Estimation of flood design hydrographs using bivariate analysis (copula) and distributed hydrological modelling

By A. Candela, G. Brigandí, and G. T. Aronica

Submitted to Natural Harzards and Earth System Sciences
MS-NR: nhessd-2-27-2014

REFEREE REPORT

I regret to say that my opinion about this paper is less positive than that of the other Reviewers. The manuscript attempts to perform a design hydrograph inference putting together a number of techniques but missing their meaning and the essential relationships between physics and statistical analyses. The paper is also badly written in terms of grammar, syntax and materials' organization. The approaches (especially the rainfall model) are described superficially or not described at all, and the literature review is outdated and overlooks relevant references. In the following, I provide some remarks that, in my opinion, highlight how the overall approach is essentially ill-posed because of an inappropriate use of statistical concepts as well as the blind use of widespread misconceptions.

General remarks

This paper can be collocated in the main stream of works supporting the hypothesis that peak, volume, average intensity and other properties of hyetographs and hydrographs are pure random variables linked by stochastic relationships which could be described by a joint distribution. Unfortunately, Serinaldi and Kilsby (2013) showed that this hypothesis is generally not valid and untenable. Therefore, the overall framework based on the construction of the joint distributions of intensity and duration of the hyetographs, and peak and volume of hydrographs is simply ill-posed and basically incorrect.

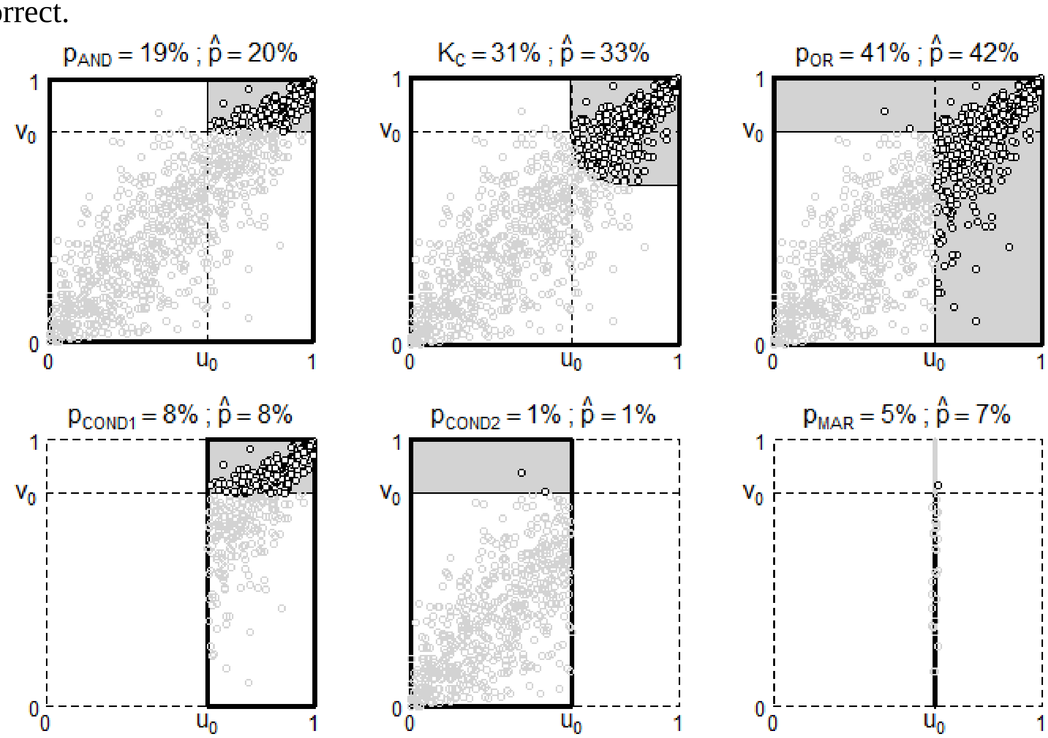
Section 3.1, which should present the stochastic generation of the rainfall events, actually does not describe a rainfall model but only the copula inference (using hold techniques and incorrect terminology) and a couple of equations describing the Beta distribution, with no description of the model structure and generating procedure. As mentioned above, the copula inference in this case is essentially incorrect for the variables at hand. However, if the Authors are not persuaded about that, they should at least recognize that such a type of approach was already proposed by Vandenberghe et al. (2010), whereas, more effective techniques were developed by Dr. Gyasi-Agyei in a recent series of papers (which can be easily found spending some time surfing Google Scholar or ISI Web Of Knowledge and reading papers before writing them).

Moreover, even more important, the proposed rainfall model was described in a previous paper submitted to Hydrological Sciences Journal on 07 Oct 2010 and withdrawn (parts of that paper have been copied and pasted in the present manuscript). In that context, I provided a detailed report, in which I discussed in depth some of the aspects mentioned above. Of course, after four years, methodologies must be updated according to new findings. However, I see that the Authors not only overlooked those suggestions but have not made any effort to update their approach. I'm also a bit surprised (actually, no very much) by the opinion of the other Reviewer Dr. Grimaldi who managed that submission and therefore is already aware of the problems affecting both that paper and the present manuscript.

