Comments on "Dung et al.: Flood hazard in the Mekong Delta a probabilistic, bivariate, and non-stationary analysis with a short-termed future perspective, Nat. Hazards Earth Syst. Sci. Discuss., 1, 275322, 2013"

Francesco Serinaldi
School of Civil Engineering and Geosciences
Newcastle University
Newcastle upon Tyne, UK
email: francesco.serinaldi@ncl.ac.uk

April 11, 2013

General comments

The paper under review introduces a probabilistic flood mapping strategy combining a quasi 2D hydrodynamic model fed by design hydrographs. These hydrographs are obtained by rescaling nondimensional typical hydrographs according to the flood peak and volume deduced from a bivariate distribution with dynamic (time varying) marginals. The paper is formally well written and the overall quality is good, however, as many other studies recently published in the literature, it suffers an original sin: the "mathematization" curse, just to use a Klemeš' definition [Klemeš, 1986]. More explicitly, as already mentioned by Reviewer 1, the statistical sophistication seems to me not justified neither theoretically or empirically based of the very limited size of the available data. Moreover, the use of multivariate probabilities and non-stationarity easily leads to theoretical inconsistencies, misleading statements and inappropriate use of concepts which are well defined in the univariate *iid* framework but lose their meaning in a multivariate nonstationary context. Since the hydrodinamic model is borrowed by an already published work, the specific comments reported below focus on the statistical apparatus.

Specific comments

I understand that the aim of the Authors is to propose a framework useful in nonstationary conditions; however, in principle, we are already able to introduce models even more complex than that proposed in this study. For example, we can incorporate dynamic copulas accounting for the time variation of the dependence structure, nonparametric link functions to allow for nonlinear variation of the copula and marginal parameters, exogenous covariates, and so forth. All these sophistications can be easily implemented and, in principle, we can push the degree of model complexity up to the exact (but uninformative) reproduction of the data. Therefore, the point is not how much we are clever in introducing complex models, but which is their correctness and usefulness when the available information (86 pairs of annual flood peaks and volumes) is just enough the reliably estimate summary statistics of central tendency. An example can help to clarify this point. Figure 1 reproduces the univariate and bivariate distributions fitted by Yue et al [1999] on 33 pairs of flood peaks and volumes. The figure also shows the confidence bands of the marginal distributions along with the sampling uncertainty areas related the 0.99 p-level curve. Since the uncertainty areas cover a large set of univariate quantiles and p-level curves, it is rather evident that the definition of the events with 0.99 "AND" and "OR" probabilities can just be an educated guess. Figure 2 shows that at least thousands *iid* points are required to reliably estimate the 0.99 p-level quantiles. Now, suppose that we have 5000 years of real-world observations from an ideal (perfectly working) stream gauge: first, we can make inference directly from the observations and do not need any model, and second, it is reasonable that the observations refer to very different historical conditions of climate, drainage basin and river network, thus making questionable the basic hypotheses required by statistical inference procedures [Klemeš, 2000a,b; Koutsoyiannis, 2004]. In other words, refined statistical tools can be useful but cannot replace data. Introducing highly parameterized models to describe small samples does not add insight if we account for the increased uncertainty related to the additional parameters. The guidance should always be the Occam's razor (*multipla non est ponenda praeter necessitatem*, which was stated before Einstein's sentence reported at the beginning of Ch. 3 of Dung [2011]), and a fair assessment of the true benefits resulting from the increased sophistication. Moreover, in real-world design problems, statistical discrimination tools such as AIC and likelihood ratio tests have a limited value, as we are interested to improvements that are significant from a physical and economical point of view rather than purely statistical.

The overall meaning of the above remarks is that we need to reconcile statistics and engineering by using strictly necessary statistical tool based on the available information, paying attention to the fulfillment of the underlying hypotheses, trying to communicate the uncertainty, and avoiding ill-posed confidence in refined techniques that can be cool and fashionable but substantially no well devised for the problem at hand.

