CrossMark

ORIGINAL PAPER

IMAGE: a multivariate multi-site stochastic weather generator for European weather and climate

Nathan J. Sparks[1] · Stephen R. Hardwick[1] · Matthias Schmid[1] · Ralf Toumi[1]

Abstract Capturing the spatial and temporal correlation of multiple variables in a weather generator is challenging. A new massively multi-site, multivariate daily stochastic weather generator called IMAGE is presented here. It models temperature and precipitation variables as latent Gaussian variables with temporal behaviour governed by an auto-regressive model whose residuals and parameters are correlated through resampling of principle component time series of empirical orthogonal function modes. A case study using European climate data demonstrates the model's ability to reproduce extreme events of temperature and precipitation. The ability to capture the spatial and temporal extent of extremes using a modified Climate Extremes Index is demonstrated. Importantly, the model generates events covering not observed temporal and spatial scales giving new insights for risk management purposes.

Keywords Weather generator · Extreme events · Climate extremes index · Multivariate · Multi-site · EOF

1 Introduction

Stochastic weather generators are used in a wide range of applications, including agricultural and water systems management (Cowden et al. 2008; Supit et al. 2012),

extreme weather risk assessment (Te Linde et al. 2010) and climate change impact studies (Steinschneider and Brown 2013). Stochastic weather generators produce realistic time series of arbitrary length of weather variables, while preserving the statistics of the input meteorological data, which can be obtained from historical observations or models. In their simplest form weather generators produce synthetic time series for a single weather variable at a single location. However, for many applications the geographic area considered is sufficiently large that weather variables, such as precipitation, can vary significantly over the domain, meaning that time series at multiple sites are desirable. The production of realistic synthetic weather data in this case requires the preservation of spatio-temporal correlation between sites, increasing the complexity of the problem significantly. Additionally, for many applications time series of multiple, correlated weather variables are needed.

A wide-range of weather generators have been developed based on various methodologies, which can be broadly categorised as parametric, non-parametric and hybrid models. Most parametric models are based on the method of Richardson (1981), who used a single-step Markov chain to simulate time series of precipitation occurrence (wet or dry days) at a single site. Precipitation amounts, as well as maximum and minimum air temperature and solar radiation were then calculated using parametric models, derived from historical data, which depended on whether it was a wet or dry day. Extension of these parametric models to multisite weather generators requires representing the spatial correlation between sites. Wilks (1999b) achieved this by driving each single site generator with correlated random numbers to determine rainfall event sequences and amounts. Baigorria and Jones (2010) generated single-step Markov chains of rainfall occurrence for two sites initially, and then

✉ Nathan J. Sparks
  n.sparks07@imperial.ac.uk

[1] Blackett Laboratory, Department of Physics, Imperial College London, London SW7 2AZ, UK

generated time series for other sites one-by-one, based on the local transition probability as well as the time series of the two most highly correlated sites that had already been generated. Serinaldi (2009) used a copula-based method to account for pairwise correlations between sites. An alternative to the Markov chain approach is to model precipitation as a transformed censored latent Gaussian process (e.g. Allard and Bourotte 2014). In this case, precipitation data is transformed so as to have a Gaussian or almost Gaussian distribution, with zero precipitation values being treated as censored values below a certain threshold. It is then modelled as a Gaussian process, spatially continuous and stationary defined by a single correlation function. Alternatively, Verdin et al. (2015) use a censored latent Gaussian process to determine rainfall occurrence only, while rainfall intensity is modelled as a gamma random variable. Uniform spatial covariance functions are used to represent spatial correlation of the residuals, and this constraint can limit the model's ability to reproduce spatial correlation. Verdin et al. (2015), Youngman and Stephenson (2016) also use geostatistics to allow the simulation of weather variables at sites without observational data. While this can be advantageous for some applications, in areas with abundant data it may not be required. Weather generators that do not use geostatistics can still be used to generate weather series at arbitrary locations, by interpolating data either prior to (Camberlin et al. 2014) or after (Camera et al. 2016) simulation.

Non-parametric models resample from historical data to produce new sequences of weather data. Resampling normally employs a k-nearest neighbour technique (Rajagopalan and Lall 1999; Leander and Buishand 2009; Caraway et al. 2014). One drawback of resampling methods is that they cannot generate events that have not been observed historically. Srivastav and Simonovic (2015) use a maximum entropy bootstrap method which preserves temporal correlation by exactly replicating the rank-ordering of historical data values. Hybrid models combine elements of parametric and non-parametric models, for example by using a Markov chain to generate a sequence of precipitation states, followed by k-nearest neighbour resampling to produce the values of the multiple weather variables (Apipattanavis et al. 2007; Steinschneider and Brown 2013). The advantage of hybrid models compared to purely non-parametric models is their ability to produce sequences of events quite different to those present in the observational data, while not encountering some of the difficulties found in defining fully parametric models.

Overdispersion is a problem discussed by many authors (e.g., Katz and Parlange 1998; Wilks 1999a; Wang and Nathan 2007; Chen et al. 2010; Steinschneider and Brown 2013) and refers to an underestimation of monthly and annual variability of simulated weather variables. This leads to an under representation of extreme events on longer time scales. Different methods of addressing this problem have been suggested including coupling course and fine scale time series (Wang and Nathan 2007), applying frequency spectrum based corrections (Chen et al. 2010) and using wavelet decomposition analysis to modulate simulation on an annual scale (Steinschneider and Brown 2013).

Here we present the first multi-site multivariate stochastic weather generator based on the use of periodically extended empirical orthogonal functions (EOFs). We model precipitation as a censored latent Gaussian process. EOFs are an attractive option for a multisite weather generator, as they have been frequently and successfully applied to analyses of spatial correlations in climate (e.g. Zhang et al. 1997; Fyfe et al. 1999). EOFs have not been widely applied to weather generators in the past, but cyclostationary EOF analysis was used by Kim et al. (2013) to model summer rainfall in Korea. Our model differs from theirs in a number of ways, including: our use of periodically extended EOFs to capture the low-frequency variability of weather variables to overcome the problem of overdispersion; our use of autoregressive models at each individual site, rather than over the whole domain; our integrated approach to the modelling of extremes as opposed to modelling extremes as events always occurring simultaneously over predefined clusters of sites.

One key application of stochastic weather generators is the synthesis of long time series of weather data for both current and future climate to allow a full risk assessment of rare events to be conducted. In most cases observations and climate projections are available for sufficient time periods to assess the mean climatology, but are not sufficient for assessing the frequency and magnitude of extreme events. Here we introduce IMAGE (Imperial College Weather Generator), our multi-site, multi-variable weather generator. It is designed to be used to assess the risk of events for which the spatial distribution of weather variables is important, such as rainfall anomalies over several months over a large watershed affecting flow rates or heatwaves affecting several regions of a country over a period of a few days. In general it is these events, with significant spatial extent, that have the most harmful impacts (e.g. Russo et al. 2015). Our generator is not especially suited to predicting extreme weather events at single sites, as there are many single-site weather generators that are already extremely accurate in this regard.

2 Modelling framework

A schematic describing the modelling process including the parameter estimation and simulation phases is shown in Fig. 1.

Fig. 1 Schematic of simulation procedure including parameter estimation and generation phases
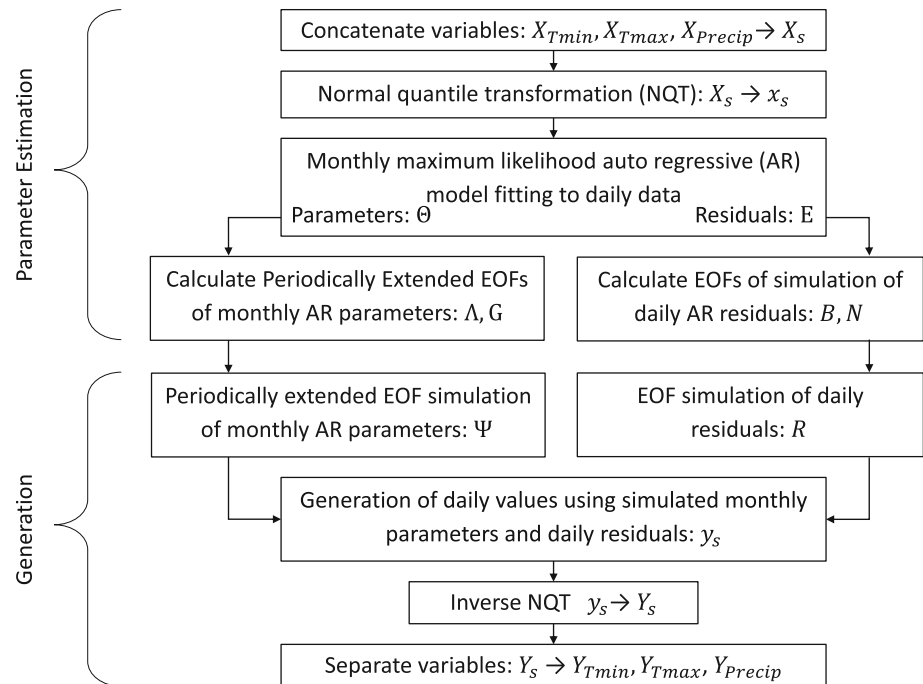
In IMAGE multiple variable types (e.g. T_{min}, T_{max}, $Precip$) across multiple locations are treated similarly, simulated simultaneously and hence are here identified by a single index, $s = \{1..S\}$ where S is the total number of variables to be modelled and here is equal to the sum of the number of T_{max}, T_{min} and $Precip$ locations. Variable s at a time t is represented by $Y_s(t)$ which is modelled as a latent Gaussian variable, y_s, such that:

$$Y_s(t) = Q_s^{-1}(y_s(t)), \qquad (1)$$

where Q_s is the normal quantile transform (NQT) (see, for example, Krzysztofowicz 1997) defined such that an arbitrarily distributed observed variable X_s is transformed to a normally distributed variable x_s by

$$x_s(t) = Q_s(X_s(t)). \qquad (2)$$

No treatment for values of y_s outside of the observed range is made here for simplicity so the maximum range of modelled values Y_s is identical to the range of observed values X_s. This approach allows us to focus on the core strength of the model in reproducing spatial and temporal correlations. For zero-inflated variables such as precipitation the inverse NQT takes the form,

$$Q_s^{-1}(y_s(t)) = \begin{cases} F_s^{-1}(\Phi(y_s(t))), & \text{if } y_s > d_s \\ 0, & \text{if } y_s <= d_s \end{cases}, \qquad (3)$$

where F_s is the cumulative distribution function (CDF) of X_s, Φ is the standard Gaussian CDF and d_s is some threshold value for the Gaussian variable below which the outcome is censored. In the case of precipitation, d_s

is equal to $\Phi^{-1}(f_s)$, where f_s is the dry day fraction. For variables which are not lower bounded (e.g. temperature) d_s effectively takes the value of negative infinity. This process enables the simulation of wet or dry days and amounts in one step which is convenient but also means the two quantities are fundamentally related in the model. We justify the use of latent Gaussian variables through their success in modelling precipitation in studies such as (Allard and Bourotte 2014; Baxevani and Lennartsson 2015). This transformation approach also means the model is general enough to allow the simultaneous modelling of different weather and climate variables.

Time evolution of y_s is modelled as a first order auto regressive (AR) process:

$$y_s(t) = c_s + \alpha_s y_s(t-1) + \epsilon_s \qquad (4)$$

where c_s and α_s are referred to as the constant and memory parameters respectively and ϵ_s is a noise term. First order AR was chosen as it is the simplest temporal model with memory. The Bayesian information criterion (BIC) was used to test the suitability of a range of ARIMA models (up to third order in both parameters) and AR1 was the optimal choice in the majority of cases.

The broad AR modelling methodology is to use simulated residual terms with variance and covariance across s derived from the observation data to drive the model on a daily scale. Low-frequency, monthly parameter values with cross covariances across s and t derived from the observation data are used to control longer term behaviour.

We proceed by first estimating parameters c and α for each variable s for each month m of the transformed observation data x using a maximum likelihood method. This process also yields the observed residual terms, ϵ, which may be correlated across s.

To simulate new realizations of these residuals which preserve the observed correlation we use an EOF resampling technique. From the fitting process we obtain the observed residuals matrix $\mathbf{E}_{(T \times S)}$ where T is the total number of observations time steps for a given month. Then for each month, an EOF decomposition of $\mathbf{E}$ is performed to give matrix $\mathbf{B}$, whose columns contain the EOF modes. The projection of $\mathbf{E}$ onto $\mathbf{B}$ gives $\mathbf{N}$, whose columns contain the principle component time series associated with each mode,

$$\mathbf{N} = \mathbf{EB}. \tag{5}$$

A new realization of residuals for a single time step, $\mathbf{R}$, is created by randomly sampling from the principle component time series for each EOF mode,

$$\mathbf{R} = \mathbf{B}\Pi, \tag{6}$$

where Π is the vector of random samples from the principle component time series defined by $\pi_i = N_{r_i,i}$, r is a random variable following the discrete uniform distribution over the set $\{1..T\}$, and $i = \{1..n\}$ is an index where n is the number of EOF modes considered. In our simulation we consider all EOF modes so we reproduce exactly the variance in the sample data. A new realization of $\mathbf{R}$ which is a vector containing S elements is created for each time step of the simulation. Element s of $\mathbf{R}$ corresponds to the residual for variable s at a given time step.

The variability of and correlation between parameters c and α over time t and variable s control longer term correlations and variability in the model. We use a technique of resampling bivariate periodically extended empirical orthogonal functions (PXEOF) to simulate these parameters. PXEOFs are traditionally used as an analytical technique for reducing a periodic dataset to its leading modes of variability across time and space and are described by Kim and Wu (1999).

Each year of observation data gives parameter matrices $\mathbf{C}_{(S \times M)}$ and $\mathbf{A}_{(S \times M)}$ where M is the number of months in a year. To calculate the PXEOFs the parameter matrices are then vectorized,

$$\begin{aligned}\mathbf{c_p} &= vec(\mathbf{C}) \\ &= [C_{1,1}, ...C_{S,1}, C_{1,2}, ...C_{S,2}, ...C_{1,M}, ...C_{S,M}]_{(p)}\end{aligned} \tag{7}$$

$$\begin{aligned}\alpha_\mathbf{p} &= vec(\mathbf{A}) \\ &= [A_{1,1}, ...A_{S,1}, A_{1,2}, ...A_{S,2}, ...A_{1,M}, ...A_{S,M}]_{(p)}\end{aligned} \tag{8}$$

where subscript p is the observation year index (period). The estimated observation parameter matrix Θ is then defined as follows:

$$\Theta = \begin{bmatrix} \mathbf{c_1} & \alpha_1 \\ \vdots & \vdots \\ \mathbf{c_P} & \alpha_\mathbf{P} \end{bmatrix} \tag{9}$$

whose each row represents the estimates of observation parameters for a given year and each column is the annual time series of a parameter for a given variable at a given month and P is the number of years of observation data. An EOF decomposition on Θ yields matrix Λ whose columns contain the EOF modes and projecting Θ onto Λ gives,

$$\mathbf{G} = \Theta\Lambda \tag{10}$$

where the columns of $\mathbf{G}$ contain the principle component time series associated with each mode. As above, a new set of coefficients Ψ can then be created via

$$\Psi = \Lambda\Gamma + \mu \tag{11}$$

where μ is a vector containing the column means of Θ, $\Gamma_i = G_{r_i,i}$ and r is a random variable following the discrete uniform distribution over the set $\{1..P\}$.

Simulated c and α parameters are obtained from Ψ by $c_j = \Psi_j$ and $\alpha_j = \Psi_{j+S \times M}$ with $j = \{1..S \times M\}$. Finally, inverting the vectorization operations from Eqs. (7) and (8) gives the simulated parameter matrices, $\mathbf{c}_{s,m}$ and $\alpha_{s,m}$, from which values are extracted for use in Eq. (4). A new realization of these parameters is created for each year of simulation.

The simulated residuals and parameters can now be used to generate daily values of latent Gaussian variable y_s using Eq. (4) before transformation to Y_s via Eq. (1). Finally Y_s can be separated into constituent variables (e.g. T_{min}, T_{max}, $Precip$).

