic weather time series under different climate conditions defined by target summary statistics $\boldsymbol{s}^\star$. Recall that our primary objective is to obtain the posterior distribution over weather generator model parameters on the left-hand side of Equation 11, conditional on a set of target summary statistics, $\boldsymbol{s}^\star$. These parameters can then be used to generate new time series following the target statistics. The procedure can be summarized as follows:

1. Obtain an estimate $q_{\text{SVI}}(\boldsymbol{\theta})$ of the calibration posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y})$ of the GAMLSS weather generator, as discussed in Section 2.2.
2. Derive a climatology matching prior $p(\boldsymbol{\theta})$ from the variational approximation, $q_{\text{SVI}}(\boldsymbol{\theta})$, to the calibration posterior.
3. Draw N parameter samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ from the proposal distribution $\tilde{p}(\boldsymbol{\theta})$ (initially set to the climatology matching prior), and use the weather generator $p(\boldsymbol{Y}_i|\boldsymbol{\theta}_i)$ in Equation 3 to simulate time series $\boldsymbol{Y}_i$ given the parameters $\boldsymbol{\theta}_i$; then compute the respective summary statistics, $\{\boldsymbol{s}_i = h(\boldsymbol{Y}_i)\}_{i=1}^N$.
4. Train SNPE via the APT algorithm on $\{(\boldsymbol{\theta}_i, \boldsymbol{s}_i)\}_{i=1}^N$ to obtain an estimator, $q_{\boldsymbol{\psi}}(\boldsymbol{\theta}|\boldsymbol{s})$, of the climatology matching posterior density and predict $q_{\boldsymbol{\psi}}(\boldsymbol{\theta}|\boldsymbol{s}^\star) \approx p(\boldsymbol{\theta}|\boldsymbol{s}^\star)$, for the given target summary statistics, $\boldsymbol{s}^\star$.
5. Repeat steps 3 and 4 for K rounds, using $q_{\boldsymbol{\psi}}$ from the last round as the new proposal distribution $\tilde{p}(\boldsymbol{\theta})$.
6. Draw N samples $\{\theta_1^*, \ldots, \theta_N^*\}$ from the final posterior approximation, $q_{\boldsymbol{\psi}}^\star(\boldsymbol{\theta}|\boldsymbol{s}^\star)$, and use the weather generator to simulate time series from each parameter sample to obtain a distribution over $\boldsymbol{Y}$. Computing summary statistics on these samples yields an estimate of the posterior predictive distribution given in Equation 13.

A general overview of the workflow is shown in Figure 2. The collection of synthetic time series produced by simulations from the (approximate) posterior samples can be interpreted as an ensemble of potential realizations

of plausible meteorological time series which will, in expectation, follow the climatological characteristics captured by the summary statistics, up to the limits of approximation errors in the posterior distribution. As is the case for most models, it is important to note that the inferred posterior distribution is conditional on the choice of weather generator and climatology matching prior.

The climatology matching prior $p(\boldsymbol{\theta})$ plays a crucial role in this workflow as it constrains the space of weather generator parameters initially explored by the SBI algorithm. One natural choice would be the variational approximation of the calibration posterior obtained from SVI, $q_{\text{SVI}}(\boldsymbol{\theta})$. Intuitively, using the calibration posterior as the prior for climatology matching means that samples from the climatology matching posterior will be more similar to the historically calibrated parameters, thereby producing more plausible weather patterns. In practice, however, the calibration posterior will tend to be too concentrated around the climate of the calibration period and will not necessarily provide good coverage over many plausible alternative climatic conditions.

To address this, we construct the climatology matching prior as a multivariate normal distribution, $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\text{CM}}, \boldsymbol{\Sigma}_{\text{CM}})$, centered on the mean of the approximated calibration posterior (i.e., $\mu_{\text{CM}} \coloneqq \mu_{\text{SVI}}$) with covariance,

$$\boldsymbol{\Sigma}_{\text{CM}} = \gamma^2 \cdot \text{diag}(\boldsymbol{\Sigma}_{\text{SVI}}), \tag{15}$$

where $\text{diag}(\boldsymbol{\Sigma}_{\text{SVI}})$ corresponds to the diagonal of the variational posterior covariance matrix with all non-diagonal entries, capturing the correlations between parameters, set to zero. The augmented diagonal covariance matrix, $\boldsymbol{\Sigma}_{\text{CM}}$, inflates the posterior variance along each dimension, thereby spreading out the prior probability mass and broadening the "search space" considered by the SBI algorithm. We also experimented with using the full covariance matrix in the case studies below and found there to be no substantial difference in the coverage of the resulting posterior predictive distribution after a fixed number of SBI rounds, compared to only using rescaled marginal variances. It is worth noting that the climatology matching prior could also be constructed from a maximum likelihood estimate. In this case, one would take the maximum likelihood estimate for the mean and estimated standard errors instead of the variational inference covariance. These standard deviations could then be rescaled in the same manner described above.

The scaling parameter γ in Equation 15 determines the spread of the prior and thus in turn how much of the parameter space is initially explored by the SBI algorithm. We found that values of γ between 2.5 and 3.5 typically produce good results based on prior predictive checks, i.e., comparing simulated summary statistics produced by samples from the climatology matching prior to the target values. These diagnostic checks are important because if γ is chosen to be too small the SBI algorithm may still converge but will typically require more rounds to do so. In contrast, if γ is chosen to be too large this can lead to unstable (i.e., divergent) or unrealistic weather generator simulations, and thus less usable training data for the SBI algorithm. As we discuss further in the next section, selectively choosing parameters for which to scale the variance can make the climatology matching prior more informative in cases where additional a priori constraints are desired. See Appendix B for more discussion on the use of diagnostic checks in the design of the climatology matching prior.

Finally, we note that one key assumption of the climatology matching method is the assumption of structural stationarity of the weather generator. The SBI method only modifies parameters of the weather generator; the model structure, however, remains the same between the historical calibration period and period with perturbed climatology. For example, in the parametric weather generator described above, the distributional forms for each variable will remain the same between the original and perturbed climatology. Furthermore, since the weather generator describes a stationary process, it will not be able to capture summary statistics representing trends in climatology matching unless nonstationary covariates are included in the model. How realistic the assumption of structural stationarity is depends on the complexity of the weather generator and the time scales considered. We discuss modes of model failure further in Appendix B.

3.4. Sources of Uncertainty in Climatology Matching

The climatology matching posterior in Equation 11 fully describes the uncertainty in the weather generator parameters, given a particular choice of summary statistics, while the posterior predictive in Equation 13 additionally accounts for the internal stochasticity of the weather generator. The resulting variability in the simulated

time series therefore reflects several different potential sources of uncertainty which are worth discussing in more detail.

Uncertainties in probabilistic modeling are often divided into two general categories: epistemic and aleatoric (Der Kiureghian & Ditlevsen, 2009; Heße et al., 2019; Hüllermeier & Waegeman, 2021). In short, epistemic uncertainty is that which arises from a lack of information, while aleatoric uncertainty can be seen as that which arises from stochasticity in the model. Uncertainty due to the internal stochastic variability of the weather generator can be considered as aleatoric; it is present even for a fixed choice of parameters and mimics the local-scale effects of the climate system's internal variability. For any given model, aleatoric uncertainty also implicitly includes any potential sources of systematic error arising from discrepancies between the model and the true data-generating process. In contrast, uncertainty in the estimated the weather generator parameters, characterized by the posterior distribution in Equation 11, can be considered epistemic. In the problem setting of climatology matching, this uncertainty arises primarily from three sources: (a) the initial uncertainty in weather generator parameters expressed by the climatology matching prior distribution, (b) the sufficiency of the chosen summary statistics to adequately constrain the uncertainty expressed by the prior, and (c) the non-uniqueness of the solution to the inverse problem.

Every possible collection of time series, $\boldsymbol{Y}$, that is within the support of the weather generator, necessarily has a nonzero likelihood under any given choice of parameters. For any particular set of target summary statistics, multiple parameter settings of the weather generator may, therefore, be (approximately) equiprobable. For example, an increase in mean annual precipitation could be explained by a decrease in the number of dry days, an increase in average intensity (amount of rain per wet day), or simply a change in the seasonal cycle. Each of these cases would lead to statistically distinct simulated time series that produce the same summary statistics. Many different possible settings of the weather generator parameters may, therefore, be relevant for downstream impact modeling applications. These uncertainties, and their interactions with the parameterizations of the underlying weather generator, can be systematically explored by (a) adding constraints through the climatology matching prior, such as restricting the prior variance for certain parameters, as well as (b) adding or modifying summary statistics to better constrain the desired climatic conditions. We explore these strategies further in the case studies presented in Section 4.1.

In our case studies below, we found it useful to also compute the *maximum* a posteriori (MAP) estimate in addition to sampling from the full climatology matching posterior. The MAP can be obtained via gradient ascent on the density estimator, q_ψ. We typically use the MAP as a starting point for further investigation. Yet, results should be treated with care since the MAP neglects epistemic uncertainty in the inferred parameters and only captures the aleatoric uncertainty arising from the internal variability of the weather generator.

3.5. Implementation Details

We use an open source Python implementation of the SNPE-C/APT algorithm (Greenberg et al., 2019) provided by the `sbi` package (Tejero-Cantero et al., 2020), which is based on `pytorch` (Paszke et al., 2019). We use the default configuration of masked autoregressive normalizing flows (Papamakarios et al., 2018) provided by `sbi` in SNPE-C. During each round, we draw $N = 1000$ samples from the proposal/prior distribution, that are used to simulate 50-year time series from which summary statistics are computed. Running SNPE-C to find weather generator parameters takes roughly 10–30 min using a single CPU, depending on the number of simulation rounds required and the complexity of the summary statistics considered. Once the posterior estimate is obtained, synthetic time series can be simulated in a matter of seconds. In practical settings, computational efficiency can be further improved by reducing both the number of training rounds as well as the number of parameters sampled in each round. Both CPU and GPU parallelization can be used to further accelerate the weather generator simulations.

4. Application to Weather Station Data

In this section we demonstrate the efficacy of the proposed weather generator (Section 2) and climatology matching method (Section 3) on weather station data from the Potsdam surface weather observation station provided by the German Meteorological Service (Deutscher Wetterdienst, DWD; Deutscher Wetterdienst, 2024; Kaspar et al., 2013). This station has a continuous record dating from 1893 up to present day. We consider data

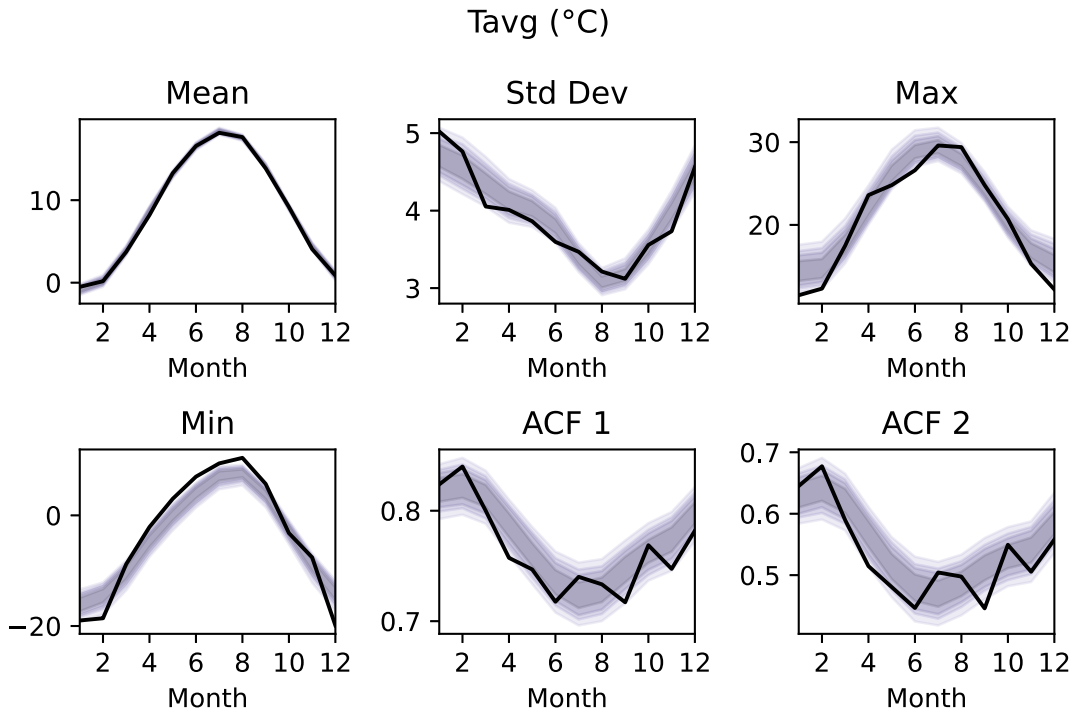

Figure 3. Monthly summary statistics for the average temperature T_{avg} in °C. Observations are given as black lines, while the shaded areas correspond to 50%, 70%, 80% and 90% confidence intervals from the simulation runs. Top row: monthly mean, monthly standard deviation, monthly maximum. Bottom row: monthly minimum, autocorrelation (ACF) at lags 1 and 2.

from 1950 to 2000 as this period exhibits relatively minimal long-term trends. We use observations of daily average, minimum and maximum temperature, as well as daily precipitation accumulation. We calibrate the weather generator using the approach outlined in Secion 2.2.

4.1. Weather Generator Validation

After calibration, we simulate 1,000 realizations of the 50-year period (50,000 years of data) and evaluate the performance of the weather generator in replicating statistics of our meteorological variables of interest. We find that our GAMLSS weather generator performs well overall, capturing all primary modes of variability. Most temperature statistics such as monthly mean, monthly variability (standard deviation), as well as monthly maxima and autocorrelation are well captured by the confidence bands obtained by the simulations (Figure 3). There seems to be some slight deviation in monthly minima of daily average temperature.