As mentioned above, I understand the general idea to describe both copulas and marginals with (linearly) time varying parameters for the sake of generality and transferability (P285L10), but the final model actually reads as a bivariate Gaussian distribution applied to preliminarily log-transformed data, that is, probably the most classical bivariate approach that have been applied in hydrology for at least 50 years. Therefore, cannot section 3.1 be synthesized in a few sentences saying that the log-transformed peak-volume pairs are described by a Gaussian bivariate distribution, which can be replaced by alternative families, if this is required based on the empirical evidence? In passing, the main advantage of copulas in not the possibility to incorporate different marginals (this is already allowed in the meta-Gaussian framework and is the basis of the hydrological applications of Box-Jenkins ARMA modeling framework, for instance), but the possibility of using dependence structures different from the meta-Gaussian (under "suitable conditions", the Sklar's theorem guarantees that plugging whatever marginals in a copula the resulting model is always a valid joint distribution).

The use of marginal distributions with time-varying parameters is an element of distinction with respect to a classical bivariate Gaussian distribution. However, this is exactly one of the above-mentioned practical cases in which statistics must meet engineering and common sense. I agree with the Koutsoyiannis [2006] statement that "*stationarity is closely associated with a model (rather than a real world system) either deterministic or stochastic, which implies an ensemble of infinite realizations.*". The Authors justify the use of dynamic marginals based on the results reported in Section 6.3.2.1 of Dung [2011] that reads

There exist several methods which can be used for detecting trends. This *study uses a simple but robust non-parametric test, the Mann-Kendall test (Mann, 1945), to test for trends in the peak and volume series. The null hypothesis is that there is no trend in the peak, resp. volume series at the significance level of 10%. The Mann-Kendall test shows that trends are present in both peak and volume series. It means that it is reasonable to apply the non-stationary flood frequency analysis for both series.*

In my opinion, this analysis is not enough for two reasons: (1) based on my experience, and looking at the time series shown in Fig. 4 of the manuscript, I am rather confident that the null hypothesis is not rejected at the 5% or 1% significance levels, and (2), more important, it is often overlooked that MK test checks for monotic trends, meaning that the alternative hypothisis is not unique (e.g. linear trend) but multiple, thus emcompassing whatever linear or nonlinear, abrupt or slowly varying monotonic patterns (this is why MK is less pawerful then its parametric competitors when the alternative matches the specific alternative embedded in the parametric techniques). Therefore, assuming a linear pattern for the functional relationships of the distribution parameters is arbitrary and not justified neither theoretically or empirically. Even though MK indicates rejection, it does not provide any information about the shape of the possible trend: for instance, it can be related to a regime shift, or an S-shaped pattern converging to some asymptotic level. The key point is however that the data are not enough to infer about the shape of the possible trends and their evolution in the future [e.g., Guerreiro et al, 2013]. Moreover, without a physical (deterministic) justification for possible trends (e.g. antropic intereventions on the basin), it is more likely that we are just dealing with fluctuations of natural processes that evolve (in an unknown manner) on time scales that go beyond the period of observation.

From a modeling point view, the large uncertainty of the model parameter estimates discussed above makes the difference between stationary and nonstationary p-level curves likely insignificant. Indeed, this is confirmed by the small difference in the final flood maps. Accounting for the sampling and parameter uncertainty, the difference is probably even less evident.

The "multivariate-nonstationary" business

Moving from simple techniques to slightly more complicated statistical tools, some concepts are not so easy to extend. In the present context, the discussion about the bivariate return periods raised in the text and review process seems to reveal some confusion about the meaning and consequences of working in a nonstationary multivariate framework. Unfortunately, it seems that this misunderstanding is more and more spread in the recent literature. The Authors correctly acknowledge that the choice between "AND" and "OR" bivariate return periods concerns the purpose of the study; however, the subsequent selection (P290L20) is based on a widespread misunderstanding which is generated by the use of the apparently friendly joint return periods instead of the joint probabilities. In more detail, the underlying joint probability of T_{OR} is $p_{OR} = \Pr[Q \geq q \cup V \geq v]$. Referring to the orthogonal blue lines in Fig. 9, they define four quadrants while p_{OR} defines a probability measure on a subset of the bi-dimensional domain corresponding the first, second and fourth quadrants (counted from the top right counter-clockwise). p_{OR} describes the probability that a realization of (Q, V) exceeds (q, v) in terms of q or v or both. The statement "*...the OR definition most of the observed data fall below the 10-yr return period, even the event of 2000 with the historically largest damage. This is not plausible, and thus the AND definition is selected.*" is therefore incorrect because even though the event of 2000 is not exceeded in terms of volume, it is exceeded by seven events in terms of peak (to visualize this, it is sufficient to trace orthogonal lines crossing in the 2000 point and focus on the points falling in the three mentioned quadrants). This means that we observed in average eight events exceeding or equaling the 2000 event in 86 years, i.e. one event every 10 years in average, which is exactly the information coherently provided by p_{OR} and thus T_{OR}. On the other hand, p_{AND} defines a probability measure on the first quadrant and describes the probability that a realization of (Q, V) simultaneously exceeds (q, v) in terms of both q and v. Focusing on the first quadrant defined by the orthogonal lines crossing in the 2000 point, it is clear that only this event occurred in 86 year, thus leading to $p_{AND} \approx 0.99$ and $T_{AND} \approx 100$.

