Kalai, Chingka; Mondal, Arpita; Griffin, Adam; Stewart, Elizabeth. 2020.
Comparison of nonstationary regional flood frequency analysis techniques based on the index-flood approach.

Supplementary Information for

Comparison of non-stationary regional flood frequency analysis techniques based on the index-flood approach

Chingka Kalai, Arpita Mondal, Adam Griffin and Elizabeth Stewart

The O'Brien and Burn Method (OBM)

O'Brien and Burn (2014) use a regional pooling based on geographic location of the catchments, trends and seasonality of the flood data. In the non-stationary index-flood procedure used by O'Brien and Burn (2014, denoted by OBM), the following steps are employed:

(i) The at-site algebraic mean $\overline{X}_i$ is used as the index-flood to normalize records $X_i(t)$ at site i, where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, n_i$. The normalized flood records are given by $Y_i(t) = X_i(t)/\overline{X}_i$.

(ii) Non-stationary GEV distribution is then fitted to the normalized records. $\mu_i^Y(t)$, $\sigma_i^Y(t)$ and $\xi_i^Y(t)$ represent the location, scale and shape parameters respectively for site i in the normalized space.

(iii) Regional parameters are obtained in the normalized space by weighted-averaging given by

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^N n_i} \quad (1)$$

where $\hat{\theta}_k^R$ is the regional average of the k^{th} parameter and $\hat{\theta}_k^{(i)}$ is the value of the k^{th} parameter for the i-th site.

(iv) The non-stationary regional growth curve $q(f,t)$ is obtained by inverting the GEV distribution with the regional parameters. From the regional growth curve, time-varying quantiles corresponding to a given return period are obtained.

(v) These regional quantiles are further multiplied by the site-specific index-flood to get the time-varying at-site quantiles $Q_i(f,t)$ as follows:

$$Q_i(f,t) = \overline{X}_i * q(f,t) \quad (2)$$

Thus, the non-stationary flood quantiles is obtained at each site i.

The Hanel Method (HM and HM*)

Hanel et al. (2009, denoted by HM) first use a regional pooling to identify homogeneous regions, based on the dispersion parameter that is the ratio of the scale and location parameters. In HM, the regional growth curve is represented directly in terms of the dispersion parameter

and the shape parameter estimated from the pooled normalized data. The time-varying location parameter $\mu_i(t)$ at each site is chosen as the index-flood. HM consists of the following steps:
(i) Stationary location parameter at each site ($\mu_{0,i}$), regional trend in location parameter (μ_1), regional dispersion (γ_R) and regional shape (ξ_R) are estimated by pooling records from all sites in the homogeneous region.
(ii) Time varying location parameter is then obtained for each site i as follows:

$$\mu_i(t) = \mu_{0,i} + (\mu_1 \times \mathrm{t}) \tag{3}$$

(iii) The growth curve is constructed directly in terms of the dispersion and the shape parameters as follows:

$$q(f) = 1 - \frac{\gamma_R}{\xi_R}\left(1 - \left(\log\left(\frac{1}{1-f}\right)\right)^{-\xi_R}\right) \tag{4}$$

Therefore, in HM, the growth curve is stationary, while the index-flood is non-stationary.
(iv) The nonstationary quantiles in the original space are then obtained by multiplying the growth curve with the index-flood.

$$Q_i(f,t) = \mu_i(t) * q(f) \tag{5}$$

Applying a similar ideology of normalization, it is possible to implement a modified non-stationary index-flood approach for regional flood frequency analysis (RFFA). This modified approach is denoted by HM*, and consists of the following steps.
(i) At-site location, scale and shape parameters for the i-th site denoted by $\mu_i(t)$, $\sigma_i(t)$ and $\xi_i(t)$ respectively are obtained from the fitted non-stationary GEV distribution. Records at each site $X_i(t)$ are normalized by the at-site location parameter $\mu_i(t)$. The normalized records are denoted by $Y_i(t)$=$X_i(t)$/$\mu_i(t)$.
Steps (ii), (iii) and (iv) are the same as OBM.
(v) Finally, the regional quantiles are multiplied by the site-specific index-flood to get the time-varying at-site quantiles $Q_i(f,t)$ as follows:

$$Q_i(f,t) = \mu_i(t) * q(f,t) \tag{6}$$

The Nam Method (NM)

In OBM and HM*, in the transformed space, the distribution is still non-stationary, thereby representing the independent but not identically distributed (i/nid) case, violating the iid assumptions of index-flood method. Nam et al. (2015) propose an extension of HM* which can circumvent this violation, where they fit stationary distribution to the transformed flood records (denoted by NM).

After following step (i) of HM*, step (ii) of NM involves estimation of stationary parameters of the at-site GEV distributions denoted by μ_i^Y, σ_i^Y and ξ_i^Y.

(iii) Regional parameters are thereafter obtained by the weighted averaging given by Eq. (1) above.

(iv) The regional growth curve $q(f)$ is then obtained by inverting the stationary GEV distribution with regional parameters.

(v) Finally, the regional quantiles are transformed back to their original space by multiplying with the index-flood as follows:

$$Q_i(f,t) = \mu_i(t) * q(f) \tag{7}$$

The Sung Method (SM)

More recently, Sung et al. (2018) proposed a method where stationary index flood method is employed to the detrended non-stationary data. The index flood procedure proposed by them (denoted as SM) consist of following steps.

(i) Following estimation of the trend parameter (β_i) at each site in the region, a detrended flood series is obtained by subtracting estimated trend from the observed flood records at that site, given by $Z_i(t) = X_i(t) - \beta_i t$).

