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

Response to reviewer comment 1

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)

We thank the anonymous reviewer for his helpful comments to our manuscript. The manuscript for GWGEN, a weather generator for precipitation, temperature, cloud fraction and wind speed using a hybrid Gamma-GP distribution, a hybrid-order Markov Chain and a cross correlation approach) has been revised and improved.

In summary, a bug has been fixed that now makes the quantile-based bias correction for the minimum temperature redundant and instead another quantile-based bias correction for the wind speeds intercept has been implemented to further improve the results. Furthermore we made several attempts to improve the manuscript text for clarity and style. This includes a schematic representation of the workflow, changes in the structure of the paper, more explanations to the figures and a fix of the notation in the equations.

The spatial autocorrelation, however, that has also been addressed by the other reviewer and the editor is, to our believe, beyond the scope of this manuscript. Although we think that it is possible, we agree with the reviewers that it is not that simple and subject to further research. Already for the technical aspect we would need a few months to fix this issue. Nevertheless we think that this does not affect the utility of the weather generator for a wide range of applications.

Detailed responses to the comments of the reviewer can be found below.

Responses

Reviewer Section 2.2.1. The first line is rather strange if one takes into account that this is actually the frequential definition of probability. The probability that a given day is wet is defined as

$$P(\text{wet}) = \frac{\#\text{wet days}}{\#\text{total days}},$$

for a specific month and station, so obviously there is a strong relationship. They represent exactly the same thing.

Response We edited the text to acknowledge this fact.

Reviewer Use of a 2nd order MC gives better fit and results when modelling precipitation, see for example the study done in Lennartsson et al. (2008). A 2nd order MC is characterised by the transition probabilities $p_{ijk}; i, j, k = 0, 1$, therefore a total of 8 transition probabilities. I understand that the authors are interested only at the event of a wet day, i.e. we need

$p_{ij1}; i; j = 0, 1$. These probabilities would be : $p_{001}, p_{101}, p_{011}, p_{111}$. Could the authors explain why instead of the last two they model p_{11}?

Response Any model development requires choices and trade-offs between absolute realism and computational demand. Following the analyses and recommendation of Wilks (1999), we use a hybrid-order model that retains first-order Markov dependence for wet spells but allows higher-order dependence for dry sequences as a compromise between effectiveness and simplicity. This approach therefore only uses the probabilities up to the last wet day, which are p_{11} and p_{101}, as well as p_{001} for a dry sequence. Using this, the MC only needs 3 probabilities instead of 4. We will include this explanation in the paper.

Reviewer In Fig.2 the fit is per station in a given month or for all months and stations together? The authors could do a better job explaining waht it is plotted in every Figure.

Response We clarify our methodology by providing the following description in the text

We perform this analysis on a station and month-wise basis, i.e., we first extract 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 with precipitation, which show the characteristic linear relationship described by Geng et al. (1986)

Furthermore, the figure captions now include a new short description for clarification:

The underlying data for the fits correspond to the means of the the multi-year series for each month for each station.

Reviewer In section 2.2.2, line 11: The strong relation... The mean of a gamma random variable equals the product of its two parameters. i.e $E(\Gamma) = \alpha\theta$

Response We include following explanation in the revised manuscript:

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. It is based upon the fact that the expected value of a gamma random variable equals the product of its two parameters. i.e $E(\Gamma) = \alpha\theta$.

Reviewer Cleary the extreme values cannot be modelled using the Gamma distribution, it is not suitable for this.

Response We agree, and as noted, we adopt the hybrid Gamma-GP approach to capture high precipitation amounts as suggested by several previous studies.

Reviewer Page 8, line 9 : I would prefer the use of the term density since distribution is usually reserved for the cumulative distribution function.

Response It has been changed to probability density function (pdf)

Reviewer How was the estimation of the Gamma distribution parameters performed? If the authors used likelihood, how did they deal with the fact that the the excesses above level are modelled as GP. Durban and Glasbey (2001) and Baxevani and Lennartsson (2015) suggest a type of modeified likelihood that treats the excesses as sensored data.

Response We used all of the data in fitting the Gamma distribution using likelihood. We acknowledge that there could have been different approaches to this problem, including censuring data above the threshold, but the final results of our model as presented are acceptable to us. We clarify this point in the text when discussing the fitting procedure.

Reviewer If α and θ are estimated by fitting the distribution to the data, then $E(\Gamma) = \bar{p} = \alpha\theta$. So what exactly is modelled by (7) and (8)? I am confused about what the authors are trying to achieve here.

Response We explain that the resulting α in our calculations ends up being a constant, effectively the slope of the relationship between the Gamma scale parameter and $\bar{p}$, and revise equation (8) to reflect this fact.

Reviewer Fig. 11 Right. The data do not seem to be in a linear relation here. I think the authors shoud try some other relation or transformation also.

Response We agree and use a third order polynomial now which significantly improves the relationship

Reviewer Section 2.2.6. I could appreciate some comments on why the matrices A and B are needed and what they actually represent or try to model. Moreover, are the matrices estimated for each station and for every month? I assume that they are estimated using all months and stations together? How is something like this justified?

Response We added additional clarification and explanation on this point at the beginning of the relevant section.

Reviewer What is the exact length of each simulation record? When we compare simulated versus observed records I assume that the simulated records are of exactly the same length as he observed ones?

Response While the lengths of the observed meteorological records differ for each station, in our model evaluation, we simulate a daily weather record that is exactly as long as the input monthly weather observations. We clarify this point in the text.

Reviewer The authors notice that the gamma distribution does not perform so well for low values. Maybe it would make sense to stich another distribution for these low values, in the same fashion they did for the high ones.

Response Ecological and hydrological significance of very low precipitation is small, also we are close to the precision of the measurements. For the sake of model parsimony, we use the current methodology, as also suggested by several other authors.

Reviewer Section 2.5 The choice of the threshold μ is a rather difficult one, see for example Frigessi et al. (2002). The problem is that the fitting of the GP by likelihood, is based on the assumption that the excesses above level μ are independent and identically distributed. A rather difficult to satisfy assumption. If the level μ is chosen too low this will result to too many excesses that will be probably dependent. If it is chosen too high that would result to too few excesses to make any kind of reasonable fitting. Moreover, I think the use of a global threshold is oversimplifying.

Response We agree to this point. However, although we did fit the GP to our parameterization data above the threshold, this information could not be used. Instead, we decided to use constant parameters for the GP shape and the threshold and make a sensitivity analysis (previous section 2.5). The reason for this is, that after extensive data analysis, we could not find any good relationship between ξ, μ and the input data for our weather generator. In fact, as stated in the text, we also tried a varying threshold such that the GP distribution is used, when the Gamma random variable exceeds a certain percentile, but we could not find any improvement.

Therefore we could not justify a varying ξ and μ although we acknowledge the fact, that this is oversimplifying and we clarified this in the discussion. At the moment this is the best we can do and the results are nevertheless better compared to using the Gamma distribution alone and, indeed, they are surprisingly good.

References

Baxevani, A. and Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent Gaussian field, Water Resources Research, 51, 4338–4358, doi:10.1002/2014WR016455, http://dx.doi.org/10.1002/2014WR016455, 2015.

Durban, M. and Glasbey, C.: Weather modelling using a multivariate latent Gaussian model, Agricultural and Forest Meteorology, 109, 187 – 201, doi:http://dx.doi.org/10.1016/S0168-1923(01)00268-4, http://www.sciencedirect.com/science/article/pii/S0168192301002684, 2001.

Frigessi, A., Haug, O., and Rue, H.: A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection, Extremes, 5, 219–235, doi:10.1023/A:1024072610684, http://dx.doi.org/10.1023/A:1024072610684, 2002.

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:10.1016/0168-1923(86)90014-6, <GotoISI>://WOS:A1986C086500007, 1986.

Lennartsson, J., Baxevani, A., and Chen, D.: Modelling precipitation in Sweden using multiple step markov chains and a composite model, Journal of Hydrology, 363, 42 – 59, doi:10.1016/j.jhydrol.2008.10.003, http://www.sciencedirect.com/science/article/pii/S0022169408004848, 2008.

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

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

Response to reviewer comment 2

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)

We thank the anonymous reviewer for his helpful comments to our manuscript. The manuscript for GWGEN, a weather generator for precipitation, temperature, cloud fraction and wind speed using a hybrid Gamma-GP distribution, a hybrid-order Markov Chain and a cross correlation approach) has been revised and improved.

In summary, a bug has been fixed that now makes the quantile-based bias correction for the minimum temperature redundant and instead another quantile-based bias correction for the wind speeds intercept has been implemented to further improve the results. Furthermore we made several attempts to improve the reading. This includes a schematic representation of the workflow, changes in the structure of the paper, more explanations to the figures and a fix of the notation in the equations.

Detailed responses to the comments of the reviewer can be found below.

Responses

Reviewer Section 2.2.1 : Why are you not interested in p011 and p111?

Respones We use the hybrid-order Markov Chain as recommended by Wilks (1999) as a tradeoff between accuracy and simplicity. This model retains first-order Markov dependence for wet spells but allows second-order dependence for dry sequences. It therefore only requires the three parameters p_{11}, p_{001} and p_{101}, i.e. the probabilities up to the last wet day. We specified it explicitly now.

Reviewer Figure 2 : There is no histogram.

Respones It is a 2D histogram, i.e. the value for each grid cell represents the sum of observations that fall into this cell. To clarify this point, we replaced the word histogram with *density plot*.

Reviewer Table 1 and Figure 11 : you should mention the fact that R2 are artificially high for models without constant because the R2 formulae is modified for such models.

Response We now acknowledge this fact in the paper.

Reviewer In equation (5), can ξ be equal to 0?

Response Yes, in this case, $g(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}}$. We added it to the equation.

Reviewer Line 19 : you should explain how you estimate the parameters.