The inequalities in Eq. 8 are also a natural consequence of the above definitions: without resorting to probabilistic reasoning, it is intuitive and evident that observing an event falling in a wider domain (three quadrants) is more probable than observing an event in a smaller domain (the top right quadrant).

Actually, as it does not make sense to say that p_{AND} is better than p_{OR} and vice versa, the same holds for the corresponding joint return periods. Joint return periods give values (in years) which appear falsely friendly and easy to be interpreted; however, they simply hide the true meaning of the underlying joint (or conditional) distributions, leading to misunderstandings and wrong statements. Unfortunately, the literature does not help shedding light on this concepts and proposes incoherent comparisons of concepts that are essentially incomparable, thus increasing the confusion. This is rather dangerous, especially if we plan to deliver results to unskilled policy-makers and administrative authorities.

Based on the above discussion, it is evident that that the joint return period cannot be chosen from a statistical reasoning but selecting the joint probability that describes the scenarios that are critical for the (hydraulic) system: if the failure of the system or device occurs when q or v are exceeded, we must assess the risk by p_{OR}; if the system collapses only when both q and v are exceeded but is not damaged when only one quantity exceeds the critical value, p_{AND} is therefore required, whereas p_{OR} does not apply at all, as it describes scenarios that do not match the system/device operation rules.

Talking about return periods, it should be also noted that further sources of confusion raise when we move from stationary to nonstationary framework. Namely, the definition of return period as the reciprocal of the probability of exceedance (not "of occurrence") holds only under *iid* conditions. Unfortunately, the derivation of this relationship seems to be forgot by too many people, thus allowing for extensions in the nonstationary framework that are actually incorrect. Some informal remarks and references on these aspects (as well as on the definition of joint return periods) can be found in Serinaldi [2012]. Such comments apply to the present manuscript as well (especially Sections 3, 4 and 5 therein).

Minor and editing remarks

P277L18: "People" perhaps is a typo.

P278L10-15: Based on the above discussion, I would avoid statements introducing nonstationarity as something taken for granted both in general and especially for the data at hand.

P280L24: "...the combination of the natural hydraulic peculiarities in combination with the large anthropogenic influence". Maybe it is better "Thus, the combination of the natural hydraulic peculiarities with the large anthropogenic influence..." or "Thus, the natural hydraulic peculiarities in combination with the large anthropogenic influence..."

P282L22: "adapted"

P284L15-25: Please, specify how the typical nondimensional hydrographs are rescaled. The text specifies that the peak is multiplied by the peak values simulated from the bivariate distributions and then the volume is adjusted to match the simulated volumes. Is this done by removing the peak value from the hydrograph? Please, add a few technical details.

P286L20-26: Tail dependence is defined as the limit of the conditional probability $\Pr[Q \geq t | V \geq t]$ as $t \to \infty$. It is defined for every copula but the value of the limit is zero for some families. Therefore it is better to say that copulas can have tail dependence equal to zero. Nonetheless, the discussion about tail dependence can be avoided as it is not applied in the analysis and is not used to discriminate between the candidates. On the other hand, the sample size is certainly insufficient to assess whatever asymptotic property.