Precipitation is similarly well captured by the weather generator, with the monthly mean (both overall and on wet days), as well as the proportion of wet days and autocorrelation statistics well contained within the confidence bands obtained through the simulations (Figure 4). There is some indication that the confidence bands for the monthly standard deviation might be too narrow (particularly of the conditional standard deviation statistics). Compared to a location-only Gamma model, however, capturing both the shape and scale of the precipitation intensity Gamma distribution in a GAMLSS approach already helped with the seasonal cycle in this statistic. The monthly maxima are well reproduced throughout most of the year; there is, however, some deviation in the month of August. This deviation can be attributed to a single recorded precipitation event in August 1978 (see Der Tagesspiegel, 2003) which is currently not reproduced by the weather generator and may indicate a need for improved modeling of the right tail of the precipitation intensity distribution.

Figures for T_{range} and T_{skew}, as well as the derived T_{min} and T_{max} can be found in the Appendix (Figures C1 and C2). Both variables, temperature range and temperature skew, are well captured using Beta regressions, and modeling both the location and dispersion parameters of the Beta distribution instead of just the location improves model performance. The present-day seasonal cycles is well captured in statistics such as mean, standard deviation, minima and maxima as well as autocorrelation. The pronounced seasonal cycle in temperature skew

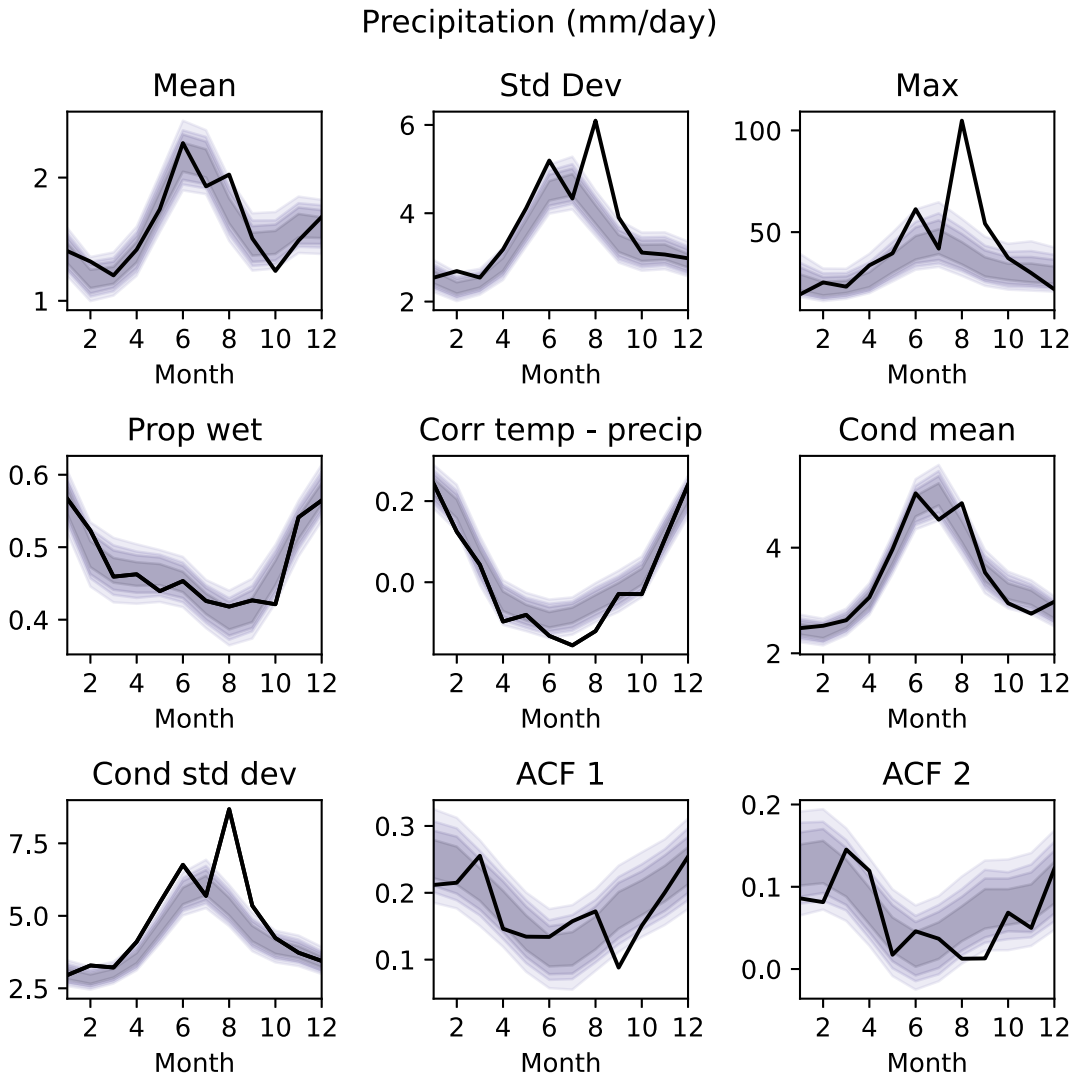

Figure 4. Monthly summary statistics for precipitation P in mm/day. Observations are shown as black lines, while the shaded areas correspond to 50%, 70%, 80% and 90% confidence intervals from the simulation runs. Top row: monthly mean, monthly standard deviation, monthly maximum. Middle row: proportion of wet days, correlation between temperature and precipitation values, mean on wet days. Bottom row: standard deviation on wet days, autocorrelation (ACF) at lags 1 and 2.

confirm our physical intuitions that the assumption of symmetric daily temperature distributions in earlier work (see e.g., Kilsby et al., 2007) may not always be appropriate. Although the daily temperature range and the monthly maxima and minima are well captured, for temperature skew they are underestimated and overestimated, respectively. This, however, appeared to be of minor importance for accurately simulating T_{min} and T_{max}.

Finally, most minimum and maximum temperature (Figure C2) statistics are well reproduced by the model using the temperature skew and range parametrization. The minima of maximum temperature and maxima of minimum temperature are slightly under- and overestimated by the model and similarly the lag-1 autocorrelation is slightly underestimated in both statistics. Whether these deviations are problematic will likely be use-case dependent. Finally, we note that capturing seasonality-dependent correlations between T_{range}, T_{skew} and T_{avg} was particularly important for simulating monthly variability in minimum and maximum temperature. This is apparent from the decomposition of maximum temperature variance, $\text{Var}(T_{\text{min}}) = \text{Var}(T_{\text{avg}}) + \text{Var}(T_{\text{range}}T_{\text{skew}}) - 2\text{Cov}(T_{\text{avg}}, T_{\text{range}}T_{\text{skew}})$, where the last term corresponds to the covariance between the range-skew product and average temperature.

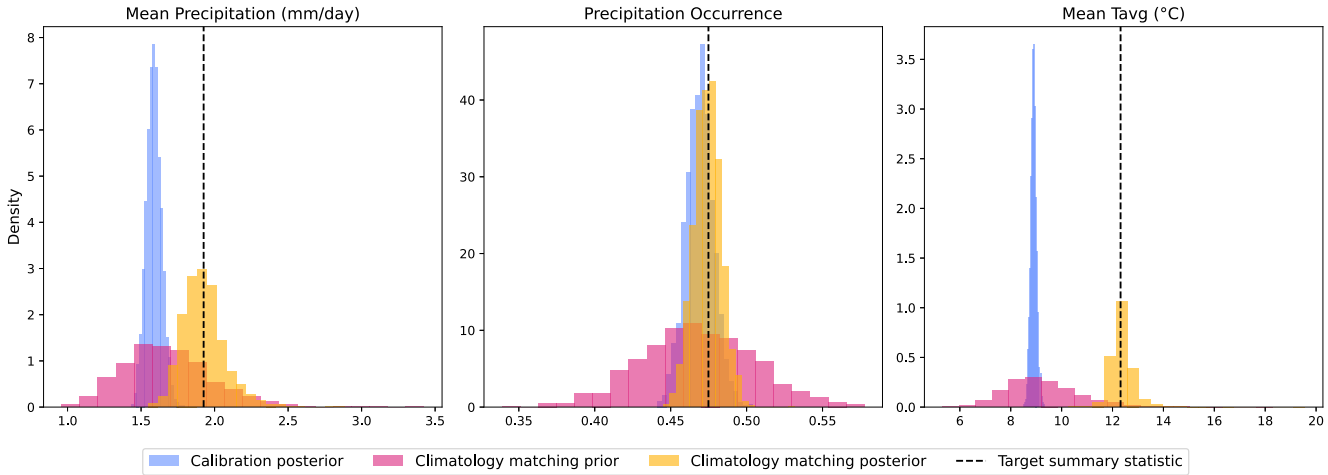

Figure 5. Case study 1: Prior and posterior predictive histograms of 2000 simulations from the calibration posterior (blue), climatology matching prior (purple) and climatology matching posterior produced by SBI (yellow) for average precipitation, precipitation occurrence and average temperature. Black lines: target summary statistics.

These results suggest that our weather generator, once calibrated, is able to produce realistic simulated time series for each of the four variables considered. In the following section, we will show how our weather generator can be used in conjunction with the SBI method outlined in Section 3.3 to produce realistic time series matching summary statistics of interest. We demonstrate our approach in a series of three case studies that reflect different levels of difficulty in terms of climatology matching. The changes in summary statistics, in particular those considered in case studies 2 and 3, cannot be adequately represented using simple scaling factor or delta change approaches. While the methods proposed by Guo et al. (2018); Culley et al. (2019) are similar in spirit, they do not address the probabilistic formulation of the climatology matching inverse problem, and employ substantially simpler weather generator models with far fewer parameters than the one considered here (8 vs. 124). Consequently, these methods would likely struggle to capture the complexity and variability represented in our case studies, particularly in cases 2 and 3.

Note that, while the perturbations of the summary statistics chosen here are purely synthetic, they are based on realistic trends reported in the Federal German Climate Risk Report (Kahlenborn et al., 2021) for a high emission scenario at the end of the century. In all cases, the perturbed climatologies are defined relative to the calibration period (1950–2000).

4.2. Case Study 1: Changes in Long-Term Averages

In this first case study we aim to generate time series with the following changes in 50-year average statistics of average temperature and precipitation:

- Change in average daily temperature: +3.5°C warming. Target value: 12.32°C.

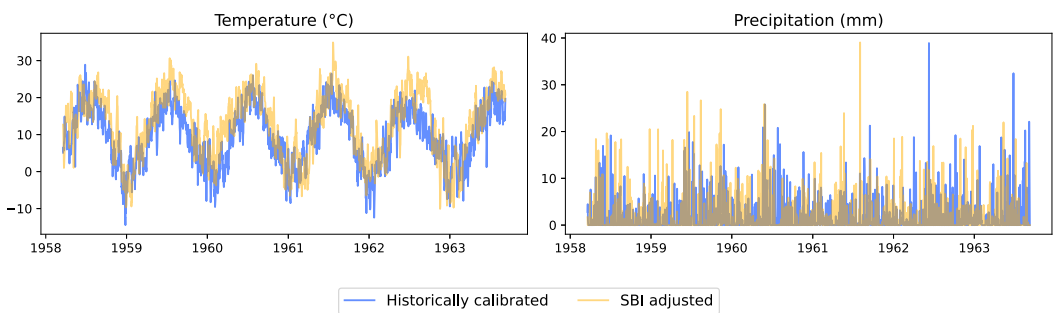

Figure 6. Case study 1: Simulated precipitation and temperature time series from the historically calibrated and climatology matched weather generator.

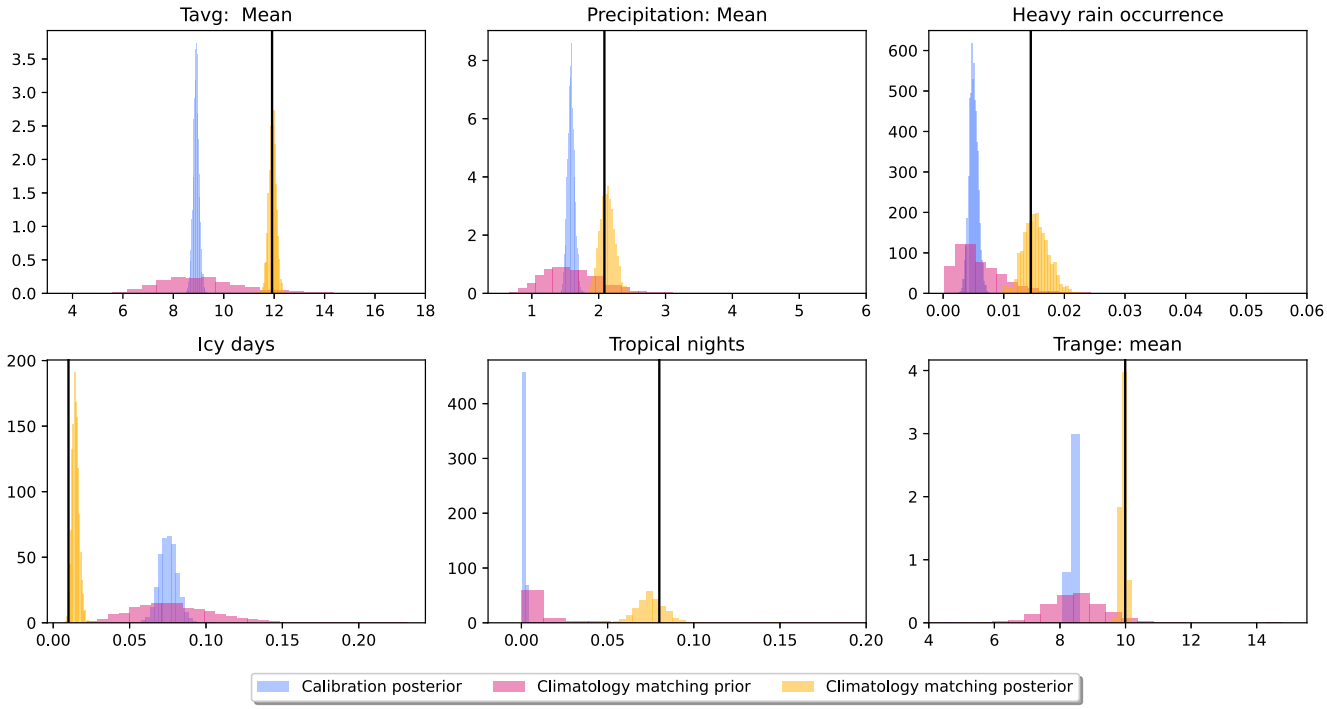

Figure 7. Case study 2: Prior and posterior predictive histograms of 2000 simulations from the calibration posterior (blue), climatology matching prior (purple) and climatology matching posterior (yellow) for average temperature, precipitation, heavy rain occurrence (top), icy days, tropical nights, temperature range (bottom). Black lines: target summary statistics.