3 Case study and data

We evaluate the performance of IMAGE using gridded daily minimum and maximum temperature (T_{min} & T_{max}) and precipitation ($Precip$). The domain covers most of Europe and extends from 10W to 18E and 36N to 60N with 22 longitudinal and 15 latitudinal divisions as shown in Fig. 2. The resolution was chosen to provide a reasonable number of data points for our model (approximately 500 in total) while covering the region of interest in the case study. The data were nearest-neighbour interpolated from the European Climate Assessment & Dataset 0.5° E-OBS product (Haylock et al. 2008). Data from this source are not available over water which reduced the number of grid points from 330 to 163. Single day gaps in precipitation were filled with linear interpolation (maximum of 12 per grid point). Grid points with longer gaps were excluded which reduced the number of grid points available for this
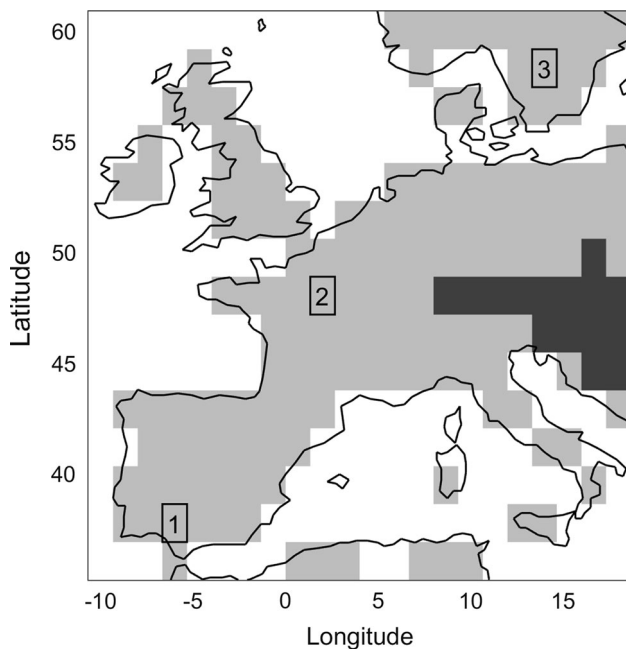
Fig. 2 *Shaded* area shows study domain with *dark gray* cells the western Danube basin case study region. *Numbered cells* are case study locations

variable to 152. There were therefore 478 time series when all the locations and variables were considered. Data were available for a 65 year period (1950–2015). IMAGE was fitted to this data and used to simulate 6500 years (100 times the available record length). Where we refer to "ensembles" we separate the simulation into 100 members each 65 years long.

We chose to examine the output of IMAGE at three points in detail as case studies: (1) southern Spain, (2) central France and (3) southern Sweden. The western Danube basin consisting of 15 cells was chosen as a region to assess spatiotemporal precipitation performance.

We aim to test the ability of the model to capture the multi-variable spatial-temporal extent of extremes. We therefore used a modified form of the annual U.S. Climate Extremes Index (CEI) documented by Gleason et al. (2008) to determine the performance of the model's ability to simulate extreme events on climatic scales. The CEI is a composite of five indices that describe the percentage of the contiguous United States subject to daily extreme minimum and maximum temperatures, extreme 1-day precipitation, extreme number of wet/dry days and extreme values of the Palmer Drought Severity Index (PDSI). We replace the PDSI with a purely meteorological index, the total precipitation in a year. Extreme thresholds are defined as 10th and 90th percentiles over the period of record except for the extreme precipitation index which only considers the upper threshold. The extreme precipitation index therefore carries twice the weight of the others in the composite index.

4 Results

Annual mean minimum (T_{min}) and maximum temperature (T_{max}) were accurately simulated by IMAGE across the domain (mean biases: $\Delta T_{min} = -0.05$ degrees C, $\Delta T_{max} = 0.002$ degrees C) (Fig. 3). Annual mean precipitation was slightly overestimated by IMAGE (mean bias: Δ *Precip* = 28.2 mm, mean percentage bias: 3.4%). There was a systematic positive bias in the standard deviation of interannual means. This was present across all variables at all sites with a mean positive bias of 42%.

IMAGE successfully reproduced the seasonal cycles of the mean T_{max}, T_{min} and *Precip* (Fig. 4). The root mean square error of simulated monthly means across all locations was 0.08K for T_{min} 0.11K for T_{max} and 0.11 mm for *Precip* respectively. The standard deviations of the daily values of T_{min} and T_{max} for each month were slightly underestimated, with mean biases of -0.11K for T_{min} and -0.10K for T_{max}. IMAGE also accurately simulated the monthly cycle of the rain day fraction, defined as the fraction of days on which *Precip* exceeded 0.1 mm (Fig. 4). The overall bias of the rain day fraction was 0.003 (mean percentage bias: 4.4%), with a root mean square error of 0.019 (root mean square percentage error: 8.3%).

The distributions of T_{min} and T_{max} simulated by IMAGE are very close to the observed distributions. Quantile-quantile plots from the three case study sites are shown in Fig. 5. Across all sites IMAGE tended to produce slightly fatter-tailed distributions than those observed, with simulated temperatures lower than observed at percentiles below 50, and simulated temperatures higher than observed at percentiles above 50. The mean biases of simulated T_{min} at a selection of percentiles were: 1st percentile: -0.79K, 10th percentile: -0.24K, 33rd percentile: -0.28, 50th percentile: -0.08K, 66th percentile: 0.23K, 90th percentile: 0.19K, 99th percentile: 0.06K. Similarly, the mean biases of simulated T_{max} at the same percentiles were: 1st percentile: -0.39K, 10th percentile: -0.20K, 33rd percentile: -0.29K, 50th percentile: 0.02K, 66th percentile: 0.31K, 90th percentile: 0.24K, 99th percentile: -0.02K. IMAGE tended to overestimate extreme *Precip* values, as illustrated in Fig. 5. The mean biases at upper percentiles were: 90th percentile: -1%, 95th percentile: 3%, 99th percentile: 11% (Figs. 5, 6). The largest bias occurred in the southwest of the domain, in Portugal and western Spain (Fig. 6).

Spatial correlation also needs to be assessed. Inter-gridcell Pearson's correlation coefficients calculated on daily data were slightly underestimated for T_{min}, T_{max} and *Precip* (Fig. 7) (mean observed ρ: T_{min}: 0.80, T_{max}: 0.85, *Precip*: 0.10; mean simulated ρ: T_{min}: 0.73, T_{max}: 0.77, *Precip*: 0.04). For T_{min} and T_{max} IMAGE performed best
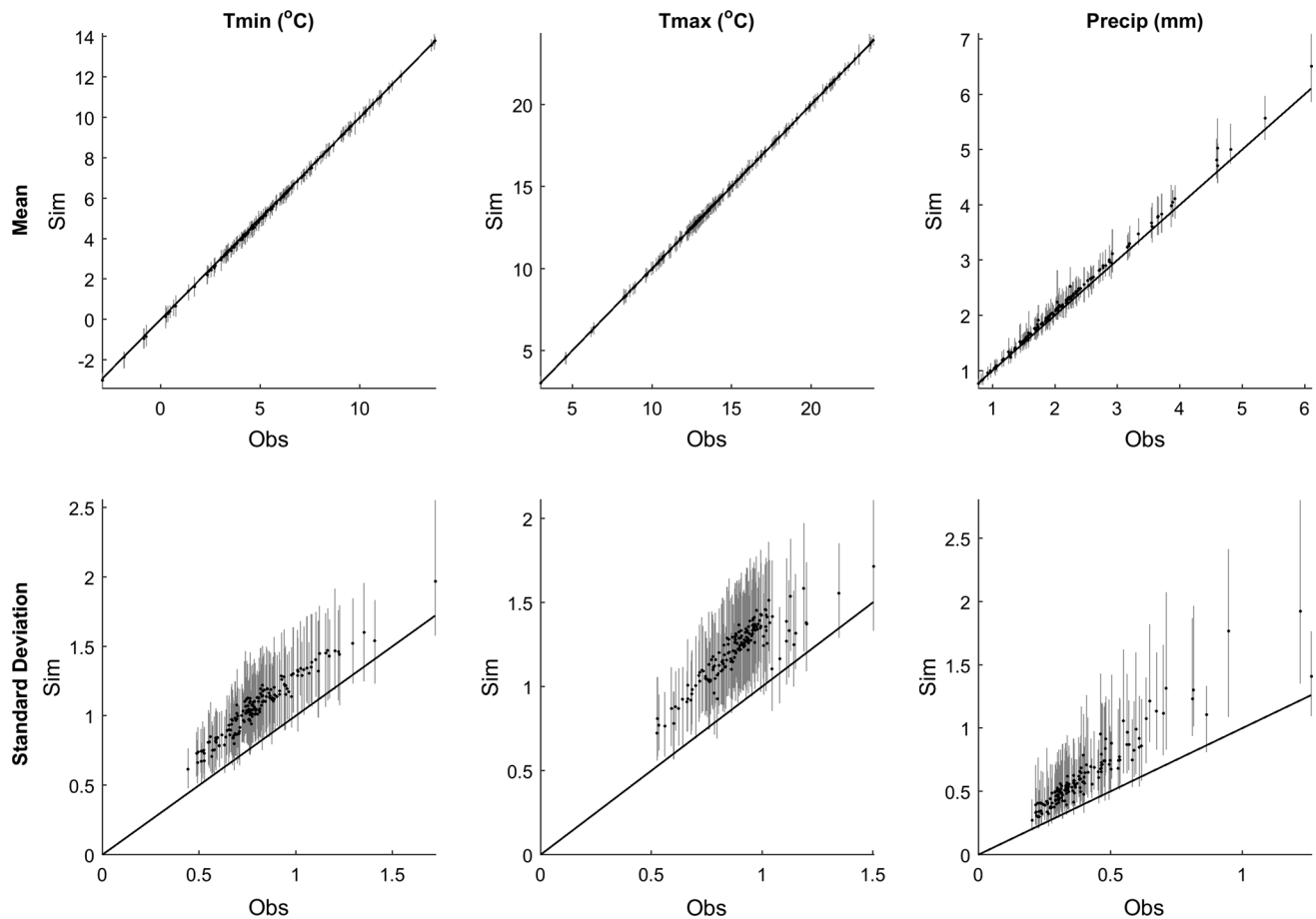
Fig. 3 Observed and simulated annual means and standard deviation of annual means for T_{min}, T_{max} and *Precip*. *Each point* represents one grid point. *Vertical bars* show the range across 100 ensemble members where each ensemble has length equal to that of the observation data (65 years)