Sections 3.2 and 3.3: AIC and its modifications are performance measures and not goodness-of-fit tests. I suggest to apply at least a likelihood ratio test or better some ECDF based test. These tests are implemented in the `R` package `copula`. Moreover the scatter plots do not provide a good graphical diagnostic. The diagram of the Kendall distribution, i.e. the distribution function of the copula is a better tool. A description can be found in the paper by Genest and Favre (2007) cited in the manuscript.

P293L1-5: It is not clear to me how the pairs are simulated. Do the Authors simulate 100 pairs from the p-level curve, i.e. from the Volpi-Fiori's conditional distribution? Please, add some technical detail.

Sections 6-7: The inundation maps are a bit difficult to read. The overall patterns seem to be coherent but does the pixel-wise calculations guarantee that the values in neighbor pixels are coherent even in light of the time of propagation along the drainage network?

P296L17-25: I do not agree with the Authors. The uncertainty of extrapolating beyond the observed frequencies is so high that every model can only provide guess estimates which are statistically indistinguishable. Assessing which extrapolation is more correct is rather difficult [Klemeš, 2000b] without further data or other sources of information. In my opinion the best strategy is to apply robust techniques complemented by a sound assessment of the sampling and (distribution) parameter uncertainty (the Bayesian procedures mentioned by Reviewer 1 are an option).

P298L5-10: I do not agree with the Authors. I believe that there is not any theoretical justification for the use of nonstationary models. They can be used if we have some evidence that they perform better than the stationary distribution according to some design criterion. By complementing the estimates of the extreme quantiles with suitable confidence intervals, it will be evident that the only inference we can do (by 86 values) about e.g. the flood with 0.999 (univariate, conditional or joint) probability of exceedance is just that it will be surely larger than the largest observation. Of course this holds if no additional data (and information) is available.

P298L13-18: I believe that these statements are incorrect. As shown by the figures reported below, the combination of the sampling and parameter uncertainty is larger than the inherent variability described by the pairs that fall over a p-level curve, which on the contrary are just a subsample of the uncertain p-level scenarios. The main source of uncertainty is exactly epistemic. Resorting to univariate distributions as an example, a distribution describes the inherent uncertainty while the confidence bands describe the epistemic uncertainty. For small sample, such as that in the present case, the uncertainty of the extrapolated quantiles is dominant. Data contain information, which is the basis of the specific knowledge (synthetic a posteriori prepositions, just to use classical Kant's terminology), which in turn provides a better understanding of the world when it is combined with the general knowledge (a priori synthetic prepositions). Thus, I think that the sentence reported at the beginning of Ch. 5 of Dung [2011]
"*Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.*" (Anonymous)
could be rephrased as follows
"*Data contain information, information is one of the bases of the knowledge, knowledge is a necessary step for a better understanding, understanding does not lead necessarily to wisdom (it depends on how one uses knowledge and understanding based on ethical values).*"

References: please, check typos throughout the reference list.

Sincerely,

Francesco Serinaldi

References

Dung NV (2011) Multi-objective automatic calibration of hydrodynamic models development of the concept and an application in the mekong delta. PhD thesis, University of Stuttgart, URL http://elib.uni-stuttgart.de/opus/volltexte/2012/6831/

Guerreiro SB, Kilsby CG, Serinaldi F (2013) Analysis of time variation of rainfall in transnational basins in iberia: abrupt changes or trends? International Journal of Climatology DOI 10.1002/joc.3669

Klemeš V (1986) Dilettantism in hydrology: Transition or destiny? Water Resources Research 22(9S):177S–188S

Klemeš V (2000a) Tall tales about tails of hydrological distributions. I. Journal of Hydrologic Engineering 5(3):227–231, DOI 10.1061/(ASCE)1084-0699(2000)5:3(227)

Klemeš V (2000b) Tall tales about tails of hydrological distributions. II. Journal of Hydrologic Engineering 5(3):232–239, DOI 10.1061/(ASCE)1084-0699(2000)5:3(232)