Response The parameters of the Gamma distribution are estimated using likelihood. We included it in the text.

Reviewer Lines 20 to 25 : it is not clear, you should explain quickly what is done in Geng et al. (1986)

Response We clarified in the text that the expected value of the Gamma distribution is the product of the shape and scale parameter, i.e. $E(\Gamma) = \alpha\theta$. This justifies equation (7) and (8).

Reviewer Equation (10) : x_{wet} and x_{dry} have to be replaced by $\bar{x}_{wet}$ and $\bar{x}_{dry}$? Same remark for equation (11).

Response They have been replaced.

Reviewer Figure 5 and 7 : As you write, the adjustement is very bad. I think you should propose another way of fixing the standard deviations. You write "we believe that the error introduced by the poor linear fit is negligible", but this is not convincing.

Response We agree with this comment and in our revised version of the model have completely re-thought the way we estimate the SD of temperature. These changes are described in the revised manuscript. In summary, instead of a linear fit, we now use a combination out two polynomials combined with a linear extrapolation at the cold and warm extremes. For minimum temperature that means, that values below $-40°C$ and above $25°C$ are linearly extrapolated, whereas $\sigma_{T_{\min},\mathrm{dry/wet}}$ between $-40°C < \bar{T}_{\min,\mathrm{dry/wet}} \leq 0°C$ and $0°C < \bar{T}_{\min,\mathrm{dry/wet}} \leq 25°C$ is modeled by two different polynomials of order 5. We use the same methodology for maximum temperature with $-30°C$ instead of $-40°C$, and $35°C$ instead of $25°C$ (see attached figures below).

Although this procedure is more complicated, results in a significant improvement in the simulation of extreme temperatures, and an overall improvement in the simulation of daily temperature.

Reviewer Equation (12) : please explain how this formulae has been chosen.

Response We wrote in the original manuscript that we chose the shapes of the curves to reflect the phenomenon that wet days should be cloudier on average than dry days. We clarify our selection of these equations based on the constraints presented by the variable, e.g., that cloud fraction on wet days must be 0 when total mean cloud fraction is 0 and 1 when the total mean cloud fraction is 1. We used a qualitative graphical analysis to develop "best guess" equations that had the desired shape.

Reviewer Page 13 line 5 : $c_{sd,dry}$ has to be replaced by $\sigma_{c,dry}$, same for "wet".

Response They have been replaced

Reviewer Equation (12) : c has to be replaced by c_{wet} or c_{dry}. Moreover, bars have to be added, since you describe mean cloud cover.

Response Bars have been added but c (or rather $\bar{c}$) should not be replaced by $\bar{c}_{wet}$ or $\bar{c}_{dry}$ since in this case we use the monthly mean cloud fraction $\bar{c}$ to calculate the mean of the wet ($\bar{c}_{wet}$) or dry ($\bar{c}_{dry}$) days in the month

Reviewer Equation (13) : same remark, bars have to be added.

Response They have been added

Reviewer Section 2.2.6 : Please describe how you add the residual noise in practice. It is described at the end of Algorithm 1 but it is not clear : residuals for one day are really computed from the residuals of the previous day as written line 18? If so, you should explain why.

Response Yes, they are computed from the previous day. We extended the explanation in the corresponding section. The procedure is based upon Matalas (1967) and preserves the serial and the cross correlation between the simulated variables. Otherwise longer periods of, e.g. warm temperatures, could not be simulated.

Reviewer Section 2.5 : Can you present/discuss some references about the estimation problem of GP parameters?

Response We acknowledge that choosing globally fixed parameters for the location parameter μ and the threshold ξ is a simplified aspect of our model (we also state that in the revised manuscript) and is generally not easy (e.g. Davison and Smith, 1990; Neykov et al., 2014; Rootzén and Tajvidi, 1997). Frigessi et al. (2002) suggest to use a dynamical mixture model instead of a fixed threshold. Rust et al. (2009) vary the parameters with seasonality.

However, to our knowledge, no global application of these methods has been published and for now, therefore we stick to the simplest methodology with fixed parameters that are based on a sensitivity analysis (described in detail in section 3.5 of the revised manuscript). We also performed extensive data analysis in an attempt to correlate the GP parameters with other input variables for our weather generator, but could not find any relationship that would allow us to perform a dynamic calculation. As we say in the discussion section, this is subject to further improvement, but, nevertheless, despite the simplicity of our parameterization of the hybrid Gamma-GP distribution, the results are excellent.

Reviewer Section 3 : I think such a global presentation of the model should be given at the beginning of the paper in order to help the reader following al steps. Maybe with a schematic description?

Response It has been moved and a schematic of the workflow has been added

Reviewer I think Sections 4 et 5 could be merged.

Response They have been merged.

Reviewer Section 5 : I am not convinced that the introduction of a spatial autocorrelation field on the sequence of random numbers would solve the problem so easily. The spatial correlation will not be the same for the whole globe and for all variables, and may be hard to fix.

Response We agree, but also continue to believe that implementation of spatial autocorrelation is beyond the scope of the current manuscript, and does not affect the utility of the weather generator for a wide range of applications. We clarify the challenge of implementing autocorrelation in the manuscript, and remove our specific recommendation, tending to agree with the reviewer that, although possible, the solution would not be **that** simple.

References

Davison, A. C. and Smith, R. L.: Models for Exceedances over High Thresholds, Journal of the Royal Statistical Society. Series B (Methodological), 52, 393–442, http://www.jstor.org/stable/2345667, 1990.

Frigessi, A., Haug, O., and Rue, H.: A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection, Extremes, 5, 219–235, doi:10.1023/A:1024072610684, http://dx.doi.org/10.1023/A:1024072610684, 2002.

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:10.1016/0168-1923(86)90014-6, <GotoISI>://WOS:A1986C086500007, 1986.

Matalas, N. C.: Mathematical assessment of synthetic hydrology, Water Resources Research, 3, 937–945, doi:10.1029/WR003i004p00937, http://dx.doi.org/10.1029/WR003i004p00937, 1967.

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.

Rootzén, H. and Tajvidi, N.: Extreme value statistics and wind storm losses: A case study, Scandinavian Actuarial Journal, 1997, 70–94, doi:10.1080/03461238.1997.10413979, http://dx.doi.org/10.1080/03461238.1997.10413979, 1997.

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.

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

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~~wind speed. 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~~ should be 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 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 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.

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 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 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. $1^\circ \times 1^\circ$ 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 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 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.

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 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 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~~ bias correcting 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 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 description

GWGEN requires the following six monthly summary values as input: 1) total monthly precipitation, 2) the number of days in the month with measurable precipitation (i.e., wet days), 3-4) monthly mean daily minimum and maximum temperature, 5) mean cloud fraction, and 6) 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 figure 1 and 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 (section 3.2.1) to decide the wet/dry state of the day. If it is a wet day, we calculate the gamma parameters using the equations (7) and (8). The resulting distribution allows us to draw a

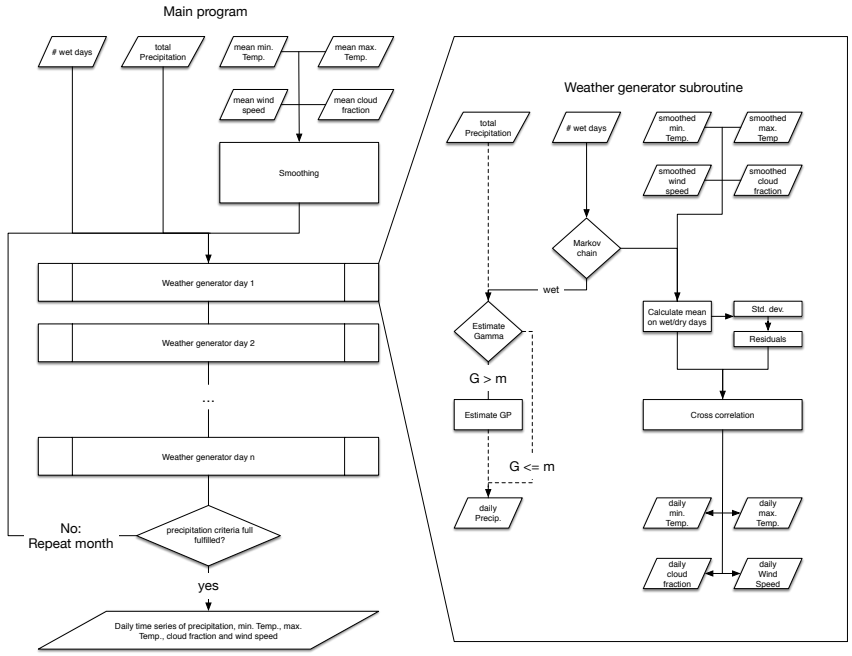


Figure 1. Schematic workflow of GWGEN. After smoothing the monthly input, the Markov Chain is used to decide, whether it is a dry or a wet day. If it is a wet day, we draw a random number from the Gamma-GP distribution. Furthermore, the other means of the variables ($\bar{T}_{\mathrm{min/max}}, \bar{c}, \bar{w}$) are adjusted and their daily values are calculated using the estimated standard deviations and residuals. The wind speed furthermore undergoes a square root transformation before applying the cross correlation and in the end is corrected using the bias correction. A quality check in the end restricts our model to be within a 5% range of the observed total precipitation and to replicate the number of wet days from the input.

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 11 and 15 in algorithm 1). After that, we use the cross-correlation approach described in Richardson (1981) (lines 18 - 20 and subsubsection 3.2.6) and calculate the daily values of these variables. Finally we use the quantile-based bias correction described in section 3.4 to correct the simulated wind speed.

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 4).