- Change in average daily precipitation: 20% increase. Target value: 1.93 mm.
- No change in the proportion of wet days (held constant). Target value: 47.49%.

All changes are defined relative to observed values and align with projected regional changes under the RCP-8.5 scenario 2070–2100 (Kahlenborn et al., 2021). We specifically examine a scenario where the increase in average precipitation occurs while maintaining a constant proportion of wet days. This represents a change in precipitation primarily driven by thermodynamics, assuming that the atmospheric moisture content increases, while the dynamics governing precipitation occurrence remain unchanged. Seasonal variation in summary statistics is neglected in this experiment but is considered in the experiments presented in Section 4.4.

We follow the methodology outlined in Section 3.3 to recalibrate the weather generator to the target climatology. Given a weather generator calibrated on historical time series data, our first step is to inflate the variance of the calibration posterior in order to construct a prior for climatology matching. Since our focus is on changes in average temperature and precipitation amounts, we do not apply the variance inflation to parameters related to temperature range, skewness, or precipitation occurrence. Instead, we leave the priors for these variables the same as the calibration posterior. The calibration posterior is relatively narrow, with prior predictive simulations closely centered around historically observed values for all three variables (Figure 5). In contrast, the variance-inflated calibration posterior, which we use as the prior for climatology matching, encompasses a much broader range of possible summary statistics (Figure 5).

We run the SBI algorithm (SNPE-C/APT) for $K = 10$ rounds (see Section 3.3). The target summary statistics (black lines) are well captured by the climatology matching posterior (green densities in Figure 5), even for average temperature, which was initially far from the historical value and assigned relatively minimal probability mass by the prior. For precipitation occurrence the SBI posterior strongly resembles the calibration posterior and is centered on the historical value. Sampling from the climatology matching posterior enables simulation of an ensemble of meteorological time series which, in expectation, follow the target summary statistics. Since each parameter sample can be seen as a separate instantiation of the weather generator model, they can be explored systematically using approaches outlined in Section 3.3. A natural first step is to consider the MAP point estimate, which represents the most likely value under the chosen prior and target summary statistics. Precipitation and

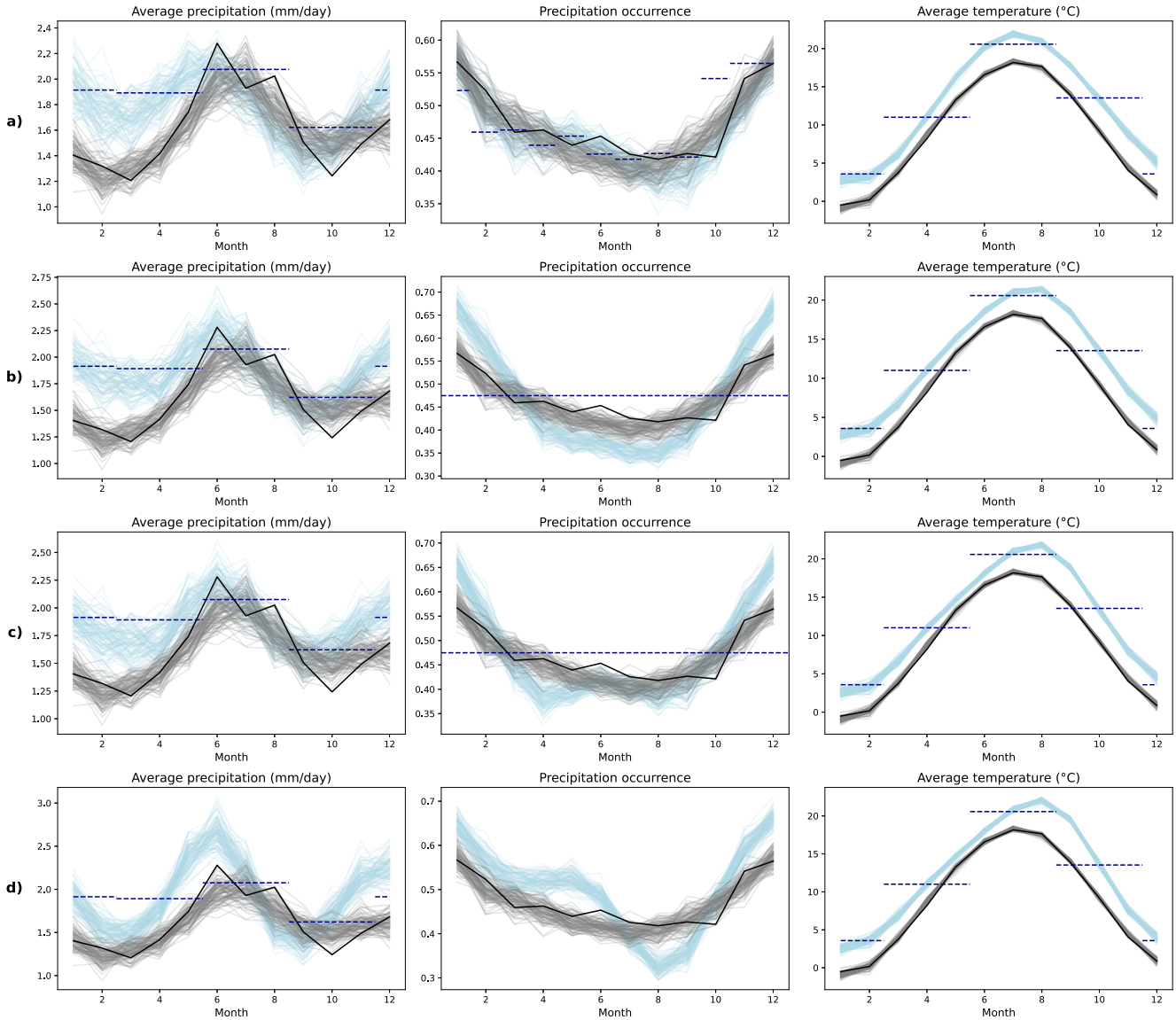

Figure 8. Case study 3: Average monthly precipitation, precipitation occurrence and temperature from 100 50 years realizations of the historically calibrated (gray) and MAP estimate of the climatology matched (blue) weather generator, together with observations (black line) for all four cases (a–d) of the case study. The dashed lines correspond to the target summary statistics on either monthly, seasonal, or yearly level.

temperature time series generated by the climatology matched weather generator shifts the temperature upwards and produces heavier precipitation events compared to the non-perturbed simulations (Figure 6).

In order to show the general applicability of the method, we repeated case study 1 at a number of different locations from the DWD observation network. The results are shown in Appendix E. In all cases considered, the SBI posterior successfully covers the target summary statistics.

4.3. Case Study 2: Changes in Threshold-Based Climate Indices

We now consider a set of summary statistics which affects all four variables simultaneously:

- Change in average daily temperature: +3.1°C warming. Target value: 11.92°C.
- Change in average daily precipitation: 30% increase. Target value: 2.09 mm.
- Change in the number of heavy rain days (days with rain over 20 mm): 100% increase. Target value: 0.0145% of days.

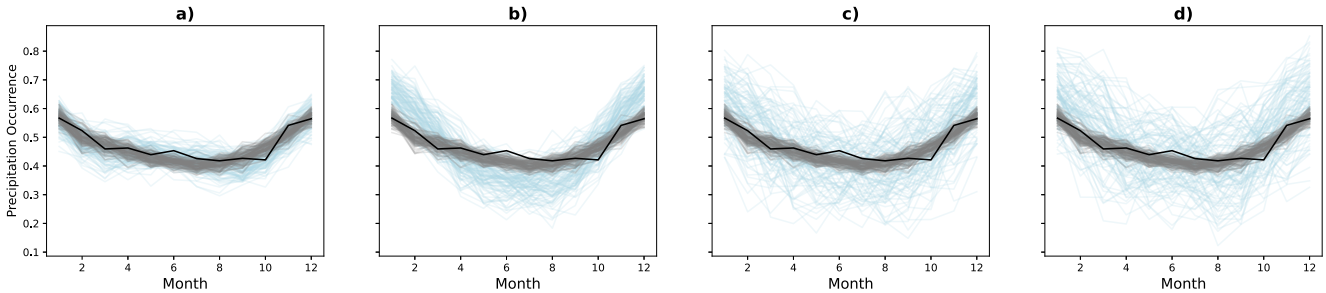

Figure 9. Case study 3: Average monthly precipitation occurrence from 100 50-year realizations of the historically calibrated weather generator (gray) and the full climatology matching posterior (blue), together with observations (black line), for all settings (a–d).

- Change in icy days ($T_{\max} < 0$°C): target of 3 days per year (on average).
- Change in tropical nights ($T_{\min} > 20$°C): target of 20 nights per year (on average).
- Change in T_{range}: 20% increase. Target value: 9.99°C.

The changes relative to the historical period are again motivated by trends reported in Kahlenborn et al. (2021) for a high-emission scenario at the end of the century. Changes in T_{range} are inspired by Liu et al. (2024). For the climatology matching prior, we apply variance inflation to all parameters of the weather generator except the Fourier features, as we are only interested in yearly changes.

Figure 7 shows histograms of all the summary statistics above generated by the calibration posterior, the climatology matching prior, and the climatology matching posterior for $K = 10$ iterations. The target summary statistics (black) are again well captured by the climatology matching posterior. This holds for all statistics considered, including those for tropical nights and icy days, which are statistics on $T_{\min}$ and $T_{\max}$. Any change imposed on the latter variables necessarily includes modifications to the $T_{\text{avg}}, T_{\text{range}}$ and T_{skew} models. However, we find that SNPE nevertheless converges, even in such cases where the target value is very far from the historically observed value and climatology matching prior.

4.4. Case Study 3: Changes in Seasonal Averages

In this final case study, we use the same summary statistics as in Case Study 1; however, instead of considering only overall changes, we incorporate seasonal changes in average temperature and precipitation. We apply the following perturbations, again inspired by Kahlenborn et al. (2021):

- Average precipitation (winter, spring, summer, autumn): +30%, +30%, +0%, +15%.
- Average temperature (winter, spring, summer, autumn): +3.4°C, +2.6°C, +3.1°C, +4.1°C.

In addition, we consider four configurations of the climatology matching prior in order to explore the effects of prior information. We also show how adding constraints through the prior allow us to explore a range of different weather generator "storylines" (Shepherd et al., 2018):

a. We constrain monthly dry days to their historical level by using the proportion of dry days per months as additional summary statistics. We inflate the variance of the prior around the occurrence model.
b. We constrain the overall dry days to their historical level by using the overall proportion of dry days as an additional summary statistic. We place a tight prior around the occurrence model.
c. We constrain the overall dry days to their historical level by using the overall proportion of dry days as an additional summary statistic. We inflate the variance of the prior around the occurrence model.
d. We apply variance inflation to all model parameters in the climatology matching prior.

Cases (a)-(c) correspond to different ways that the occurrence of precipitation can be constrained in the climatology matched weather generators. We can either use additional target summary statistics—on a granular monthly level (case a) or overall level (case c)—or we can instead use more informative priors for model components that we want to remain close to their calibrated values (case b). Finally, case (d) corresponds to the least constrained case where the SBI algorithm has the most freedom to adjust the weather generator parameters.

Figures showing the calibration posterior, climatology matching prior, and climatology matching posterior, similar to Figure 5, can be found in Appendix C (Figures C3–C6). We generally find that the target summary statistics are well covered by the climatology matching posterior. We once again inspect the simulated time series by starting with the MAP estimate, that is, the parameter setting most likely under the prior and likelihood. This allows us to focus on the uncertainty arising only from the natural variability of the weather generator. Figure 8 shows plots of average monthly precipitation, precipitation occurrence and temperature for 100 50-year realizations of the historically calibrated and climatology matched weather generators for all four cases. The average temperature time series are quite similar across cases (a)-(d) and generally agree quite well with the target lines in blue. Similarly, the average precipitation time series mostly fall close to the target values; however, they differ slightly in some periods such as during the summer months.

The annual cycle of precipitation occurrence, on the other hand, differs substantially between models. Although in cases (a), (b), and (c) the occurrence stays relatively close to the historical values due to the constraints imposed by the tight prior and monthly (b) or annual (c) summary statistics. In case (d), which corresponds to the least constrained case, the seasonal cycle changes substantially and the higher average precipitation targets in winter and spring are partially met by a change in the occurrence seasonal cycle.

This can also be seen in Figure 9 where we show annual cycles of precipitation occurrence from 100 50-year realizations sampled from the full climatology matching posterior, rather than the MAP estimate, along with realizations from the calibration posterior. In cases (a) and (b), the posterior simulations are well constrained by the additional summary statistics and informative priors, leading to less variance in the posterior predictive. In cases (c) and (d), however, the uncertainty is much larger because the NPE is less well constrained by the prior and the summary statistics. In case (c), it is only constrained by the average precipitation throughout the period, but can vary the seasonal cycle, while in (d) no aspect of precipitation occurrence is fixed.