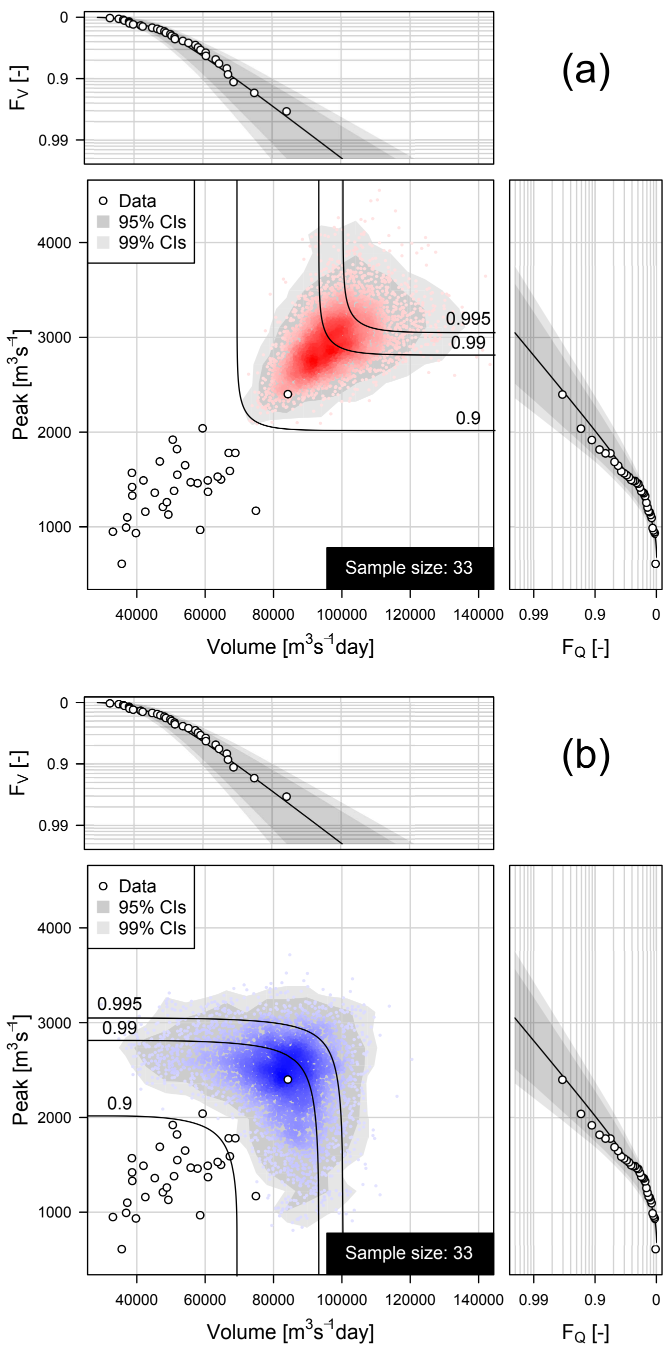
Koutsoyiannis D (2004) Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrological Sciences Journal 49(4):575–590, DOI 10.1623/hysj.49.4.575.54430

Koutsoyiannis D (2006) Nonstationarity versus scaling in hydrology. Journal of Hydrology 324(1–4):239–254

Serinaldi F (2012) 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:C3721C3721, URL http://www.hydrol-earth-syst-sci-discuss.net/9/C3721/2012/

Yue S, Ouarda TBMJ, Bobée B, Legendre P, Bruneau P (1999) The Gumbel mixed model for flood frequency analysis. Journal of Hydrology 226(1–2):88–100

List of Figures



(a)
F_V [-]
0
0.9
0.99
Data
95% CIs
99% CIs
Peak [m^{3}s^{-1}]
4000
3000
2000
1000
0.995
0.99
0.9
Sample size: 33
40000
60000
80000
100000
120000
140000
Volume [m^{3}s^{-1}day]
0.99
0.9
0
F_Q [-]
(b)
F_V [-]
0
0.9
0.99
Data
95% CIs
99% CIs
Peak [m^{3}s^{-1}]
4000
3000
2000
1000
0.995
0.99
0.9
Sample size: 33
40000
60000
80000
100000
120000
140000
Volume [m^{3}s^{-1}day]
0.99
0.9
0
F_Q [-]