for very highly correlated gridcell pairs ($\rho > 0.95$) and for gridcell pairs with the lowest observed correlations ($\rho < 0.7$). Overall this spatial correlation performance is very similar to that seen in Verdin et al. (2015). For *Precip* IMAGE performed best for gridcell pairs with little or no observed correlation and underestimated ρ for more highly correlated gridcells. Pairwise gridcell correlations between T_{min} and T_{max} were consistently slightly underestimated by IMAGE (mean observed ρ: 0.80, mean simulated ρ: 0.73). Similarly, there was a tendency for IMAGE to slightly underestimate the pairwise gridcell correlations of T_{min} and *Precip* and T_{max} and *Precip* for pairs with observed correlation $|\rho| > 0.15$, however, in general there was very little observed correlation between these variables. The simulated pairwise correlation of annual rainfall amounts was closely to correlated to the observed values ($\rho = 0.91$). Similarly, the simulated pairwise correlation of annual rain days was very closely correlated with the observed values ($\rho = 0.94$).

The Danube basin was chosen as a case study as it is an important catchment in Europe. The simulated return

period of extreme rainfall events aggregated across the Danube basin was realistic (Fig. 8). Observed events with a return period of 1 year had a mean *Precip* of 18.0 mm, while simulated events with a return period of 1 year had a mean *Precip* of 17.1 mm (error: −5%). IMAGE demonstrated it was capable of simulating extreme multi-gridcell rainfall events of greater magnitude than those observed and at a realistic frequency. These events are not simply an extrapolation of historical data. An interesting application of IMAGE is to more confidently predict the return period of events. For example, the highest observed mean rainfall across gridcells in the Danube basin (33 mm) in the 65 year E-OBS data set corresponded to an event with a return period of approximately 89 years in the IMAGE simulation.

Cold spells are important extreme events with impacts on health, energy and transport. Cold spells were defined as consecutive days with T_{min} below the 5th percentile of historical values of T_{min} (Fig. 9). IMAGE tended to slightly underestimate the occurrence rates of short cold spells and slightly overestimate the occurrence of longer cold spells.

Fig. 4 Observed and simulated annual cycles of monthly mean and standard deviation of T_{min} and T_{max} and monthly mean and rain day fraction of *Precip*. *Vertical bars* show ± 1 inter-annual standard deviation
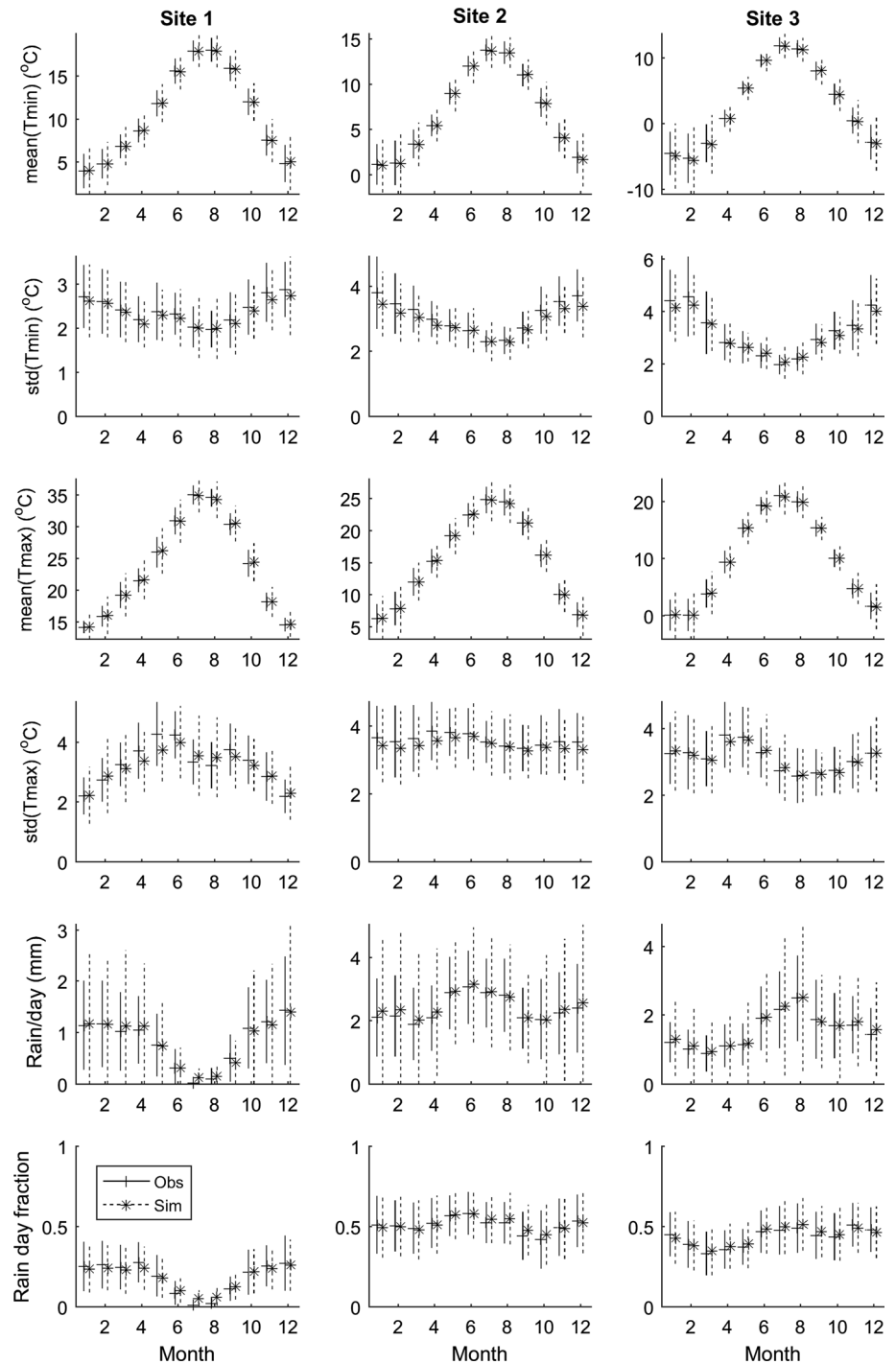
The cumulative probability density of cold spells less than 5 days in duration is very slightly overestimated, while the cumulative probability density of cold spells of between 5 and 20 days is generally underestimated across all sites. There were similar results for hot spells, which can have a profound impact on health. Hot spells were defined as consecutive days with T_{max} above the 95th percentile of historical values of T_{max}. As with cold spells, the cumulative probability density of hot spells less than 5 days in duration is very slightly overestimated, while the cumulative probability density of hot spells of between 5 and 20 days is generally underestimated across all sites. IMAGE was also able to simulate cold and hot spells which were longer than any that occurred in the observational record.
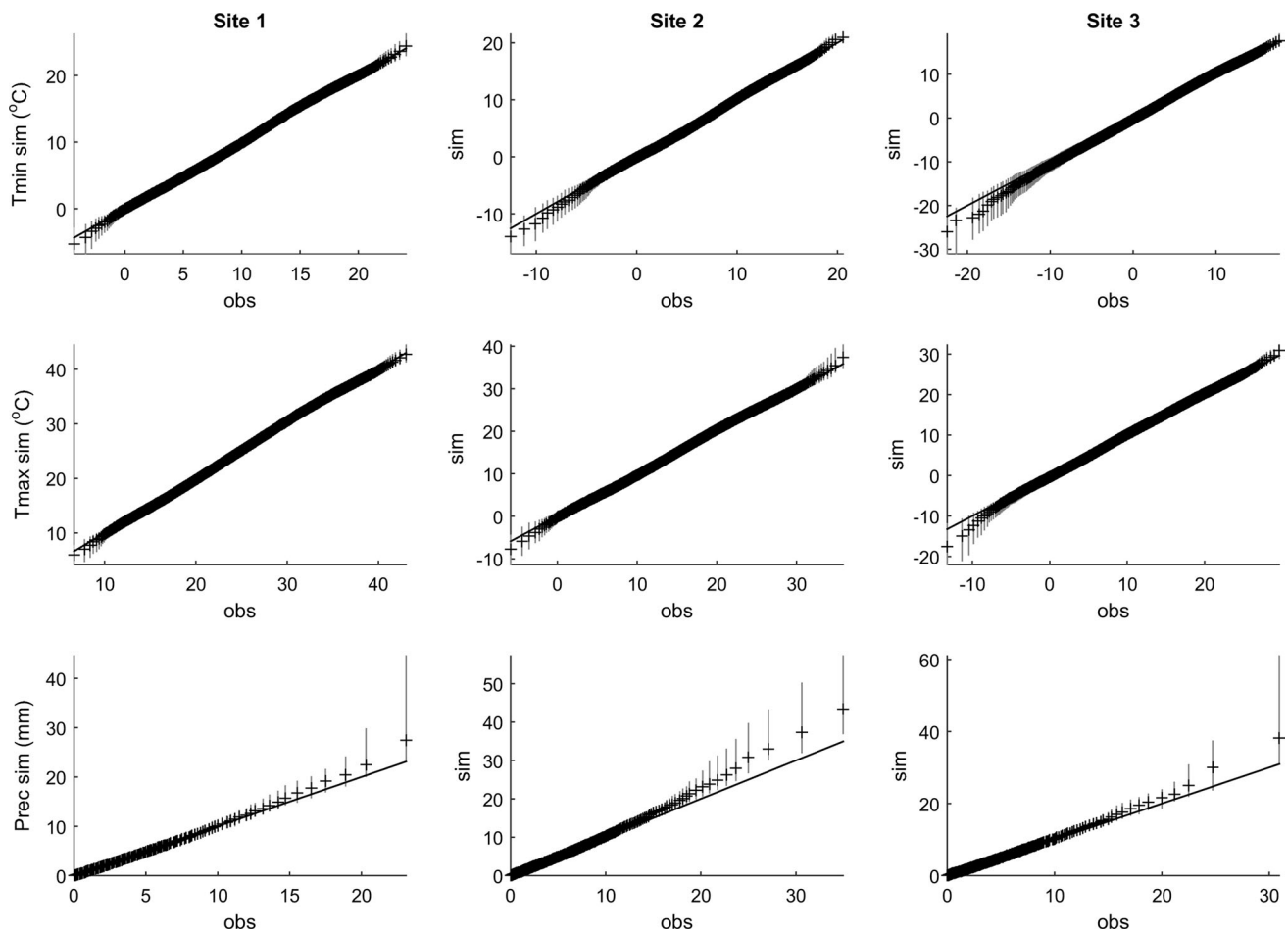
Fig. 5 QQ plots of T_{min}, T_{max} and *Precip* at the three sample locations. *Vertical bars* show the range across 100 ensemble members where each ensemble has length equal to that of the observation data (65 years)

