A globally calibrated scheme for generating daily meteorology from monthly statistics: Global-WGEN (GWGEN) v1.0

Philipp S. Sommer[1] and Jed O. Kaplan[1]

[1]Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, 1015 Lausanne, Switzerland

Correspondence to: Philipp S. Sommer (philipp.sommer@unil.ch)

Abstract. While a wide range of earth system processes occur at daily and even sub-daily timescales, many global vegetation and other terrestrial dynamics models historically used monthly meteorological forcing, both to reduce computational demand and because global datasets were lacking. Recently, dynamic land surface modeling has moved towards resolving daily and subdaily processes, and global datasets containg daily or sub-daily meteorology have become available. These meteorological datasets, however, cover only the instrumental era of the last ca. 120 years at best, are subject to considerable uncertainty, and represent extremely large data files with associated computational costs of data input/output and file transfer. For periods before the recent past or into the future, global meteorological forcing can be provided by climate model output, but the quality of these data at high temporal resolution is low, particularly for daily precipitation frequency and amount. Here we present GWGEN, a globally applicable statistical weather generator for the temporal downscaling of monthly climatology to daily meteorology. Our weather generator is parameterized using a global meteorological database and simulates daily values of five common variables: minimum and maximum temperature, precipitation, cloud cover, and windspeed. GWGEN is lightweight, modular, and requires a minimal set of monthly mean variables as input. The weather generator may be used in a range of applications, for example, in global vegetation, crop, soil erosion, or hydrological models. While GWGEN does not perform spatially autocorrelated multi-point downscaling of daily weather, this additional functionality could be easily implemented in future versions.

1 Introduction

The development of the first global vegetation models in the 1970's (e.g., Lieth, 1975) brought about the demand for meteorological forcing datasets with global extent and relatively high spatial resolution, e.g., 1°x1°. While a global weather station-based monthly climate dataset was available at this time (Walter and Lieth, 1967), limitations in computers and storage allowed only the simplest treatment of these data. The first global simulations of the net primary productivity of the terrestrial biosphere (Lieth, 1975), thus used rasterized polygons of annual meteorological variables that had been crudely interpolated from the station-based climatology. A decade later saw the development of better computers and more sophisticated global vegetation models (Prentice et al., 1992; Prentice, 1989) that recognized the need for forcing at a sub-annual timestep and development of these models was done in parallel with the first global, gridded high resolution (0.5°) monthly climatology (Leemans and Cramer, 1991). At the time, monthly meteorological data was the only feasible global data that could be pro-

duced, in terms of the raw station data available to feed the interpolation process, the processing time required to produce gridded maps, and the data storage and transfer capabilities of contemporary computer systems and networks. Global gridded monthly climate data thus became the standard for not only large-extent vegetation modeling (Haxeltine and Prentice, 1996; Haxeltine et al., 1996; Kaplan et al., 2003; Kucharik et al., 2000; Woodward et al., 1995), but also for a wide range of studies
5 on biodiversity and species distribution (e.g., Elith et al., 2006), vegetation trace gas emissions (e.g., Guenther et al., 1995), and even the geographic distribution of human diseases (e.g., Bhatt et al., 2013)

Over subsequent years, the global gridded monthly climate datasets were improved (New et al., 1999, 2002), developed with very high spatial resolution (Hijmans et al., 2005), and expanded beyond climatological mean climate to cover continuous timeseries over decades (Harris et al., 2014; Mitchell and Jones, 2005; New et al., 2000). The latter was an essential requirement
10 for forcing dynamic global vegetation models (DGVMs) (e.g., Sitch et al., 2003). However despite increasing quality, spatial resolution, and temporal extent in these datasets, the basic time step remained monthly, partly for legacy reasons — models had been developed in an earlier era subject to computational limitations and therefore used a monthly timestep for efficiency even if this was no longer strictly a constraint — and partly because of the challenge in developing a global, high-resolution climate dataset with a daily or shorter timestep still presented a major data management challenge.

15 On the other hand, there was increasing awareness that accurate simulation of many earth surface processes required representation of processes at a shorter-than-monthly timestep. Global simulation of surface hydrology (Gerten et al., 2004), crop growth (Bondeau et al., 2007), or biogeophysical processes (Krinner et al., 2005) needed sub-monthly forcing to produce reliable results. To address this need for better forcing data, two main approaches were taken: either monthly climate data were downscaled online using a stochastic weather generator (e.g., Pfeiffer et al., 2013), or a sub-daily, high-resolution, gridded
20 climate timeseries was generated directly by merging high-temporal-resolution reanalysis data (e.g., NCEP, 6h, 2.5°) with high-spatial-resolution monthly climate data (e.g., CRU, 0.5°). The latter process resulted in the CRUNCEP dataset (Viovy and Ciais, 2016; Wei et al., 2014), which, while global, is large even by modern standards (ca. 350 Gb), is not available at spatial resolution greater than 0.5°, and covers only the period 1901-2014.

Forcing data for global vegetation and other models with shorter than monthly resolution at higher spatial resolutions than
25 0.5°, or for any other period than the last ca. 120 years, e.g., for the future or the more distant past, may therefore only be available through downscaling techniques. One approach to overcome the limitations of currently available datasets could be to use GCM output directly, however, most GCM output currently available does not have greater than 0.5° spatial resolution, with the current generation of GCMs typically approaching ca. 1° × 1° degree. Furthermore, there is a general observation that daily meteorology produced by GCMs is not realistic, particularly for precipitation (Dai, 2006; Stephens et al., 2010; Sun
30 et al., 2006). An alternative approach is, therefore, to perform temporal downscaling on monthly meteorological data using a statistical weather generator.

Statistical weather generators were first developed primarily for crop and hydrological modeling at the field to catchment scale (Richardson, 1981; Woolhiser and Pegram, 1979; Woolhiser and Roldan, 1982). The weather generator was parameterized using daily meteorological observations at one or more weather stations close to the area of interest, although some attempts
35 were made to generalize the parameterization over larger, sub-continental regions (e.g., Wilks, 1999b, 1998; Woolhiser and

Roldán, 1986). Locally parameterized weather generators have been applied to a very wide range of studies (Wilks, 2010; Wilks and Wilby, 1999), and enhanced to include additional meteorological variables beyond the original precipitation, temperature, and solar radiation (e.g., Parlange and Katz, 2000). Applications of a weather generator at continental to global scales was still limited, however, because of the need to perform local parameterization.

5 The need to simulate daily meteorology in regions of the world with short, unreliable, or unavailable daily meteorological timeseries brought about the realization that certain features of weather generator parameterization might be generalized across a range of climates (Geng and Auburn, 1987; Geng et al., 1986). This ultimately led to the development of globally applicable weather generators (Friend, 1998), and their incorporation in DGVMs (Bondeau et al., 2007; Gerten et al., 2004; Pfeiffer et al., 2013). The original global parameterization (Geng et al., 1986) of these weather generators was, however, limited to seven

10 weather stations, mostly in the temperate latitudes. Friend (1998) does not publish the parameters used in his global weather generator, but we assume these were the same as the original Geng and Auburn (1987) and Geng et al. (1986) models. Given the availability of 1) large datasets of daily meteorology, and 2) computers powerful enough to process these data, we therefore decided that it would be valuable to revisit these parameterizations, perform a systematic and quantitative evaluation of the resulting downscaled meteorology, and potentially improve our ability to perform monthly-to-daily downscaling of common

15 meteorological variables with a single, globally applicable parameterization.

  In the following sections we describe Global-WGEN (GWGEN), a weather generator parameterized using more than 50 million daily weather observations from all continents and latitudes. We demonstrate how updated schemes for simulating precipitation occurrence and amount, and for correcting bias in temperature and wind speed, further improve the quality of the model simulations. We perform an extensive model evaluation and parameter uncertainty analysis in order to settle on a parameter

20 set that provides the most accurate, globally applicable results. We comment on the limitations of the model and priorities for future research. GWGEN is an open-source, stand-alone model that may be incorporated into any number of models designed to work at global scale, including, e.g., vegetation, hydrology, climatology, and animal distribution models.

2 Model development

GWGEN is based on the WGEN weather generator (Richardson, 1981), using the method of defining the model parameters

25 based on monthly summaries described by Geng et al. (1986) and Geng and Auburn (1987). GWGEN diverges from the original WGEN by using a second-order Markov chain to simulate precipitation occurrence (Wilks, 1999a), and a hybrid Gamma-GP distribution (Furrer and Katz, 2008; Neykov et al., 2014) to estimate precipitation amount. Temperature, cloud cover, and wind speed are calculated following (Richardson, 1981), using cross correlation and depending on the wet/dry-state of the day. We further add a quantile-based bias correction for wind speed and minimum temperature, which improves the simulation results

30 significantly.

