Supplement of Hydrol. Earth Syst. Sci., 29, 5031–5047, 2025
https://doi.org/10.5194/hess-29-5031-2025-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Informativeness of teleconnections in frequency analysis of rainfall extremes

Andrea Magnini et al.

Correspondence to: Andrea Magnini (andrea.magnini@unibo.it)

S1. Statistical significance of empirical RML values: Monte Carlo simulations

The present Appendix describes the methodology adopted to empirically assess the statistical significance of RML values obtained in the study (i.e., ratio of models' likelihood, see Section 2.2). In particular, the statistical test consists of comparing RML with lower threshold values (th_{RML}) associated with a 5% significance level (see Section 2.2). We determine these threshold values by means of Monte Carlo (MC) simulations that are designed to empirically identify the RML value associated with a p-value of 95% when comparing the likelihood of a doubly stochastic GEV (i.e., GEV_{DS} with DS=1, 2, 3, see also Section 2.2) with that of a stationary distribution (i.e., GEV_0) when the parent distribution of the generated annual sequences is GEV_0. Specifically, our simulations generate stationary synthetic annual sequences (AMS), but we do not fit new frequency distributions on the synthetic AMS. Instead, the original frequency models are used to compute the likelihood of the synthetic timeseries. The procedure to define values of th_{RML} can be described as follows:

1. For each station, the GEV_0 model (i.e., the stationary distribution fitted to the at-site mean, and regional L-CV and L-CS) is used to generate 1000 annual maxima time series with the same length of the original AMS.

2. For each simulation, RML is computed with reference to the GEV_0 model and, in turn, one of the considered doubly-stochastic models GEV_{DS} (i.e., GEV_1, GEV_2, and GEV_3). As stated above, we consider only the frequency models fitted to the original data, but RML refers to the synthetic time series. Since the parent distribution is GEV_0, RML is expected to be low (i.e., GEV_0 should provide the best fit in the majority of cases).

3. For each station (st) in the study region and each doubly stochastic GEV model, we identify the 95th percentile from the series of simulated RML. This step produces 680 percentiles for each doubly stochastic model, meaning that each station has its own percentiles ($th_{RML,st}$).

4. The values of th_{RML} for each doubly-stochastic GEV model for the entire region are defined as 95th percentiles of the 680 at-site values $th_{RML,st}$ from the previous step. Thus, we obtain a single RML threshold for a specific GEV_{DS} model that is valid for the whole study region.

We repeat this procedure for both durations, 1h and 24h, and all of the teleconnections considered for the test, namely NAO, EA-WR, PDO, WeMOI and MOI (see Section 4.1). We therefore obtain several values of th_{RML}, each one corresponding to a specific duration, teleconnection and GEV type.

The procedure described above preserves, or reproduces, all the main elements of our specific application case that may impact RML's diagnostic power: the study area morphoclimatic variability is considered by applying the MC experiments to the entire dataset; synthetic time series preserve the original length at each site (step 1). Furthermore, considering the original frequency distributions (step 2) is a key point, as it allows us to test the predictive power of RML in the real case, avoiding us to make additional assumptions on synthetic distributions.

Having the MC reference threshold values of RML, we can test the doubly-stochastic hypothesis; if the RML value we obtain for a specific case in our study region (i.e., for a given site duration, teleconnection and GEV model) is greater than the corresponding simulated th_{RML}, then the frequency regime of rainfall extremes for that site can be considered doubly-stochastic at 5% significance level.

S2: Field significance of doubly-stochasticity signals: spatial bootstrap resampling

This Appendix describes the bootstrap experiments adopted to test the field significance of doubly-stochasticicty spatial signals in presence of cross-correlation, or intersite spatial dependence (see Sections 2.2 and 3.2 and 4.2). In particular, we test the statistical significance of the number of doubly stochastic regimes that is observed in a specific region (n_{DS}), that is the number of sites where, according to the RML analysis, a given GEV_{DS} for a specific teleconnection provides a better fit to the data relative to GEV_0. The test adopts a threshold number of stations, th_{nDS}, that represents the 95th percentile of 1000 n_{DS} realizations for 1000 stationary (i.e. non doubly-stochastic) synthetic regions, which are generated from the original regional sample of annual maxima through spatial bootstrap resampling (see also Castellarin et al., 2024; Vogel et al., 2001). The bootstrap procedure to define values of th_{nDS} is the following:

1. We generate 1000 synthetic regions by randomly reshuffling year-wise 1000 times the dataset of 680 AMS of rainfall depths with duration 1h (or 24h) year-wise. Each reshuffling produces a synthetic AMS dataset that is distinct from the original one, while preserving the distribution of observations among all AMS in any given year. Since the reshuffling is random, the 1000 synthetic datasets are realizations of a stationary world.

2. For each synthethic reshuffled dataset, we repeat the same hierarchical regional frequency analysis (RFA) adopted for the original dataset (see Sections 2.2 and 3.2). First, the stationary model, GEV_0, is fitted to the syntehtic data. Second, the synthetic time series of at-site μ and 30km tile L-CV are computed. Third, the polynomial functions are fitted to the synthetic μ and L-CV timeseries and the original (i.e. non-reshuffled) teleconnection timeseries; it is worth underlining here that any correlation, or relationship, between the original teleconnection and the synthetic μ and L-CV timeseries is spurious, as, by construction, these synthetic μ and L-CV timeseries are obtained from stationary realizations. Fourth, the parameters of GEV_1, GEV_2 and GEV_3 are obtained for each synthetic time series.

3. For each synthethic reshuffled dataset, RML values (ratio of models' likelihood, see Section **??**) are computed for GEV_1, GEV_2 and GEV_3

4. For each synthethic reshuffled dataset, n_{DS} is computed for GEV_1, GEV_2 and GEV_3 (see also Supporting material A and Section 3.2)

5. The values of th_{nDS} are then obtained as the 95th percentiles of 1000 synthetic n_{DS} obtained for each considered case, that is given duration (i.e., 1h or 24h), teleconnection (i.e., EA-WR and WeMOI, see Section 4.2), doubly stochastic model GEV_{DS} (i.e., GEV_1, GEV_2 or GEV_3), and RML threshold (i.e., the theoretical one, and the empirical obtained using the procedure illustrated in Supporting material A).

The field significance of the doubly-stochasticity signals can then be assessed at the 5% significance level by testing whether the empirical n_{DS} values obtained for the original dataset are larger than the empirical th_{nDS} obtained as described above.

References

Castellarin, A., Magnini, A., Kyaw, K.K., Ciavaglia, F., Bertola, M., Blöschl, G., Volpi, E., Claps, P., Viglione, A., Marinelli, A., Vogel, R.M., 2024. Frequency of Italian Record-Breaking Floods over the Last Century (1911–2020). Atmosphere 15, 865. https://doi.org/10.3390/atmos15070865

Vogel, R.M., Zafirakou-Koulouris, A., Matalas, N.C., 2001. Frequency of record-breaking floods in the United States. Water Resources Research 37, 1723–1731. https://doi.org/10.1029/2001WR900019