IMAGE realistically simulated the occurrence frequency of simultaneous hot and cold events across multiple sites in the domain (Fig. 10). The size of simultaneous cold and hot events in the observed data and the data simulated by IMAGE were compared, without considering whether events of the same size occurred in the same geographic region in both datasets. Small cold spells, taken here to be days with T_{min} less than 0 °C across between 10 and 50% of the domain were simulated 5% too frequently by IMAGE. Similarly, days with T_{min} less than 0 °C at over 50% of sites in the domain were simulated 5% too frequently by IMAGE. Days with T_{max} greater than 30 °C at between 10 and 50 % of sites in the domain were simulated 15% too frequently by IMAGE. The largest errors occurred for heat waves with large spatial footprints. Days with T_{max} greater than 30 °C at greater than 50% of sites in the domain were simulated 57% less frequently by IMAGE than they occur in the observed data.

The final model validation test considers extremes in multiple variables. The frequency and magnitude of large annual values of the modified climate extreme index (CEI) across the entire domain were realistically simulated by IMAGE (Fig. 11). The CEI of events with a 5 year return period estimated from 65 years of observations was 0.26, while the CEI of events with a 5 year return period based on 6500 years of IMAGE simulations was 0.27. The most extreme observed event corresponds to an event with a return period of approximately 135 years in the simulation. Contributions of individual components to the CEI suggest that domain-wide extreme values of T_{min} and T_{max} and the number of dry (or wet) days were least realistically simulated by IMAGE. For example, for T_{min} the most extreme evens in the 65 year observational record had a magnitude equivalent to an event with a 310 year return period in the simulation. For T_{max} and the number of dry (or wet) days the two most extreme observed events had magnitudes equivalent to events with return periods of 810 and 380 years, respectively.

5 Discussion

IMAGE successfully reproduces basic weather and climate data phenomenon such as seasonal cycles in variable mean and standard deviations, rain day fractions. However the
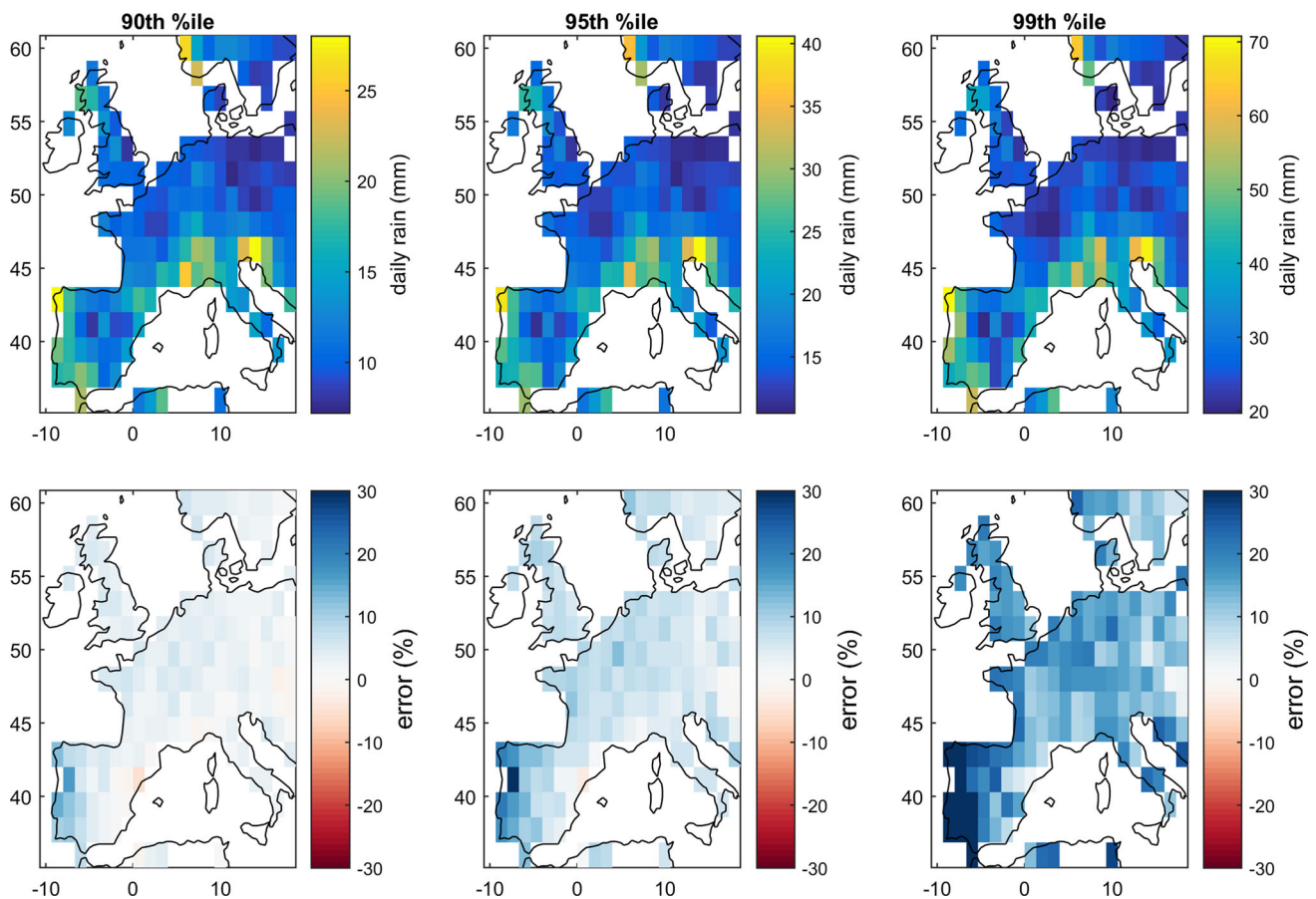
Fig. 6 Maps of upper percentiles (90, 95, 99) precipitation and bias

major achievement is in simultaneously reproducing extreme multivariate spatial, temporal and spatiotemporal events as demonstrated by the at-site temperature event durations, the basin-scale return values and the modified climate extreme index analysis. These large temporal and spatial scale events are crucial for many applications and have the highest social and economic impact.

