1 **Closure to “Probability Distribution of Waiting Time of the k th Extreme**

2 **Event Under Serial Dependence” by F. Serinaldi and F. Lombardo**

3 **June 2020, Vol. 25, No. 6, pp. 04020025(1–11)**

4 **DOI: 10.1061/(ASCE)HE.1943-5584.0001923**

5 Francesco Serinaldi^{1,2} and Federico Lombardo³

6 ¹School of Engineering, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK. Email:

7 francesco.serinaldi@ncl.ac.uk

8 ²Willis Research Network, 51 Lime St., London, EC3M 7DQ, UK

9 ³Corpo Nazionale dei Vigili del Fuoco, Ministero dell’Interno, Piazza del Viminale, 1, Rome

10 00184, Italy

11 We thank very much the Discussor for his interest in the original article. He makes some points
12 of practical importance that stem from a widespread approach to statistical analysis characterizing
13 a significant part of the hydro-climatological literature. However, such an approach suffers from
14 two problems that may usually lead to incorrect conclusions and statements, namely: (i) merging
15 and confusing theoretical aspects and application settings, and (ii) overlooking/missing formal
16 definitions, assumptions, and properties of the statistical models when they are applied to real-world
17 problems/analysis. Some examples and detailed discussions of these issues and their consequences
18 can be found for instance in Lombardo et al. (2012), Lombardo et al. (2014), Serinaldi (2015),
19 Serinaldi et al. (2015), Volpi et al. (2015), Serinaldi (2016), Lombardo et al. (2017), Serinaldi
20 and Kilsby (2018), Serinaldi et al. (2018), Lombardo et al. (2019), Serinaldi et al. (2020a), and
21 Serinaldi et al. (2020b), among others.

22 Concerning the specific case under discussion, the mathematical expressions of the beta negative
23 binomial $\beta\mathcal{NB}$ distribution (Eqs. 1-6 of the original article) indicate that this model describes the

number of "trials"/"events" W necessary to obtain k failures for a given rate of failure p , and not the waiting time of "*at least k extreme events*".

Therefore, the application of the $\beta\mathcal{NB}$ distribution requires a preliminary clear and precise definition of two elements: (i) the process of interest enabling an unambiguous identification of what "trials" or "events" mean, and (ii) the rule defining successes/failures in each trial. These theoretical concepts need to be univocally specialized for the case study of interest before analyzing data, and results must be interpreted accordingly. Therefore, for the "peak flow" example reported in the original study, (i) the process considered for illustration purposes is that of the (single and unique) annual maximum peak flows, X_t , and a "trail"/"event" is the occurrence of an annual peak flow (and nothing else), while (ii) the "success/failure" rule is $X_t \leq x_d$, where x_d is a given design value. With this setting, the $\beta\mathcal{NB}$ distribution describes the probability of the number of events (i.e. annual maximum peak flows) before observing the k th extreme event, i.e. the annual maximum peak flow exceeding x_d . Similarly, for the "temperature" example, (i) X_t specializes as the (single and unique) monthly maximum temperature anomalies, and a "trail"/"event" is the occurrence of a monthly maximum temperature anomaly (and nothing else), while the "success/failure" rule is again $X_t \leq x_d$. As for the "peak flow" example, with this setting, the $\beta\mathcal{NB}$ distribution describes the probability of the number of events (i.e. monthly maximum temperature anomalies) before observing the k th extreme event, i.e. the k th monthly maximum temperature anomaly exceeding x_d .

Under the above mentioned settings, it is clear that the "peak flow" analysis does not (and cannot) deal in any respect with the occurrence of possible high values lower than the annual maximum flow peaks or monthly maximum temperature anomaly. In this context, Discusser's comment seems to be motivated by the practical interest in the occurrence of every flow value exceeding a given threshold rather than the occurrence of annual maximum peak flow exceeding that threshold. Of course, the $\beta\mathcal{NB}$ distribution can be applied to study the occurrence of the k th failure/success in Discusser's setting; however, in this case, X_t specializes as a different process, e.g. daily/hourly peak flows, and the corresponding "events" are not annual maximum peak flows (occurring once

a year) but e.g. daily/hourly flows occurring every day/hour. In this new setting (which requires the availability of daily/hourly sequences), the $\beta\mathcal{NB}$ distribution describes the probability of the number of events (now, daily/hourly flows) before observing the k th extreme event, i.e. the k th daily/hourly flows exceeding x_d .