As discussed in Section 3.4, uncertainty in the SBI methodology comes from multiple sources. All of these sources of uncertainty are relevant to the case study shown here. However, while in Figure 9d the uncertainty is fairly large, in settings (a) and (b) we show how expert knowledge—here incorporated through the climatology matching prior—can reduce epistemic uncertainty and lead to more well-constrained posterior (Figure 9). If no a priori judgments about model components can be made, then the unconstrained climatology matching prior (d) can still be used to obtain a more diverse set of samples. Alternatively, the full climatology matching posterior can be explored using different ways. Expert judgment can be used to define what a realistic weather generator within the posterior is: for example, some draws from the posterior might lead to seasonal cycles that are deemed unrealistic and can therefore be excluded through rejection sampling. Alternatively, different generated weather "storylines" within the posterior can be explored such as grouping all weather generator models that include a certain change in precipitation seasonality. Finally, posterior summaries, such as the MAP estimate can be considered in cases where comprehensive uncertainty quantification is not necessary.

5. Discussion and Conclusion

In this work, we introduced a new method for "climatology matching", that is, the problem of generating synthetic time series of meteorological variables following an arbitrary set of low-dimensional summary statistics, using simulation-based inference. Our contributions can be summarized in two parts: First, we proposed a novel approach to multivariate stochastic weather generation that jointly models daily average, maximum and minimum temperature as well as daily precipitation, together with their interdependencies. The weather generator is constructed and fitted within a Bayesian framework and enables highly efficient Monte Carlo simulations over long time periods spanning multiple decades or centuries. We validated our model against observed station data and showed that it is able to produce synthetic time series that are in good agreement with the observations. Second, we showed how the climatology matching inverse problem can be made well-posed by casting it as a problem of probabilistic inference. We then proposed a method for solving this problem using recently developed methods from the field of simulation-based inference. We empirically validated the methodology through three case studies of varying complexity where we use SBI to adapt our weather generator to simulate meteorological time series under scenario-neutral climate change conditions based on in situ meteorological data from Potsdam, Germany.

The methods presented here are broadly applicable and allow users to easily generate time series for a wide variety of potential use cases. The weather generator model is, to the best of our knowledge, the first GAMLSS-

based weather generator capable of modeling all four variables, that is, mean/min/max daily temperature and precipitation, autoregressively while also explicitly modeling their interdependencies. The proposed weather generator includes many common models in the literature as special cases, but has additional flexibility to represent many aspects of the data-generating process, such as seasonal variations in temperature or precipitation. It also has the benefit of being fully Bayesian, allowing for comprehensive uncertainty quantification through the use of modern software tools for probabilistic programming. As a result, it is fully autodifferentiable and massively parallelizable on GPUs, allowing for highly scalable inference and simulation. On both CPU and GPU, it is capable of simulating thousands of years worth of data in a matter of seconds on widely available commercial hardware—a considerable improvement over most existing weather generators.

Our SBI-based method for climatology matching allows users to quickly generate synthetic time series with perturbed summary statistics. This could be used to drive downstream impact models, such as hydrological, forest or agricultural models. The methodology could also be used to support the probing of adaptation options in a scenario-neutral framework. Furthermore, it could be applied to stress-testing studies which aim to investigate how climate-sensitive systems react to changes in certain input parameters, such as the occurrence of extreme rainfall, thereby identifying relevant vulnerabilities of the system. For example, given certain "storylines" (Shepherd et al., 2018) or an "exposure space" (Guo et al., 2018) of climatic changes, encoded as changes in summary statistics of extreme rainfall and temperature, the methodology could subsequently be used to generate synthetic time series to drive a hydrological model. From this, return periods of certain runoff events could be calculated, allowing users to investigate how rainfall changes influence flood probabilities, and give realistic estimates of runoff to consider in resilience planning. The methodology is also potentially applicable to other commonly used weather generator or statistical downscaling algorithms, so long as they are both parametric and stochastic. Furthermore, it could be used to make tools for scenario-neutral impact analysis, such as the *fore*-SIGHT package (Bennett et al., 2021), more efficient and applicable to a wider range of problems. The methodology can also be interpreted as an extension of so-called "change-factor" weather generators whereby global or regional climate model-based change factors are applied to parameters of a weather generator in order to generate synthetic precipitation time series under climate change conditions (Bartholy et al., 1995; Jones et al., 2011; Kilsby et al., 2007; Maraun & Widmann, 2018; Wilby et al., 1998). The probabilistic inverse approach taken here allows for an explicit treatment of weather generator parameter perturbations in order to incorporate large-scale trends, which enables applications to more complex summary statistics and weather generators.

We identify several avenues for future work: First, there are a number of possible ways in which our weather generator model could be further developed. The model does not currently capture the upper tail of precipitation or the lower tail of average temperature very well and this might be addressed by considering extreme value theoretic approaches such as in Vrac and Naveau (2007). Other natural extensions would be to consider multi-site or "full field" models (Maraun & Widmann, 2018) for example, by capturing spatial dependence using latent Gaussian fields or copulas such as in Kleiber et al. (2012); Ambrosino et al. (2014); Wessel and Chandler (2025). Copulas could also be explored as an alternative to capture inter-variable dependence (see e.g. Li & Babovic, 2019). Further, it would be useful to explore alternative distributional specifications, such as for T_{range}. One possible drawback of modeling T_{range} as a double-bounded variable is that no diurnal temperature ranges outside the pre-specified temperature range can be generated. We found this not to be a significant problem in practice due to the relatively narrow tails of the observed T_{range} distributions, implying that large ranges are exceedingly rare. The upper bound can also be justified statistically, as we found that fitting a generalized Pareto distribution to observed temperature ranges over a number of thresholds led to a negative shape parameter, which indicates that the distribution of temperature range extremes is bounded from above (cf. Coles, 2001). Nonetheless, a more rigorous method of accounting for potential uncertainty in this upper bound would be beneficial. Finally, the weather generator currently models all variables at daily resolution. If values at higher temporal resolution are important the weather generator could be coupled with temporal disaggregation schemes (e.g., Kossieris et al., 2018; Zabel & Poschlod, 2023); or the SBI method could be applied to weather generators modeling variables directly at subdaily scales (e.g., Kim & Onof, 2020).

Second, it would be interesting to explore the amortization properties of the neural posterior estimates. Once trained on a sufficiently large set of parameters and summary statistics, NPE algorithms are capable of predicting

the posterior distribution for other values of the same set of summary statistics. This makes it possible to easily generate time series for many different climate change scenarios in the case studies without retraining the neural density estimator. Our initial experiments with these amortized inferences were encouraging; however, further investigation is needed to verify the robustness of the amortized posterior inferences (see Appendix B for discussion). Similarly, it would be interesting to explore and contrast the SNPE algorithm used in this work with other SBI methods such as sequential neural likelihood (Papamakarios et al., 2019) or likelihood ratio (Hermans et al., 2020) estimation, as well as classical ABC methods. It would also be potentially beneficial to consider alternative neural density estimators to the normalizing flows used here, such as mixture density networks (Bishop, 1994) or flow matching models (Lipman et al., 2023).

Third, for some applications, conditioning on intervals rather than point values of target summary statistics might be useful, such as in cases where the target statistics aren't known precisely. It is worth noting that this is already to some extent possible in the proposed framework by incorporating a stochastic noise model into the simulator (see Equation 12); this would have a similar effect as conditioning on intervals but with the added benefit of avoiding the need to specify hard bounds. However, there are also recent works allowing for explicit conditioning on intervals in SBI (Chang et al., 2025; Gloeckler et al., 2024), which would also be interesting to explore.

Fourth, although our method for constructing the climatology matching prior is sufficient for the case studies considered in this work, our approach of applying variance inflation to the calibration posterior is largely ad hoc and may be inefficient for some use cases. This is apparent, for example, in Case Study 2, where the target for the tropical nights summary statistic is far outside of the range of the prior predictive distribution, and thus requires more iterations of the SNPE algorithm in order to converge. This is sometimes referred to as a *prior-data-conflict* and can potentially be remedied by a more judicious method of deriving the prior. One interesting avenue to explore in this direction would be the application of so-called "prior elicitation" methods (Mikkola et al., 2024) to construct climatology matching priors that better capture the full range of plausible climate conditions. Alternatively, Bayesian methods that are robust under prior misspecification, as proposed by Knoblauch et al. (2022), could also be explored.

Finally, it would be interesting to apply the SBI methodology to more complex weather generators, such as those which simulate additional meteorological variables like solar radiation and cloud cover, or multi-site methods that model spatial variability. This would allow for the investigation of changes in not only temporal but also spatial statistics, such as areal sums or spatial autocorrelation, since the SBI-approach has proven to be very flexible. Furthermore, while the current formulation the model assumes a stationary process, non-stationarity could be potentially accounted for by including additional covariates in the model, such as time, global mean temperature trends, or indices of large-scale circulation. Future work could also explore changes in complex summary statistics such as those describing compound events (Zscheischler et al., 2018), or explicitly conditioning on trends, if the corresponding weather generator model is flexible enough to represent this. In a similar vein, it would potentially be possible to couple the weather generator with an impact model in order to identify climatic changes that lead to certain impacts. The SBI methodology has proven very flexible even when considering larger amounts of summary statistics as in case study 2 or 3a. However, when considering larger sets of more complex statistics, alternative NPE architectures or additional simulation rounds might be necessary. Generally, we find in the present study that the number of summary statistics is less important than the "difficulty" of individual statistics, that is, how likely they are under the prior and the degree of nonlinearity in their relationships to the weather generator parameters. This should, however, be further investigated in future work.

While there are many areas for improvement, we believe our work makes a valuable contribution to both the literature on stochastic weather generation and scenario-neutral impact analysis. This study constitutes one of the first applications of SBI to the domain of weather and climate simulation. Furthermore, our proposed method for climatology matching addresses long-standing limitations in scenario-neutral climate impact analysis and enables a much broader exploration of scenarios to assess system vulnerabilities. We believe that this work will help to enable a wide range of applications of scenario-neutral methods to support impact and adaptation planning.

Appendix A: Weather Generator Model Details

The GLM models for T_{range} and T_{skew} are based on a Beta distribution with density:

$$f_B(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, x \in (0,1) \tag{A1}$$

here $B(\alpha,\beta) \coloneqq \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ with the Gamma function Γ. α,β are both shape parameters. Instead of modeling α,β directly we reparameterize this default form of the Beta distribution through mean μ and inverse dispersion ϕ, which are related to α,β as:

$$\alpha = \frac{\mu}{\phi} \tag{A2}$$

$$\beta = \frac{1-\mu}{\phi} \tag{A3}$$

The precipitation amounts model uses a reparameterized Gamma distribution with density:

$$f(y|\mu,\sigma) = \frac{y^{(1/\sigma^2-1)}\exp\left[-y/\left(\sigma^2\mu\right)\right]}{\left(\sigma^2\mu\right)^{(1/\sigma^2)}\Gamma(1/\sigma^2)}, \tag{A4}$$

here μ,σ correspond to location and scale parameters.

For completeness, we provide here a detailed accounting of the covariates used as predictors for each distribution parameter in the GAMLSS weather generator model. The notation $\boldsymbol{x} \times \mathbf{y}$ here refers to a set of linear interaction terms, that is, the outer product between the two vectors. $\mathbf{f}^{(t)}$ refers to the Fourier-features.

$$\begin{aligned}
\boldsymbol{x}^{(t)}_{T,\mu} &= \left(T^{(t-1)}_{\text{avg}}, T^{(t-2)}_{\text{avg}}, \mathbf{f}^{(t)}, T^{(t-1)}_{\text{avg}} \times \mathbf{f}^{(t)}, T^{(t-2)}_{\text{avg}} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{T,\sigma} &= \left(\mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{\tau,\mu} &= \left(\tau^{(t-1)}, \tau^{(t-2)}, \mathbf{f}^{(t)}, \tau^{(t-1)} \times \mathbf{f}^{(t)}, \tau^{(t-2)} \times \mathbf{f}^{(t)}, T^{(t)}_{\text{avg}}, T^{(t)}_{\text{avg}} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{\tau,\phi} &= \left(\tau^{(t-1)}, \tau^{(t-2)}, \mathbf{f}^{(t)}, \tau^{(t-1)} \times \mathbf{f}^{(t)}, \tau^{(t-2)} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{s,\mu} &= \left(T^{(t)}_{\text{avg}}, \mathbf{f}^{(t)}, T^{(t)}_{\text{avg}} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{s,\phi} &= \left(T^{(t)}_{\text{avg}}, \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{O,\mu} &= \left(O^{(t-1)}, O^{(t-2)}, \mathbf{f}^{(t)}, O^{(t-1)} \times \mathbf{f}^{(t)}, O^{(t-2)} \times \mathbf{f}^{(t)}, R^{(t-1)}, R^{(t-2)}, \mathbf{f}^{(t)},\right.\\
&\qquad \left. R^{(t-1)} \times \mathbf{f}^{(t)}, R^{(t-2)} \times \mathbf{f}^{(t)}, T^{(t)}_{\text{avg}}, T^{(t)}_{\text{avg}} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{R,\mu} &= \left(R^{(t-1)}, R^{(t-2)}, \mathbf{f}^{(t)}, R^{(t-1)} \times \mathbf{f}^{(t)}, R^{(t-2)} \times \mathbf{f}^{(t)}, T^{(t)}_{\text{avg}}, T^{(t)}_{\text{avg}} \times \mathbf{f}^{(t)}, 1\right)^\top,\\
\boldsymbol{x}^{(t)}_{R,\lambda} &= \left(\mathbf{f}^{(t)}, 1\right)^\top.
\end{aligned}$$

Autoregressive simulation from the weather generator is outlined in Algorithm 1.