While the model is relatively successfully in reproducing the observed return values of the CEI we acknowledge that the combined index hides some discrepancies in the components, for example, in the extremes of the minimum temperature index. This may be related to the deficiency in pairwise correlation exhibited in the model output. The underestimation of pairwise correlation in for example T_{min} and T_{max}, with a mean error of 9%, appears very similar to that seen in Verdin et al. (2015) who state an error on the order of 5%. The use of latent Gaussian variables in both models may be the cause of this similar loss of correlation.

The reproduction of observed distributions and rain day fractions confirm that that approach of modelling precipitation as a latent Gaussian variable demonstrated by Allard and Bourotte (2014) amongst others is valid in this extended multi-site multivariate context. We also tested

precipitation metrics such as pairwise correlations of number of wet days per year and annual rainfall which improve upon, for example, those shown in Mehrotra et al. (2015).

However, one limitation of this model is the simple normal quantile transform and its inverse which means that daily values exceeding those in the observation sample cannot be obtained for a given grid point. Out of observation range values can therefore only be generated by spatial aggregation (i.e. over the Danube basin) or temporal averaging/accumulation (5-day precipitation totals) or threshold exceedence type events (hot and cold spell durations). It would be a relatively simple modification to incorporate a more advanced distribution model and transform method such as those discussed by Bogner et al. (2012) but we felt that using the simplest method available demonstrated the strength at the core of the model in spatiotemporal event simulation and since the most impacting weather and climate metrics are spatially or temporally aggregated we do not consider this to be a critical shortcoming.

The under-representation of low-frequency variability is a common problem in weather generators (e.g., Katz and

Fig. 7 Intra- and inter-variable pairwise Pearson correlation coefficients
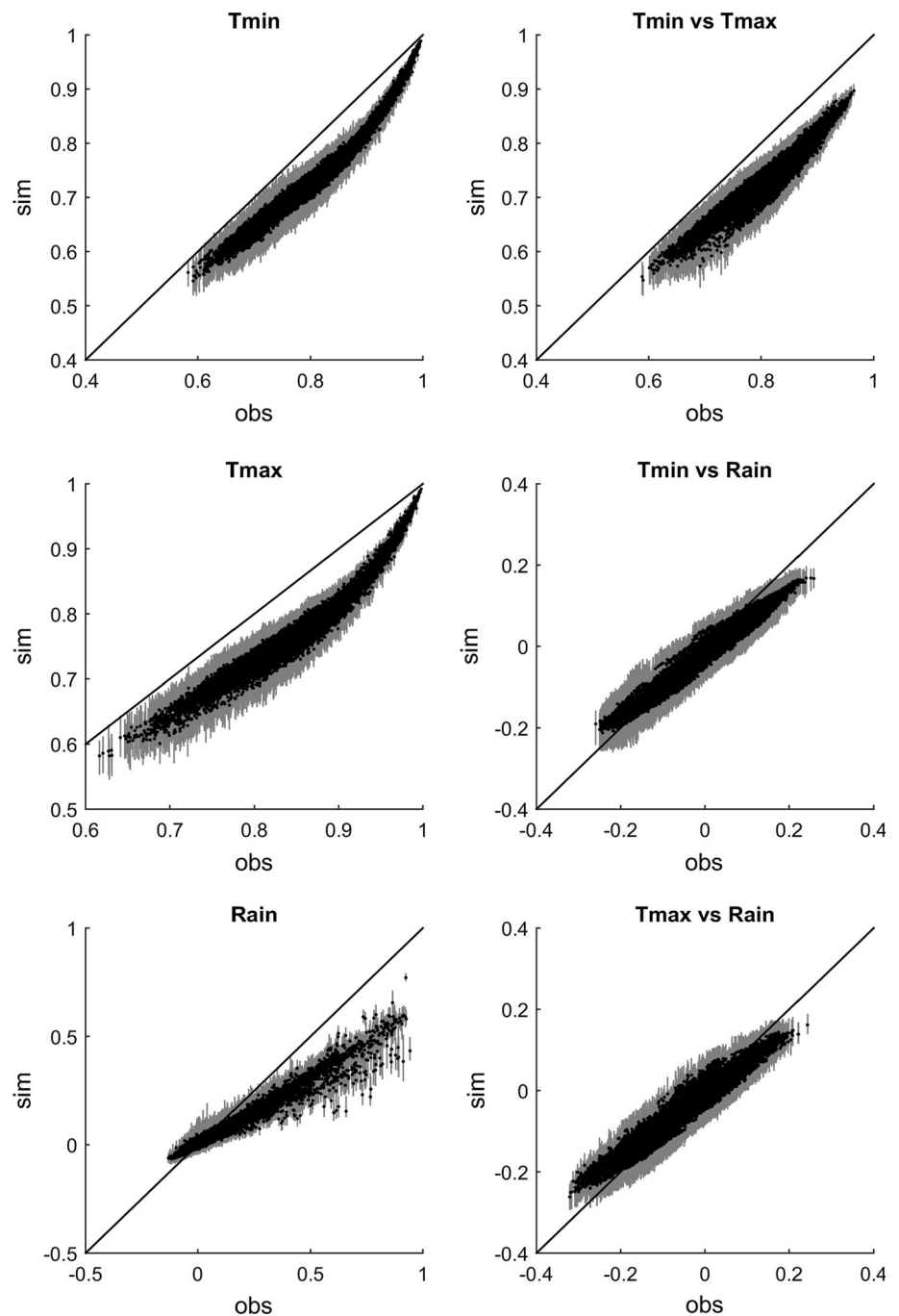
Parlange 1998; Wilks 1999a; Wang and Nathan 2007; Chen et al. 2010; Steinschneider and Brown 2013) and generally referred to as overdispersion. IMAGE attempts to preserve this variability by simulating a monthly time series of AR coefficients using PXEOFs. However Fig. 3 reveals a systematic positive bias in the standard deviation of annual means of the three variables across all grid points which indicates underdispersion. This is the opposite problem to, for example, Steinschneider and Brown (2013) whose model still suffers from a lack of variablity at annual

scales. The over estimation of the standard deviation of interannual means in IMAGE may be attributed to an excess of low-frequency variability in the AR parameters produced by the PXEOF simulation technique. We note this bias could be reduced by relaxing the simulated AR parameters towards climatological means per month, per location (not shown). However this adjustment also affected the magnitude of simulated extreme events. We therefore decided against adjusting the AR parameters and accepting the bias in interannual standard deviations as a
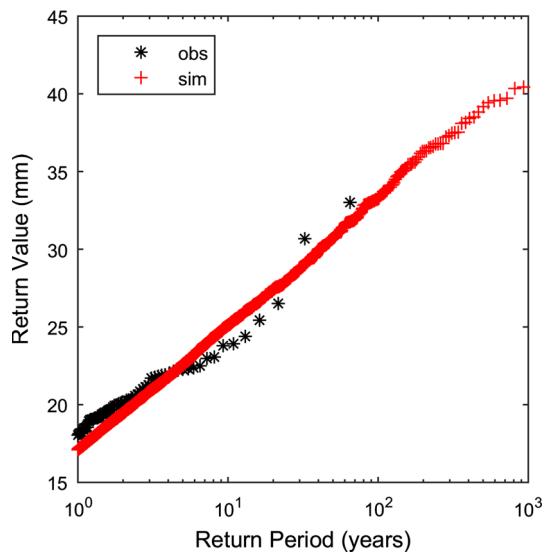
Fig. 8 Western Danube basin daily *Precip* return values against return periods

trade-off against improved representations of extreme events. This positive bias may counteract an expected detrimental effect of weaker simulated spatial correlations on extreme event footprint simulation.