Figure 1

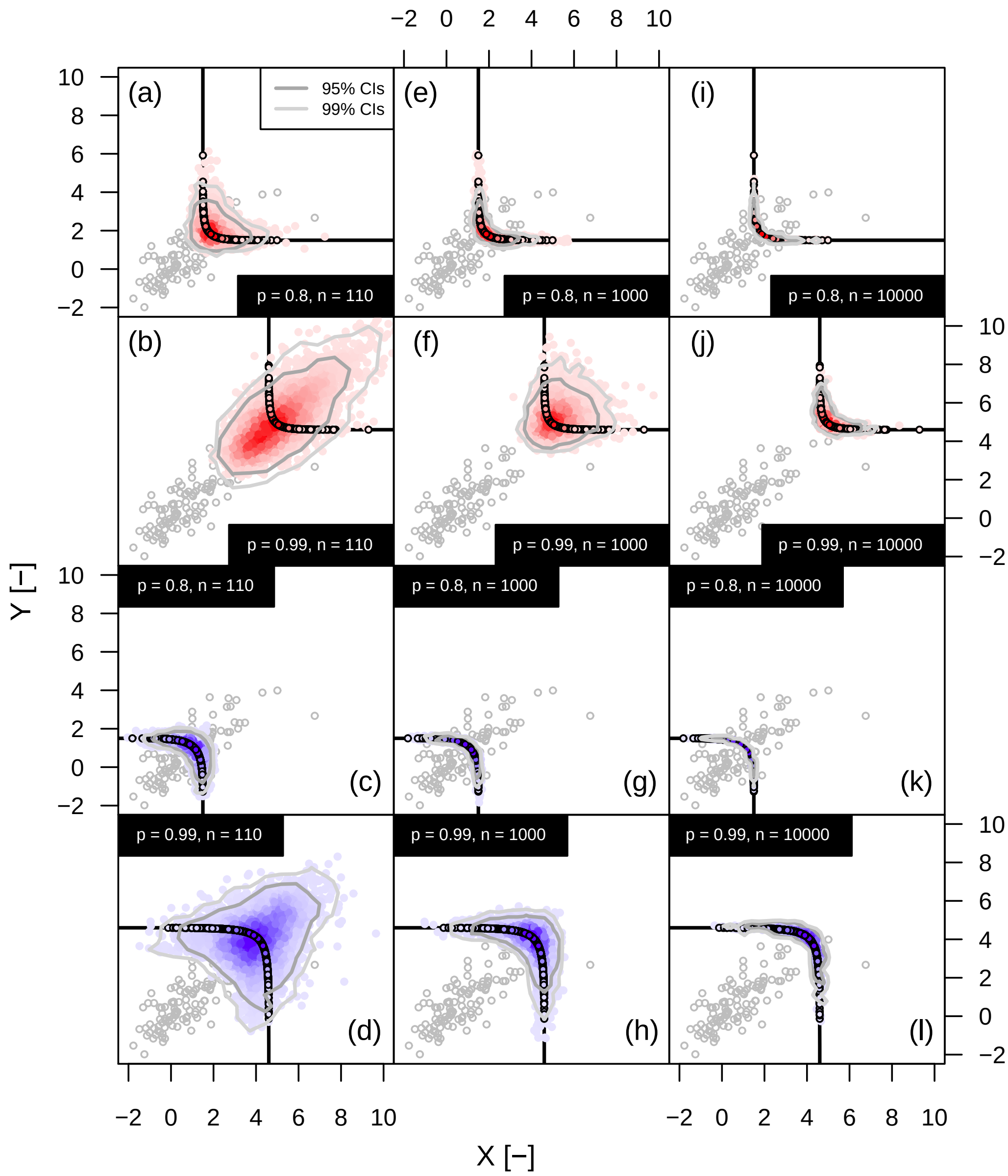

95% CIs
99% CIs
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
p = 0.8, n = 110
p = 0.8, n = 1000
p = 0.8, n = 10000
p = 0.99, n = 110
p = 0.99, n = 1000
p = 0.99, n = 10000
p = 0.8, n = 110
p = 0.8, n = 1000
p = 0.8, n = 10000
p = 0.99, n = 110
p = 0.99, n = 1000
p = 0.99, n = 10000
Y [−]
X [−]

Figure 2