Algorithm 1. Autoregressive simulation from the weather generator

Require: Initial lagged values $T^{(1)}_{\text{avg}}, P^{(1)}, T^{(1)}_{\text{range}}, T^{(2)}_{\text{avg}}, P^{(2)}, T^{(2)}_{\text{range}}$

Require: Posterior over model parameters $p(\boldsymbol{\theta}|\boldsymbol{Y})$

for $t = 3, \ldots, N_t$ **do**

Sample $T^{(t)}_{\text{avg}}|T^{(t-1)}_{\text{avg}}, T^{(t-2)}_{\text{avg}} \sim \mathcal{N}\left(\mu^{(t)}_T, \sigma^{(t)^2}_T\right)$

Sample $T^{(t)}_{\text{skew}}|T^{(t)}_{\text{avg}} \sim \mathcal{B}\left(\mu^{(t)}_{T_{\text{skew}}}, \phi^{(t)}_{T_{\text{skew}}}\right)$

Sample $\tau^{(t)}|T^{(t)}_{\text{avg}}, \tau^{(t-1)}, \tau^{(t-2)} \sim \mathcal{B}\left(\mu^{(t)}_\tau, \phi^{(t)}_\tau\right)$

Sample precipitation occurrence: $O^{(t)}|T^{(t)}_{\text{avg}}, P^{(t-1)}, P^{(t-2)} \sim \text{Bernoulli}\left(p^{(t)}_O\right)$

if $O^{(t)} = 1$ **then**

Sample precipitation amount: $P^{(t)}|T^{(t)}_{\text{avg}}, P^{(t-1)}, P^{(t-2)} \sim \text{Gamma}\left(\mu^{(t)}_R, \lambda^{(t)}_R\right)$

else

Set $P^{(t)} \leftarrow 0$

end if

Compute $T^{(t)}_{\text{range}} \leftarrow \tau^{(t)} \cdot T^{\max}_{\text{range}}$

Compute $T^{(t)}_{\min} \leftarrow T^{(t)}_{\text{avg}} - T^{(t)}_{\text{range}} \cdot T^{(t)}_{\text{skew}}$

Compute $T^{(t)}_{\max} \leftarrow T^{(t)}_{\text{avg}} + \left(1 - T^{(t)}_{\text{skew}}\right) \cdot T^{(t)}_{\text{range}}$

Update lagged variables with current simulated values

end for

return Simulated time series $\left\{T^{(t)}_{\text{avg}}, T^{(t)}_{\min}, T^{(t)}_{\max}, P^{(t)}\right\}_{t=3}^{N_t}$

Appendix B: Diagnostics for Climatology Matching With SBI

It is worth noting some practical considerations and diagnostics for the interpretation and evaluation of climatology matching with SBI. First, we emphasize that the success of climatology matching depends crucially on the summary statistics chosen and their corresponding physical assumptions. If the summary statistics are misspecified in the sense that they lie beyond the range of the climatology matching prior, define a set of targets that either the model cannot represent or are physically impossible, then inference with SBI (or any algorithm) will necessarily fail. Evaluation of both the climatology matching prior and posterior are crucial. We strongly recommend the use of both prior and posterior predictive checks (Gelman et al., 2020), that is, sampling weather generator parameters from the prior and posterior and running simulations with those parameters to generate distributions of the selected summary statistics, to critically assess the model. If the target summary statistics lie outside the range of realizations from the prior predictive, SBI algorithms will typically struggle to extrapolate beyond this initial subspace. In such cases, users might consider increasing the coverage of the climatology matching prior by increasing the variance scaling factor (see Section 3.3) or considering alternative methods of constructing it, for example, using prior elicitation techniques (Mikkola et al., 2024).

After inference, the posterior predictive should cover (and ideally be centered on) on the target summary statistics. Furthermore, the posterior predictive density should be reasonably concentrated around them. In practice a good rule for the concentration, or "sharpness", is that the variance of the SBI posterior predictive should not be much larger than the variance of the predictive distribution when conditioning on a point estimate, such as the MAP. In other words, the variance should not be much larger than that arising from the internal variability of the weather generator. If the posterior predictive is very diffuse (high variance), or does not cover the target summary statistics, this could be due either to a failure of the inference algorithm or a failure of the weather generator to simulate time series matching the target summary statistics. Unfortunately, distinguishing between these two failure modes can be quite difficult.

To identify and address failures of SBI, one can consider running the SBI algorithm (e.g., SNPE) for more rounds, increasing the number of simulations per round, experimenting with alternative neural density estimators, or using more informative priors for model components where relevant prior information is available. Simulation-based

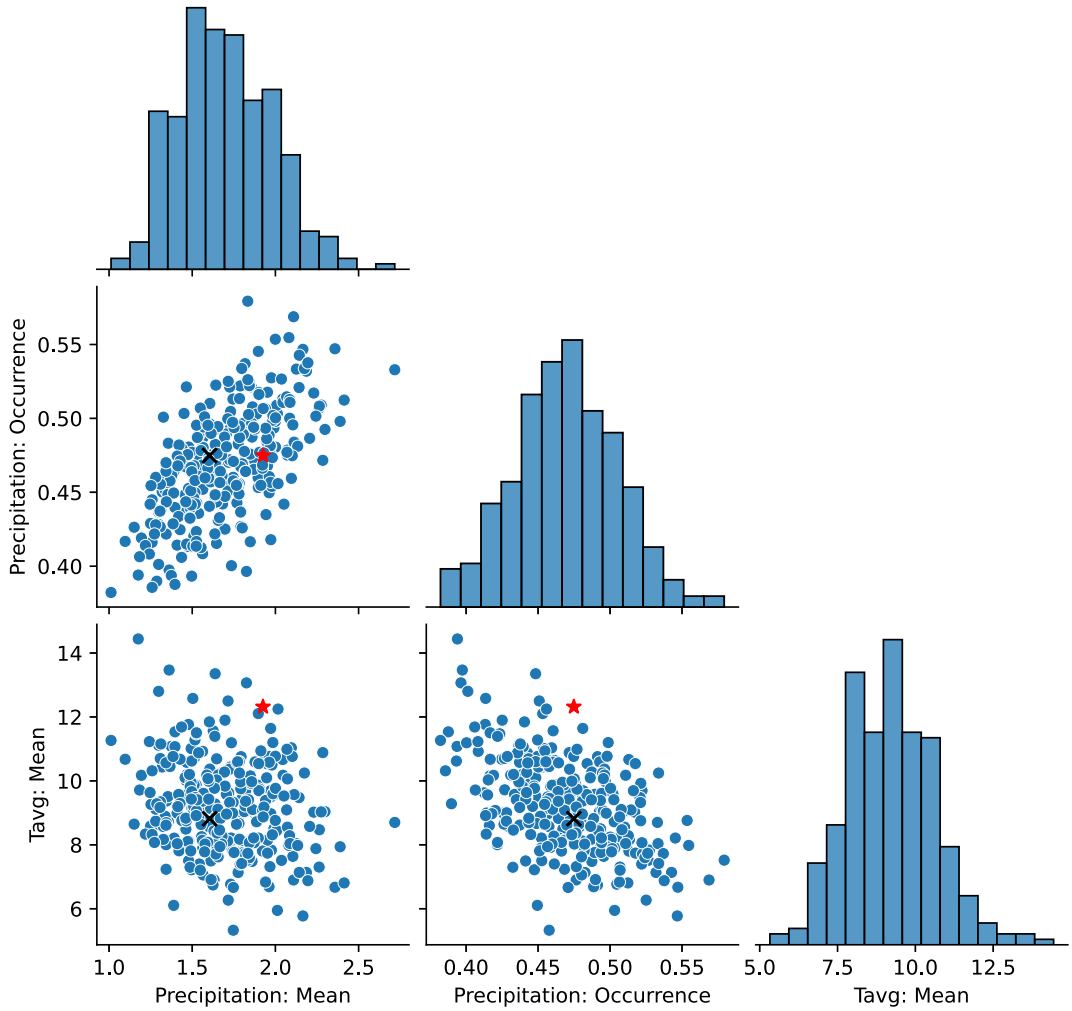

Figure B1. Prior predictive diagnostic plot for Case Study 1. The black cross corresponds to the "observed" value of the summary statistics computed from the calibration data and the red star to the target used in climatology matching.

calibration (Talts et al., 2020) can also be used to check the validity of the inferred posterior distribution; this is particularly helpful when making amortized inferences using the trained NPE (Deistler et al., 2022).

A failure of the model can occur if the weather generator is incapable of producing the target summary statistics, or if the variance of summary statistics for any fixed choice of parameters is large. The latter can particularly be an issue for summary statistics describing the tail of the climatological distribution, and it can be useful to assess how much the summary statistics vary across multiple simulation runs with the same parameters. To identify whether a weather generator might not be capable of reproducing target summary statistics, one diagnostic is the use of prior predictive pair or corner plots. Figure B1 shows an example of such a plot for Case Study 1. We can see that, in this case, the target values are well covered by the prior predictive, also in the (bivariate) joint distributions. In contrast, consider instead a case where we also include the cumulative dry days (CDD), that is, average length of dry spells, as an additional summary statistic. We then consider a target with an increase of 30% in both CDD and precipitation frequency. This constitutes a highly implausible scenario since these statistics are negatively correlated; more frequent precipitation should imply shorter dry spells. In the corresponding prior predictive corner plot (Figure B2), we can clearly see the pathological behavior in the scatter plot for CDD and precipitation frequency (occurrence), where the target value appears highly anomalous. Although this is only a visual diagnostic, future work could explore more rigorous quantitative measures that, for example, estimate the likelihood of the target statistics under the joint distribution over summary statistics.

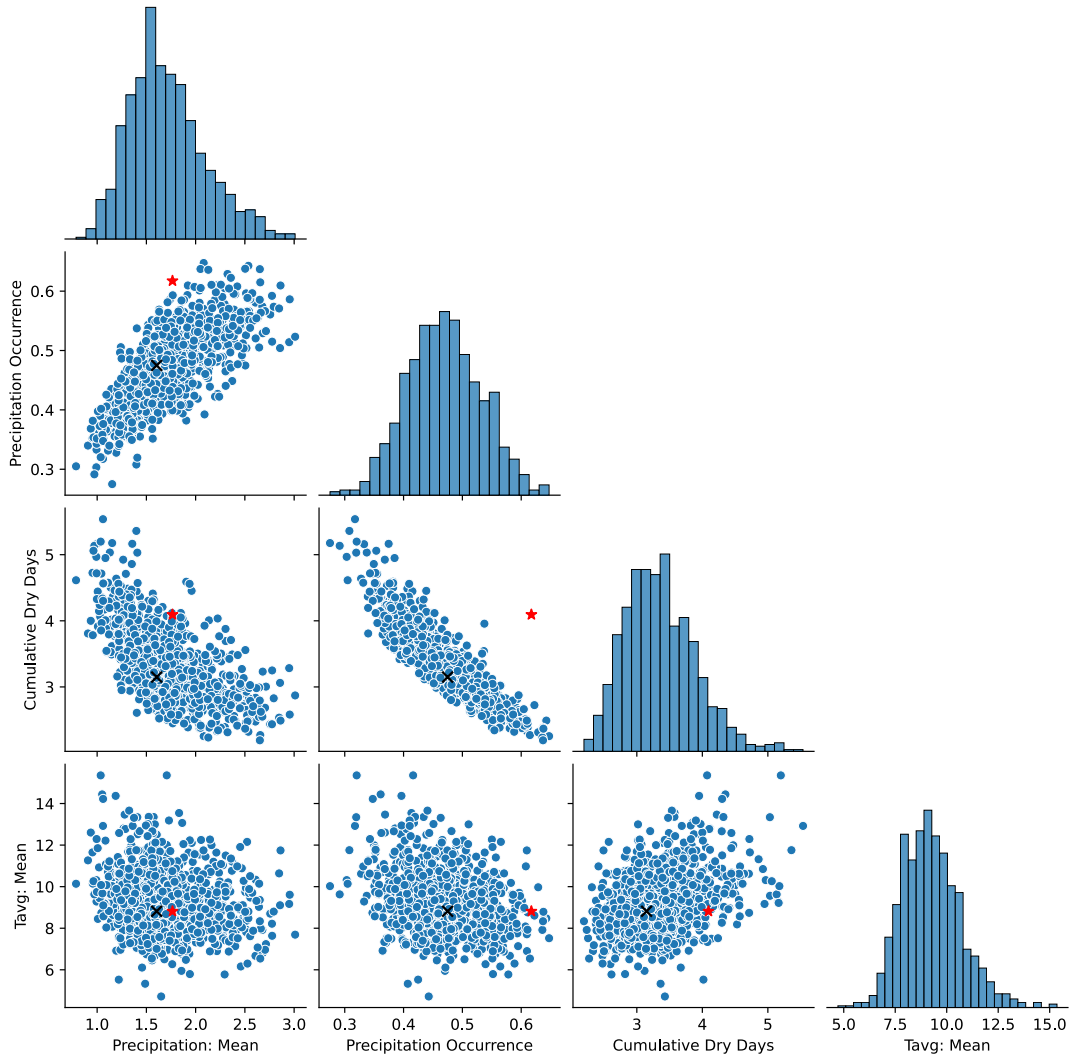

Figure B2. Same as Figure B1 but for an artificial test case where the target summary statistics for precipitation occurrence/frequency and the cumulative dry days (CDD) are intentionally set to implausible values that are physically contradictory. The joint density for precipitation occurrence and CDD shows the correlation between the summary statistics and the anomalous target.

It should be emphasized that, given weather generators of reasonable complexity and a suitably well-crafted and physically coherent set of summary statistics, the SBI approach used in this work (SNPE) tends to be fairly robust.