Algorithm 1 Basic workflow of GWGEN

Require: monthly precipitation $P_{\text{in}}\,[\text{mm}]$, cloud cover fraction c_{in}, minimum ($T_{\text{min,in}}\,[^\circ\text{C}]$) and maximum ($T_{\text{max,in}}\,[^\circ\text{C}]$) 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 $s_i \in \{0,1\}$

for month m in $input$ **do**
  smooth the monthly data using Rymes and Myers (2001)
  Set $j = 0, \chi = 0$
  while $j \equiv 0$ or $\left|\sum_{d_i \in m} P_i - P_{\text{in}}\right| > \min\left(5\% \cdot P_{\text{in}}, 0.5mm\right)$ or $|n_{\text{sim}} - n_{\text{in}}| > 1$ **do**
    for day d_i in m **do**
      Calculate p_{11}, p_{101}, p_{001} after equations (1) - (3) using n {Precipitation occurence after Wilks (1999a)}
      Use the Markov chain to determine whether d_i is wet ($s_i = 1$) or dry ($s_i = 0$)
      if $s_i = 1$ **then**
        Calculate θ, α and σ via eq. (7)-(9) {Precipitation amount after Neykov et al. (2014)}
        Draw a random number P_i from the Gamma-GP distribution, eq. (6)
        Set $T_{\text{min},i} = T_{\text{min,wet}}, T_{\text{max},i} = T_{\text{max,wet}}, c_i = c_{\text{wet}}, w_i = w_{\text{wet}}$ from eq. (10) and (12) and tables 1, 3
        Set $\sigma_{T_{\text{min}},i} = \sigma_{T_{\text{min}},\text{wet}}, \sigma_{T_{\text{max}},i} = \sigma_{T_{\text{max}},\text{wet}}, \sigma_{w,i} = \sigma_{w,\text{wet}}, \sigma_{c,i} = \sigma_{c,\text{wet}}$ from eq. (11), (13) and (14) and tables 1, 2, 3
      else
        Set $P_i = 0\,\text{mm/d}$
        Set $T_{\text{min},i} = T_{\text{min,dry}}, T_{\text{max},i} = T_{\text{max,dry}}, c_i = c_{\text{dry}}, w_i = w_{\text{dry}}$ from eq. (10) and (12) and tables 1, 3
        Set $\sigma_{T_{\text{min}},i} = \sigma_{T_{\text{min}},\text{dry}}, \sigma_{T_{\text{max}},i} = \sigma_{T_{\text{max}},\text{dry}}, \sigma_{w,i} = \sigma_{w,\text{dry}}, \sigma_{c,i} = \sigma_{c,\text{dry}}$ from eq. (11), (13) and (14) and tables 1, 2, 3
      end if
      Draw 4 normally distributed random numbers $\epsilon \in \mathbb{R}^4$ {Cross correlation after Richardson (1981)}
      Set the residuals $\chi_i = \begin{pmatrix}\chi_{T_{\text{min}}} & \chi_{T_{\text{max}}} & \chi_c & \chi_w\end{pmatrix} = A\chi_{i-1} + B\epsilon \in \mathbb{R}^4$ with A and B from eq. (17)
      Calculate daily variables via

$$T_{\text{min},i} = \chi_{T_{\text{min}}} \cdot \sigma_{T_{\text{min}},i} + T_{\text{min},i} \qquad c_i = \chi_c \cdot \sigma_{c,i} + c_i$$

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

      Apply bias correction w (eq. (23))
      $j = j + 1$
    end for
  end while
end for

3 Model development

GWGEN is based on the WGEN weather generator (Richardson, 1981), using the method of defining the model parameters 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~~ hybrid-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 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 2.~~

3.1 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~~GHCN-Daily) database (Menne et al., 2012b, a). The ~~GHCN-daily~~ GHCN-Daily consists of observations collected at ca. 100'000 weather stations on all continents and 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 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~~ and 1.0 mm d^{-1} were reported in the record. This resulted in ~~8345~~ 9508 stations covering all continents, although the distribution is strongly heterogenous, with the majority of the stations in North America, despite our geographic filter (Figure 2, top panel). For cloud cover, windspeed, and to calculate cross-correlations between temperature, cloud cover, and windspeed, we used the Extended Edited Cloud Report Archive (EECRA) database (Hahn and Warren, 1999). The geographic distribution of the ~~8535~~ 6978 EECRA stations we selected is different than the GHCN-Daily, with more stations in Europe (Figure 2, 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 observations used in the model parameterization contains ca. ~~50~~69 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 vari-

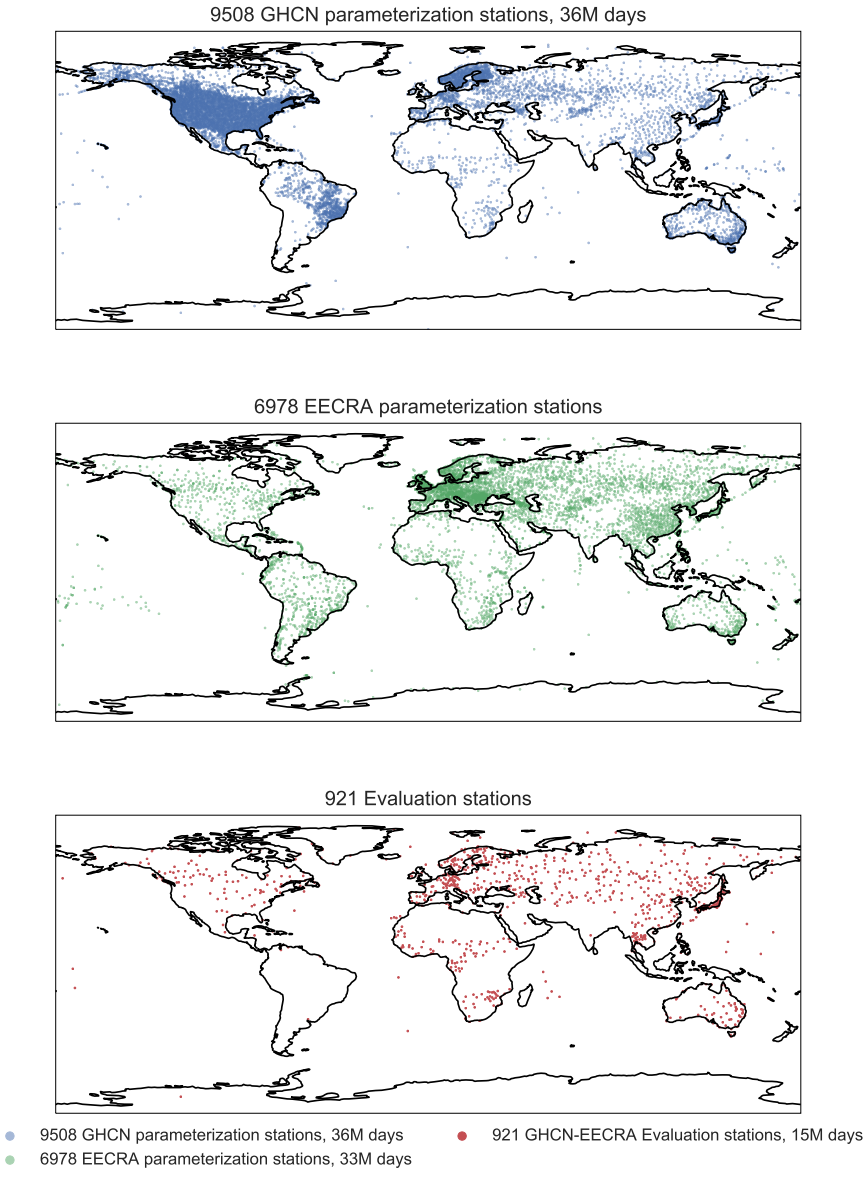


Figure 2. Weather stations used for parameterization and evaluation of the weather generator. The uppermost panel shows the locations of the stations used for parameterizing precipitation and temperature, the middle panel shows the stations for cloud ~~cover~~ fraction and ~~windspeed~~wind speed, as well as for calculating the cross correlations between temperature, ~~cloudiness~~cloud fraction, and ~~windspeed~~wind speed. The lower plot shows the location of the stations used to evaluate the model, which were excluded from the parameterization stations.

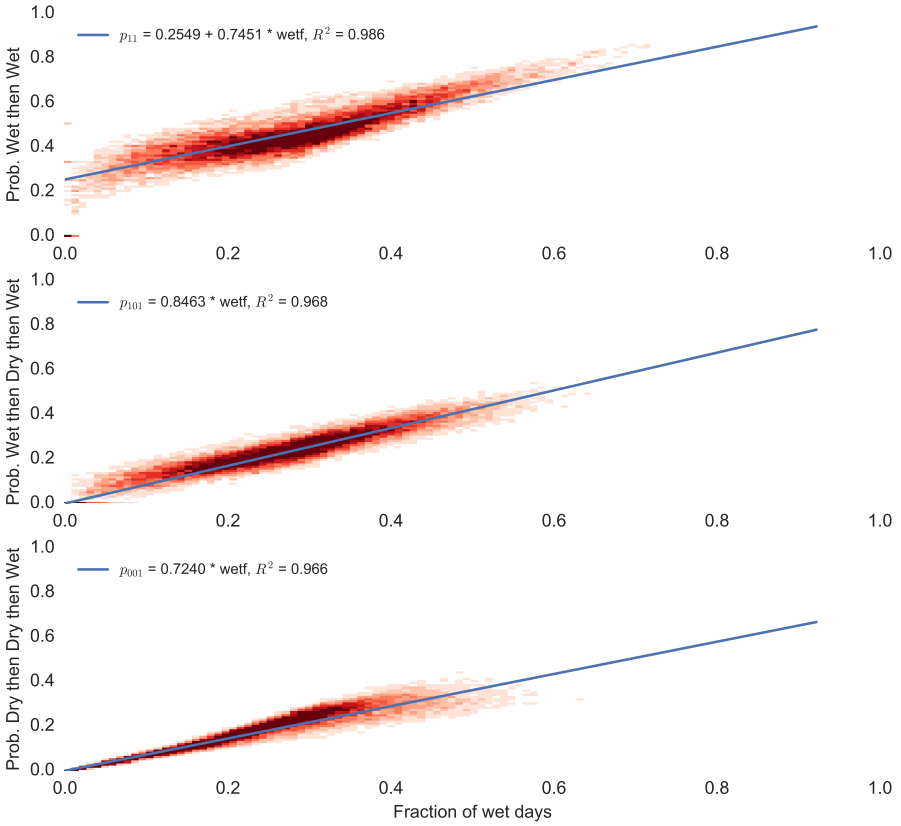


Figure 3. Transition probabilities vs. wet fraction. The red ~~2D histogram~~ density plot in the background shows the density of the observations, and the blue lines are the linear regression line of the probability against the wet fraction. The fit for the p_{11} transition probability was forced to the point $(1,1)$, the others were forced to $(0,0)$. The underlying data for the fits correspond to the means of the the multi-year series for each month for each station.