The inference procedure for the joint distribution of intensity and duration of the hyetographs is based on 80 pairs (rainfall events; P40L3). This small sample size not only conceals the actual dependence structure (which cannot be described by commonly used copulas; see Serinaldi and Kilsby, 2013), but also makes the inference output unreliable because of the large intrinsic uncertainty (see Serinaldi, 2013a).

Section 5 highlights how the Authors completely missed the purpose of using statistical models. Making inference means trying to understand the key aspects of a sample (taking for granted that it really represents a realization of random variables) and to build a model to make interpolation and extrapolations (with care!). Now, if the Authors are able to simulate an arbitrary number of Q_{max}-V pairs from the rainfall-runoff model, why do they fit a copula on these samples? The rainfall generation is exactly devised to allow us to make calculations directly on the simulated output (Q_{max}, V, damage, losses, or whatever else output variable of interest) without any further statistical model. If we have 1000 Q_{max}-V pairs we can obtain empirical joint probabilities of exceedance (or RPs, if we like them) directly from the simulated sample at any specified degree of accuracy (playing with the simulated sample size). Fitting another statistical model (which introduces unavoidably a source of error and data compression) is simply useless and meaningless, because we have already all the required information and we do not need to reduce it: we have to use it! Moreover, in this specific case, we already know that all copulas used by the Authors are incorrect, because the dependence structure of Q_{max}-V pairs (Fig. 13) is identical to that shown for instance by Gräler et al. (2013). As explained by Serinaldi and Kilsby (2013) such a dependence structure is general and cannot be described by any commonly used copula. Please note that Gräler et al. (2013) made the same error: they fitted a model on simulated Q_{max}-V pairs when no model was required, as every statistic of interest can be estimated directly (with the required accuracy) on the simulated samples coming from the rainfall-runoff model. The importance of rainfall modelling and the big effort people make to do this accurately (see e.g. Serinaldi 2009, 2010; Serinaldi and Kilsby 2012a, 2014, as well as references therein describing effective approaches at several space-time scales) rely on the fact that this input variable is the main requirement for every subsequent analysis. Simulating a sample of Q_{max}-V pairs of arbitrary size and then fitting a statistical model on this sample (which already contains all the information concerning the process under study, with the required accuracy) means not to understand the rationale of these procedures.

The use of joint return period deserves a further comment. Unfortunately, misconceptions spread more widely and fast than good practices, probably because of a general superficiality characterizing this time of decadence. Joint return periods (in the *iid* context) simply define the reciprocal of *a* probability of exceedance. Which one? Simple: such a probability can be joint, marginal, conditional or defined on a support such as iso-probability curves or structure functions (see Volpi and Fiori, 2014), but the key issue is that the choice is not free and cannot be done without describing the physics of the dangerous events or the failure mechanism of a device. Marginal, conditional, and joint probabilities are not purely statistical concepts but are devised to describe specific phenomena and dynamics via probabilistic reasoning: if one does not specify the mechanism of failure, it is not possible do determine the probabilistic description and the suitable distribution to be used. In this respect, works proposing ill-posed comparisons such as Gräler at al. (2013) do not make a good service. These issues are discussed informally in Serinaldi (2012b, 2013b) and are summarized in the figure below. Leaving out details, which can be found in the literature, the figure shows that different definitions of joint and conditional probabilities impart a probability measure on different sub-spaces and are all correct for the specific phenomenon they describe (empirical estimates based on simulated samples are equal to the expected theoretical values apart from sampling uncertainty). Therefore, the so-called “OR” JRP used in the manuscript is an arbitrary choice with no link to design requirements. Unfortunately there is a widespread tendency to talk about meaningless comparisons and arbitrary choices among JRPs that are simply incorrect.



Specific comments

As mentioned before, there are too many language errors and typos, even in the abstract (P28L16 and 28), which should be the business card of a paper.

P30L13: “to statistically value flood volume and duration”, maybe “evaluate” or “assess”

P30L13-15: As mentioned above, this is incorrect. The relationship between these variables is not purely stochastic, and commonly applied copulas could be suitable only after data pre-processing (see Serinaldi and Kilsby, 2013).

P30L17-21: The sentence is incorrect. Consistent joint laws with suitable pairwise correlations and arbitrary marginals have been used for decades via multivariate Gaussian models and normal quantiles transformation. The advantage of the copulas is the choice of a variety of dependence structures different from the Gaussian one.

P30L23: Nelsen (1999)… I already suggested updating the reference four years ago!

P30L24: “Balistrocchi and Bacchi”!

P32L14-17: as for P30L13-15.