Appendix C: Additional Results

Figure C1

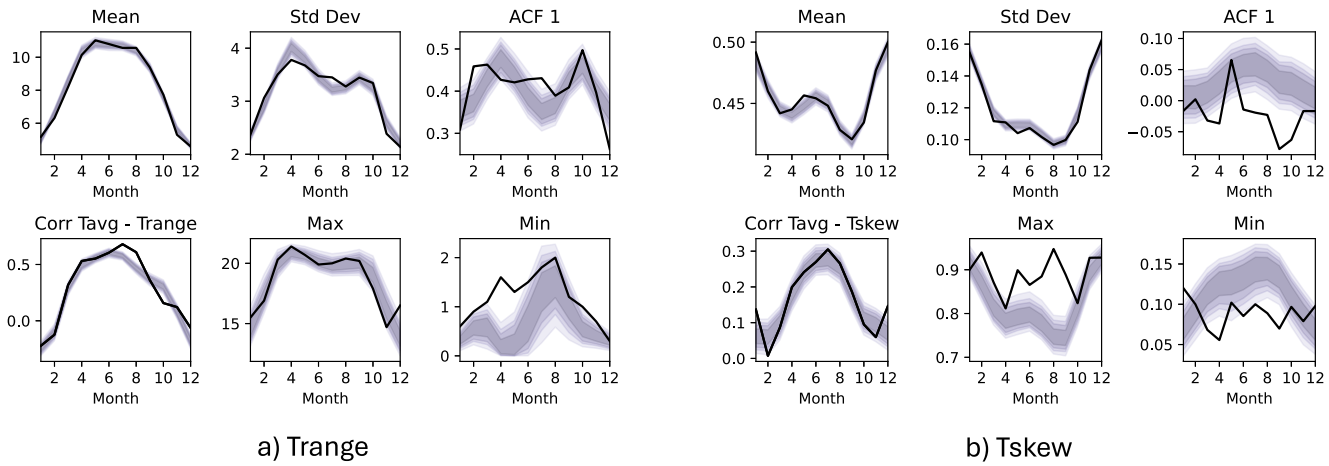

Figure C1. Same as Figure 3 but for daily temperature range (left, a) and daily temperature skew (right, b). Top row: monthly mean, standard deviation and autocorrelation at lag 1. Bottom row: autocorrelation at lag 2, monthly maximum and minimum.

Figure C2

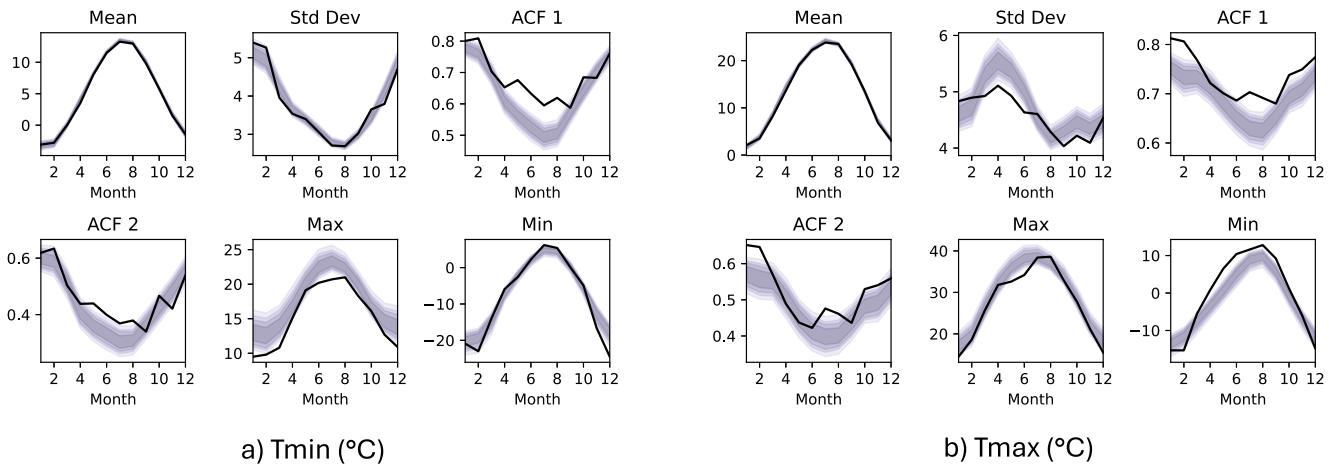

Figure C2. Same as Figure 3 but for daily temperature minimum (left, a) and daily temperature maximum (right, b). Top row: monthly mean, standard deviation and autocorrelation at lag 1. Bottom row: autocorrelation at lag 2, monthly maximum and minimum.

Figure C3

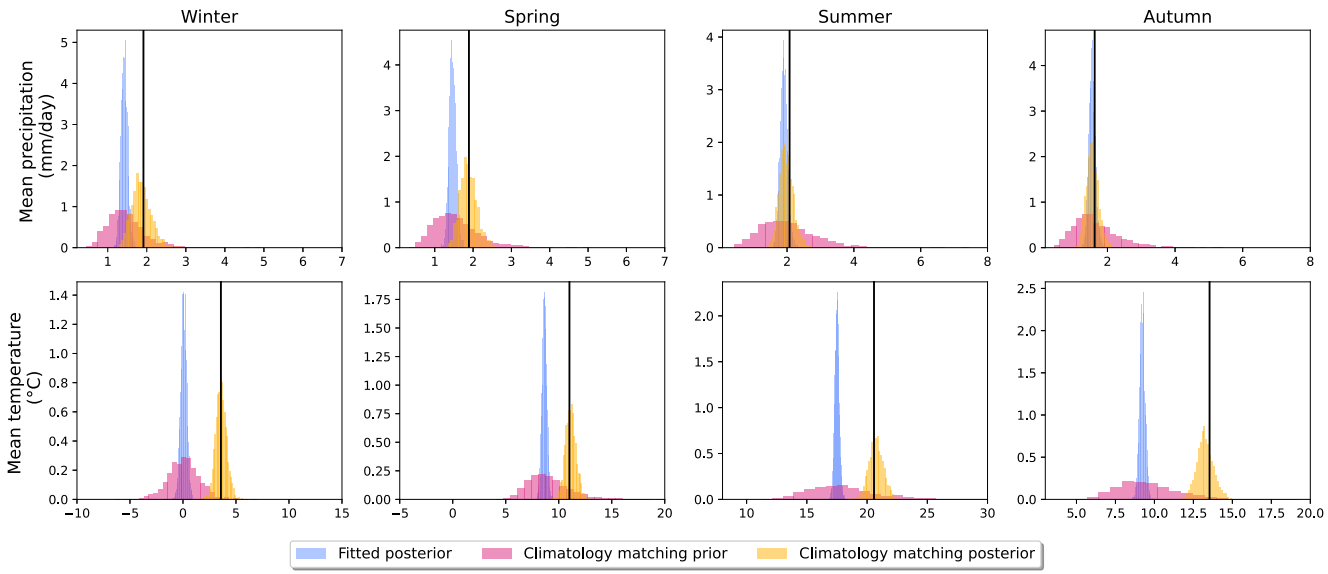

Figure C3. Case study 3a: Histograms of 2000 simulations from the calibration posterior (blue), climatology matching prior (purple) and climatology matching posterior (yellow) for average precipitation (top row) and average temperature (bottom row) for each season (columns). Black lines: target summary statistics.

Figure C4

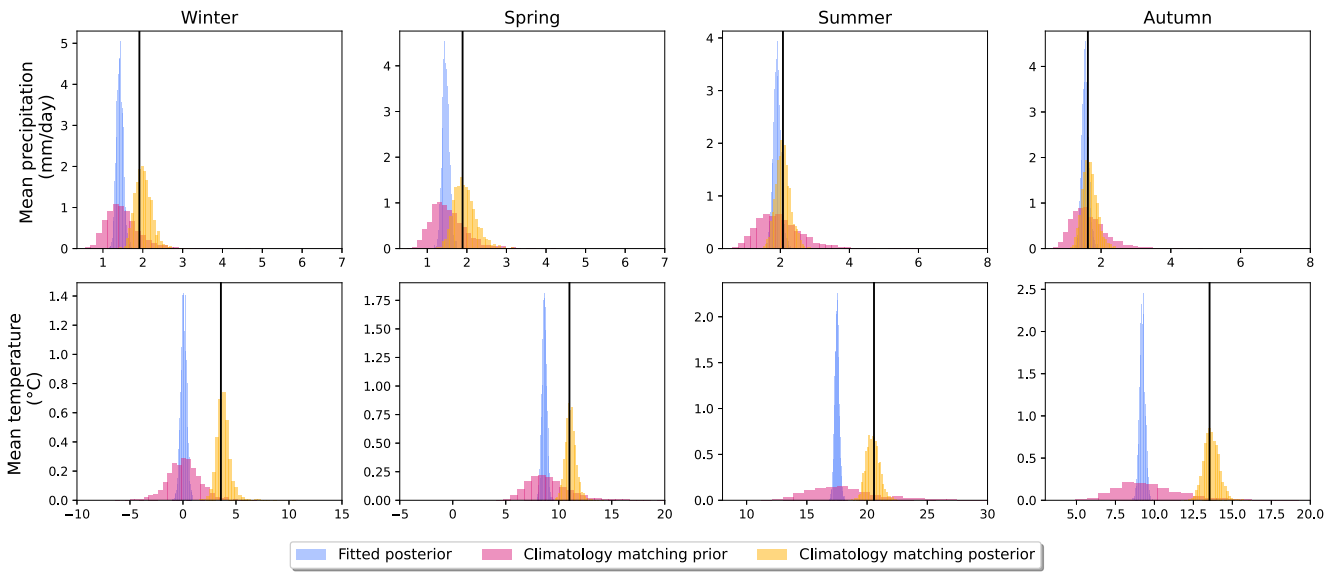

Figure C4. Same as Figure C3 but for case study 3b.

Figure C5

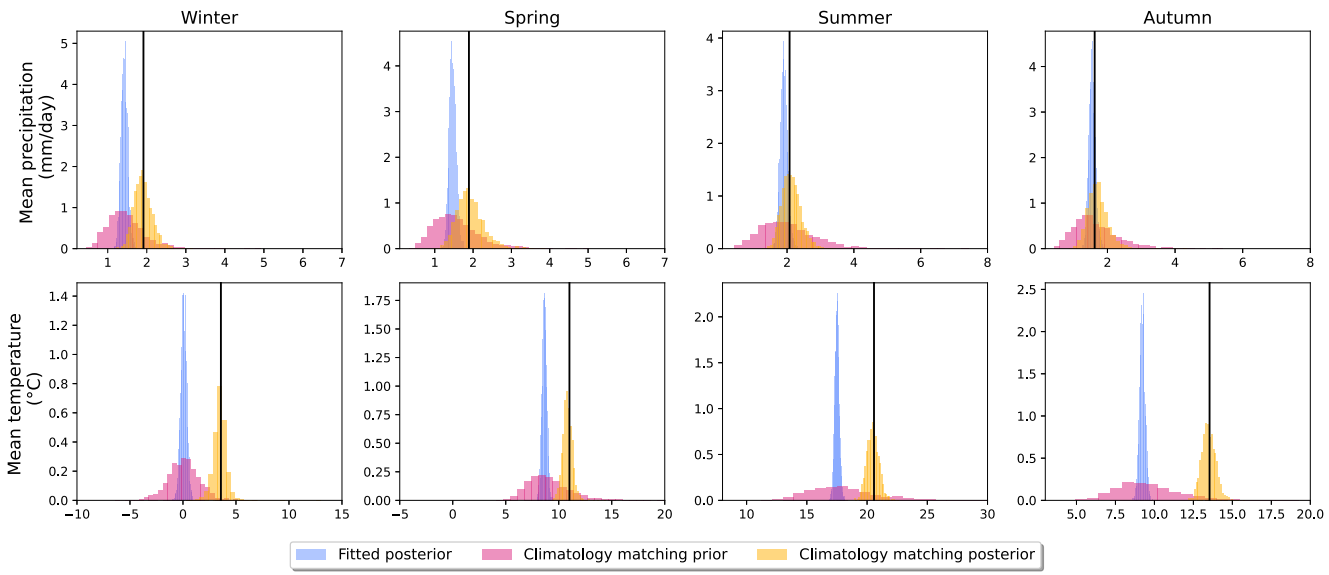

Figure C5. Same as Figure C3 but for case study 3c.

Figure C6

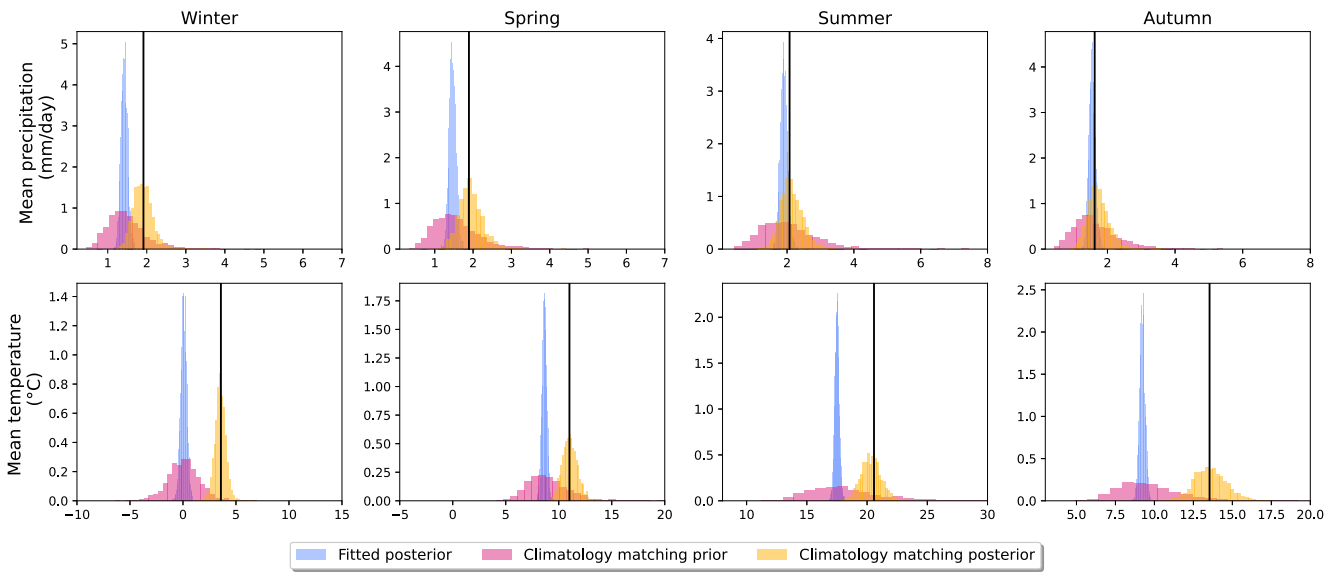

Figure C6. Same as Figure C3 but for case study 3d.