ables that GWGEN simulates (P, $T_{\min}$, $T_{\max}$, c, w) were recorded on the same dates ~~. This selection~~ in the EECRA database. This merged selection from EECRA and GHCN resulted in a set of ~~489~~ 921 stations representing ca. ~~6~~15 million daily records, with observations on all continents, although the geographic distribution is once again highly heterogenous, with a particularly high density of stations in Japan and Germany (Figure 2, bottom panel).

3.2 Parameterization

3.2.1 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 more accurate than the first order chain used in the original WGEN and many subsequent models, so we adopt this approach in~~

GWGENFollowing Wilks (1999a) we use a hybrid-order model that retains first-order Markov dependence for wet spells but allows second-order dependence for dry sequences; this hybrid-order scheme has been shown to be a good compromise between performance and simplicity. 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 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 with precipitation, which show the characteristic linear relationship described by Geng et al. (1986) (Figure 3).

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. One has to note, however, that this methodology artificially increases the R^2 coefficient for the fit because we fix the intercept (see for example Gordon, 1981).

The analysis results in the the following relationships:

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

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

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

To In the weather generator (see line 6 in algorithm 1) we determine if any given day will have precipitation , by calculating 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.

3.2.2 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. It is based upon the fact that the expected value of a gamma random variable equals the product of its two parameters, i.e $E(\Gamma) = \alpha\theta$. 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

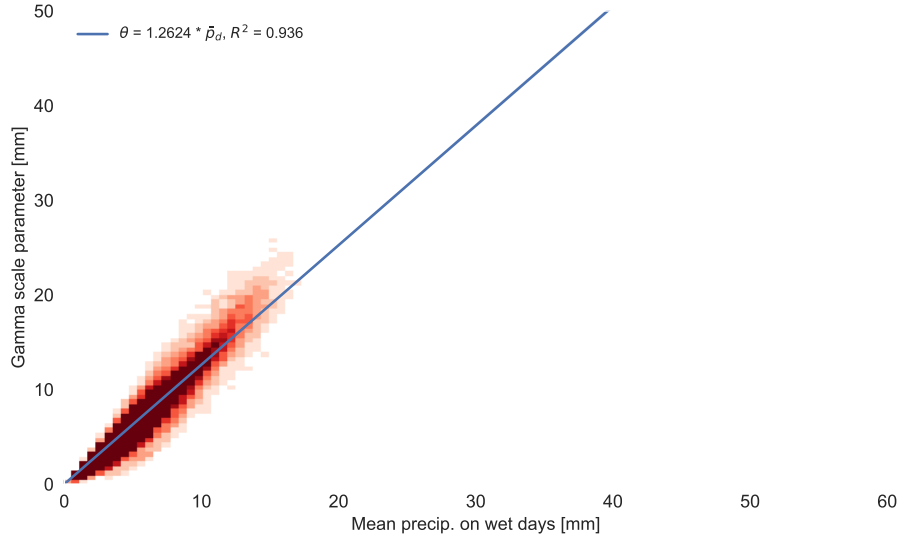


Figure 4. Mean precipitation - Gamma scale relationship. The blue line represents the best fit line of the mean precipitation on wet days to the estimated gamma scale parameter of the corresponding distribution. Each data point corresponds to one multi-year series of one month for one station.

and the generalized pareto (GP) distribution, has superior accuracy in ~~cimulating~~ simulating 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~~ probability density function (pdf) of the gamma distribution is defined as

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

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

$$g(x) = \frac{1}{\sigma}\left(1 + \frac{\xi\,(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} \begin{cases} \frac{1}{\sigma}\left(1+\frac{\xi\,(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} & \text{for } \xi \neq 0 \\ \frac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}} & \text{for } \xi = 0 \end{cases} \tag{5}$$

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~~ pdf 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

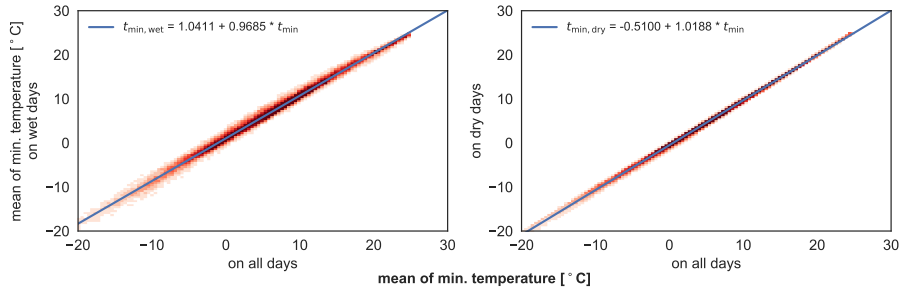


Figure 5. Correlation of minimum temperature on wet and dry days to the monthly mean. The y-axes show the mean minimum temperature on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

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~~ using maximum likelihood to it in order to 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 4) and found the relationship

$$\theta = 1.2681.262\,\bar{p}_d. \tag{7}$$

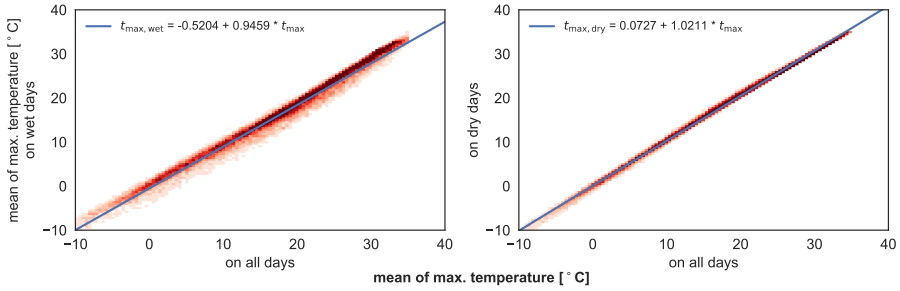
As proposed by Geng et al. (1986), we use this relationship in our model to estimate the scale parameter of the distribution. ~~The~~ Using this approach, the gamma shape parameter α is ~~calculated dynamically in the weather generator via~~ a constant, given via

$$\alpha = \frac{\bar{p}_d}{\theta} = \frac{1}{1.262}. \tag{8}$$

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 3.5.



~~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.~~

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.

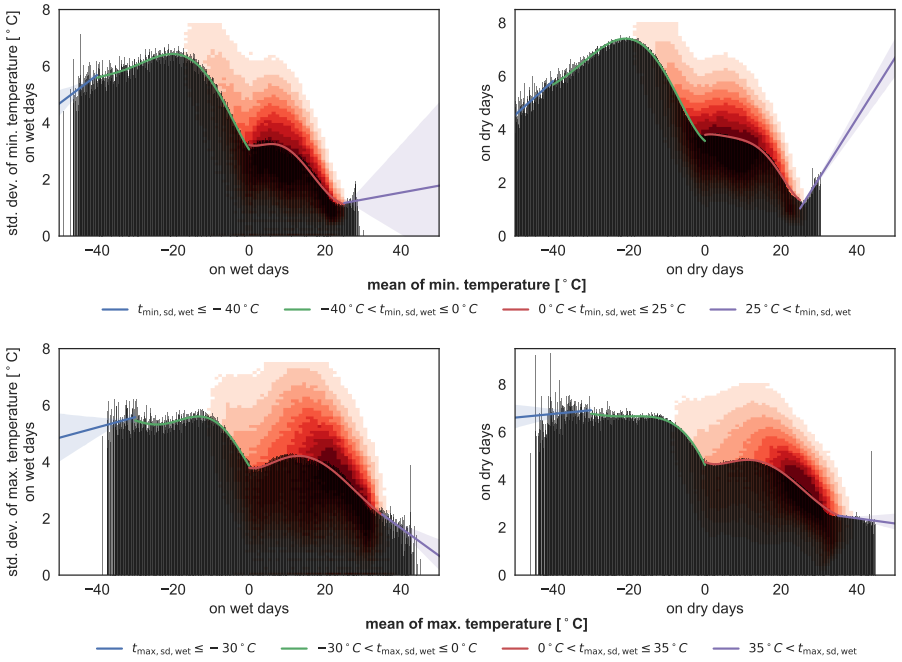


Figure 7. Correlation of standard deviation of the minimum and maximum temperature on wet and dry days to the monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The ~~blue line corresponds~~ bars have a width of $0.1^{\circ}C$ (the data accuracy) and indicate the mean standard deviation for a given mean minimum temperature in one month. The lines are fitted to these bars, where the ~~best fit line~~green and red polynomials of order 5 are the use all the data below or above $0^{\circ}C$ respectively and the blue and violet lines are a linear extrapolation of the data below $-40^{\circ}C$ (or $-30^{\circ}C$ for T_{max}) or above $25^{\circ}C$ (or $35^{\circ}C$) respectively. The red density plot in the background indicates the spread of the data. The bars and the density plot are based on the single month for each station (i.e. not the multi-year monthly series as for, e.g. mean temperature (figure 5 and 6)). Parameters of the fits are also shown in table 1.

