Hydrol. Earth Syst. Sci. Discuss., 6, C1794–C1802, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C1794/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Copula based multisite model for daily precipitation simulation" *by* A. Bárdossy and G. Pegram

F. Serinaldi (Referee)

francesco.serinaldi@uniroma1.it

Received and published: 11 August 2009

General comments

The work by A. Bárdossy and G. Pegram proposes (i) a new multisite approach for modeling and simulating daily rainfall series, and (ii) an entropy-based criterion for assessing the spatial clustering exhibited by rainfall fields. In particular, the multisite rainfall model accounts for spatial association of the at-site rainfall sequences by using multivariate copulas (de facto bivariate) to jointly model the spatial structure of rainfall amounts and occurrences. The entropy-based criterion provides an interesting alternative to bivariate measures of association (e.g., Pearson's or Kendall's correlations) for evaluating the spatial structure of rainfall fields. In my opinion, the ideas developed

in this paper are attractive, and the overall quality of the work is good. However, I have some remarks about the partitioned copula model, which I discuss in the next section. Moreover, some inferential aspects are not completely clear to me, and the comparison between rainfall observations and simulations can be improved in order to better highlight the potentialities of the model.

Specific comments

In Sect. 1.1, the authors state that their model generalizes the meta-Gaussian approach introduced by Herr and Krzysztofowicz (2005) (henceforth HK05) moving from a bivariate to a multivariate framework. Nevertheless, the model is described and applied in a bivariate framework, even though it can be extended to many dimensions (> 2). However, in my opinion, the most important concern is that the rationale of the HK05 approach is different from the bivariate approach suggested by the authors. The HK05 mixed distribution for a bivariate zero-inflated vector (X, Y) can be written in terms of copulas as (Serinaldi, 2008; 2009a):

$$\Omega(x,y) = p_{00} + p_{10}H_X(x) + p_{01}H_Y(y) + p_{11}C(F(x),G(y)), \tag{1}$$

with marginals:

$$\Omega_X(x) = (1 - p_{10} - p_{11}) + p_{10}H_X(x) + p_{11}F(x), \tag{2}$$

$$\Omega_Y(y) = (1 - p_{01} - p_{11}) + p_{01}H_Y(y) + p_{11}G(y), \tag{3}$$

where p_{00}, p_{10}, p_{01}, p_{11} are the probabilities $P(X = 0, Y = 0)$, $P(X > 0, Y = 0)$, $P(X = 0, Y > 0)$, $P(X > 0, Y > 0)$, respectively; $H_X(x) = P(X \leq x | X > 0, Y = 0)$, $H_Y(y) = P(Y \leq y | X = 0, Y > 0)$, $F(x) = P(X \leq x | X > 0, Y > 0)$, $G(y) = P(Y \leq$

$y|X > 0, Y > 0)$, and C is the copula of positive pairs. In Eq. 1, the copula describes the dependence structure of contemporaneous positive pairs at both sites, when X and Y denote vectors of contemporaneous observations at two sites. On the other hand, the partitioned copula introduced by the authors seems to implicitly assume that the full sample (zero and non zero values) comes from a continuous bivariate distribution describing a continuous process, and the ties (that is, the poles of probability in the unit square) related to pairs $(X = 0, Y = 0)$, $(X > 0, Y = 0)$, $(X = 0, Y > 0)$ arise as a kind of rounding-off process. In other words, the model described by the authors seems to be the generalization of the bivariate distribution implied by the Guillot's rainfall model (Guillot, 1999). As discussed by HK05 (in Section 7.4), the Guillot's distribution looks like a meta-Gaussian distribution with discrete-continuous marginals. It can be written in term of copulas as:

$$\Omega(x, y) = C(\Omega_X(x), \Omega_Y(y)), \tag{4}$$

which is the model proposed by the authors. HK05 state that this distribution "is neither structurally correct nor empirically valid" (except when $p_{11} = 1$ so that $\Omega(x, y) = C(F(x), G(y))$), and "It may be useful, though, as an approximate simulator when p_{11} is near one (Guillot and Lebel, 1999)". Even though HK05 refer to a meta-Gaussian distribution, their remarks about the structure of the model apply to copula-based models as well, since the meta-Gaussian distribution is simply a Gaussian-copula-based distribution. Figure 1 provides an example which highlights the differences between the two models. An explanation of graphical details is omitted as the figure is similar to Fig. 2 in the paper in discussion, and to Fig. 13 in HK05. The panel on the upper left shows the scatter plot of contemporaneous daily rainfall pseudo-observations measured at two sites from a dataset described in Serinaldi (2009b). The density of positive pairs (lower left panel) displays a behavior similar to that shown in Fig. 2 of the paper in discussion. As an example, a Gumbel copula was fitted to the positive pairs and introduced in the two models to simulate a sample with the same size and probabilities p_{00}, p_{10}, p_{01}, p_{11}