Appendix D: Amortized Climatology Matching With SBI

The NPE estimate obtained from the algorithm described in Sections 3 is amortized. This means that, given a trained NPE, valid posteriors $p(\boldsymbol{\theta}|\boldsymbol{s}^\star)$ can be obtained for other values of the target summary statistics $\boldsymbol{s}^\star$. We demonstrate this on the NPE obtained in case study 1. Recall that the target summary statistic were a change in average temperature of +3.5°C and a change in average daily precipitation of 20%, corresponding to the historical values, while keeping the proportion of wet days constant on the historical value (see Section 4.2. We consider two additional scenarios to showcase the amortization:

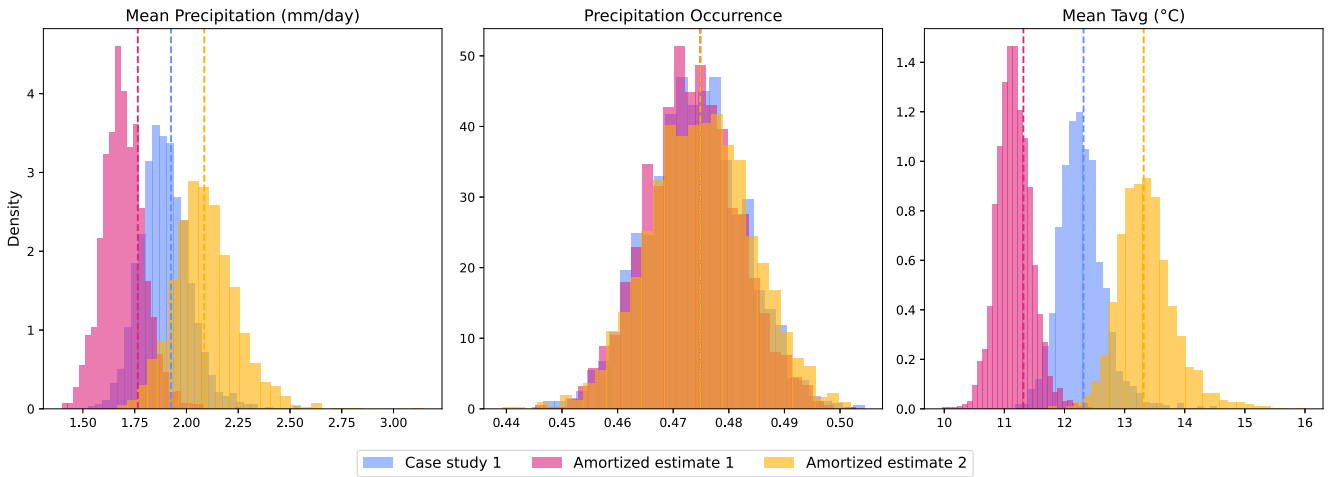

Figure D1. Posterior predictive distributions from case study 1, together with amortized estimates for two different levels of the target summary statistics.

1. Change in average temperature of +2.5°C, change in average daily precipitation of 10%.
2. Change in average temperature of +4.5°C, change in average daily precipitation of 30%.

In both cases again the overall proportion of wet days is kept constant.

Figure D1 contains the corresponding posterior predictive distributions for both scenarios considered, as well as the original target values in case study 1. We can see that, in both cases, the amortized estimates cover the corresponding target summary statistics, similar to original NPE estimate from case study 1. The advantage of amortization is that these posterior densities are obtain near-instantaneously, without the need to re-run the SNPE training procedure described in Section 3.3. We note that when using the amortized estimates in multi-round training we find some evidence of miscalibration in prior simulation-based calibration checks (Talts et al., 2020). This does not occur when using single round SNPE-C/APT. However, when checking posterior simulation-based calibration, arguably a more relevant diagnostic for climatology matching, the amortized estimates appear well calibrated in both cases.

Appendix E: Analysis at Additional Stations

In order to demonstrate the ability of our method to generalize to a broader range of conditions, we separately calibrated the weather generator and replicated case study 1 at 10 randomly chosen stations from the DWD observation network. We note that weather generator performance appeared generally satisfactory across all stations, although for sake of brevity these Figures are omitted here. They are, however, available in the supplementary material published on Zenodo (Wessel & Groenke, 2025).

Figure E1 contains the results of the climatology matching case study 1 at all additional locations. The method seems to perform well across locations with the SBI posterior covering the station-specific target summary statistics in all cases. Interestingly, at station 5,419 the calibration posterior showed a bias for the modeled precipitation occurrence, but this bias seems to be partly removed in the SBI posterior, highlighting that the SBI method could also be used to improve weather generator performance over the historical period.

10.1029/2025JH000902

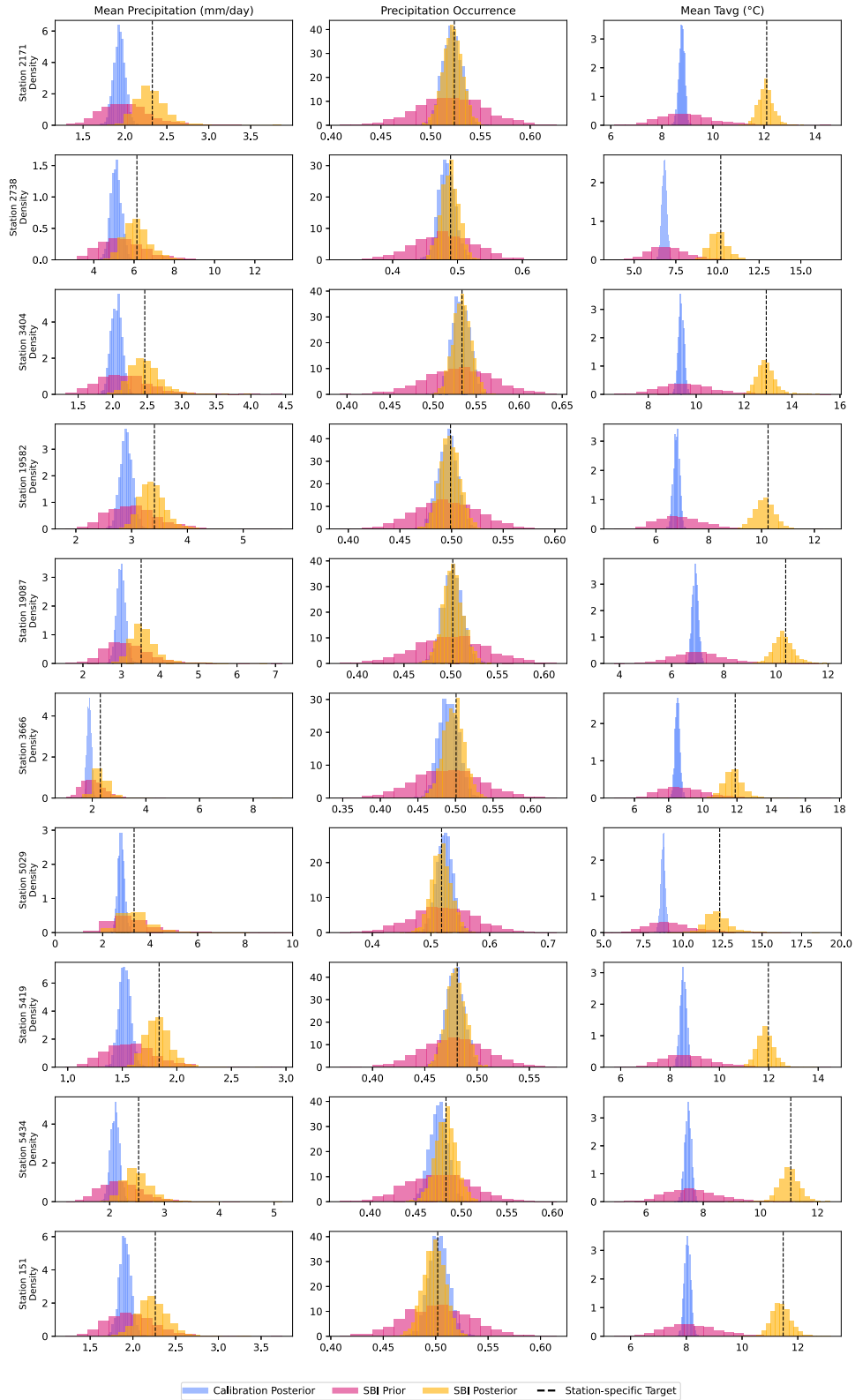

Figure E1. Same as Figure 5 but for 10 randomly selected stations from the DWD station network.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The live code repository for the weather generator and climatology matching framework is available on GitHub (https://github.com/bgroenks96/wxsbi, Groenke & Wessel, 2025). The version of the code and the data used to generate the results from this paper is permanently archived on Zenodo (https://doi.org/10.5281/zenodo.15448998, Wessel & Groenke, 2025). The weather station data used in Secion 4.1 is available from the DWD open data server (https://opendata.dwd.de, Deutscher Wetterdienst, 2024).

Acknowledgments
B.G. acknowledges the funding and support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS) under Grant HIDSS-0001. J.W. was supported during this work by the Helmholtz Visiting Researcher Grant of the Helmholtz Information and Data Science Academy (HIDA). He also acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) under Grant award number 2696930. P.M. and J.Z. acknowledge the Helmholtz Initiative and Networking Fund (Young Investigator Group COMPOUNDX, Grant Agreement VH-NG-1537). The authors would also like to thank the anonymous reviewers whose comments and feedback substantially improved this manuscript. Open Access funding enabled and organized by Projekt DEAL.

References

Ailliot, P., Allard, D., Monbet, V., & Naveau, P. (2015). Stochastic weather generators: An overview of weather type models. *156*(1).
Ailliot, P., Thompson, C., & Thomson, P. (2009). Space-time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society—Series C: Applied Statistics*, *58*(3), 405–426. https://doi.org/10.1111/j.1467-9876.2008.00654.x
Ambrosino, C., Chandler, R. E., & Todd, M. C. (2014). Rainfall-derived growing season characteristics for agricultural impact assessments in South Africa. *Theoretical and Applied Climatology*, *115*(3), 411–426. https://doi.org/10.1007/s00704-013-0896-y
Anand, M., Bohn, F. J., Camps-Valls, G., Fischer, R., Huth, A., Sweet, L.-b., & Zscheischler, J. (2024). 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
Asong, Z. E., Khaliq, M. N., & Wheater, H. S. (2016). Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models. *Climate Dynamics*, *47*(9), 2901–2921. https://doi.org/10.1007/s00382-016-3004-z
Bárdossy, A., & Plate, E. (1992). Space-time model for daily rainfall using. *Atmospheric Circulation Patterns*, *28*(5), 1247–1259. https://doi.org/10.1029/91wr02589
Bartholy, J., Bogárdi, I., & Matyasovszky, I. (1995). Effect of climate change on regional precipitation in Lake Balaton watershed. *Theoretical and Applied Climatology*, *51*(4), 237–250. https://doi.org/10.1007/BF00867282
Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. https://doi.org/10.1093/genetics/162.4.2025
Bennett, B., Devanand, A., Culley, S., Westra, S., Guo, D., & Maier, H. R. (2021). A modelling framework and R-package for evaluating system performance under hydroclimate variability and change. *Environmental Modelling & Software*, *139*, 104999. https://doi.org/10.1016/j.envsoft.2021.104999
Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., et al. (2019). Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, *20*(28), 1–6. http://jmlr.org/papers/v20/18-403.html
Bishop, C. M. (1994). *Mixture density networks*. (Tech. Rep.). Aston University. Retrieved from https://publications.aston.ac.uk/id/eprint/373/
Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., et al. (2018). JAX: Composable transformations of Python+NumPy programs [Software]. http://github.com/jax-ml/jax
Chandler, R. E. (2020). Multisite, multivariate weather generation based on generalised linear models. *Environmental Modelling & Software*, *134*, 104867. https://doi.org/10.1016/j.envsoft.2020.104867
Chandler, R. E., & Wheater, H. S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, *38*(10). https://doi.org/10.1029/2001wr000906
Chang, P. E., Loka, N., Huang, D., Remes, U., Kaski, S., & Acerbi, L. (2025). Amortized probabilistic conditioning for optimization, simulation and inference. *arXiv*. https://doi.org/10.48550/arXiv.2410.15320
Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0
Cranmer, K., Brehmer, J., & Louppe, G. (2020). The Frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117
Culley, S., Bennett, B., Westra, S., & Maier, H. R. (2019). Generating realistic perturbed hydrometeorological time series to inform scenario-neutral climate impact assessments. *Journal of Hydrology*, *576*, 111–122. https://doi.org/10.1016/j.jhydrol.2019.06.005
Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H., et al. (2016). A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate. *Water Resources Research*, *52*(9), 6751–6768. https://doi.org/10.1002/2015wr018253
Deistler, M., Gonçalves, P. J., & Macke, J. H. (2022). Truncated proposals for scalable and hassle-free simulation-based inference. In *Proceedings of the 36th Conference on neural information processing systems*.
Der Tagesspiegel. (2003). Berlin: 11. August 1978: Der Tag nach dem großen Regen. Der Tagesspiegel Online. Retrieved from https://www.tagesspiegel.de/berlin/11-august-1978-der-tag-nach-dem-grossen-regen-1032997.html
Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020
Deutscher Wetterdienst. (2024). Daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany [Dataset]. Retrieved from https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/
Diggle, P. J., & Gratton, R. J. (1984). Monte Carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B*, *46*(2), 193–212. https://doi.org/10.1111/j.2517-6161.1984.tb01290.x
Fatichi, S., & Ivanov, V. Y. (2014). Interannual variability of evapotranspiration and vegetation productivity. *Water Resources Research*, *50*(4), 3275–3294. https://doi.org/10.1002/2013wr015044
Fatichi, S., Ivanov, V. Y., & Caporali, E. (2011). Simulation of future climate scenarios with a weather generator. *Advances in Water Resources*, *34*(4), 448–467. https://doi.org/10.1016/j.advwatres.2010.12.013
Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Furrer, E. M., & Katz, R. W. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2008WR007316

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., et al. (2020). Bayesian workflow (No. arXiv:2011.01808). *arXiv*. https://doi.org/10.48550/arXiv.2011.01808

Gloeckler, M., Deistler, M., Weilbach, C., Wood, F., & Macke, J. H. (2024). All-in-one simulation-based inference. In *Proceedings of the 41st International Conference on machine learning* (Vol. 235, pp. 15735–15766). JMLR.org.

