Hydrol. Earth Syst. Sci. Discuss., 11, C3423–C3427, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C3423/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Stochastic modelling of spatially and temporally consistent daily precipitation time-series over complex topography" *by* D. E. Keller et al.

Anonymous Referee #1

Received and published: 28 August 2014

REVIEW OF THE PAPER ENTITLED "Stochastic modelling of spatially and temporally consistent daily precipitation time-series over complex topography" BY D.E. Keller, A.M. Fischer, C. Frei, M.A. Liniger, C. Appenzeller, and R. Knutti

In this paper, the authors develop a stochastic weather generator (WG), which aims at mimicking precipitation (mean behavior, variability, space-time properties, etc.) over the catchment of the river Thur in the Swiss Alps. The proposed multi-site WG is calibrated using daily data collected at 8 monitoring stations during 51 years, and is shown to reproduce more or less adequately several statistics of interest, such as annual cycles, multi-day spells, areal daily sums, etc.

The core of this multi-site model is based on a Richardson-type WG, describing precipitation intermittence and amounts using 1st-order Markov chains with spatially correlation parameters, and simulating non-zero precipitation amounts with a mixture of exponential variables. This temporal and spatial dependence structure permits to capture space-time interactions reasonably well.

I start with a few general considerations about the paper. Below are more detailed comments about specific aspects that, in my point of view, should be improved (maybe in a future work), or at least discussed in the paper. This is followed by some final technical comments.

GENERAL CONSIDERATIONS:

(1) I would like to congratulate the authors for this very nicely and clearly written paper. The latter is very easy to read, which makes it accessible to statisticians/non-statisticians and hydrologists/non-hydrologists. Furthermore, the authors have done a great effort in producing beautiful and clear graphs, which are useful to validate (or invalidate) various features of the model.

(2) As already mentioned in the paper, the proposed model is not novel (see the references Richardson, 1981 and Wilks, 1998), but the novelty resides in the application (Swiss alps data). The paper is also interesting for its the detailed analysis, and the general discussion of the WG implemented.

SPECIFIC COMMENTS:

(1) This model captures the mean behavior relatively correctly, but does not capture extreme events properly. The mixture of exponential variables, which describe precipitation amounts, is not only unable to reproduce observed extremes (see Fig. 9), but it is certainly even worse at predicting future (non-observed) extremes, owing to the light tail of the assumed exponential variable. . . The model advocated by Vrac and Naveau (2007), which uses a generalized Pareto distribution (GPD) for the tail, has an

asymptotic justification and thus much better at estimating probabilities for rare events. If the goal of the proposed WG is to be used in impact and risk assessment studies (as claimed several times in the paper), possibly in a climate change context, I think that extreme events should be better captured.

(2) The proposed WG simulates stationary time series (month by month). Does this make sense over a period of 51 years (and for the near future)? A non-stationary model could perhaps explain part of the inter-annual variability, that is not captured in Fig 4. And also, again, if the goal is to use this WG for risk and impact assessment, does it make sense to assume that climate is stationary? Can the WG be used to extrapolate in the future?

(3) The model captures spatial coherency between monitoring stations, but does not describe spatial dependence at ungauged locations. Therefore, it is impossible to simulate precipitation data over the whole catchment, which may be essential for risk assessment (e.g., if the simulated data are needed as input of a hydrological rainfall-runoff model). All pairwise correlations between the different stations are computed empirically, although there is a large geostatistics literature about Gaussian processes, (stationary or non-stationary) correlation functions, etc. Why not fit a correlation function to the data, which would: - allow simulation over the whole catchment, - decrease the number of parameters drastically (therefore also the uncertainty) by exploiting the inherent spatial structure, - automatically yield positive definite correlation matrices (without any adjustment), - easily generalize to many more time series?

(4) The number of parameters in the model is very large, so simplifications should be considered, e.g., by - fitting a correlation function (see point (3)). This would decrease the number of spatial dependence parameters, though spatial heterogeneity might be difficult to take into account. - fitting a global yearly model, for example using splines or sines/cosines (instead of fitting separately one model per month). This would decrease the number of temporal dependence parameters, and yield a coherent model throughout the whole year.

(5) The Markov chain of order 1 does not capture well multi-day spells (see Fig. 7, for example). Maybe, a 2nd-order Markov chain (AR(2)) would do a better job... Or maybe a model of type 'ARMA(1,1)' would fit better? Of course, the number of parameters would increase if a more complex model is considered, but this would also better capture long-range dependencies...

(6) In Section 4, the authors validate their WG by looking at different temporal or spatial statistics, such as long-term mean, inter-annual variance, PDF of non-zero precipitation, dry and wet spells, annual maximum sums of consecutive days, etc. However, I guess that there is no validation of space-time interactions. For example, if Z(s,t) denotes the precipitation amount at station s and time t, a possibility would be to see if the model and the data agree on statistics of the type 'X = P(Z(s2,t+k)>z | Z(s1,t)>z)' for increasing values of z? Here, the statistic X represents the probability that it rains at least z mm at station s2, given that it rained similarly at station s1, k days earlier. Have the authors checked this kind of space-time dependencies?

(7) The title of the paper is 'Stochastic modeling of [...] over complex topography', but the topography information does not appear anywhere in the model. Hence, how could the model be modified in order to incorporate information about altitudes, slopes, etc., and therefore hopefully better predict precipitation at unobserved locations?

TECHNICAL COMMENTS:

(1) p. 8742, L.7: A reference about space-time modeling of rainfall extremes in Switzerland is the following: Huser, R. and Davison, A.C. (2014, JRSS B), "Space-time modelling of extreme events".

(2) p. 8743, L.25: The BIC is used to select the order of the Markov chain. However, it is known that it typically over-penalizes complicated models (which might explain why a 2nd order model was not retained...). What does the AIC say?

(3) p. 8744, L.2: A threshold of 1mm day^{-1} was used. How sensitive are the conclu-

sions with respect to this threshold? How many zeros are there?

(4) p. 8745, L.22-23, 'they underestimate [. . .] (gamma distribution)': This sentence is misleading, I think, because the exponential and gamma densities decay at the same rate at infinity, so they are likely to give similar probabilities to extreme events.

(5) p. 8746, L.7-8, 'The parameters [. . .] maximum-likelihood': As already mentioned above, it would be better to have a spatial model linking the parameters, and estimate everything simultaneously (instead of estimating a lot of parameters separately from station to station).

(6) p. 8747, L.7-8: The point 4.2.2 is not very clear to me...

(7) p. 8747-8748, S3.3.2: How are the correlation matrices estimated? Empirically? If so, this might induce problems if the number of stations is large, and also it does not ensure that the correlation matrices are positive definite. A better solution is, as explained above, to assume and estimate a correlation function.

(8) p. 8750, 19-20, 'In absence of [. . .] matrix was chosen': How was this fall-back solution implemented? By minimizing a certain norm? If so, which one?

(9) p. 8751, S3.4.1: This section shows that the model has a lot of parameters, and that it is crucial to reduce this number to avoid huge uncertainties and optimization issues. . .

(10) p. 8752, L.18, 'roughly 19': In fact the theoretical reduction is (asymptotically) sqrt{10000/30} = 18.3. . .

(11) p. 8754, L.27, '33%': This percentage seems very low! This might be due to non-stationarities that the model is unable to capture, or simply because the model strongly underestimates probabilities of extreme events.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 8737, 2014.