of the observed series. The panels in the middle show the scatter plot and the density of positive pairs obtained by the mixed model in Eq. 1, while the panels on the right refer to the partitioned copula model. As shown by the density plots, the copula-based HK05 distribution reproduces the upper right quadrant better than the partitioned copula model (note that the Gumbel copula is not the "best possible" one, as it was chosen without performing any comparative analysis with other families). The difference between the two models is related to their structures and not to the chosen copula: in the partitioned copula model, only the upper tail of the copula describes the joint behavior of positive pairs, showing the same shortcomings of the Gaussian copula discussed by the authors; while, in the copula-based HK05 distribution, the whole copula models positive pairs. Thus, the introduction of the V-copula could not solve modeling problems which are due to the model structure. In other words, the mixed model introduced by HK05 assumes that the four types of possible pairs from a bivariate rainfall vector (namely, $(X = 0, Y = 0)$, $(X > 0, Y = 0)$, $(X = 0, Y > 0)$, $(X > 0, Y > 0)$) explicitly represent the four components of a really intermittent process (where a proper continuous bivariate distribution is only associated to positive pairs), while the partitioned copula model implicitly describes the rainfall as a continuous process where the intermittency arises when the rainfall amount is smaller than the measurement resolution of the instrument. In this sense, the zero values assume the same meaning of the other tied values resulting from the instrument rounding-off. Since the two models imply two very different ways of considering the rainfall structure, in my opinion, these issues should be discussed in depth by the authors.

At pp. 4888-4489, the authors state that an element of novelty in the proposed copula-based model is that the occurrence and amount processes are jointly modeled without splitting the analysis (as it is usual in the traditional Markov chain approaches). This advance in the rainfall modeling was already proposed by Serinaldi (2009) who describes a multisite daily rainfall generator based on a bivariate mixed model which exploits copulas. The approach suggested by Serinaldi (2009b) is different from that proposed by the authors, and focuses more in detail on the intermittent temporal rainfall structure. However, it already avoids the split between occurrence and amount processes. The paper was published few days before the submission of the work in discussion. Perhaps it can be of interest to the authors.

At p. 4493, the authors state that "In contrast to the work by (Serinaldi, 2008) who concentrated on the upper-tail dependence structure of 2-copulas, we are concerned about the joint interdependence of clusters of rainfall stations.". Perhaps, it can be more appropriate to say that the work by Serinaldi (2008) focuses on pairwise association properties, as it investigates the Kendall's correlation, pairwise occurrence probabilities of rainfall as well as the upper-tail dependence, and their variability with the inter-gauge distances and the aggregation time scale.

The structure of the model described in Section 3.1 is quite clear from the simulation viewpoint. However, it is not very clear to me how the maximum likelihood method allows estimating space-time correlations of the continuous hidden y-variate so that the corresponding space-time correlations of the simulated rainfall match the observed ones. Perhaps, Sect. 3.3 can be extended (or an appendix can be added) to further explain the inference procedure. Moreover, even if the model is validated by the proposed entropy-based criterion, it can be worth showing the observed and simulated spatial correlations (e.g., Kendall's or Pearson's coefficients) of the complete rainfall sequences.

In Section 3.2, the authors fit the exponential and Weibull distributions to the rainfall amounts. Given the structure of their model, they do not distinguish between the four marginals which appear in Eq. 1. In particular, the distributions $F_g(u_i, 0)$ and $F_g(0, u_j)$ can be very different from the marginal distributions of positive pairs and from the unconditional marginals (e.g., HK05; Serinaldi, 2009a), and their correct modeling can

be important as, in some climates, $F_g(u_i, 0)$ and $F_g(0, u_j)$ are mainly related to intense and short convective storms that hit one site but not the neighbors.

At p. 4502, the authors discuss the behavior of the serial and spatial correlations driving the y-variate of the hidden AR(1) model. As mentioned above, it can be interesting to show a comparison between the serial and spatial correlations of the simulated rainfall sequences (obtained by Eq. 11 in the paper) and the measured series, as the final aim is to reproduce these ones. Similarly, at p. 4503, instead of comparing the spatial correlation of the hidden y-variate with those of the rainfall amounts and occurrences process resulting from the Srikantan and Pegram (2009) model, it can be better to compare the final rainfall sequences. Moreover, nothing is said about the ability of the model to preserve the temporal structure of the rainfall. It can be worth showing the performances in terms of capability to preserve the distributions of wet/dry spells and storm volumes.