Table 1. Fit results of temperature correlation for wet and dry days for ~~figure~~ figures 5~~and~~ , 6, 10 and 11. ~~N is the total number of measurements used for the fit.~~ The ~~values of a and b~~ coefficients c_0 to c_3 correspond to the ~~values~~ coefficients used in ~~equation~~ equations (10) and (14).

plot	variable	~~Intercept~~ R^2	~~Slope~~ c_0	c_1	c_2	c_3
6	$T_{\mathrm{max,dry}}$	~~0.3865~~ 0.9969	~~0.9907~~ 0.0727	~~1.0061~~ 1.0211	0	0
6	$T_{\mathrm{max,wet}}$	~~-0.5863~~ 0.9752	~~0.9539~~ -0.5204	~~0.9487~~ 0.9459	0	0
~~??~~ 5	$T_{\mathrm{min,dry}}$	0.9972	-0.5100	1.0188	0	0
5	$T_{\mathrm{min,wet}}$	0.9840	1.0411	0.9685	0	0
11	$w_{\mathrm{sd,dry}}$	0.4243	0	1.0860	-0.2407	0.0222
11	$w_{\mathrm{sd,wet}}$	0.5003	0	0.8184	-0.1263	0.0093
10	w_{dry}	0.9930	0	0.9437	0	0
10	w_{wet}	0.9723	0	1.0937	0	0

Table 2. Fit results of the correlation of temperature standard deviation with the corresponding mean on wet/dry days for figure 7. The underlying equations are shown in equation (11).

variable	interval	R^2	c_0	c_1	c_2	c_3	c_4	c_5
$T_{\mathrm{max,sd,dry}}$	~~4.4700~~ (-∞, -30]	~~0.0722~~ 0.0125	~~-0.0387~~ 7.3746	0.0154	0	0	0	0
~~??~~	(-30, 0.0]	0.6721	4.6170	-0.3387	-0.0188	-0.0003	0.000003	0.0000001
	(0.0, 35]	0.9744	4.7455	-0.0761	0.0189	-0.0013	0.00003	-0.0000002
	(35, ∞)	0.0390	3.2554	-0.0218	0	0	0	0
$T_{\mathrm{max,sd,wet}}$	~~3.9159~~ (-∞, -30]	~~0.0407~~ 0.0366	~~-0.0294~~ 6.6720	0.0364	0	0	0	0
~~5~~	~~$T_{\mathrm{min,dry}}$~~ (-30, 0.0]	~~-0.5283~~ 0.7362	~~0.9940~~ 3.8601	~~1.0210~~ -0.2186	0.0039	0.0015	0.00006	0.0000007
~~5~~	~~$T_{\mathrm{min,wet}}$~~ (0.0, 35]	~~1.1647~~ 0.9508	~~0.9733~~ 3.7919	~~0.9558~~ -0.0313	0.0161	-0.0012	0.00003	-0.0000002
~~??~~	(35, ∞)	0.2530	5.5529	-0.0973	0	0	0	0
$T_{\mathrm{min,sd,dry}}$	~~3.5449~~ (-∞, -40]	~~0.1044~~ 0.6006	~~-0.0428~~ 10.8990	0.1271	0	0	0	0
~~??~~	~~$T_{\mathrm{min,sd,wet}}$~~ (-40, 0.0]	~~3.0388~~ 0.9509	~~0.1386~~ 3.5676	~~-0.0505~~ -0.1154	0.0282	0.0020	0.00004	0.0000003
~~11~~	~~$w_{\mathrm{sd,dry}}$~~ (0.0, 25]	~~0~~ 0.9825	~~0.8959~~ 3.7941	~~0.5120~~ 0.0330	-0.0150	0.0019	-0.0001	0.000002
~~11~~	~~$w_{\mathrm{sd,wet}}$~~ (25, ∞)	0.7784	-4.6194	0.2261	0	~~0.8994~~ 0	~~0.4404~~ 0	0
~~10~~ ~~$T_{\mathrm{min,sd,wet}}$~~	~~w_{dry}~~ (-∞, -40]	0.1661	9.7272	0.1011	0	~~0.9933~~ 0	~~0.9452~~ 0	0
~~10~~	~~w_{wet}~~ (-40, 0.0]	0.9285	3.0550	-0.2116	0.0137	0.0014	0.00004	0.0000003
	(0.0, 25]	0.9633	3.2187	-0.0451	0.0209	-0.0026	0.00010	-0.000001
	(25, ∞)	0.0089	0.5571	0.0244	0	~~0.9720~~ 0	~~1.0929~~ 0	0

3.2.3 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 3.2.6). ~~For minimum temperature, we additionally perform a bias correction ().~~

In the weather generator, we know from the Markov chain (subsubsection 3.2.1), whether the current simulated day is a wet or dry day. Based upon the simple linear relationships

$$
\begin{aligned}
x\bar{x}_{\text{wet}} &= a_{x,\text{wet}} c_{0,x,\text{wet}} + b_{x,\text{wet}} c_{1,x,\text{wet}} \cdot \bar{x} \\
x\bar{x}_{\text{dry}} &= a_{x,\text{dry}} c_{0,x,\text{dry}} + b_{x,\text{dry}} c_{1,x,\text{dry}} \cdot \bar{x}
\end{aligned}
\tag{10}
$$

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 $x \in \{T_{\min}, T_{\max}\}$:~~

$$
\begin{aligned}
\sigma_{x,\text{wet}} &= a_{\sigma,x,\text{wet}} + b_{\sigma,x,\text{wet}} \cdot x_{\text{wet}} \\
\sigma_{x,\text{dry}} &= a_{\sigma,x,\text{dry}} + b_{\sigma,x,\text{dry}} \cdot x_{\text{dry}}.
\end{aligned}
$$

To estimate the values of the parameters ~~a and b~~ c_0 and c_1 in the above equations, we follow the same procedure as for the parameters of the Markov chain (subsubsection 3.2.1). 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 5 and 6). ~~We further regressed the standard deviation on wet and dry days against its corresponding mean (Figures ?? and ??).~~ Through this procedure, we estimate the parameters necessary for equations (10) ~~and~~ (see table 1). ~~As can be seen in figures ?? and ?? the linear model does not really satisfy the complex behaviour~~

To estimate residual noise, we also need an estimate of the standard deviation of the variable (see subsubsection 3.2.6). Figure 7 shows the correlation between standard deviation on wet and dry days and the corresponding mean. The means of the standard ~~deviation. However, since this value is only used to add random noise (see section 2~~deviations (black bars in figure 7) indicate a strong but non-linear relationship between the standard deviation and the corresponding mean. The correlation changes particularly at $0°\,C$. We therefore use two different polynomials of order 5 for the values below and above the freezing point. Furthermore, to account for the sparse data below $-40°\,C$ and above $25°\,C$ for minimum temperature (or $-30°\,C$ and $35°\,C$ for maximum temperature), we ~~believe that the error introduced by the poor linear fit is negligible~~use an extrapolation for the extremes as indicated by the blue and violet lines in figure 7. The formulae for the standard deviations σ of minimum and maximum temperature are therefore a combination of 4 polynomials:

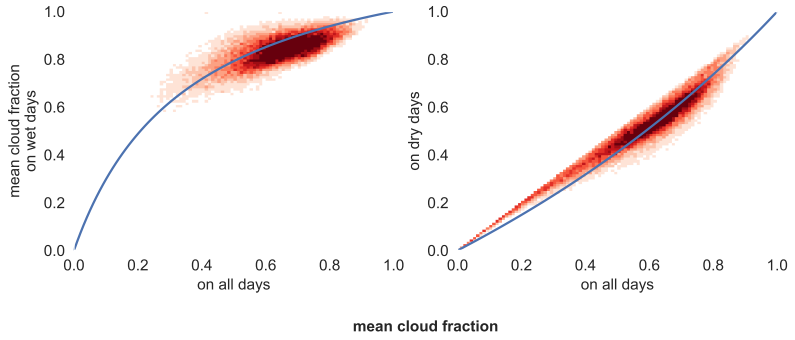


Figure 8. Correlation of cloud fraction on wet and dry days to the monthly mean.The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 3.

$$\sigma_{T_{\min,\mathrm{wet/dry}}} = \begin{cases} p_1(\bar{T}_{\min,\mathrm{wet/dry}}), & \text{for } \bar{T}_{\min,\mathrm{wet/dry}} \leq -40^\circ C \\ p_5(\bar{T}_{\min,\mathrm{wet/dry}}), & \text{for } -40^\circ C < \bar{T}_{\min,\mathrm{wet/dry}} \leq 0^\circ C \\ p_5(\bar{T}_{\min,\mathrm{wet/dry}}), & \text{for } 0^\circ C < \bar{T}_{\min,\mathrm{wet/dry}} \leq 25^\circ C \\ p_1(\bar{T}_{\min,\mathrm{wet/dry}}), & \text{for } 25^\circ C < \bar{T}_{\min,\mathrm{wet/dry}} \end{cases}$$

$$\sigma_{T_{\max,\mathrm{wet/dry}}} = \begin{cases} p_1(\bar{T}_{\max,\mathrm{wet/dry}}), & \text{for } \bar{T}_{\max,\mathrm{wet/dry}} \leq -30^\circ C \\ p_5(\bar{T}_{\max,\mathrm{wet/dry}}), & \text{for } -30^\circ C < \bar{T}_{\max,\mathrm{wet/dry}} \leq 0^\circ C \\ p_5(\bar{T}_{\max,\mathrm{wet/dry}}), & \text{for } 0^\circ C < \bar{T}_{\max,\mathrm{wet/dry}} \leq 35^\circ C \\ p_1(\bar{T}_{\max,\mathrm{wet/dry}}), & \text{for } 35^\circ C < \bar{T}_{\max,\mathrm{wet/dry}} \end{cases} . \quad (11)$$

p_1 in eq. (11) denotes a polynomial of order 1, p_5 a polynomial of order 5. The coefficients of the different polynomials are shown in table 2.