  In the following subsections, we first describe the global weather station database used to develop and evaluate the model, then describe the underlying relationships that we use to define GWGEN's parameters. A complete description of the final model is in section 3.
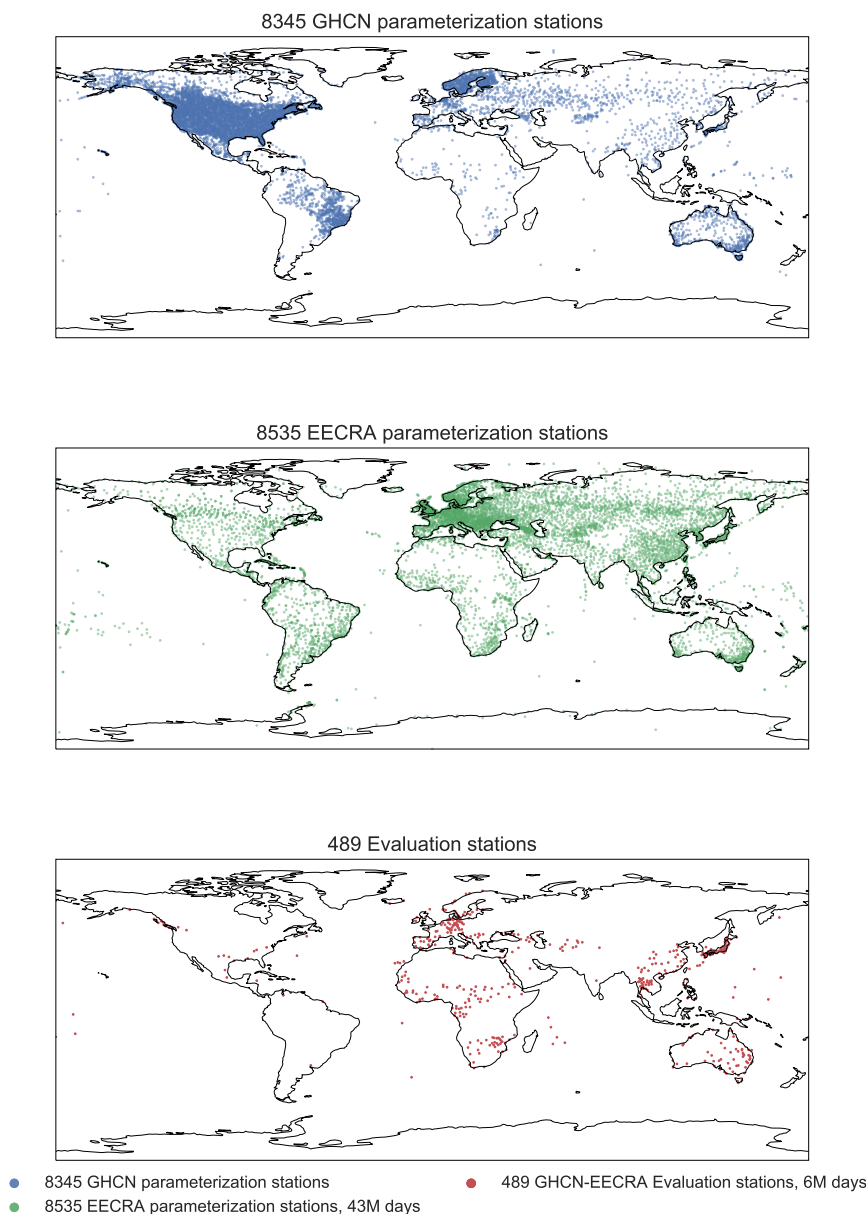
Figure 1. Weather stations used for parameterization and evaluation of the weather generator. The uppermost panel shows the locations of the stations used for parameterizing precipitation and temperature, the middle panel shows the stations for cloud cover and windspeed, as well as for calculating the cross correlations between temperature, cloudiness, and windspeed. The lower plot shows the location of the stations used to evaluate the model, which were excluded from the parameterization stations.

2.1 Development of a global weather station database

To parameterize GWGEN, we assembled a global dataset of daily meteorological observations. Precipitation and minimum and maximum daily temperature come from the daily Global Historical Climatology Network (GHCN) database (Menne et al., 2012b, a). The GHCN-daily consists of observations collected at ca. 100'000 weather stations on all continents and

5 many oceanic islands. As the GHCN-Daily stations are highly concentrated in some parts of the world, particularly in the conterminous United States, we selected stations for our study using a geographic anti-aliasing filter to avoid an especially strong geographic bias in the generation of the model parameters. Dividing the world up into a 0.5° grid, we selected the single station with the longest record in each cell, if one was present. While the GHCN-Daily units for precipitation have a nominal precision of 0.1 mm, several of the stations in the United States reported precipitation in fractions of an inch, which were later

10 converted to mm. To ensure uniform precision across all of our calibration stations — this was particularly important when generating the probability density functions for precipitation amount — we selected only those GHCN-Daily stations where all precipitation amounts between 0.1 to 1.0 mm d^{-1} were reported in the record. This resulted in 8345 stations covering all continents, although the distribution is strongly heterogenous, with the majority of the stations in North America, despite our geographic filter (Figure 1, top panel). For cloud cover, windspeed, and to calculate cross-correlations between temperature,

15 cloud cover, and windspeed, we used the Extended Edited Cloud Report Archive (EECRA) database (Hahn and Warren, 1999). The geographic distribution of the 8535 EECRA stations we selected is different than the GHCN-Daily, with more stations in Europe (Figure 1, middle panel), but overall a relatively similar number of stations were used from both datasets. For the observations from both GHCN-Daily and EECRA, we made one additional filtering step, selecting only complete months, i.e., months with no days having missing observations, for further processing. In total, our database of daily meteorological

20 observations used in the model parameterization contains ca. 50 million individual records.

Finally, we reserved some weather station records for model evaluation that were not used for model parameterization. These were individual stations, or two stations separated by a maximum distance of 1 km, where all of the daily meteorological variables that GWGEN simulates ($P, T_{\min}, T_{\max}, c, w$) were recorded on the same dates. This selection resulted in a set of 489 stations representing ca. 6 million daily records, with observations on all continents, although the geographic distribution

25 is once again highly heterogenous, with a particularly high density of stations in Japan and Germany (Figure 1, bottom panel).

2.2 Parameterization

2.2.1 Precipitation occurrence

Following Geng et al. (1986), we expect to find a good relationship between the fraction of days in a month with measurable precipitation and the probability that any given day will be wet. Wilks (1999a) suggest that a second-order Markov chain is

30 more accurate than the first order chain used in the original WGEN and many subsequent models, so we adopt this approach in GWGEN. To parameterize the precipitation occurrence part of the model, we thus calculated transition probabilities for a wet day being followed by a wet day (p_{11}), for a wet day being followed by a dry day being followed by a wet day (p_{101}) and for two dry days being followed by a wet day (p_{001}). We perform this analysis on a station and month-wise basis, i.e., we first extract
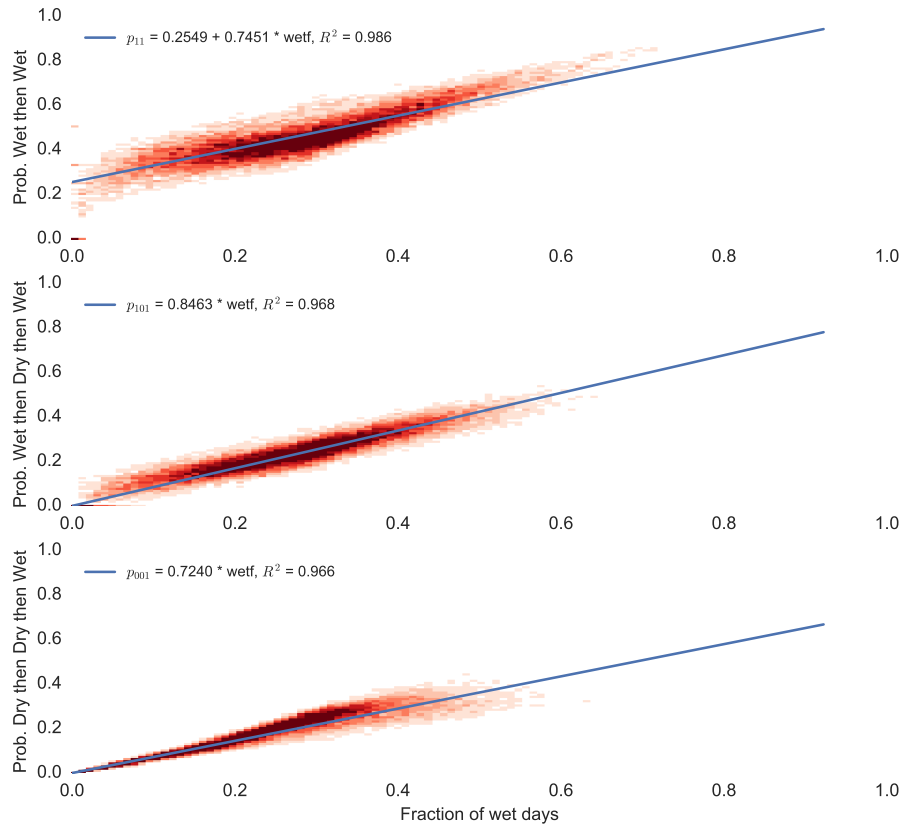
Figure 2. Transition probabilities vs. wet fraction. The red 2D histogram shows the density of the observations, and the blue lines are the linear regression line of the probability against the wet fraction. The fit for the p_{11} transition probability was forced to the point $(1, 1)$, the others were forced to $(0, 0)$.

each of the (complete) Januaries, Februaries, etc. for a given station, and then merge all of the Januaries (Februaries, Marches, etc...) for this station into a single series representing each month. Merging months over several years is particularly important for stations that have relatively little precipitation in a given month; for example, it could take several years of observations to observe a single (p_{101}) event. The final transition probabilities were then regressed against the fraction of days in the month

5 with precipitation, which show the characteristic linear relationship described by Geng et al. (1986) (Figure 2).