Figures 5 and 6 show a systematic error in the model's reproduction of extreme precipitation. We believe this is likely a result of using a single transformation function for each variable for all seasons. This could lead to subsets (e.g. by season or month) of a transformed time series (x_s) having non-Gaussian and perhaps skewed distributions. This effect will be more prominent for skewed data such as precipitation than for approximately symmetrically distributed data such as temperature. The principal component resampling technique allows for some flexibility in the distribution shape but perhaps not enough to cover the seasonal variability displayed in the transformed precipitation data. A possible route to solving this problem is to implement a seasonally dependent transformation function.
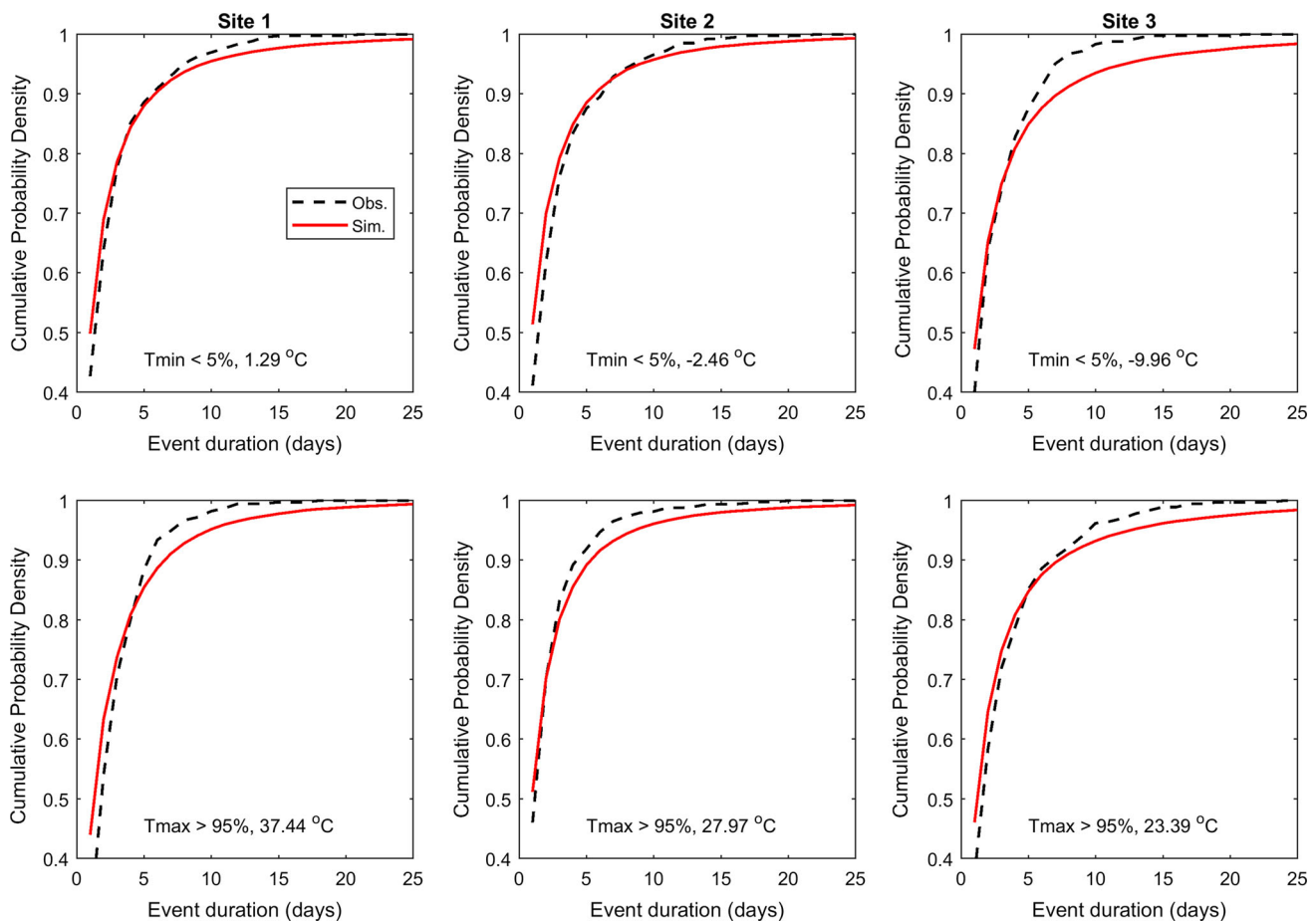
Fig. 9 Cold and hot spell event duration at the three sample sites. Cold event duration defined as number of consecutive days with T_{min} below the local 5th percentile. Hot event duration defined as number of consecutive days with T_{max} above the local 5th percentile

Fig. 10 Hot and cold event occurrence rate against spatial extent defined as fraction of domain on a given day with T_{min} below 0 °C or T_{max} above 30 °C
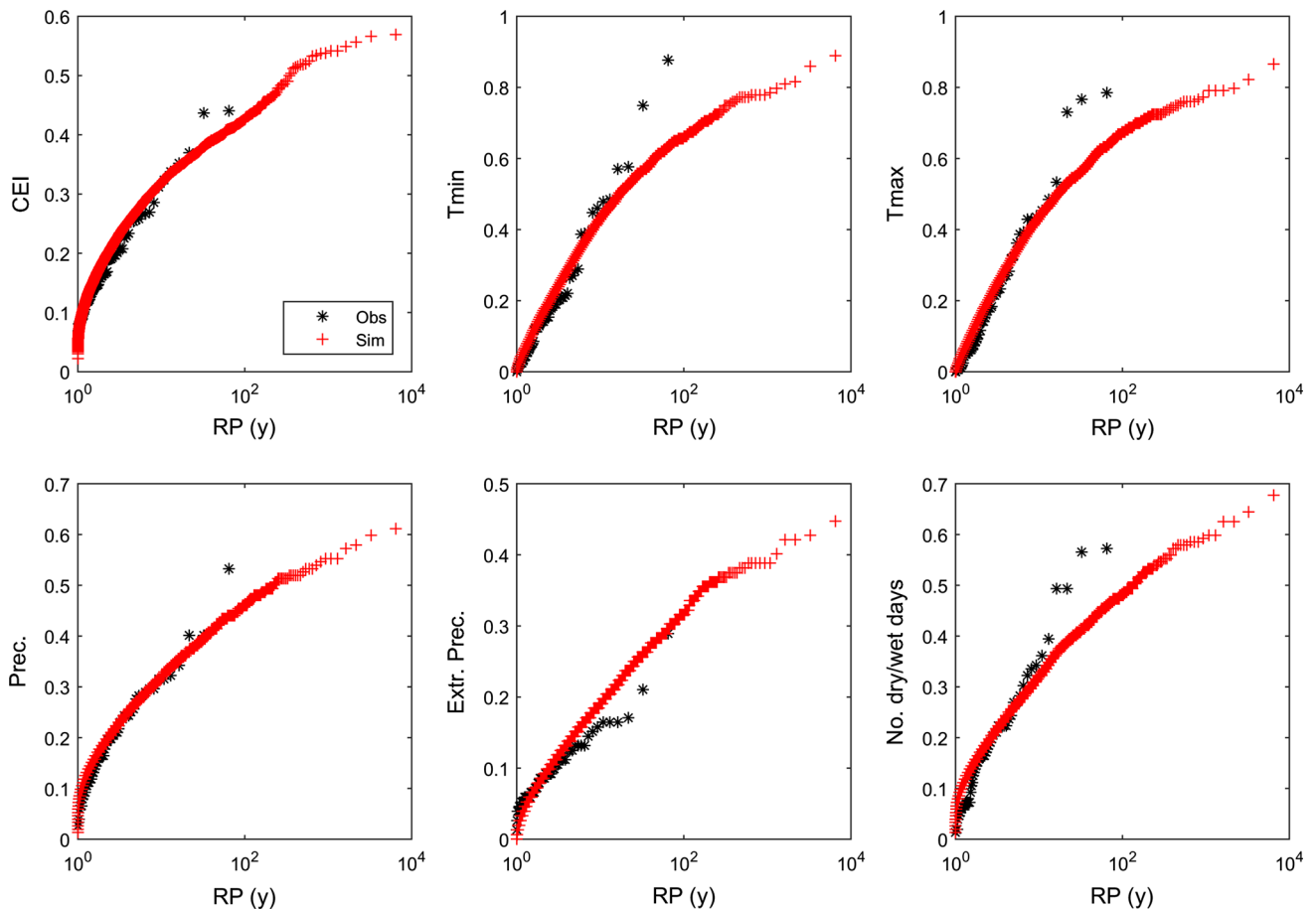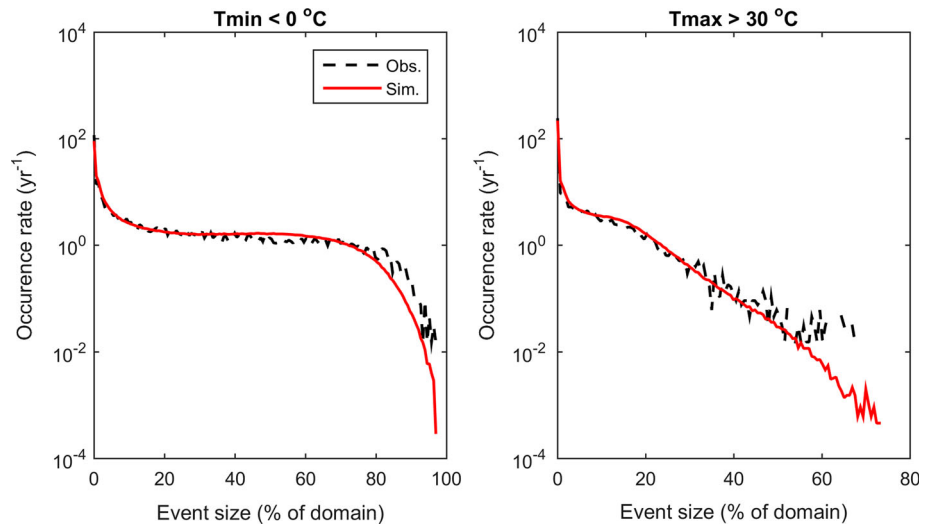
Fig. 11 *Top left* shows modified climate extreme index (CEI) return values against return periods (RP). Return values and periods of indices contributing to CEI are shown in remaining plots

The problems described above may also have some impact on the the hot and cold spell cumulative probability densities shown in Fig. 9. Although these are reasonably well reproduced by the model there is a small systematic bias towards spells of longer duration. However this is an improvement over the spell lengths exhibited by Caraway et al. (2014) whose model under represents spell lengths.