These coefficients are based on the means of the standard deviation (black bars in figure 7). We chose this procedure to give the same weight to all temperatures. Otherwise the fit would be dominated by the temperature values around the freezing points.

3.2.4 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

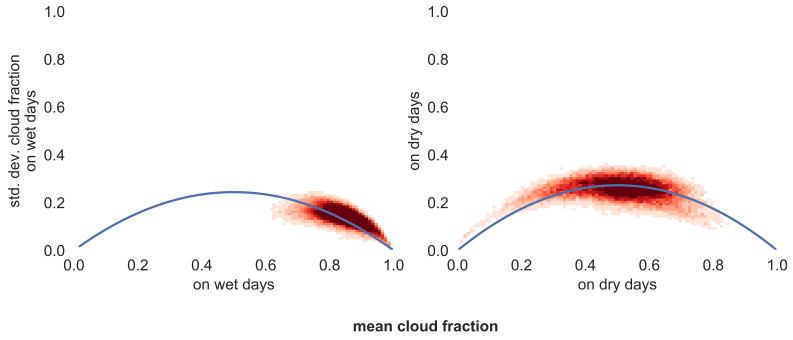


Figure 9. Correlation of standard deviation of the cloud fraction on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line. Parameters of the fits are also shown in table 3.

Table 3. Fit results of cloud correlation for wet and dry days for figure 8

plot	variable	~~a~~ a	std. ~~dev. of a~~ err. of a	R^2
8	c_{dry}	~~0.4205~~ 0.4302	~~0.0011~~ 0.0013	0.8745
8	c_{wet}	~~-0.7383~~ -0.7376	~~0.0005~~ 0.0006	0.3881
9	$c_{\mathrm{sd,dry}}$	~~1.0417~~ 1.0448	~~0.0003~~ 0.0004	0.2803
9	$c_{\mathrm{sd,wet}}$	~~0.9819~~ 0.9881	~~0.0005~~ 0.0006	0.0802

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.

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~~ used a qualitative graphical analysis to develop "best guess" equations that had the desired shape and propose the following formulae for the regression linking cloud cover on wet or dry days to the overall mean:

$$
\begin{aligned}
\underline{c}\bar{c}_{\mathrm{wet}} &= \frac{-a_{\mathrm{c,wet}}-1}{a_{\mathrm{c,wet}}^2 * c - a_{\mathrm{c,wet}}^2 - a_{\mathrm{c,wet}}} \frac{-a_{\mathrm{c,wet}}-1}{a_{\mathrm{c,wet}}^2 \cdot \bar{c} - a_{\mathrm{c,wet}}^2 - a_{\mathrm{c,wet}}} - \frac{1}{a_{\mathrm{c,wet}}} \\
\underline{c}\bar{c}_{\mathrm{dry}} &= \frac{-a_{\mathrm{c,dry}}-1}{a_{\mathrm{c,dry}}^2 * c - a_{\mathrm{c,dry}}^2 - a_{\mathrm{c,dry}}} \frac{-a_{\mathrm{c,dry}}-1}{a_{\mathrm{c,dry}}^2 \cdot \bar{c} - a_{\mathrm{c,dry}}^2 - a_{\mathrm{c,dry}}} - \frac{1}{a_{\mathrm{c,dry}}}
\end{aligned}
\tag{12}
$$

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

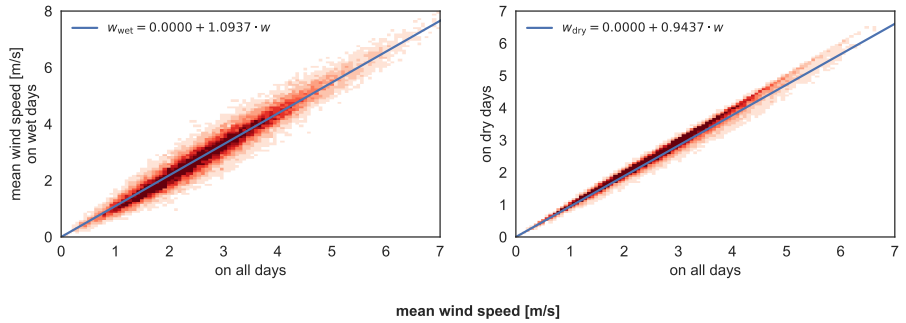


Figure 10. Correlation of wind speed on wet and dry days to the monthly mean.The y-axes show the mean cloud fraction on wet or dry days respectively, the blue line corresponds to the best fit line. Parameters of the fits are also shown in table 1.

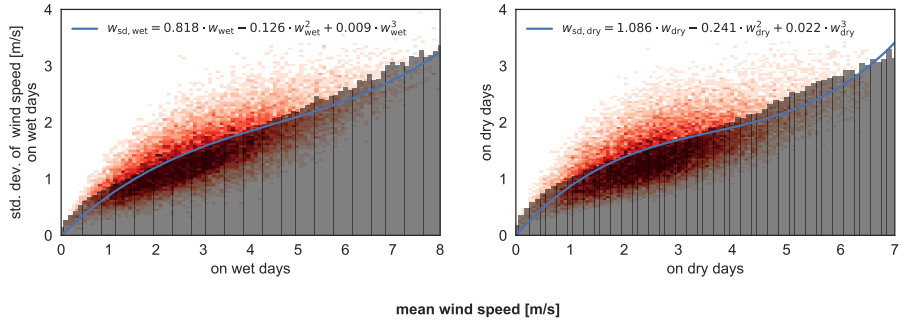


Figure 11. Correlation of standard deviation of ~~the~~ wind speed on wet and dry days to the corresponding monthly mean. The y-axes show the standard deviation, the x-axes the mean on wet or dry days respectively. The blue line corresponds to the best fit line, a third order polynomial to the underlying red density plot. The black bars have a width of $0.1\,\mathrm{m\,s^{-1}}$, the accuracy of the input data, and indicate the mean standard deviations for the given interval range. Parameters of the fits are also shown in table 1.

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

$$\sigma_{\mathrm{c,wet}} = a_{\mathrm{c,wet}}^2 \cdot \underline{\mathrm{c}}\bar{\mathrm{c}}_{\mathrm{wet}} \cdot (1 - \underline{\mathrm{c}}\bar{\mathrm{c}}_{\mathrm{wet}})$$
$$\sigma_{\mathrm{c,dry}} = a_{\mathrm{c,dry}}^2 \cdot \underline{\mathrm{c}}\bar{\mathrm{c}}_{\mathrm{dry}} \cdot (1 - \underline{\mathrm{c}}\bar{\mathrm{c}}_{\mathrm{dry}}) \quad (13)$$

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 3.

3.2.5 Wind speed

The parameterization of the mean wind speed is based upon the same ~~equations~~ linear equation (10) ~~and as temperatureand~~ as temperature. For the standard deviation however, we use a third-order polynomial given that is forced through the origin, given

via

$$\sigma_{w,\mathrm{wet}}(\bar{w}_{\mathrm{wet}}) = c_{1,w,\mathrm{wet}}\,\bar{w}_{\mathrm{wet}} + c_{2,w,\mathrm{wet}}\,\bar{w}^2_{\mathrm{wet}} + c_{3,w,\mathrm{wet}}\,\bar{w}^3_{\mathrm{wet}}$$

$$\sigma_{w,\mathrm{dry}}(\bar{w}_{\mathrm{dry}}) = c_{1,w,\mathrm{dry}}\,\bar{w}_{\mathrm{dry}} + c_{2,w,\mathrm{dry}}\,\bar{w}^2_{\mathrm{dry}} + c_{3,w,\mathrm{dry}}\,\bar{w}^3_{\mathrm{dry}}. \quad (14)$$

This better resolves the complex behavior close to $0\,\mathrm{m\,s^{-1}}$ compared to a linear fit. The plots are 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 a square root transformation to the monthly mean windspeed values before performing the regression fitting (see line 20 in the model algorithm 1)~~and the parameters for the fits are shown in table 1.

3.2.6 Cross correlation

Following Richardson (1981) we use cross correlation to add additional residual noise to the simulated meteorological variables, which provides more realism in the daily weather result. ~~As above, windspeed is subject to a square-root transformation before calculating the cross correlations.~~ This methodology, based on Matalas (1967) preserves the serial and the cross correlation between the simulated variables. It implies that the serial correlation of each variable may be described by a first-order linear autoregressive model

~~Using the EECRA database (see figure 2) we calculate~~

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

Given the cross correlation matrix $M_0 \in \mathbb{R}^4 \times \mathbb{R}^4$ and the lag-1 correlation matrix $M_1 \in \mathbb{R}^4 \times \mathbb{R}^4$, we calculate

$$A = M_1 M_0^{-1} \qquad BB^T = M_0 - M_1 M_0^{-1} M_1^T. \quad (15)$$

~~where the~~ The matrices A, B, M_0 and M_1 are calculated using the stations from the EECRA database in figure 2. The results are

$$M_0 = \begin{pmatrix} 1. & 0.565 & 0.041 & 0.035 \\ 0.565 & 1. & -0.089 & -0.043 \\ 0.041 & -0.089 & 1. & 0.114 \\ 0.035 & -0.043 & 0.114 & 1. \end{pmatrix} \qquad M_1 = \begin{pmatrix} 0.933 & 0.55 & 0.016 & 0.03 \\ 0.557 & 0.417 & -0.066 & -0.043 \\ 0.004 & -0.095 & 0.599 & 0.093 \\ 0.011 & -0.063 & 0.061 & 0.672 \end{pmatrix}. \quad (16)$$

leading to

$$A = \begin{pmatrix} 0.916 & 0.031 & -0.018 & 0.001 \\ 0.485 & 0.135 & -0.069 & -0.047 \\ 0.004 & -0.043 & 0.592 & 0.023 \\ 0.012 & -0.043 & -0.02 & 0.672 \end{pmatrix} \qquad B = \begin{pmatrix} 0.358 & 0. & 0. & 0. \\ 0.112 & 0.809 & 0. & 0. \\ 0.142 & -0.06 & 0.785 & 0. \\ 0.077 & -0.016 & 0.061 & 0.733 \end{pmatrix}. \tag{17}$$

The columns and rows in the two matrices correspond to min. and max. temperature, cloud fraction and square root of wind speed, respectively.