Therefore, Discusser's misunderstanding and consequent incorrect conclusions result from merging the practical problem of his interest and the $\beta\mathcal{NB}$ statistical analysis in a discussion that misses or overlooks the importance of a precise formal definition of the key elements entering the statistical analysis for the correct and unambiguous interpretation of the results. Indeed, when stating that "*in a failure year, the threshold might be exceeded more than once*", the Discusser keeps the "success/failure" rule but suddenly changes the target process, which is no longer that of the unique annual maximum peak flows or monthly maximum temperature anomalies (occurring once a year or once a month by definition) but the process of e.g. daily/hourly records, which implies a different definition and meaning of what one calls "event". An analysis that is set up to study a specified process cannot be used to draw conclusions about another process without updating its assumptions.

To summarize, contrary to Discusser's statements, the computed waiting times throughout the original articles are those of the k th extreme "event" (as reported correctly in the title), i.e. the k th extreme "annual maximum peak flow" (exceeding a specified value), the k th extreme "monthly maximum temperature anomaly", or the k th extreme "daily/hourly peak flow", if one performs the same analysis on daily/hourly peak flow data. The same holds for any other resolution or any process in any applied field (as discussed in the conclusion of the original article), under a preliminary clear and unambiguous definition of the process of interest and the corresponding "trials"/"events".

Finally, we are indebted to the Discusser for his comment on our paper, as it allows us to stress once again the importance of an accurate consideration of the assumptions, concepts, and meaning of the terminology behind statistical models before their application.

REFERENCES

- Lombardo, F., Napolitano, F., Russo, F., and Koutsoyiannis, D. (2019). “On the exact distribution of correlated extremes in hydrology.” *Water Resources Research*, 55(12), 10405–10423.
- Lombardo, F., Volpi, E., and Koutsoyiannis, D. (2012). “Rainfall downscaling in time: theoretical and empirical comparison between multifractal and hurst-kolmogorov discrete random cascades.” *Hydrological Sciences Journal*, 57(6), 1052–1066.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S. M. (2014). “Just two moments! a cautionary note against use of high-order moments in multifractal models in hydrology.” *Hydrology and Earth System Sciences*, 18(1), 243–255.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Serinaldi, F. (2017). “A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall.” *Water Resources Research*, 53(6), 4586–4605.
- Serinaldi, F. (2015). “Dismissing return periods!.” *Stochastic environmental research and risk assessment*, 29(4), 1179–1189.
- Serinaldi, F. (2016). “Can we tell more than we can know? the limits of bivariate drought analyses in the united states.” *Stochastic Environmental Research and Risk Assessment*, 30(6), 1691–1704.
- Serinaldi, F., Bárdossy, A., and Kilsby, C. G. (2015). “Upper tail dependence in rainfall extremes: would we know it if we saw it?.” *Stochastic environmental research and risk assessment*, 29(4), 1211–1233.
- Serinaldi, F., Chebana, F., and Kilsby, C. G. (2020a). “Dissecting innovative trend analysis.” *Stochastic Environmental Research and Risk Assessment*, 34(5), 733–754.
- Serinaldi, F. and Kilsby, C. G. (2018). “Unsurprising surprises: The frequency of record-breaking and overthreshold hydrological extremes under spatial and temporal dependence.” *Water Resources Research*, 54(9), 6460–6487.
- Serinaldi, F., Kilsby, C. G., and Lombardo, F. (2018). “Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology.” *Advances in Water Resources*, 111, 132–155.
- Serinaldi, F., Lombardo, F., and Kilsby, C. G. (2020b). “All in order: Distribution of serially correlated order statistics with applications to hydrological extremes.” *Advances in Water Resources*,

104 144, 103686.
105 Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., and Koutsoyiannis, D. (2015). “One hundred years
106 of return period: Strengths and limitations.” *Water Resources Research*, 51(10), 8570–8585.