Many applications require multi-annual contiguous simulations. However, the current formulation of the model does not explicitly replicate any inter-annual correlation or multi-annual trends which may be present in the observation data set. A possible extension to the present model would be to look for correlations between the principle component time series of the leading EOF modes and large scale exogenous controls, for example the North Atlantic Oscillation, El NioSouthern Oscillation or a climate change index. The exogenous controls could then be allowed to affect the contribution from the associated EOF modes to the simulated output.

We have presented here a case study using a gridded data product. This was chosen partly because it provided a flexible number of sites to cover our study region with almost no missing data points. We also expect that fitting models such as this to future climate scenario simulation data is an important potential application for the risk management industry and for governments. Therefore, demonstrating our model's performance on data of this type is valuable. We have tested the model performance on data at different resolutions (i.e. number of grid cells) over the same region and its performance is not sensitive to changes in this parameter. We have also tested the model using direct station data and the results are similar (see supplementary material).

Finally, it is worth noting that IMAGE is scalable in that the number of variables included could be increased (e.g. pressure, humidity) or the number of grid points increased to suit fitting to a higher resolution data set. In the present case study, the biggest computational effort is spent on estimating the AR parameters for each month of available data but as the number of variables increases the computational bottleneck would likely become the eigenvector decomposition for which there are many available documented techniques.

6 Conclusions

We have presented IMAGE, a novel multi-site multivariate stochastic weather generator. We have demonstrated IMAGE's ability to accurately reproduce climatology at multiple sites across a large domain, as well as the spatial and temporal correlation of weather variables. Importantly, IMAGE was able to accurately generate the frequency and magnitude of both univariate and multivariate extreme events over multiple sites and over extended time periods. These events include heat waves and cold spells, droughts and excess rainfall, which have large social and economic impact. To our knowledge, this is the first time a stochastic weather generator has been demonstrated to produce such events.

References

Allard D, Bourotte M (2014) Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. Stoch Environ Res Risk Assess 29(2):453–462. doi:10.1007/s00477-014-0913-4

Apipattanavis S, Podestá G, Rajagopalan B, Katz RW, (2007) A semiparametric multivariate and multisite weather generator. Water Resour Res 43(11). doi:10.1029/2006WR005714

Baigorria GA, Jones JW (2010) GiST: a stochastic model for generating spatially and temporally correlated daily rainfall data. J Clim 23(22):5990–6008. doi:10.1175/2010JCLI3537.1

Baxevani A, Lennartsson J (2015) A spatiotemporal precipitation generator based on a censored latent Gaussian field. Water Resour Res 51:4338–4358. doi:10.1002/2014WR016259 2014WR016527

Bogner K, Pappenberger F, Cloke HL (2012) Technical note: the normal quantile transformation and its application in a flood forecasting system. Hydrol Earth Syst Sci 16(4):1085–1094. doi:10.5194/hess-16-1085-2012

Camberlin P, Gitau W, Oettli P, Ogallo L, Bois B (2014) Spatial interpolation of daily rainfall stochastic generation parameters over east Africa. Clim Res 59(1):39–60. doi:10.3354/cr01198

Camera C, Bruggeman A, Hadjinicolaou P, Michaelides S, Lange MA (2016) Evaluation of a spatial rainfall generator for generating high resolution precipitation projections over orographically complex terrain. Stoch Environ Res Risk Assess 31(3):1–17. doi:10.1007/s00477-016-1239-1

Caraway NM, McCreight JL, Rajagopalan B (2014) Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. J Hydrol 508:197–213. doi:10.1016/j.jhydrol.2013.10.054

Chen J, Brissette FP, Leconte R (2010) A daily stochastic weather generator for preserving low-frequency of climate variability. J Hydrol 388(3–4):480–490. doi:10.1016/j.jhydrol.2010.05.032

Cowden JR, Watkins DW, Mihelcic JR (2008) Stochastic rainfall modeling in West Africa: Parsimonious approaches for domestic rainwater harvesting assessment. J Hydrol 361(1–2):64–77. doi:10.1016/j.jhydrol.2008.07.025

Fyfe JC, Boer GJ, Flato GM (1999) The Arctic and Antarctic Oscillations and their projected changes under global warming. Geophys Res Lett 26(11):1601–1604. doi:10.1029/1999GL900317

Gleason KL, Lawrimore JH, Levinson DH, Karl TR, Karoly DJ (2008) A revised U.S. climate extremes index. J Clim 21(10):2124–2137. doi:10.1175/2007JCLI1883.1

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res Atmos 113(20). doi:10.1029/2008JD010201

Katz RW, Parlange MB (1998) Overdispersion phenomenon in stochastic modeling of precipitation. J Clim 11(4):591–601. doi:10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2

Kim JW, Kim KY, Kim MK, Cho CH, Lee Y, Lee J (2013) Statistical multisite simulations of summertime precipitation over South Korea and its future change based on observational data. Asia-Pac J Atmos Sci 49(5):687–702. doi:10.1007/s13143-013-0061-7

Kim KY, Wu Q (1999) A comparison study of EOF techniques: analysis of nonstationary data with periodic statistics. J Clim 12(1):185–199. doi:10.1175/1520-0442-12.1.185

Krzysztofowicz R (1997) Transformation and normalization of variates with specified distributions. J Hydrol 197(1–4):286–292. doi:10.1016/S0022-1694(96)03276-3

Leander R, Buishand TA (2009) A daily weather generator based on a two-stage resampling algorithm. J Hydrol 374(3–4):185–195. doi:10.1016/j.jhydrol.2009.06.010

Mehrotra R, Li J, Westra S, Sharma A (2015) A programming tool to generate multi-site daily rainfall using a two-stage semi parametric model. Environ Modell Softw 63:230–239. doi:10.1016/j.envsoft.2014.10.016

Rajagopalan B, Lall U (1999) A k-nearest-neighbor simulator for daily precipitation and other weather variables. Water Resour Res 35(10):3089–3101. doi:10.1029/1999WR900028

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17(1):182–190. doi:10.1029/WR017i001p00182

Russo S, Sillmann J, Fischer EM (2015) Top ten European heatwaves since 1950 and their occurrence in the coming decades. Environ Res Lett 10(12):124003. doi:10.1088/1748-9326/10/12/124003

Serinaldi F (2009) A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. J Geophys Res Atmos 114(10). doi:10.1029/2008JD011258

Srivastav RK, Simonovic SP (2015) Multi-site, multivariate weather generator using maximum entropy bootstrap. Clim Dyn 44(11–12):3431–3448. doi:10.1007/s00382-014-2157-x

Steinschneider S, Brown C (2013) A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments. Water Resour Res 49(11):7205–7220. doi:10.1002/wrcr.20528

Supit I, van Diepen CA, de Wit AJW, Wolf J, Kabat P, Baruth B, Ludwig F (2012) Assessing climate change effects on European crop yields using the Crop Growth Monitoring System and a weather generator. Agric For Meteorol 164: 96–111. doi:10.1016/j.agrformet.2012.05.005

Te Linde AH, Aerts JCJH, Bakker AMR, Kwadijk JCJ (2010) Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data. Water Resour Res 46(3). doi:10.1029/2009WR007707

Verdin A, Rajagopalan B, Kleiber W, Katz RW (2015) Coupled stochastic weather generation using spatial and generalized linear models. Stoch Environ Res Risk Assess 29(2):347–356. doi:10.1007/s00477-014-0911-6

Wang Q, Nathan R (2007) A method for coupling daily and monthly time scales in stochastic generation of rainfall series. J Hydrol 346(3):122–130. doi:10.1016/j.jhydrol.2007.09.003

Wilks DS (1999a) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93(3):153–169. doi:10.1016/S0168-1923(98)00125-7

Wilks DS (1999b) Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. Agric For Meteorol 96(1–3):85–101. doi:10.1016/S0168-1923(99)00037-4

Youngman BD, Stephenson DB (2016) A geostatistical extreme-value framework for fast simulation of natural hazard events. ProcRSocA 472. doi:10.1098/rspa.2015.0855

Zhang Y, Wallace JM, Battisti DS (1997) ENSO-like interdecadal variability : 1900–93. J Clim 10:1004–1020. doi:10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2