Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-42-AC1, 2017


Geoscientific Model Development Discussions

# *Interactive comment on* "A globally calibrated scheme for generating daily meteorology from monthly statistics: Global-WGEN (GWGEN) v1.0" *by* Philipp S. Sommer and Jed O. Kaplan

**Philipp S. Sommer and Jed O. Kaplan**

philipp.sommer@unil.ch

Received and published: 30 June 2017

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 clar-



ity 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

$$P(\text{wet}) = \frac{\#\text{wet days}}{\#\text{total days}}, \tag{1}$$

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 $\alpha$ and $\theta$ 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 $\alpha$ 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 $\mu$ 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 $\mu$ are independent and identically distributed. A rather difficult to satisfy assumption. If the level $\mu$ 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 $\xi, \mu$ 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 $\xi$ and $\mu$ 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.

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-42, 2017.