Because the transition probabilities (p_{001}) and (p_{101}) must be zero by definition when the fraction of wet days (f_{wet}) is zero, i.e., a completely dry month, we force the linear regression between these quantities to pass through the origin. Likewise, we require the regression line for (p_{11}) to equal 1 when f_{wet} is 1. The analysis results in the the following relationships:
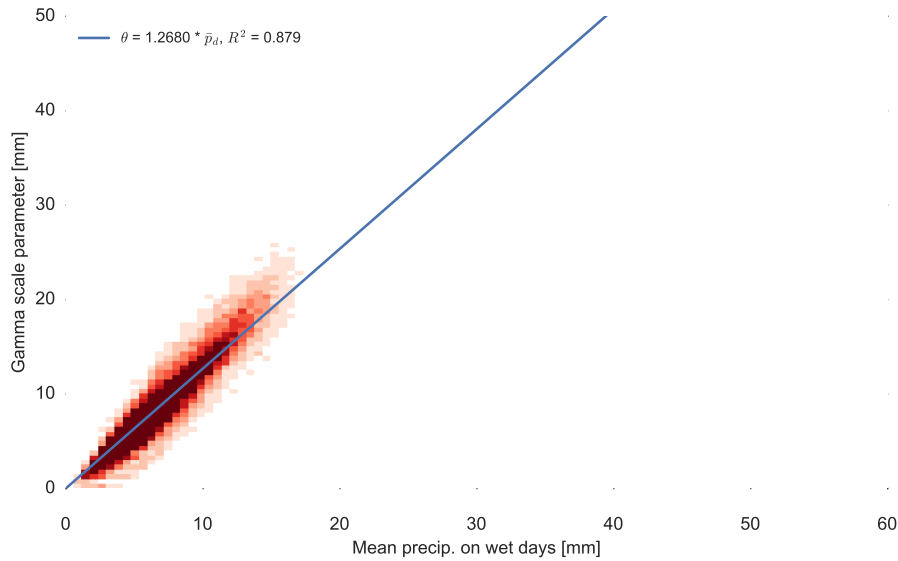
Figure 3. Mean precipitation - Gamma scale relationship. The blue line represents the best fit line of the mean precipitation on wet days to the estimated gamma scale parameter of the corresponding distribution. Each data point corresponds to one multi-year series of one month for one station.

$$p_{11} = 0.2549 + 0.7451 \cdot f_{\text{wet}} \tag{1}$$

$$p_{101} = 0.8463 \cdot f_{\text{wet}} \tag{2}$$

$$p_{001} = 0.7240 \cdot f_{\text{wet}}. \tag{3}$$

To determine if any given day will have precipitation, the appropriate probability density function is selected from equations (1)-(3) on the basis of the precipitation state of the previous day (or two). Comparing the calculated probability from the selected equation with a random number $u \in [0,1]$, a precipitation day is simulated if u is greater than its corresponding probability.

2.2.2 Precipitation amount

Following the original WGEN (Richardson, 1981), GWGEN disaggregates precipitation amount using a statistical distribution. A number of different probability density functions have been used to estimate precipitation amount in weather generators including, e.g., single exponential or mixed exponential, one or two parameter gamma, or Weibull distribution (Wilks and Wilby, 1999). The strong relationship between the gamma scale parameter and the mean precipitation on wet days noted by Geng et al. (1986) makes generation of precipitation amounts with only monthly input data feasible. The gamma distribution, however, shows poor performance in simulating high-precipitation events consistent with observations. Furrer and Katz (2008) and Neykov et al. (2014) suggest that a hybrid probability density function, based on both gamma and the generalized pareto

(GP) distribution, has superior accuracy in cimulating extreme precipitation events when compared to gamma alone. Because of its superior accuracy and ease of implementation, we therefore adopt the hybrid gamma-GP distribution for simulating precipitation amount in GWGEN.

The gamma distribution function is defined as

$$f(x) = \begin{cases} \frac{x^{\alpha-1}\exp^{-\frac{x}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{for } x = 0 \end{cases} \tag{4}$$

where $\alpha > 0$ is the shape, and $\theta > 0$ the scale parameter. The generalized pareto (GP) distribution is defined via

$$g(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} \tag{5}$$

with $\sigma > 0$ being the scale parameter and $\xi \in \mathbb{R}$ the shape parameter. μ is the location parameter.

Following Furrer and Katz (2008), we define the hybrid gamma-GP distribution as

$$h(x) = \begin{cases} f(x) & \text{for } x \leq \mu, \\ (1 - F(\mu))\,g(x) & \text{for } x > \mu \end{cases}, \tag{6}$$

where $F(\mu)$ describes the cumulative gamma distribution function at the threshold μ. In our weather generator however, we first draw a random number from the gamma distribution and, if we are above the threshold, we draw another random number from the GP distribution. Thus, the frequency of precipitation events larger than μ is determined by the gamma distribution, but the actual amount of precipitation simulated when above the threshold μ is determined by the GP distribution (Furrer and Katz, 2008).

To determine the parameters of the hybrid distribution for precipitation, we started with the simple strategy by Geng et al. (1986). As above when calculating the Markov chain parameters, we created multi-year series for each of the parameterization stations for each month and extracted the days with precipitation. If a series contained more than 100 entries, we fit a gamma distribution to it and estimated the α and θ parameters.

Following Geng et al. (1986), we then fit a regression line of the gamma scale parameter against the mean precipitation on wet days $\bar{p}_d$ (see figure 3) and found the relationship

$$\theta = 1.268\,\bar{p}_d. \tag{7}$$

As proposed by Geng et al. (1986), we use this relationship in our model to estimate the scale parameter of the distribution. The gamma shape parameter α is calculated dynamically in the weather generator via

$$\alpha = \frac{\bar{p}_d}{\theta}. \tag{8}$$

The GP scale parameter σ on the other hand is calculated during the simulation following Neykov et al. (2014) via

$$\sigma = \frac{1 - F(\mu)}{f(\mu)}. \tag{9}$$

The other parameters of the GP distribution are obtained through a sensitivity analysis described in section 2.5.
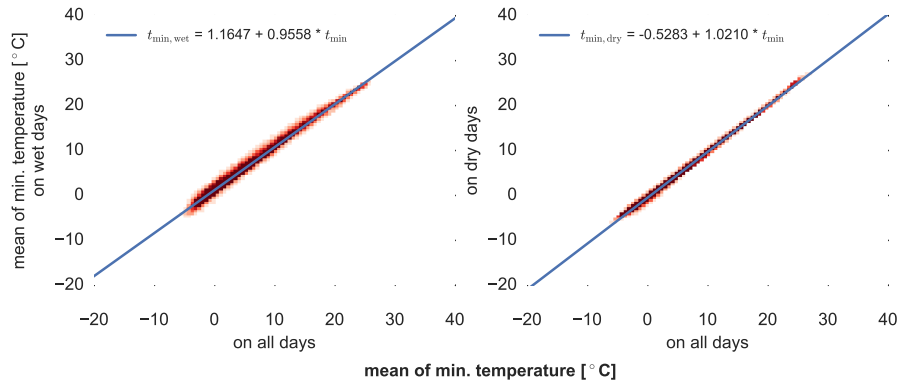
Figure 4. Correlation of minimum temperature on wet and dry days to the monthly mean. The y-axes show the mean minimum temperature on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.
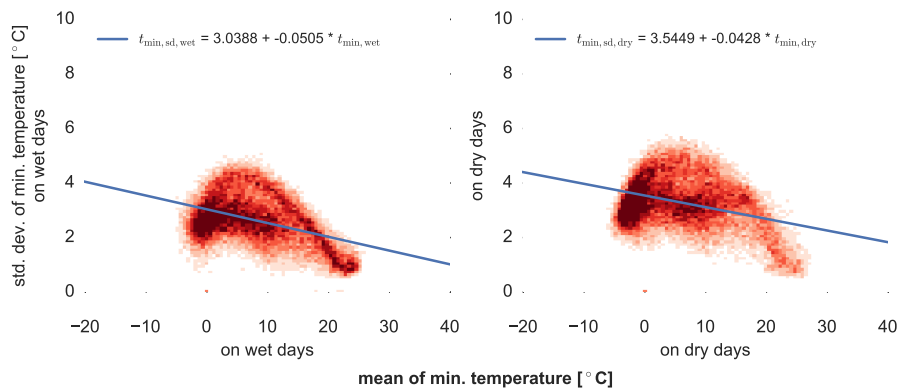
Figure 5. Correlation of standard deviation of the minimum temperature on wet and dry days to the monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

2.2.3 Temperature

Following the standard WGEN methodology (Richardson, 1981) and Geng et al. (1986), daily temperature is determined through 2 processes: First, the wet/dry state of the day, and second the cross correlation (subsubsection 2.2.6). For minimum temperature, we additionally perform a bias correction (subsection 2.4).

5   In the weather generator, we know from the Markov chain (subsubsection 2.2.1), whether the current simulated day is a wet or dry day. Based upon the simple linear relationships
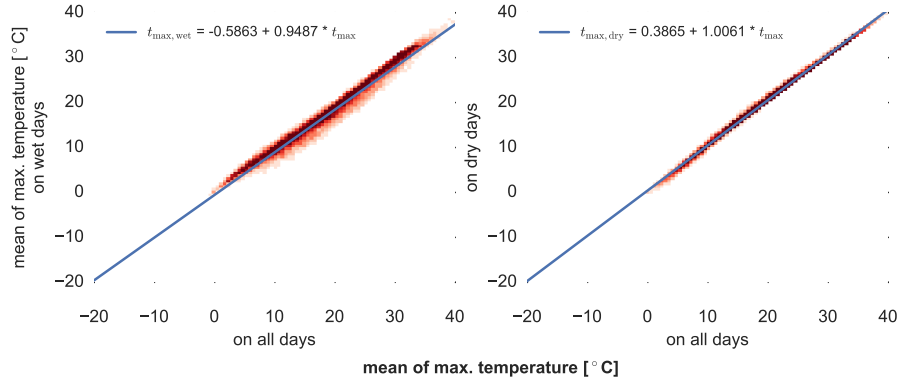
Figure 6. Correlation of maximum temperature on wet and dry days to the monthly mean. The y-axes show the mean maximum temperature on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.
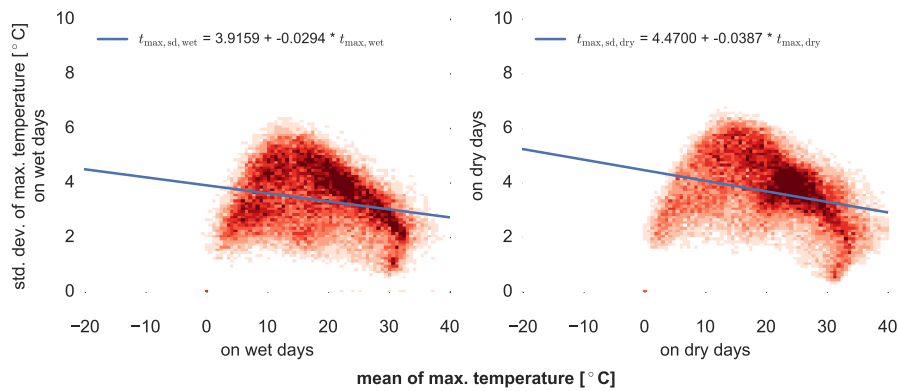
Figure 7. Correlation of standard deviation of the minimum temperature on wet and dry days to the monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

$$x_{\text{wet}} = a_{x,\text{wet}} + b_{x,\text{wet}} \cdot \bar{x}$$
$$x_{\text{dry}} = a_{x,\text{dry}} + b_{x,\text{dry}} \cdot \bar{x} \tag{10}$$

we adjust the monthly mean $\bar{x}$ of the variable $x \in \{T_{\min}, T_{\max}\}$. The same procedure adjustment is used for the standard deviation σ_x of variable $x \in \{T_{\min}, T_{\max}\}$:

Table 1. Fit results of temperature correlation for wet and dry days for figure 4 and 6. N is the total number of measurements used for the fit. The values of a and b correspond to the values in equation (10) and (11).

plot	variable	Intercept	R^2	Slope
6	$T_{\mathrm{max,dry}}$	0.3865	0.9907	1.0061
6	$T_{\mathrm{max,wet}}$	-0.5863	0.9539	0.9487
7	$T_{\mathrm{max,sd,dry}}$	4.4700	0.0722	-0.0387
7	$T_{\mathrm{max,sd,wet}}$	3.9159	0.0407	-0.0294
4	$T_{\mathrm{min,dry}}$	-0.5283	0.9940	1.0210
4	$T_{\mathrm{min,wet}}$	1.1647	0.9733	0.9558
5	$T_{\mathrm{min,sd,dry}}$	3.5449	0.1044	-0.0428
5	$T_{\mathrm{min,sd,wet}}$	3.0388	0.1386	-0.0505
11	$w_{\mathrm{sd,dry}}$	0	0.8959	0.5120
11	$w_{\mathrm{sd,wet}}$	0	0.8994	0.4404
10	w_{dry}	0	0.9933	0.9452
10	w_{wet}	0	0.9720	1.0929

$$\sigma_{x,\mathrm{wet}} = a_{\sigma,x,\mathrm{wet}} + b_{\sigma,x,\mathrm{wet}} \cdot x_{\mathrm{wet}}$$

$$\sigma_{x,\mathrm{dry}} = a_{\sigma,x,\mathrm{dry}} + b_{\sigma,x,\mathrm{dry}} \cdot x_{\mathrm{dry}}. \tag{11}$$

To estimate the values of the parameters a and b in the above equations, we follow the same procedure as for the parameters of the Markov chain (subsubsection 2.2.1). We extracted the complete months for $T_{\min}$ and $T_{\max}$ from the GHCN-Daily dataset and created a multi-year series for each month and station. We then regressed the mean on wet and dry days separated against the overall mean of each month (Figures 4 and 6). We further regressed the standard deviation on wet and dry days against its corresponding mean (Figures 5 and 7). Through this procedure, we estimate the parameters necessary for equations (10) and (11) (see table 1). As can be seen in figures 5 and 7 the linear model does not really satisfy the complex behaviour of the standard deviation. However, since this value is only used to add random noise (see section 3), we believe that the error introduced by the poor linear fit is negligible.

2.2.4 Cloud fraction

Monthly mean cloud fraction is disaggregated, as for temperature, using the standard WGEN procedure of adding statistical noise to a wet- or dry-day mean and accounting for cross-correlation among the different weather variables. For the parameterization of the cloud fraction equations, we used the EECRA dataset. The original dataset contains eight measurements per day of the total cloud cover in units of octas, i.e., values ranging from 0 (clear sky) to 8 (overcast). Hence, to calculate the daily cloud fraction, those values were averaged and divided by 8 to produce a daily mean.
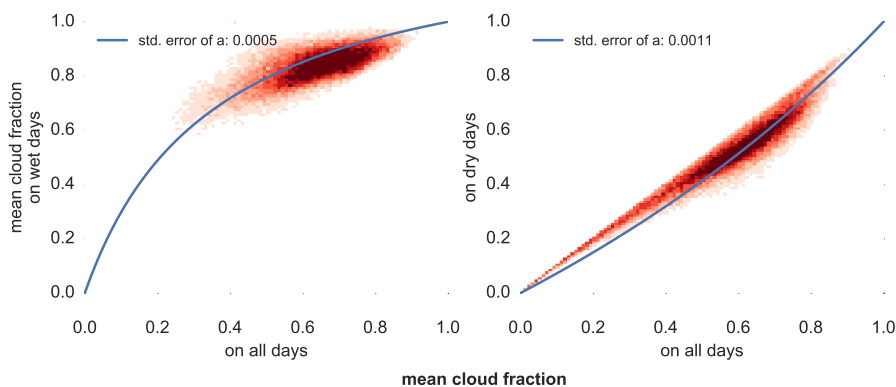
Figure 8. Correlation of cloud fraction on wet and dry days to the monthly mean.The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 2.
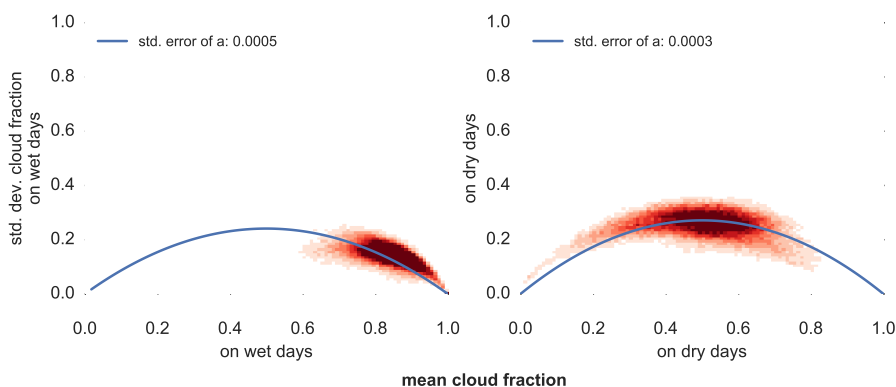
Figure 9. Correlation of standard deviation of the cloud fraction on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 2.

To adjust the monthly mean depending on the wet/dry state of the day, we could not use a simple linear relationship as we used for temperature because cloud fraction is bounded by a lower limit 0 and an upper limit of 1. Furthermore, we observed that cloud cover on wet days is usually greater or equal to the monthly mean cloud cover, whereas the cloud cover on dry days is usually less or equal to the monthly mean cloud cover. This results in a concave curve for the wet case and a convex curve for dry days. We therefore propose the following formulae for the regression linking cloud cover on wet or dry days to the overall mean:

Table 2. Fit results of cloud correlation for wet and dry days for figure 8

plot	variable	a	std. dev. of a
8	c_{dry}	0.4205	0.0011
8	c_{wet}	-0.7383	0.0005
9	$c_{\mathrm{sd,dry}}$	1.0417	0.0003
9	$c_{\mathrm{sd,wet}}$	0.9819	0.0005
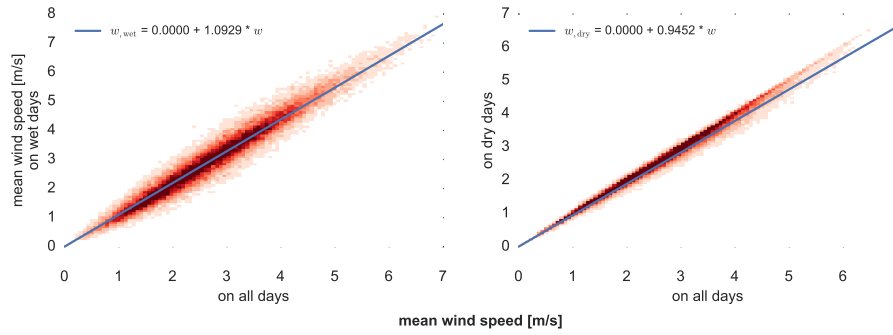
Figure 10. Correlation of wind speed on wet and dry days to the monthly mean. The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

$$c_{\mathrm{wet}} = \frac{-a_{\mathrm{c,wet}} - 1}{a_{\mathrm{c,wet}}^2 * c - a_{\mathrm{c,wet}}^2 - a_{\mathrm{c,wet}}} - \frac{1}{a_{\mathrm{c,wet}}}$$

$$c_{\mathrm{dry}} = \frac{-a_{\mathrm{c,dry}} - 1}{a_{\mathrm{c,dry}}^2 * c - a_{\mathrm{c,dry}}^2 - a_{\mathrm{c,dry}}} - \frac{1}{a_{\mathrm{c,dry}}} \tag{12}$$

with $a_{\mathrm{c,wet}} < 0$ and $a_{\mathrm{c,dry}} > 0$.

The standard deviation of cloud cover fraction becomes 0 when the mean monthly cloud fraction reaches both the minimum or maximum limits of 0 and 1. Hence, for $c_{\mathrm{sd,dry}}$ and $c_{\mathrm{sd,wet}}$ we have an concave parabola with the formula

$$\sigma_{\mathrm{c,wet}} = a_{\mathrm{c,wet}}^2 \cdot c_{\mathrm{wet}} \cdot (1 - c_{\mathrm{wet}})$$

$$\sigma_{\mathrm{c,dry}} = a_{\mathrm{c,dry}}^2 \cdot c_{\mathrm{dry}} \cdot (1 - c_{\mathrm{dry}}) \tag{13}$$

with $a_{\mathrm{c,wet}}, a_{\mathrm{c,dry}} \geq 0$. Results of the fits can be seen in figure 8, 9 and the parameters in table 2.

2.2.5 Wind speed

The parameterization of wind speed is based upon the same equations (10) and (11) as temperature and shown in the figures 10 and 11. Following Parlange and Katz (2000), and because of the non-uniform distribution of windspeeds in time, we apply
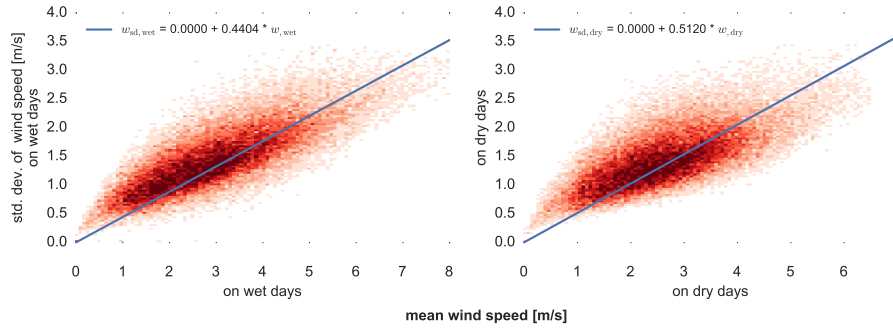
Figure 11. Correlation of standard deviation of the wind speed on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

a square root transformation to the monthly mean windspeed values before performing the regression fitting (see line 19 in the model algorithm 1).