~~The above matrices where~~ In the weather generator, the variables $T_{\min}, T_{\max}, c$ and w are then calculated using a combination of residual noise χ_i (where i denotes the current simulated day) and the mean of the variables. χ_i is determined by the other variables and the previous day using A and B from above (Richardson, 1981; Matalas, 1967). Hence, χ_i is given via

$$\chi_i = \begin{pmatrix} \chi_{T_{\min}} & \chi_{T_{\max}} & \chi_c & \chi_w \end{pmatrix} = A\chi_{i-1} + B\epsilon \in \mathbb{R}^4. \tag{18}$$

The daily values for the variables are then calculated via

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

$$T_{\min,i} = \chi_{T_{\min}} \cdot \sigma_{T_{\min},\text{wet/dry}} + \bar{T}_{\min,\text{wet/dry}} \qquad c_i = \chi_c \cdot \sigma_{c,\text{wet/dry}} + \bar{c}_{\text{wet/dry}} \tag{19}$$

$$T_{\max,i} = \chi_{T_{\max}} \cdot \sigma_{T_{\max},\text{wet/dry}} + \bar{T}_{\max,\text{wet/dry}} \qquad w_i = \left(\chi_w \cdot \sqrt{\sigma_{w,\text{wet/dry}}} + \sqrt{\bar{w}_{\text{wet/dry}}}\right)^2 \tag{20}$$

~~from the lag-0 and lag-1 covariance matrices~~ with $\sigma_{T_{\min},\text{wet/dry}}, \sigma_{T_{\max},\text{wet/dry}}$ from eq. (11), $\sigma_{c,\text{wet/dry}}$ from eq. (13), $\sigma_{w,\text{wet/dry}}$ from eq. (14), $\bar{T}_{\min,\text{wet/dry}}, \bar{T}_{\max,\text{wet/dry}}, \bar{w}_{\text{wet/dry}}$ from eq. (10) and $\bar{c}_{\text{wet/dry}}$ from eq. (12).

Since this procedure always requires the residuals from the previous day, χ_{i-1}, we initialize χ_0 with 0, simulate the month and then simulate it again.

Note that, through the entire procedure, wind speed is subject to a square-root transformation (also when calculating M_0 and M_1 ~~with~~) to account for the fact that it is not normally distributed.

$$M_0 = \begin{pmatrix} 1. & 0.572 & 0.025 & 0.032 \\ 0.572 & 1. & -0.101 & -0.045 \\ 0.025 & -0.101 & 1. & 0.127 \\ 0.032 & -0.045 & 0.127 & 1. \end{pmatrix} \qquad M_1 = \begin{pmatrix} 0.932 & 0.557 & -0.001 & 0.025 \\ 0.564 & 0.426 & -0.08 & -0.046 \\ -0.012 & -0.108 & 0.6 & 0.104 \\ 0.006 & -0.066 & 0.071 & 0.667 \end{pmatrix}.$$

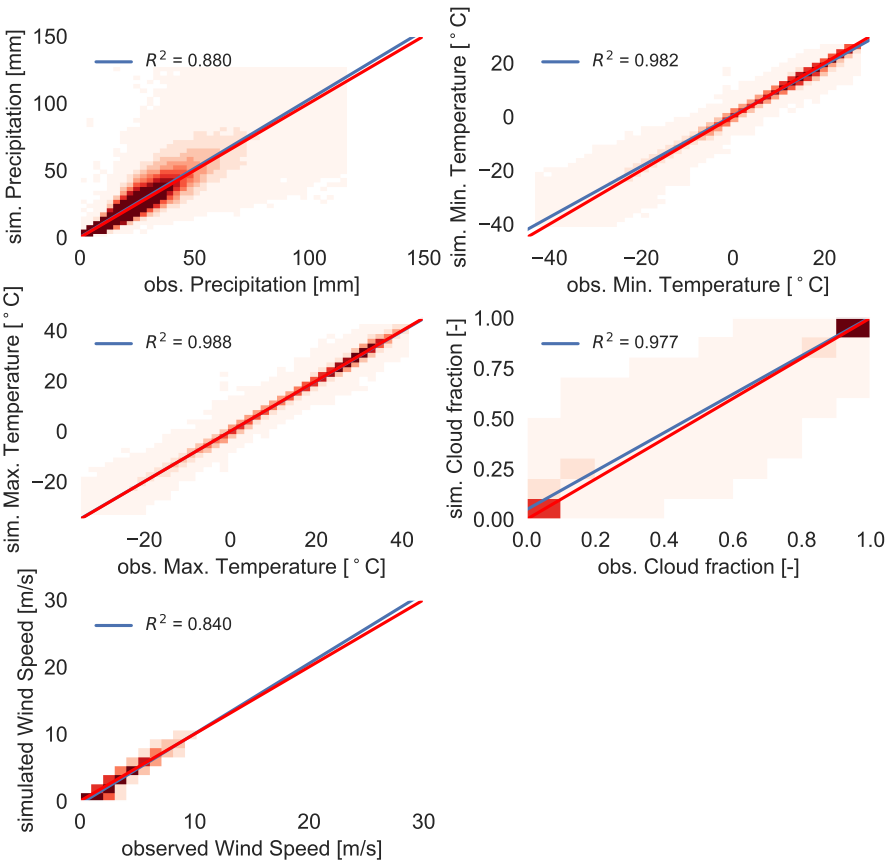


Figure 12. QQ-plots for all variables with all quantiles (1, 5, 10, 25, 50, 75, 90, 95 and 99) for ~~$\mu = \ldots$ mm~~ $\mu = 5.0\,\mathrm{mm}\,\mathrm{mm}$, ~~$\xi = \ldots$~~ $\xi = 1.5$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval. ~~Plots~~ The plots compares the simulated quantile from the list above of one year of one station to the corresponding observed quantile of the same year and station. The plot for wind speed ~~and minimum temperature~~ underwent used the bias correction from subsection 3.4.

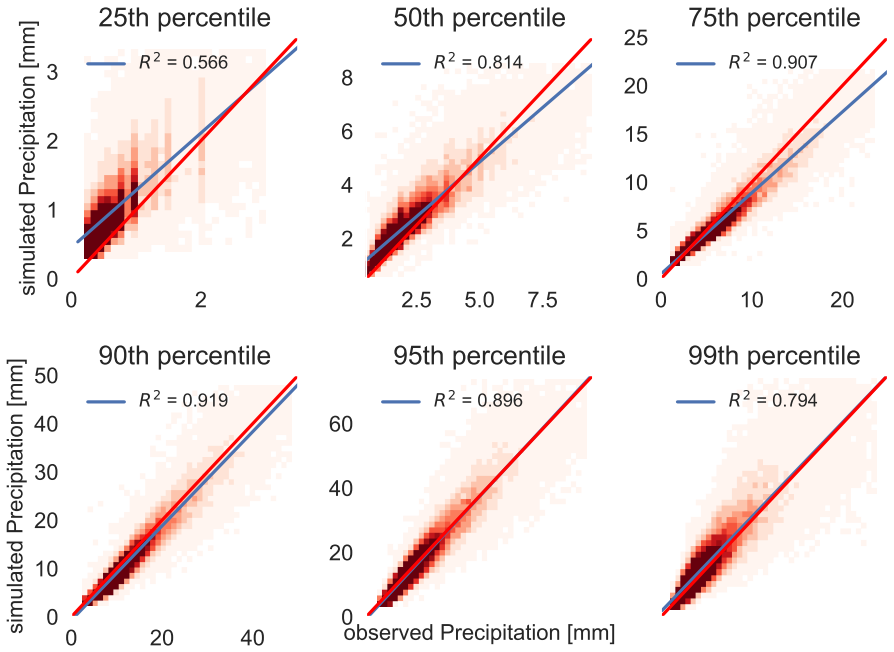


Figure 13. QQ-plot for different quantiles for precipitation for ~~$\mu = 15\,\mathrm{mm}$~~$\mu = 5.0\mathrm{mm}$, ~~$\xi = 0.08303$~~$\xi = 1.5$. The blue lines are linear regression from simulation to observation. The red line shows the ideal fit (the identity line). Blue shaded areas represent the 95% confidence interval. The plots compare the simulated quantile of one year of one station to the corresponding observed quantile of the same year and station.

3.3 Model Evaluation

To evaluate GWGEN, we started with the daily meteorology at the evaluation stations described above and calculated monthly summaries. We used ~~these~~ this monthly data to drive the model and simulate daily meteorology. The resulting daily series now has the same length as the observed meteorology from the GHCN and EECRA database. 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 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

[1]Note that the ~~plots~~ plot for wind speed ~~and minimum temperature were~~ has been bias corrected using the approach in subsection 3.4.