Finally, Section 5.1 could be slightly improved by adding the information about the variability around the simulated mean values reported in Figs. 13 to 19. Moreover, the unconditional probabilities in Figs. 14 to 17 do not provide an effective illustration of the model performance in the multivariate sense, as they only describe the marginal behavior, so that similar results can be easily obtained by simulating from univariate mixed distributions at each site without introducing spatial and temporal dependences. The distributions of wet/dry spells and storm volumes as well as the log-odds ratios can be more representative.

To summarize, the paper is an interesting work. The model and the validation criterion proposed by the authors are promising. In my opinion, some theoretical aspects of the model need to be clarified; a more detailed discussion of the inference procedures and the introduction of additional statistics, able to highlight the spatio-temporal properties of the simulations, can increase the confidence of the reader in the proposed methods.

Once the issues raised in the review have been addressed to the Editor's satisfaction, I recommend publication in HESS.

Technical corrections

Fig. 8, caption: "Sect. 3.4"
"References" section: "Serinaldi, F.: Analysis of inter-gauge dependence by Kendall's τ_K, upper tail dependence coefficient, and 2-copulas with application to rainfall fields, Stoch. Env. Res. Risk A., 22, 671–688, 2008."

Acknowledgments

The analyses in this report were performed in `R` (R Development Core Team 2009) by `copula` (Yan, 2007) and `fCopulae` (Wuertz et al., 2009) packages.

References

Herr, H. D. and Krzysztofowicz, R.: Generic probability distribution of rainfall in space: The bivariate model, J. Hydrol., 306, 234– 263, 2005.

R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2009. URL http://www.R-project.org

Serinaldi, F.: Analysis of inter-gauge dependence by Kendall's τ_K, upper tail dependence coefficient, and 2-copulas with application to rainfall fields. Stochastic Environmental Research and Risk Assessment, 22(6), 671– 688, 2008.

Serinaldi, F.: Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. Stochastic Environmental Research and Risk Assessment, 23(5), 677-693, 2009a.

Serinaldi, F.: A multisite daily rainfall generator driven by bivariate copula-based

mixed distributions. Journal of Geophysical Research - Atmospheres, 114, D10103, doi:10.1029/2008JD011258, 2009b.

Srikanthan, R. and Pegram, G.: A nested multisite daily rainfall stochastic generation model, J. Hydrol.,371(1-4), 142-153, 2009.

Wuertz,D., et al.: fCopulae: Rmetrics - Dependence Structures with Copulas. R package version 2100.76, 2009. http://CRAN.R-project.org/package=fCopulae

Yan, J.: Enjoy the joy of copulas: with a package copula, Journal of Statistical Software, 21(4), 1-21, 2007.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 4485, 2009.

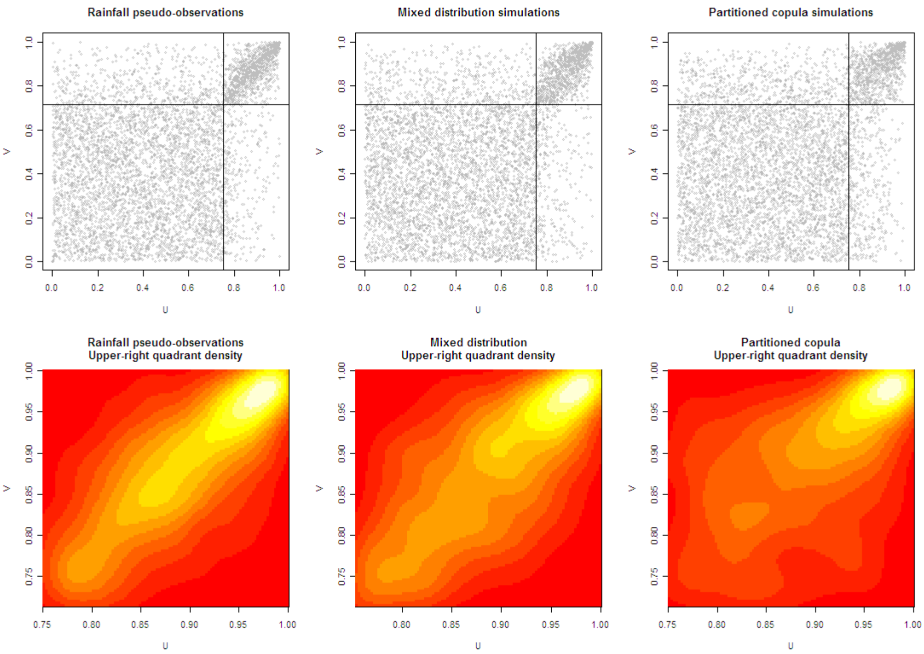

Fig. 1. Scatter plots and upper-right quadrant densities of daily rainfall pseudo-observations (left), simulations from a copula-based mixed distribution (middle), and from a partitioned copula model (right).