2.2.6 Cross correlation

Following Richardson (1981) we use cross correlation to add additional residual noise to the simulated meteorological vari-
5 ables, which provides more realism in the daily weather result. As above, windspeed is subject to a square-root transformation before calculating the cross correlations.

Using the EECRA database (see figure 1) we calculate

$$
A = \begin{pmatrix}
0.913 & 0.033 & -0.021 & 0.001 \\
0.489 & 0.137 & -0.073 & -0.046 \\
-0.002 & -0.046 & 0.592 & 0.026 \\
0.011 & -0.044 & -0.019 & 0.667
\end{pmatrix}
\qquad
B = \begin{pmatrix}
0.362 & 0. & 0. & 0. \\
0.114 & 0.803 & 0. & 0. \\
0.145 & -0.061 & 0.783 & 0. \\
0.081 & -0.016 & 0.066 & 0.737
\end{pmatrix}
\tag{14}
$$

where the columns and rows correspond to min. and max. temperature, cloud fraction and square root of wind speed.
10 The above matrices (14) where calculated via

$$
A = M_1 M_0^{-1} \qquad BB^T = M_0 - M_1 M_0^{-1} M_1^T
\tag{15}
$$

from the lag-0 and lag-1 covariance matrices M_0 and M_1 with

$$
M_0 = \begin{pmatrix}
1. & 0.572 & 0.025 & 0.032 \\
0.572 & 1. & -0.101 & -0.045 \\
0.025 & -0.101 & 1. & 0.127 \\
0.032 & -0.045 & 0.127 & 1.
\end{pmatrix}
\qquad
M_1 = \begin{pmatrix}
0.932 & 0.557 & -0.001 & 0.025 \\
0.564 & 0.426 & -0.08 & -0.046 \\
-0.012 & -0.108 & 0.6 & 0.104 \\
0.006 & -0.066 & 0.071 & 0.667
\end{pmatrix}.
\tag{16}
$$
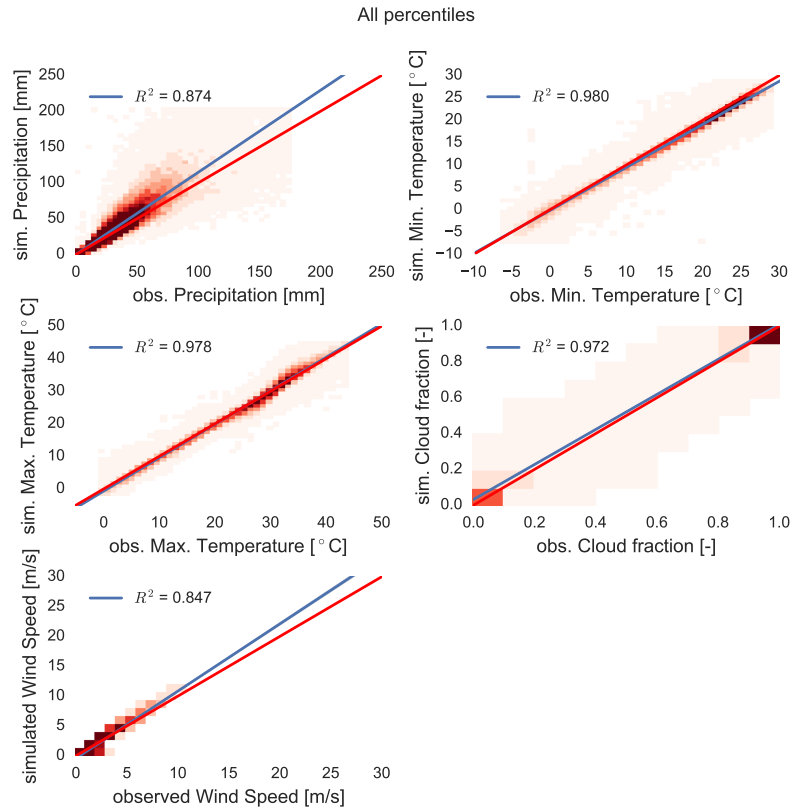
Figure 12. QQ-plots for all variables with all quantiles (1, 5, 10, 25, 50, 75, 90, 95 and 99) for $\mu = \ldots$ mm, $\xi = \ldots$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval. Plots for wind speed and minimum temperature used the bias correction from subsection 2.4.

2.3 Evaluation

To evaluate GWGEN, we started with the daily meteorology at the evaluation stations described above and calculated monthly summaries. We used these monthly data to drive the model and simulate daily meteorology. Because we cannot expect the weather generator to reproduce the weather exactly as observed, for example the number of rainy days in a month may be the same as observed but they may not occur in precisely the same order, our evaluation is restricted to comparing the statistical properties of the input observed versus the output simulated daily meteorology.

Figure 12 shows the comparison of simulated versus observed values for each of the five meteorological variables handled by GWGEN. For temperature, wind, and cloud fraction, the model does an excellent job of downscaling monthly input to daily resolution[1]. The comparison between precipitation amounts looks good when considering all of the data, however a closer look

---

[1]Note that the plots for wind speed and minimum temperature were bias corrected using the approach in subsection 2.4.
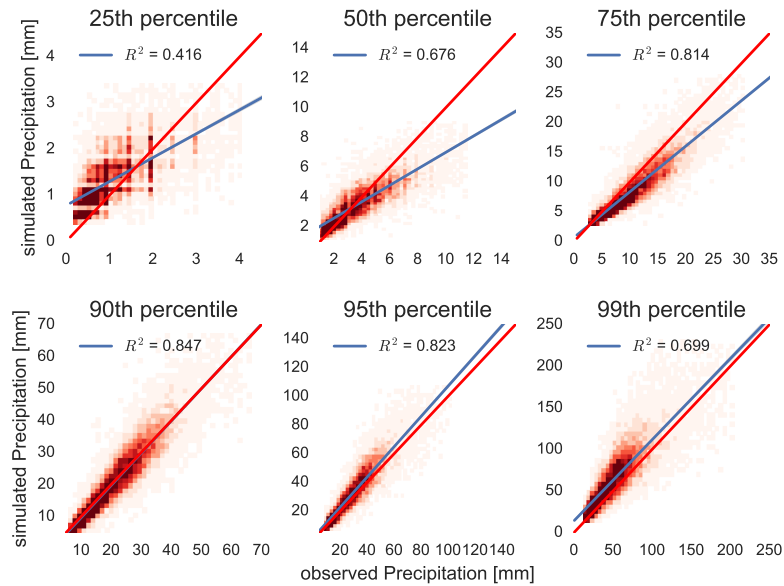
Figure 13. QQ-plot for different quantiles for precipitation for $\mu = 15\,\text{mm}$, $\xi = 0.08303$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval.

into the results (Fig. 13) shows that while the higher precipitation percentiles are well captured using the hybrid Gamma-GP distribution, the lower percentiles show somewhat worse results. This observation of poor performance for very low values also holds true for wind speed (not shown here). The lower values of the two variables, however, are very close to the precision of the observation ($0.1\,\text{mm}$ for precipitation and $0.1\,\text{m/s}$ for wind speed). Very small precipitation amounts and low windspeeds are also ecologically not as important as the higher percentiles. We therefore consider the results of the evaluation largely acceptable.

In table 3 we also compare the simulated versus the observed frequencies. For very light rain (<=1mm), light rain (1-10mm), heavy rain (10-20mm) and very heavy rain (>20mm). As we can see, our model underestimates the occurence of very light rain events (21.6% instead of 27.0%) and overestimates the light rain events (55% instead of 47%) but generally performs much better than GCMs (Dai, 2006; Sun et al., 2006), especially when it comes to the heavy rain events.

2.4 Bias correction

After evaluating the results of GWGEN for wind speed and minimum temperature for the different quantiles (see previous subsection 2.3) we found a strong, systematic bias between the simulated and the observed values. This observation led us to adopt a further measure to improve the quality of the model output, by implementing a quantile-based bias correction.

For minimum temperature, we use an empirical distribution correction approach (quantile-mapping) (Lafon et al., 2012) with the third-order polynomial shown in figure 14 to transform the simulated minimum temperature data based upon the modeled

Table 3. Simulated and observed precipitation frequencies for certain ranges. The frequency is defined as the number of precipitation occurences in the specified range, divided by the total number of precipitation occurences.

Precipitation [mm]	Simulated	Observed
(0, 1]	0.216	0.270
(1, 10]	0.550	0.470
(10, 20]	0.115	0.133
(20, ∞)	0.120	0.127
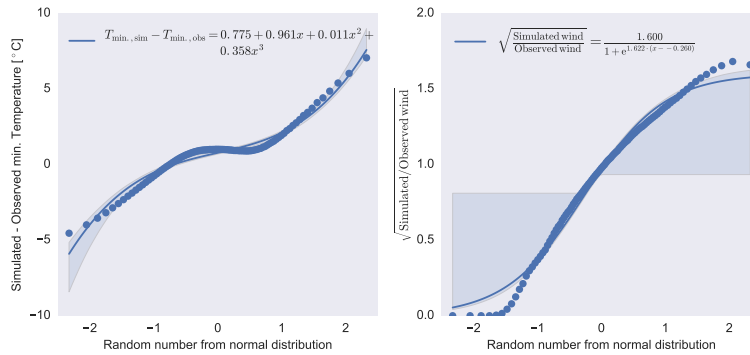
Figure 14. Basis for the minimum temperature (left) and wind bias correction (right). For the left plot (min. Temperature), each data point corresponds to the difference of a simulated percentile to the observed percentile. For the right plot (wind speed), each data point corresponds to the fraction of simulated to the observed square root of the wind speed for a given percentile. The random number on the x-axis represents the residual value from a normal distribution centered at 0 with standard deviation of unity, as it is used in the cross correlation approach (Richardson, 1981).

quantile. The coefficients of this polynomial are calculated a posteriori based upon the bias from simulated to the observed quantile for all percentiles between 1 and 99. The resulting function $f_{T_{\min}}(u)$ is then used in the weather generator to correct the minimum temperature $T_{\min}$ via

$$T_{\min} = T_{\min,biased} - f_{T_{\min}}(u) \qquad (17)$$

where $u \in \mathbb{R}$ corresponds to the number from the normal distribution drawn for the cross correlation (section 2.2.6 and Richardson (1981)).