P33L3-4: as for P30L17-21

P33L7-13: “Archimedean” is a class of copulas not a family, and Frank family is not a general model for hydrologic variables and storm average intensity and duration. Results by De Michele and Salvadori (2003) rely on a very specific and small sample and cannot be generalized, especially if one accounts for the recent findings mentioned above.

P33L17: “generation function”. The “generator” should be defined before being introduced.

P34L5: what is the “nonparametric method”? As mentioned before, time is passed since Genest and Rivest (1993). Again, these aspects were already highlighted in my 2010’s report.

P34L10: “$C(h)$”. What is the new variable h?

P35L6: The use of standardized rainfall patterns and distribution functions with finite support is not new. More effective applications can be found in Hingray et al. (2002), Kottegoda et al. (2003) (already mentioned four years ago), and Vandenberghe et al. (2010), among others.

Eq. 13 and 18: maybe “$P_{i,j,N}$” and “$H_{i,j,N}$” as I cannot see the third dimension in this notation.

P39L10: 7 hours as inter-event time should be better justified. I'm not sure that Sicilian rainfall is so similar to Ligurian rainfall dynamics studied by De Michele and Salvadori (2003).

Summary

As mentioned in the introduction, this paper attempts to put together several models and procedures. This results in an apparent complexity that however does not hide the overall weakness and incoherence of the procedure and its bases. In particular, everything relies on a few assumptions that are not correct. Spending some time to think about them could have avoided to waste time to build such a house of cards. I was negatively impressed by the lack of attention in conceiving and developing an effective rainfall model (overlooking almost all the existing literature on the topic), and by the misuse of statistical concepts and methods (building joint distributions with no rationale about their meaning and purpose).

Finally, as the Authors can see from this report and the previous 2010's one, colleagues spend time to make reviews and provide recommendations; thus, I found rather annoying to see that none of the suggestions and references I provided four years ago was accounted for, and the Authors persist to propose the same things without corrections and updates.

I realize that the Authors probably will not like my opinion, which is actually a minority report. However, I really think that this manuscript is an example of a superficial approach to research.

Sincerely,

Francesco Serinaldi

PS: The 2010's HSJ manuscript, referees's reports and HSJ Editor's decision letter are available under NHESS Editor's request.

References

Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C.: Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation, Hydrol. Earth Syst. Sci., 17, 1281-1296, doi:10.5194/hess-17-1281-2013, 2013.

Hingray B, Monbaron E, Jarrar I, Favre AC, Consuegra D, Musy A (2002) Stochastic generation and disaggregation of hourly rainfall series for continuous hydrological modelling and flood control reservoir design, Water Science and Technology, 45(2), 113-119

Kottegoda, N.T., Natale, L., Raiteri, E., 2003. A parsimonious approach to stochastic modelling and disaggregation of daily rainfall. Journal of Hydrology 274, 47-61.

Serinaldi, F. (2009), A multisite daily rainfall generator driven by bivariate copula-based mixed distributions, J. Geophys. Res., 114, D10103, doi:10.1029/2008JD011258.

Serinaldi F. (2010) Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. Nonlinear Processes in Geophysics, 17(6), 697-714.

Serinaldi F, Kilsby CG. (2012a) A modular class of multisite monthly rainfall generators for water resource management and impact studies. Journal of Hydrology, 464-465, 528-540.

Serinaldi F. (2012b) Interactive comment on "Joint return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation" by S. Vandenberghe et al. Hydrol. Earth Syst. Sci. Discuss., 9, C3721–C3721, 2012 www.hydrol-earth-syst-sci-discuss.net/9/C3721/2012/

Serinaldi F. (2013a) An uncertain journey around the tails of multivariate hydrological distributions. Water Resources Research, 49(10), 6527-6547.

Serinaldi F. (2013b) Interactive comment on "Flood hazard in the Mekong Delta – a probabilistic, bivariate, and non-stationary analysis with a short-termed future perspective" by N. V. Dung et al.Nat. Hazards Earth Syst. Sci. Discuss., 1, C90–C90, 2013 www.nat-hazards-earth-syst-sci-discuss.net/1/C90/2013/

Serinaldi F, Kilsby CG (2013) The intrinsic dependence structure of peak, volume, duration and average intensity of hyetographs and hydrographs. Water Resources Research 2013, 49(6), 3423-3442.

Serinaldi F, Kilsby CG. (2014) Simulating daily rainfall fields over large areas for collective risk estimation. Journal of Hydrology, 512, 285-302.

Vandenberghe, S., Verhoest, N. E. C., Buyse, E., and De Baets, B. (2010) A stochastic design rainfall generator based on copulas and mass curves, Hydrol. Earth Syst. Sci., 14, 2429-2442, doi:10.5194/hess-14-2429-2010.

Volpi, E., and A. Fiori (2014), Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment, Water Resour. Res., 50, 885–897, doi:10.1002/2013WR014214.