Greenberg, D. S., Nonnenmacher, M., & Macke, J. H. (2019). Automatic posterior transformation for likelihood-free inference. *arXiv*. https://doi.org/10.48550/arXiv.1905.07488

Groenke, B., & Wessel, J. (2025). wxsbi (Version 2) [Software]. Retrieved from https://github.com/bgroenks96/wxsbi

Grunwald, G. K., & Jones, R. H. (2000). Markov models for time series with mixed distribution. *Environmetrics*, *11*(3), 327–339. https://doi.org/10.1002/(SICI)1099-095X(200005/06)11:3⟨327::AID-ENV412⟩3.0.CO;2-R

Guo, D., Westra, S., & Maier, H. R. (2018). An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies. *Journal of Hydrology*, *556*, 877–890. https://doi.org/10.1016/j.jhydrol.2016.03.025

Hermans, J., Begy, V., & Louppe, G. (2020). Likelihood-free MCMC with amortized approximate ratio estimators. In *Proceedings of the 37th International Conference on machine learning* (pp. 4239–4248). PMLR. Retrieved from https://proceedings.mlr.press/v119/hermans20a.html (ISSN:2640-3498

Heße, F., Comunian, A., & Attinger, S. (2019). What we talk about when we talk about uncertainty. Toward a unified, data-driven framework for uncertainty characterization in hydrogeology. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00118

Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, *14*(1), 1303–1347.

Hughes, J. P., Guttorp, P., & Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society—Series C: Applied Statistics*, *48*(1), 15–30. https://doi.org/10.1111/1467-9876.00136

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Hyndman, R. J., & Grunwald, G. K. (2000). Applications: Generalized additive modelling of mixed distribution Markov models with application to Melbourne's rainfall. *Australian & New Zealand Journal of Statistics*, *42*(2), 145–158. https://doi.org/10.1111/1467-842X.00115

Jones, P., Harpham, C., Kilsby, C., Glenis, V., & Burton, A. (2010). *UK climate projections science report: Projections of future daily climate for the UK from the weather generator*. (Tech. Rep.). Newcastle University: UK Climate Projections science report. Retrieved from https://crudata.uea.ac.uk/cru/data/ukcp/ukcp09_wgen.pdf

Jones, P. D., Harpham, C., Goodess, C. M., & Kilsby, C. G. (2011). Perturbing a weather generator using change factors derived from regional climate model simulations. *Nonlinear Processes in Geophysics*, *18*(4), 503–511. https://doi.org/10.5194/npg-18-503-2011

Kahlenborn, W., Porst, L., Voß, M., Fritsch, U., Renner, K., Zebisch, M., et al. (2021). *Klimawirkungs-und Risikoanalyse für Deutschland 2021 (Kurzfassung)*. Umweltbundesamt. Retrieved from https://www.umweltbundesamt.de/en/publikationen/KWRA-Zusammenfassung

Kaspar, F., Müller-Westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-Weiss, A., & Deutschländer, T. (2013). Monitoring of climate change in Germany – Data, products and services of Germany's National climate data centre. In *Advances in Science and Research* (Vol. 10(1), pp. 99–106). Copernicus GmbH. https://doi.org/10.5194/asr-10-99-2013

Katz, R. W., & Parlange, M. B. (1998). Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*. Retrieved from https://journals.ametsoc.org/view/journals/clim/11/4/1520-0442_1998_011_0591_opismo_2.0.co_2.xml(Section:JournalofClimate

Katz, R. W., & Zheng, X. (1999). Mixture model for overdispersion of precipitation. *Journal of Climate*. Retrieved from https://journals.ametsoc.org/view/journals/clim/12/8/1520-0442_1999_012_2528_mmfoop_2.0.co_2.xml(Section:JournalofClimate

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., & Knutti, R. (2015). Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical alpine river catchment. *Hydrology and Earth System Sciences*, *19*(5), 2163–2177. https://doi.org/10.5194/hess-19-2163-2015

Keller, L., Rössler, O., Martius, O., & Weingartner, R. (2019). Comparison of scenario-neutral approaches for estimation of climate change impacts on flood characteristics. *Hydrological Processes*, *33*(4), 535–550. https://doi.org/10.1002/hyp.13341

Kilsby, C. G., Jones, P. D., Burton, A., Ford, A. C., Fowler, H. J., Harpham, C., et al. (2007). A daily weather generator for use in climate change studies. *Environmental Modelling & Software*, *22*(12), 1705–1719. https://doi.org/10.1016/j.envsoft.2007.02.005

Kim, D., & Onof, C. (2020). A stochastic rainfall model that can reproduce important rainfall properties across the timescales from several minutes to a decade. *Journal of Hydrology*, *589*, 125150. https://doi.org/10.1016/j.jhydrol.2020.125150

Kleiber, W., Katz, R. W., & Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011WR011105

Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application*, *11*(1), 321–346. https://doi.org/10.1146/annurev-statistics-040722-053607

Klein, N., Kneib, T., Lang, S., & Sohn, A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *Annals of Applied Statistics*, *9*(2), 1024–1052. https://doi.org/10.1214/15-AOAS823

Knoblauch, J., Jewson, J., & Damoulas, T. (2022). An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *Journal of Machine Learning Research*, *23*(132), 1–109. Retrieved from http://jmlr.org/papers/v23/19-1047.html

Kossieris, P., Makropoulos, C., Onof, C., & Koutsoyiannis, D. (2018). A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures. *Journal of Hydrology*, *556*, 980–992. https://doi.org/10.1016/j.jhydrol.2016.07.015

Li, X., & Babovic, V. (2019). A new scheme for multivariate, multisite weather generator with inter-variable, inter-site dependence and inter-annual variability based on empirical copula approach. *Climate Dynamics*, *52*(3–4), 2247–2267. https://doi.org/10.1007/s00382-018-4249-5

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2023). Flow matching for generative modeling. *arXiv*. https://doi.org/10.48550/arXiv.2210.02747

Liu, G., Guo, Y., Xia, H., Liu, X., Song, H., Yang, J., & Zhang, Y. (2024). Increase asymmetric warming rates between daytime and nighttime temperatures over global land during recent decades. *Geophysical Research Letters*, *51*(24), e2024GL112832. https://doi.org/10.1029/2024GL112832

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., & Macke, J. H. (2017). Flexible statistical inference for mechanistic models of neural dynamics. In *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3). https://doi.org/10.1029/2009RG000314

Maraun, D., & Widmann, M. (2018). Weather generators. In *Statistical downscaling and Bias correction for climate research* (pp. 201–219). Cambridge University Press. https://doi.org/10.1017/9781107588783.014

Marcolongo, A., Vladymyrov, M., Lienert, S., Peleg, N., Haug, S., & Zscheischler, J. (2022). Predicting years with extremely low gross primary production from daily weather data using convolutional neural networks. *Environmental Data Science*, *1*, e2. https://doi.org/10.1017/eds.2022.1

McInerney, D., Westra, S., Leonard, M., Bennett, B., Thyer, M., & Maier, H. R. (2023). A climate stress testing method for changes in spatially variable rainfall. *Journal of Hydrology*, *625*, 129876. https://doi.org/10.1016/j.jhydrol.2023.129876

Mikkola, P., Martin, O. A., Chandramouli, S., Hartmann, M., Pla, O. A., Thomas, O., et al. (2024). Prior knowledge elicitation: The past, present, and future. *Bayesian Analysis*, *19*(4), 1129–1161. https://doi.org/10.1214/23-BA1381

Nguyen, V. D., Vorogushyn, S., Nissen, K., Brunner, L., & Merz, B. (2024). A non-stationary climate-informed weather generator for assessing future flood risks. *Advances in Statistical Climatology, Meteorology and Oceanography*, *10*(2), 195–216. https://doi.org/10.5194/ascmo-10-195-2024

Northrop, P. J. (2024). Stochastic models of rainfall. *Annual Review of Statistics and Its Application*, *11*(11), 51–74. https://doi.org/10.1146/annurev-statistics-040622-023838

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., & Isham, V. (2000). Rainfall modelling using Poisson-cluster processes: A review of developments. *Stochastic Environmental Research and Risk Assessment*, *14*(6), 0384–0411. https://doi.org/10.1007/s004770000043

Papamakarios, G., & Murray, I. (2016). *Fast ϵ-free inference of simulation models with Bayesian conditional density estimation*. In (Vol. 29). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2016/hash/6aca97005c68f1206823815f66102863-Abstract.html

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*(1), 572617–572680.

Papamakarios, G., Pavlakou, T., & Murray, I. (2018). Masked autoregressive flow for density estimation. *arXiv*. https://doi.org/10.48550/arXiv.1705.07057

Papamakarios, G., Sterratt, D., & Murray, I. (2019). Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *Proceedings of the twenty-second International Conference on artificial intelligence and statistics* (pp. 837–848). PMLR.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *arXiv*. https://doi.org/10.48550/arXiv.1912.01703

Phan, D., Pradhan, N., & Jankowiak, M. (2019). Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv*. https://doi.org/10.48550/arXiv.1912.11554

Prudhomme, C., Wilby, R. L., Crooks, S., Kay, A. L., & Reynard, N. S. (2010). Scenario-neutral approach to climate change impact studies: Application to flood risk. *Journal of Hydrology*, *390*(3), 198–209. https://doi.org/10.1016/j.jhydrol.2010.06.043

Racsko, P., Szeidl, L., & Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modelling*, *57*(1), 27–41. https://doi.org/10.1016/0304-3800(91)90053-4

Rezende, D. J., & Mohamed, S. (2016). Variational inference with normalizing flows. *arXiv*. https://doi.org/10.48550/arXiv.1505.05770

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, *17*(1), 182–190. https://doi.org/10.1029/WR017i001p00182

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., et al. (2018). Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, *151*(3–4), 555–571. https://doi.org/10.1007/s10584-018-2317-9

Sisson, S., Fan, Y., & Beaumont, M. (2018). Handbook of approximate Bayesian computation (1st ed.). https://doi.org/10.1201/9781315117195

Sommer, P. S., & Kaplan, J. O. (2017). A globally calibrated scheme for generating daily meteorology from monthly statistics: Global-WGEN (GWGEN) v1.0. *Geoscientific Model Development*, *10*(10), 3771–3791. https://doi.org/10.5194/gmd-10-3771-2017

Stasinopoulos, M., Kneib, T., Klein, N., Mayr, A., & Heller, G. (2024). *Generalized additive models for location, scale and shape—A distributional regression approach and its application* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009410076

Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society: Series A*, *147*(1), 1–18. https://doi.org/10.2307/2981736

Stoner, O., & Economou, T. (2020). An advanced hidden Markov model for hourly rainfall time series. *Computational Statistics & Data Analysis*, *152*, 107045. https://doi.org/10.1016/j.csda.2020.107045

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., & Gelman, A. (2020). Validating Bayesian inference algorithms with simulation-based calibration. *arXiv*. https://doi.org/10.48550/arXiv.1804.06788

Tavaré, S., Balding, D. J., Griffiths, R. C., & Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, *145*(2), 505–518. https://doi.org/10.1093/genetics/145.2.505

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., et al. (2020). SBI: A toolkit for simulation-based inference. *Journal of Open Source Software*, *5*(52), 2505. https://doi.org/10.21105/joss.02505

Underwood, F. M. (2009). Describing long-term trends in precipitation using generalized additive models. *Journal of Hydrology*, *364*(3), 285–297. https://doi.org/10.1016/j.jhydrol.2008.11.003

Vrac, M., & Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005308

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The inter-sectoral impact model intercomparison project (ISI–MIP): Project framework. *Proceedings of the National Academy of Sciences*, *111*(9), 3228–3232. https://doi.org/10.1073/pnas.1312330110

Wessel, J., & Chandler, R. E. (2025). *Improving multisite precipitation generators based on generalised linear models*. ESS Open Archive. https://doi.org/10.22541/essoar.176169599.97682958/v1

Wessel, J., & Groenke, B. (2025). wxsbi: Code and data files for the manuscript "Stochastic weather generation for scenario-neutral impact assessments using simulation-based inference" (version v2) [Collection]. https://doi.org/10.5281/zenodo.15448998

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather*, *65*(7), 180–185. https://doi.org/10.1002/wea.543

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., & Wilks, D. S. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, *34*(11), 2995–3008. https://doi.org/10.1029/98WR02577

Wilks, D. S. (2010). Use of stochastic weather generators for precipitation downscaling. *WIREs Climate Change*, *1*(6), 898–907. https://doi.org/10.1002/wcc.85

Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: A review of stochastic weather models. *Progress in Physical Geography: Earth and Environment*, *23*(3), 329–357. https://doi.org/10.1177/030913339902300302

Yang, C., Chandler, R. E., Isham, V. S., & Wheater, H. S. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, *41*(11). https://doi.org/10.1029/2004WR003739

Zabel, F., & Poschlod, B. (2023). The Teddy tool v1.1: Temporal disaggregation of daily climate model data for climate impact analysis. *Geoscientific Model Development*, *16*(18), 5383–5399. https://doi.org/10.5194/gmd-16-5383-2023

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3