Similarly to $T_{\min}$, we found that the residual of simulated to observed windspeed at a given percentile strongly followed a logistic function or third order polynomial (Figure 14). Hence, as for minimum temperature, we use a quantile mapping approach and a posteriori correlate the square root of the fraction of simulated to observed wind speed with the percentile in the normal distribution. The resulting function $f(u)$, is then used inside the weather generator to correct the the wind speed w

via

$$w = w_{\text{biased}} \; f(u). \tag{18}$$

.

2.5 Sensitivity analysis

The Generalized-Pareto part of the hybrid Gamma-GP distribution we used to simulate precipitation amount has two parameters, the GP shape and the threshold parameter. Unlike the gamma parameters, we were unable to relate these GP parameters to any of the monthly summary data we use as input to GWGEN. Hence, we decided to set fixed values for these parameters, and determined them through a sensitivity analysis.

To select the "best" values of the GP parameters, we compared simulated with observed precipitation amounts, running GWGEN with a wide range of realistic parameter values. To quantitatively assess the model performance, we used two metrics: 1) direct comparison of the quantiles (see previous section), and 2) a Kolmogorov-Smirnov (KS) test that evaluates whether two data samples come from significantly different distributions. Our criteria were

1. The R^2 correlation coefficient between simulated and observed quantiles

2. The fraction $\frac{\text{simulated precipitation}}{\text{observed precipitation}}$ from the slopes in figure zz13 and it's deviation from unity

3. the fraction of simulated (station specific) years that are significantly different (KS test) from the observation

4. The mean of the above values

We tried two different approaches to select the gamma-GP crossover threshold: first we tried a fixed crossover point, second we used a quantile-based crossover point. For the latter, the model chooses to use the GP distribution if the quantile of the random number drawn from the gamma distribution is above a certain quantile threshold. This introduces a flexible crossover point in our hybrid distribution which, however, did not improve the results significantly. We therefore show here only the results using the fixed crossover point.

The values of the crossover point for our sensitivity analysis were 2, 2.5, 3, 4 and from 5 to 100 in steps of 5. Furthermore we varied the GP shape parameter from 0 to 3 in steps of 0.1 (810 experiments in total). The results of this sensitivity analysis are shown in the supplementary material, figure 15.

In general we found that the three criteria 1, 2 and 3 could not be optimized all together at the same time. The R^2 is best for high thresholds and low GP shape parameters, the slope is best for intermediate threshold and a high GP shape and the KS statistic is best for low threshold and intermediate GP shape parameters.

However, R^2 did not vary that much (from 0.68 to 0.74) and from a visual evaluation of the corresponding quantile plots we saw that the higher quantiles (>90) were much better represented for a better KS result. Hence we chose to follow the KS test criteria, which is also the strictest of our evaluation methods but again compared the different quantile plots to get good results

for the higher quantiles. Finally, we chose a threshold of $5\,\mathrm{mm}$ and a GP shape parameter of 1.5. For this setting, 83.6% of the simulated years do not show a significant difference compared to the observation, the mean R^2 of the plots in figure 13 is 0.70 and the mean deviation of the slope from unity is 0.18 and for the upper quantiles (90 to 100), 0.1.

Nevertheless, in total the results seem to be fairly independent of the two parameters since even the amount of years without significant differences vary from 70% to only 86%. It is however better than the gamma distribution alone with only 76.2% of station years not differing significantly and a slope deviation from unity for the upper quantiles of 0.26.

3   Description of the final model

The parameterization described in section 2.2 is incorporated into a stand-alone model. It requires as input total monthly precipitation, the number of days in the month with measurable precipitation (i.e., wet days), monthly mean daily minimum and maximum temperature, mean cloud fraction, and wind speed. The model outputs are the same variables with daily resolution. This section summarizes the basic workflow in the model which is also shown schematically in algorithm 1.

The first approximation of the daily variables comes from smoothing the monthly time series using a mean-preserving algorithm (Rymes and Myers, 2001).

For precipitation we then first use the markov chain approach from section 2.2.1 to decide the wet/dry state of the day. If the day is a wet day, we calculate the gamma parameters using the equations (7) and (8). The resulting distribution allows us to draw a random number, the precipitation amount of the currently simulated day. If we are above the threshold μ, we draw a second random number from the GP distribution parameterized via equation (9) and the chosen GP shape.

The next step modifies the means of temperature, wind speed and cloud fraction depending on the wet/dry state of the day (lines 10 and 14 in algorithm 1). After that, we use the cross-correlation approach described in Richardson (1981) and lines 17 - 19 and calculate the daily values of these variables. Finally we use the quantile-based bias correction described in section 2.4 to correct the simulated wind speed and minimum temperature.

We restrict the weather generator to reproduce the exact number of wet days (±1) as the input and to be within a 5% range of the total monthly precipitation (with a maximum allowed deviation of $0.5\,\mathrm{mm}$). If the program cannot produce these results, the procedure described above is repeated (see line 3).

4   Discussion

GWGEN successfully downscales monthly to daily meteorology, for any point on the globe, in any climate, in any season, and in any time in recent earth history and into the near future (e.g., next century). It extends the original Richardson-type weather generators to simulate wind speed along with precipitation, temperature, and cloud cover. The model requires only monthly values of the meteorological variables to be downscaled, and does not rely on any other spatial information, e.g., whether or not the location is in the tropics. In general, the results of our downscaled meteorology are excellent, with all simulated variables showing both very high correlation and limited bias when compared to observations. We improved the simulation of

---

Algorithm 1 Basic workflow of GWGEN

Require: monthly precipitation $P_\text{in}\,[\text{mm}]$, cloud cover fraction c_in, minimum ($T_{\min,\text{in}}\,[^\circ\text{C}]$) and maximum ($T_{\max,\text{in}}\,[^\circ\text{C}]$) temperature, wind speed $w_\text{in}\,[\text{m/s}]$, number of wet days n_in

Output: daily $P_i\,[\text{mm/d}], c_i, T_i\,[^\circ\text{C}], w_i\,[\text{m/s}]$ and the wet/dry state $s_i \in \{0,1\}$

1: **for** month m in $input$ **do**

2:     smooth the monthly data using Rymes and Myers (2001)

3:     **while** $\left| \sum_{d_i \in m} P_i - P_\text{in} \right| > \min\left(5\% \cdot P_\text{in}, 0.5mm\right)$ or $|n_\text{sim} - n_\text{in}| > 2$ **do**

4:        **for** day d_i in m **do**

5:           Calculate p_{11}, p_{101}, p_{001} after equations (1) - (3) using n {Precipitation occurence after Wilks (1999a)}

6:           Use the Markov chain to determine whether d_i is wet ($s_i = 1$) or dry ($s_i = 0$)

7:           **if** $s_i = 1$ **then**

8:               Calculate θ, α and σ via eq. (7)-(9) {Precipitation amount after Neykov et al. (2014)}

9:               Draw a random number P_i from the Gamma-GP distribution, eq. (6)

10:           Set $T_{\min,i} = T_{\min,\text{wet}}, T_{\max,i} = T_{\max,\text{wet}}, c_i = c_\text{wet}, w_i = w_\text{wet}$ from eq. (10) and (12) and tables 1, 2

11:           Set $\sigma_{T_{\min},i} = \sigma_{T_{\min},\text{wet}}, \sigma_{T_{\max},i} = \sigma_{T_{\max},\text{wet}}, \sigma_{w,i} = \sigma_{w,\text{wet}}, \sigma_{c,i} = \sigma_{c,\text{wet}}$ from eq. (11) and (13) and tables 1, 2

12:           **else**

13:               Set $P_i = 0\,\text{mm/d}$

14:           Set $T_{\min,i} = T_{\min,\text{dry}}, T_{\max,i} = T_{\max,\text{dry}}, c_i = c_\text{dry}, w_i = w_\text{dry}$ from eq. (10) and (12) and tables 1, 2

15:           Set $\sigma_{T_{\min},i} = \sigma_{T_{\min},\text{dry}}, \sigma_{T_{\max},i} = \sigma_{T_{\max},\text{dry}}, \sigma_{w,i} = \sigma_{w,\text{dry}}, \sigma_{c,i} = \sigma_{c,\text{dry}}$ from eq. (11) and (13) and tables 1, 2

16:           **end if**

17:           Draw 4 normally distributed random numbers $\epsilon \in \mathbb{R}^4$ {Cross correlation after Richardson (1981)}

18:           Set the residuals $\chi_i = \begin{pmatrix} \chi_{T_{\min}} & \chi_{T_{\max}} & \chi_c & \chi_w \end{pmatrix} = A\chi_{i-1} + B\epsilon \in \mathbb{R}^4$ with A and B from eq. (14)

19:           Calculate daily variables via

$$T_{\min,i} = \chi_{T_{\min}} \cdot \sigma_{T_{\min},i} + T_{\min,i} \qquad c_i = \chi_c \cdot \sigma_{c,i} + c_i$$
$$T_{\max,i} = \chi_{T_{\max}} \cdot \sigma_{T_{\max},i} + T_{\max,i} \qquad w_i = \left(\chi_w \cdot \sqrt{\sigma_{w,i}} + \sqrt{w_i}\right)^2$$

20:           Apply bias correction for $T_{\min}$ (eq. (17)) and w (eq. (18))

21:        **end for**

22:     **end while**

23: **end for**

---

daily precipitation amount by replacing the Gamma distribution used in the original Richardson-type weather generators with a hybrid Gamma-GP distribution, which results in the improved simulation of heavy precipitation events. While we performed an extensive sensitivity analysis to determine the best coefficients for the shape and location parameters of the GP distribution, further improvements might come through correlating the GP parameters to geographic region and/or seasonality (Maraun et al., 2009; Rust et al., 2009). Finally, we introduced a step to correct for systematic bias in the downscaling of temperature and wind speed.