Table 4. 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~~ Precip. range [mm]	Simulated	Observed
(0, 1]	~~0.216~~ 0.285688	~~0.270~~ 0.364014
(1, 10]	~~0.550~~ 0.583330	~~0.470~~ 0.486415
(10, 20]	~~0.115~~ 0.074063	~~0.133~~ 0.090178
(20, ∞~~)~~]	~~0.120~~ 0.056920	~~0.127~~ 0.059392

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 ($0.1\,\mathrm{mm}$ for precipitation and ~~$0.1\,\mathrm{m/s}$~~ $0.1\,\mathrm{ms}^{-1}$ for wind speed). Very small precipitation amounts and low ~~windspeeds are also ecologically not as important as~~ wind speeds are also less biophysically and ecologically important compared to the higher percentiles. We therefore consider the results of the evaluation largely acceptable.

In table 4 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~~ occurrence of very light rain events (~~21.6% instead of 27.0%~~28.6% instead of 36.4%) and overestimates the light rain events (~~55% instead of 47%~~58.3% instead of 48.6%) but generally performs much better than GCMs (Dai, 2006; Sun et al., 2006), especially when it comes to the heavy rain events.

3.4 Bias correction

After evaluating the results of GWGEN for wind speed ~~and minimum temperature~~ for the different quantiles (see previous subsection 3.3) 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~~ 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 databased upon the modeled 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~~

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

~~where $u \in \mathbb{R}$ corresponds to the number from the normal distribution drawn~~ to a posteriori correct the simulated data. In the quantile evaluation (previous subsection 3.3) we saw that the simulated wind speed is a linear function of the observed wind speed, i.e. $w_{sim} = \text{intercept} + \text{slope} \cdot w_{obs}$ (best fit line in figure 12). Therefore, we use two steps here, one is for the difference between simulation and observation (ideally 0), the other one is the fraction of observation and simulation (ideally 1). The first

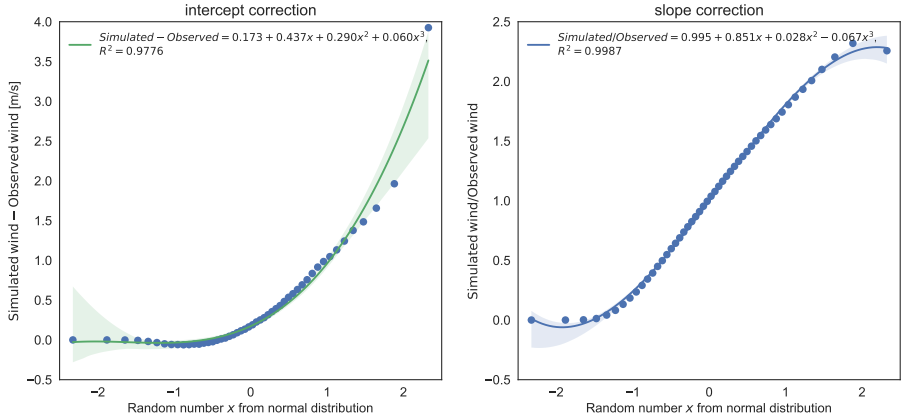


Figure 14. Basis for the ~~minimum temperature (left) and~~ wind bias correction~~(right)~~. For the left plot~~(min. Temperature)~~, each data point corresponds to the difference of a simulated percentile to the observed percentile. For the right plot (wind speed), each data point corresponds to the 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).

one corresponds to the intercept with the y-axis in figure 12, the second one to the slope of the best fit line. The analysis is based on every second percentile between 1 and 100 (i.e. $1, 3, 5, \ldots$) and mapped to it's corresponding random number $u \in \mathbb{R}$ from a normal distribution as it is used for the cross correlation in the weather generator (section 3.2.6, x-axis in figure 14 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~~ Regarding the intercept (fig. 14, left) we see that it strongly follows an exponential function given through

$$f_{exp}(u) = \mathrm{e}^{a\,u+b}, \qquad a, b, u \in \mathbb{R}. \tag{21}$$

The slope (fig. 14, right) on the other hand can be described by a simple third-order polynomial given by

$$p3(u) = c_0 + c_1\,u + c_2\,u^2 + c_3\,u^3, \qquad c_0, c_1, c_2, c3, u \in \mathbb{R} \tag{22}$$

Hence, given the best fit lines in figure 14, the simulated wind speed is corrected via

$$w'_{sim} = w_{\text{biased}}\ f(u).\frac{w_{sim} - f_{exp}(u)}{p3(u)} \tag{23}$$

with $a = 1.1582, b = -1.3359, c_0 = 0.9954, c_1 = 0.8508, c_2 = 0.0278, c_3 = -0.0671$.

3.5 Sensitivity analysis

The Generalized-Pareto part of the hybrid Gamma-GP distribution, which 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~~ determine 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 ~~zz~~13 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 20 in steps of 2.5 and 20 to 100 in steps of 5. Furthermore we varied the GP shape parameter from ~~0~~ 0.1 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~~ low to intermediate thresholds and a low 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%~~ 81.7% 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~~ 0.81 and the mean deviation of the slope from unity is ~~0.18~~ 0.10 and for the upper quantiles (90 to 100), ~~0.1~~0.017.

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%~~73% to only 83%. It is however better than the gamma distribution alone

with only 76.2% which still has 78.6% of station years not differing significantly and but with a slope deviation from unity for the upper quantiles of 0.26.

4 Description of the final model

The parameterization described in section 3.2 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 .

The first approximation of the daily variables comes from smoothing the monthly time series using 0.16. Thus using the hybrid Gamma-GP distribution improves the simulation of high-amount precipitation events by roughly factor 10 compared to a mean-preserving algorithm (Rymes and Myers, 2001). standard Gamma approach.

For precipitation we then first use the markov chain approach from section 3.2.1 to decide the wet/dry state of the day. If the day is a wet day, we calculate the gamma parameters using the equations and . 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 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 11 and 15 in algorithm 1). After that, we use the cross-correlation approach described in Richardson (1981) and lines 18 - 20 and calculate the daily values of these variables. Finally we use the quantile-based bias correction described in section 3.4 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 mm). If the program cannot produce these results, the procedure described above is repeated (see line 4).

Basic workflow of GWGEN

4 Discussion, applications, limitations, and outlook

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 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~~ The GP distribution is based upon a globally fixed shape and location parameter, which may be an oversimplification, but is still ten times more accurate than traditional methods that used Gamma alone. Our extensive sensitivity analysis to determine the best coefficients for the shape and location parameters of the GP distribution ~~,~~ suggest that further improvements might come through correlating the GP parameters to geographic region and/or seasonality (Maraun et al., 2009; Rust et al., 2009) or by introducing a dynamical location parameter (Frigessi et al., 2002). 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., regional high-resolution 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).

5 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.

6 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~~ https://github.com/ARVE-Research/gwgen/releases/tag/v1.0.1[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

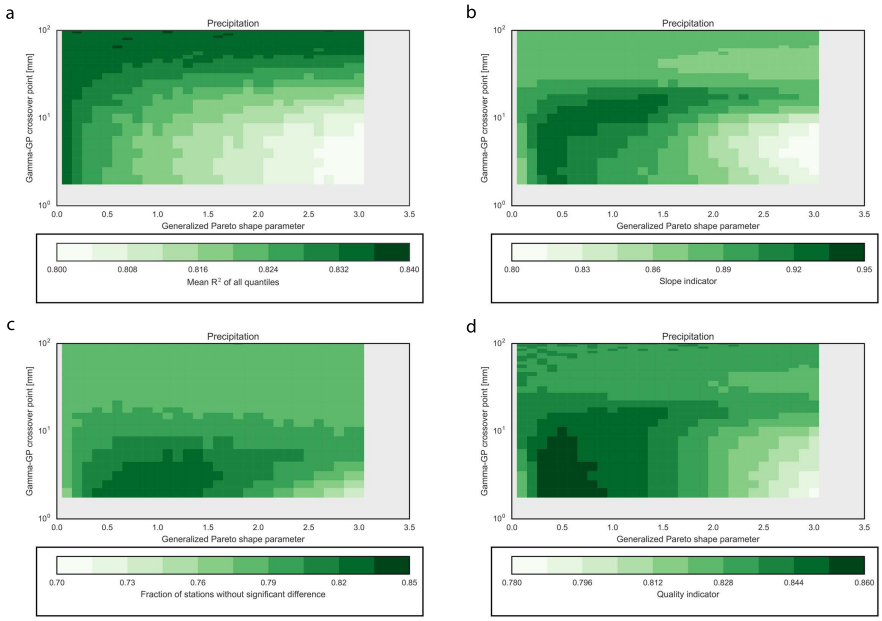


Figure 15. Results of the sensitivity analysis for the (a) correlation coefficient R^2, (b) deviation from a slope of unity, (c) the fraction of significant different station years, (d) the mean of (a) - (c). For the plots in (a) and (b) we used the means of the 25th, 50th, 75th, 90th, 95th and 99th 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.

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.

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, 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.

Frigessi, A., Haug, O., and Rue, H.: A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection, Extremes, 5, 219–235, doi:10.1023/A:1024072610684, http://dx.doi.org/10.1023/A:1024072610684, 2002.

Furrer, E. M. and Katz, R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators, Water Resources 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:10.1016/0168-1923(86)90014-6, <GotoISI>://WOS:A1986C086500007, 1986.

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.

Gordon, H. A.: Errors in Computer Packages. Least Squares Regression Through the Origin, Journal of the Royal Statistical Society. Series D (The Statistician), 30, 23–29, http://www.jstor.org/stable/2987701, 1981.

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: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.

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.

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 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.

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.

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

Matalas, N. C.: Mathematical assessment of synthetic hydrology, Water Resources Research, 3, 937–945, doi:10.1029/WR003i004p00937, http://dx.doi.org/10.1029/WR003i004p00937, 1967.

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 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 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: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: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: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: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: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.