5   Applications, limitations, and outlook

GWGEN will be useful for a wide range of applications, from global vegetation and crop modeling, to large-scale hydrologic analyses, to understanding animal behavior, to forecasting of fire, insect outbreaks, and other ecosystem disturbances. The weather generator is particularly suited for the incorporation into a model that runs on a spatial grid, for example, GWGEN can readily be incorporated into existing DGVMs such as LPJ-LMfire (Pfeiffer et al., 2013) or LPJ-ML (Bondeau et al., 2007) that already rely on a weather generator to provide daily meteorology for certain processes. Furthermore, GWGEN may be envisaged as a potential replacement for very large and cumbersome gridded datasets of high-temporal resolution meteorology such as CRUNCEP (Viovy and Ciais, 2016), particularly for models that need forcing at a daily timestep.

One limitation of this version of GWGEN is that it does not consider spatial autocorrelation in the modeled daily meteorological variables. For certain applications that require synchronous multi-point downscaling of meteorology, e.g., hydrological modeling in small catchments ($< ca.\ 2500\ km^2$), this limitation might make the model unsuitable, because, e.g., the the weather generator could simulate rainfall on different days in different parts of the catchment, where in reality storm events would be highly autocorrelated in space and controlled by mesoscale meteorological conditions. Methods for simultaneous multisite weather generation exist (Wilks, 1998, 1999b, c) and could be adapted to GWGEN. However, even simpler methods to approximate spatial autocorrelation could be possible. Running GWGEN with gridded monthly meteorology — this is the primary application we foresee for the current version of the model — means that the input variables are already highly correlated in space, i.e., the monthly climate in one gridcell generally closely resembles neighboring cells, outside of complex terrain containing sharp, monotonic climate gradients, e.g., rain shadows. Thus, one simple way of achieving a measure of spatial autocorrelation in GWGEN would be to impose a spatial autocorrelation field on the sequence of random numbers used to impose stochastic noise in the downscaling functions. If the random number sequence is similar between gridcells, then, e.g., rain is likely to fall on the same day, given that the transition probabilities will likely also be similar. Over moderate distances, e.g., <50's of km, it might even be sufficient to use the same random seed across all gridcells in a neighborhood. This would have the effect of producing strongly autocorrelated daily meteorology in space, with the only variations being imposed by the underlying input monthly climatology.

GWGEN does not downscale all of the meteorological variables that could conceivably be needed to drive certain models, for example, it does not currently handle relative humidity or downwelling radiation. These and other meteorological variables would be straightforward to include in the model framework, however, handling them in the same way that temperatures, cloud

cover, and wind speed are downscaled. Other variables, such as pressure and wind direction, might be more difficult using the basic GWGEN structure because of the importance of autocorrelation, particularly at high spatial resolution, and might benefit from a different approach towards weather generation. Finally, GWGEN only downscales meteorology from monthly to daily values; for models that require an even shorter timestep, e.g., 6-hourly, some extension of the model functionality would be required. For certain variables, e.g., temperatures, sub-daily downscaling could be easily implemented (Cesaraccio et al., 2001), for others, more detailed analysis of a sub-daily meteorological dataset would be necessary (e.g., the Integrated Surface Database).

6   Conclusions

Compiling a global database of daily precipitation, temperature, cloud cover, and wind speed measurements, we explored the relationship between daily meteorology and monthly summaries first described in the context of weather downscaling by Geng and Auburn (1987). Our analysis of more than 50 million individual records showed that daily-to-monthly relationships are relatively stable in space and time, and constant across a very wide range of stations from all latitudes and climate zones. With the resulting relationships, we parameterized a WGEN/SIMMETEO-type weather generator, with the intention of creating a generic scheme that could be applied anywhere over the earth's land surface for the past, present, and (near) future.

7   Code availability

GWGEN, is open source software, and the code, utility programs for parameterization, evaluation and manipulating the raw weather station data, and complete documentation are available at https://github.com/ARVE-Research/gwgen/releases/tag/v1. 0.0[2]. The original weather station database can be made available upon request to the authors or downloaded from Hahn and Warren (1999) and Menne et al. (2012b). The weather generator module is programmed in FORTRAN, the parameterization, evaluation and other supplementary tools are written in Python mainly using the numerical python libraries numpy and scipy (Jones et al., 2001), statsmodels (Seabold and Perktold, 2010), as well as matplotlib (Hunter, 2007) and psyplot (Sommer, 2017) for the visualization. Detailed installation instructions can be found in the user manual https://arve-research.github.io/gwgen/.

---

[2]A DOI will be provided when the paper is accepted

Appendix A: Supplementary material

A1 Sensitivity analysis
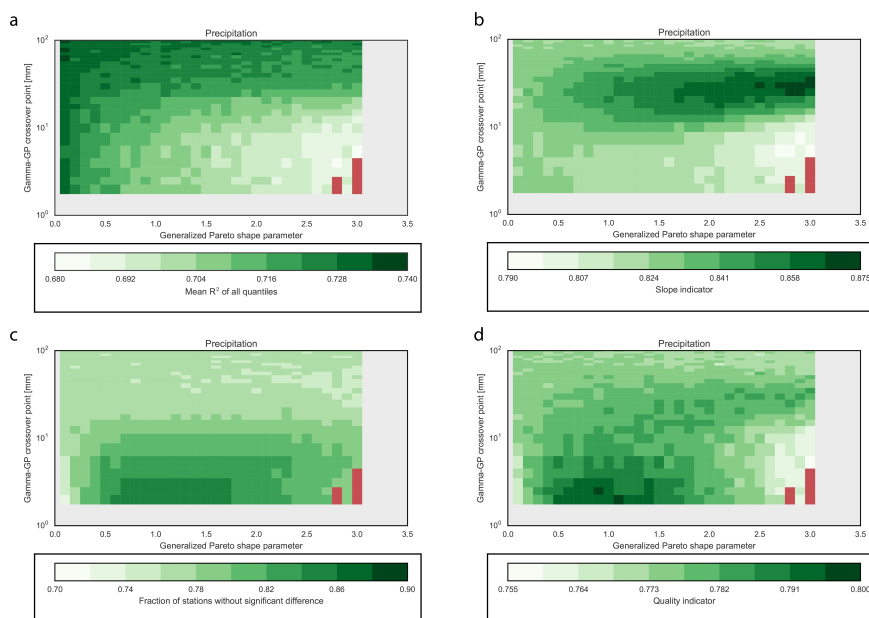
Figure 15. Results of the sensitivity analysis for the (a) correlation coefficient R^2, (b) deviation from a slope of unity, (c) the fraction of significant different station years, (d) the mean of (a) - (c). For the plots in (a) and (b) we used the means of the 25th, 50th, 75th, 90th, 95th and 99th percentiles. In general, 1 (dark green) is best, 0 (white) is worst. The dark red fields indicate experiments that failed because of a too low threshold and too high GP shape parameter. Note the logarithmic scale on the y-axis.

Author contributions. JOK conceived the model and analyses, wrote the prototype code and performed preliminary analyses, PS developed and documented the final version of the code (including parameterization and evaluation), performed all of the final analyses, and created the graphical output. Both authors contributed to the writing of the manuscript

Acknowledgements. This work was supported by the European Research Council (COEVOLVE, 313797) and the Swiss National Science Foundation (ACACIA, CR10I2_146314). We thank Shawn Koppenhoefer for assistance compiling and querying the weather databases and Alexis Berne and Grégoire Mariéthoz for helpful suggestions on the analyses. We are grateful to NOAA NCDC and the University of Washington for providing free of charge the GHCN-Daily and EECRA databases, respectively.

References

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I.: The global distribution and burden of dengue, Nature, 496, 504–507, http://dx.doi.org/10.1038/nature12060, 2013.

5 Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., MÜLler, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Global Change Biology, 13, 679–706, doi:10.1111/j.1365-2486.2006.01305.x, 2007.

Cesaraccio, C., Spano, D., Duce, P., and Snyder, R. L.: An improved model for determining degree-day values from daily temperature data, Int J Biometeorol, 45, 161–9, doi:10.1007/s004840100104, http://www.ncbi.nlm.nih.gov/pubmed/11769315, 2001.

10 Dai, A.: Precipitation Characteristics in Eighteen Coupled Climate Models, Journal of Climate, 19, 4605–4630, doi:10.1175/jcli3884.1, http://dx.doi.org/10.1175/JCLI3884.1, 2006.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., and E. Zimmermann,

15 N.: Novel methods improve prediction of species' distributions from occurrence data, Ecography, 29, 129–151, doi:10.1111/j.2006.0906-7590.04596.x, http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x, 2006.

Friend, A. D.: Parameterisation of a global daily weather generator for terrestrial ecosystem modelling, Ecological Modelling, 109, 121–140, doi:Doi 10.1016/S0304-3800(98)00036-2, <GotoISI>://WOS:000074332200001, 1998.

Furrer, E. M. and Katz, R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators, Water Resources

20 Research, 44, n/a–n/a, doi:10.1029/2008wr007316, 2008.

Geng, S. and Auburn, J. S.: Weather simulation models based on summaries of long-term data, pp. 237–254, International Rice Research Institute, Los Baños, Philippines, 1987.

Geng, S., Devries, F. W. T. P., and Supit, I.: A Simple Method for Generating Daily Rainfall Data, Agricultural and Forest Meteorology, 36, 363–376, doi:Doi 10.1016/0168-1923(86)90014-6, <GotoISI>://WOS:A1986C086500007, 1986.

25 Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., and Sitch, S.: Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model, Journal of Hydrology, 286, 249–270, doi:10.1016/j.jhydrol.2003.09.029, 2004.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P.: A Global-Model of Natural Volatile Organic-Compound Emissions, Journal of Geophysical Research-Atmospheres, 100, 8873–8892, doi:Doi 10.1029/94jd02950, <GotoISI>://WOS:

30 A1995QZ72400003, 1995.

Hahn, C. and Warren, S.: Extended Edited Synoptic Cloud Reports from Ships and Land Stations Over the Globe, 1952-1996 (with Ship data updated through 2008), doi:10.3334/CDIAC/cli.ndp026c, http://dx.doi.org/10.3334/CDIAC/cli.ndp026c, 1999.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, doi:10.1002/joc.3711, 2014.

35 Haxeltine, A. and Prentice, I. C.: BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, Global Biogeochemical Cycles, 10, 693–709, doi:Doi 10.1029/96gb02344, <GotoISI>://WOS:A1996VV17500011, 1996.

Haxeltine, A., Prentice, I. C., and Creswell, I. D.: A coupled carbon and water flux model to predict vegetation structure, Journal of Vegetation Science, 7, 651–666, doi:10.2307/3236377, http://dx.doi.org/10.2307/3236377, 1996.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, International Journal of Climatology, 25, 1965–1978, doi:10.1002/joc.1276, 2005.

5   Hunter, J. D.: Matplotlib: A 2D graphics environment, Computing In Science & Engineering, 9, 90–95, 2007.

Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python, http://www.scipy.org/, [Online; accessed 2017-02-18], 2001.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., and

10   Lozhkin, A. V.: Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections, Journal of Geophysical Research-Atmospheres, 108, doi:Artn 8171 10.1029/2002jd002559, <GotoISI>://WOS:000185928300002, 2003.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, n/a–n/a, doi:10.1029/2003gb002199, 2005.

15   Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., Ramankutty, N., Norman, J. M., and Gower, S. T.: Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure, Global Biogeochemical Cycles, 14, 795–825, doi:10.1029/1999GB001138, http://dx.doi.org/10.1029/1999GB001138, 2000.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, International Journal of Climatology, 33, 1367–1381, doi:10.1002/joc.3518, http://dx.doi.org/10.1002/joc.3518, 2012.

20   Leemans, R. and Cramer, W. P.: The IIASA database for mean monthly values of temperature, precipitation, and cloudiness on a global terrestrial grid, International Institute for Applied Systems Analysis, Laxenburg, Austria., 1991.

Lieth, H.: Modeling the Primary Productivity of the World, pp. 237–263, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-80913-2_12, http://dx.doi.org/10.1007/978-3-642-80913-2_12, 1975.

Maraun, D., Rust, H. W., and Osborn, T. J.: The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme

25   value statistics, International Journal of Climatology, 29, 1731–1744, doi:10.1002/joc.1811, http://dx.doi.org/10.1002/joc.1811, 2009.

Menne, M. J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R. S., Gleason, B. E., and Houston, T. G.: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.22, doi:10.7289/V5D21VHZ, http://dx.doi.org/10.7289/V5D21VHZ, 2012a.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An Overview of the Global Historical Climatology Network-Daily

30   Database, J. Atmos. Oceanic Technol., 29, 897–910, doi:10.1175/jtech-d-11-00103.1, http://dx.doi.org/10.1175/JTECH-D-11-00103.1, 2012b.

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, International Journal of Climatology, 25, 693–712, doi:10.1002/joc.1181, 2005.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90

35   mean monthly terrestrial climatology, Journal of Climate, 12, 829–856, doi:Doi 10.1175/1520-0442(1999)012<0829:Rtcstc>2.0.Co;2, <GotoISI>://WOS:000079181700010, 1999.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate, Journal of Climate, 13, 2217–2238, doi:Doi 10.1175/1520-0442(2000)013<2217:Rtcstc>2.0.Co;2, <GotoISI>://WOS:000088489900006, 2000.

New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas, Climate Research, 21, 1–25, doi:DOI 10.3354/cr021001, <GotoISI>://WOS:000177521400001, 2002.

Neykov, N. M., Neytchev, P. N., and Zucchini, W.: Stochastic daily precipitation model with a heavy-tailed component, Natural Hazards and Earth System Science, 14, 2321–2335, doi:10.5194/nhess-14-2321-2014, 2014.

Parlange, M. B. and Katz, R. W.: An Extended Version of the Richardson Model for Simulating Daily Weather Variables, Journal of Applied Meteorology, 39, 610–622, doi:10.1175/1520-0450-39.5.610, http://dx.doi.org/10.1175/1520-0450-39.5.610, 2000.

Pfeiffer, M., Spessa, A., and Kaplan, J. O.: A model for global biomass burning in preindustrial time: LPJ-LMfire (v1.0), Geoscientific Model Development, 6, 643–685, doi:10.5194/gmd-6-643-2013, <GotoISI>://WOS:000321137700004http://www.geosci-model-dev.net/6/643/2013/gmd-6-643-2013.pdf, 2013.

Prentice, I.: Developing a Global Vegetation Dynamics Model: Results of an IIASA Summer Workshop, Iiasa research report, IIASA, Laxenburg, Austria, http://pure.iiasa.ac.at/3223/, 1989.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M.: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate, Journal of Biogeography, 19, 117–134, doi:Doi 10.2307/2845499, <GotoISI>://WOS:A1992HY34900002, 1992.

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

Rust, H. W., Maraun, D., and Osborn, T. J.: Modelling seasonality in extreme precipitation, The European Physical Journal Special Topics, 174, 99–111, doi:10.1140/epjst/e2009-01093-7, http://dx.doi.org/10.1140/epjst/e2009-01093-7, 2009.

Rymes, M. and Myers, D.: Mean preserving algorithm for smoothly interpolating averaged data, Solar Energy, 71, 225–231, doi:10.1016/s0038-092x(01)00052-4, http://dx.doi.org/10.1016/S0038-092X(01)00052-4, 2001.

Seabold, S. and Perktold, J.: Statsmodels: Econometric and statistical modeling with python, 2010.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Change Biology, 9, 161–185, doi:10.1046/j.1365-2486.2003.00569.x, <GotoISI>://WOS:000180852800005http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2486.2003.00569.x/abstract, 2003.

Sommer, P. S.: psyplot: Python package for interactive data visualization - First official release - unstable, doi:10.5281/zenodo.293079, https://doi.org/10.5281/zenodo.293079, 2017.

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., and Haynes, J.: Dreary state of precipitation in global models, Journal of Geophysical Research: Atmospheres, 115, n/a–n/a, doi:10.1029/2010JD014532, http://dx.doi.org/10.1029/2010JD014532http://onlinelibrary.wiley.com/store/10.1029/2010JD014532/asset/jgrd16654.pdf?v=1&t=iz72qw9n&s=1a6f09a12c3ed0513483f485c675db7f5feea5a7, 2010.

Sun, Y., Solomon, S., Dai, A., and Portmann, R. W.: How Often Does It Rain?, Journal of Climate, 19, 916–934, doi:10.1175/jcli3672.1, http://dx.doi.org/10.1175/JCLI3672.1, 2006.

Viovy, N. and Ciais, P.: Online database, http://dods.extra.cea.fr/data/p529viov/cruncep, 2016.

Walter, H. and Lieth, H.: Climate diagram world atlas, VEB Gustav Fischer Verlag Jena, Jena, 1967.

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., Schwalm, C. R., Schaefer, K., Jacobson, A. R., Lu, C., Tian, H., Ricciuto, D. M., Cook, R. B., Mao, J., and Shi, X.: The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data, Geosci. Model Dev., 7, 2875–2893, doi:10.5194/gmd-7-2875-2014, http://www.geosci-model-dev.net/7/2875/2014/, 2014.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, Journal of Hydrology, 210, 178–191, doi:Doi 10.1016/S0022-1694(98)00186-3, <GotoISI>://WOS:000076861100013, 1998.

Wilks, D. S.: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models, Agricultural and Forest Meteorology, 93, 153–169, doi:Doi 10.1016/S0168-1923(98)00125-7, <GotoISI>://WOS:000079269800001, 1999a.

Wilks, D. S.: Multisite downscaling of daily precipitation with a stochastic weather generator, Climate Research, 11, 125–136, doi:DOI 10.3354/cr011125, <GotoISI>://WOS:000080067600003, 1999b.

Wilks, D. S.: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain, Agricultural and Forest Meteorology, 96, 85–101, doi:Doi 10.1016/S0168-1923(99)00037-4, <GotoISI>://WOS:000082480400008, 1999c.

Wilks, D. S.: Use of stochastic weathergenerators for precipitation downscaling, Wiley Interdisciplinary Reviews: Climate Change, 1, 898–907, doi:10.1002/wcc.85, 2010.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Progress in Physical Geography, 23, 329–357, doi:Doi 10.1177/030913339902300302, <GotoISI>://WOS:000082557200002, 1999.

Woodward, F. I., Smith, T. M., and Emanuel, W. R.: A global land primary productivity and phytogeography model, Global Biogeochemical Cycles, 9, 471–490, doi:10.1029/95GB02432, http://dx.doi.org/10.1029/95GB02432, 1995.

Woolhiser, D. A. and Pegram, G. G. S.: Maximum Likelihood Estimation of Fourier Coefficients to Describe Seasonal-Variations of Parameters in Stochastic Daily Precipitation Models, Journal of Applied Meteorology, 18, 34–42, doi:Doi 10.1175/1520-0450(1979)018<0034:Mleofc>2.0.Co;2, <GotoISI>://WOS:A1979GK32500004, 1979.

Woolhiser, D. A. and Roldan, J.: Stochastic Daily Precipitation Models: 2. A Comparison of Distributions of Amounts, Water Resources Research, 18, 1461–1468, doi:DOI 10.1029/WR018i005p01461, <GotoISI>://WOS:A1982PP69200014, 1982.

Woolhiser, D. A. and Roldán, J.: Seasonal and Regional Variability of Parameters for Stochastic Daily Precipitation Models: South Dakota, U.S.A, Water Resources Research, 22, 965–978, doi:10.1029/WR022i006p00965, 1986.