(ii) The detrended at-site algebraic mean $\overline{Z}_i$ is used as the index-flood to normalize detrended records $Z_i(t)$ at site i. The normalized flood records are given by $Y_i(t) = Z_i(t)/\bar{Z}_i$.

(iii) Stationary GEV distribution is then fitted to the normalized records to obtain μ_i, σ_i and ξ_i for site i.

(iv) Regional parameters are obtained using by the weighted averaging given by Eq. (1).

(v) The regional growth curve in the detrended space $\tilde{q}(f)$ is obtained by inverting the stationary GEV cumulative distribution function (cdf) with regional parameters.

(vi) Nonstationary quantiles in the original space are finally obtained by multiplying the regional quantile with the index-flood and adding the trend component as follows:

$$Q_i(f,t) = (\overline{Z}_i \times \tilde{q}(f)) + (\beta_i \times t) \tag{8}$$

The Modified Basu and Srinivas Method (BSM*)

Basu and Srinivas (2013) propose a mathematical transformation-based approach for stationary RFFA as an alternative to the population index-flood method. Here, a non-stationary index-flood approach is proposed extending the concept of Basu and Srinivas (2013) that is hereinafter denoted by BSM*. The following steps are carried out in BSM*.

(i) Non-stationary GEV distribution is fitted to the at-site records where the location, scale and shape parameters of the GEV distribution at the i-th site are denoted by $\mu_i(t)$, $\sigma_i(t)$ and $\xi_i(t)$ respectively.

(ii) At-site flood records $X_i(t)$ are transformed into dimensionless standardized residuals using the following transformation considering the at-site parameters.

$$Y_i(t) = \frac{1}{\xi_i(t)} \ln \{1 + \frac{\xi_i(t)(X_i(t) - \mu_i(t))}{\sigma_i(t)}\} \quad (9)$$

This transformation converts the random variable to another dimension where the location, scale and shape parameters are less biased compared to the regional parameters. The transformed records are independent and identically distributed, thereby satisfying the primary iid assumption of the index-flood method.

(iii) The stationary parameters obtained from the transformed realizations are thereafter used to compute the regional average dimensionless parameters in the transformed space by weighting the parameters with the number of records available for each site, similar to the other methods, given by Eq. (1).

(iv) Regional average parameters are used to obtain the stationary regional growth curve $q(f)$ by inverting the GEV cdf.

(v) Finally, the at-site non-stationary quantiles in the original space are obtained by the back transformation given by

$$Q_i(f,t) = \mu_i(t) - \frac{\sigma_i(t)}{\xi_i(t)}[1 - exp\{\xi_i(t) * q(f)\}] \ . \quad (10)$$

Bootstrap vector resampling approach for uncertainty estimation

To estimate the uncertainty associated with the time-varying flood quantiles, modified version of the bootstrap vector resampling approach (Burn, 2003; O'Brien and Burn, 2014) is used in this study. This approach is adopted to avoid the bias in parameter estimation due to small sample size. The bootstrap resampling approach can result in asymmetric confidence intervals thereby representing more realistic conditions (Obeysekera and Salas 2014). To preserve the spatial structure, the method considers a vector of flood values at all the sites in a region as a 'record', and such records are drawn repeatedly. The following are the steps to obtain the confidence intervals for time-varying flood quantiles corresponding to a given probability of exceedance.

(i) Standardized at-site residuals $Y_i(t)$ obtained from Eq. 9 are considered for the period of record (say, n years) at all the sites in a region with *N* sites.
(ii) This *n* x *N* residual data matrix is replicated 999 times, and concatenated one after the other, leading to a large data matrix of dimensions (999*n*) x *N*. This large data matrix is reshuffled randomly to create a permuted large data matrix of dimensions (999*n*) x *N*. The permuted large data matrix is thereafter broken down into 999 bootstrap samples each of dimensions *n* x *N*.
(iii) For each of these 999 bootstrap samples of standardized residuals, at-site stationary GEV parameters are estimated. These at-site parameters are averaged to obtain the regional parameter estimates which are further used to get the regional growth curve.
(iv) The regional quantiles are converted back to the original dimension by the back transformation proposed in Eq. (10). Thus, the time-varying flood quantiles for each of the 999 bootstrap samples are obtained.
(v) At each time step, the empirical quantiles of the 999 bootstrap samples for a given confidence level gives the upper and lower uncertainty bounds of the estimated flood quantiles at that time step.

Supplementary Table

Table S1: Test results (p-values) for significance of non-stationarity in the seven sites of the synthetically generated region. Low p-values imply rejection of the null hypothesis of no-trend.

p-value	**Site#1**	**Site#2**	**Site#3**	**Site#4**	**Site#5**	**Site#6**	**Site#7**
Mann-Kendall Trend Test	0.019	0.011	0.039	0.003	0.009	0.016	0.0002
Likelihood Ratio Test	0.0017	0.0002	0.0051	2.48×10^{-6}	0.0002	0.0068	0.0038

References

Basu, B., and Srinivas, V. V. (2013). "Formulation of a mathematical approach to regional frequency analysis." *Water Resources Research*, 49, 6810–6833.
Burn, D. H. (2003). "The use of resampling for estimating confidence intervals for single site and pooled frequency analysis." *Hydrological Sciences Journal*, 48(1), 25–38.
Hanel, M., Buishand, T. A., and Ferro, C. A. T. (2009). "A nonstationary index flood model for precipitation extremes in transient regional climate model simulations." *Journal of*

Geophysical Research, 114(August), 1–16.

Nam, W., Kim, S., Kim, H., Joo, K., and Heo, J.-H. (2015). “The evaluation of regional frequency analyses methods for nonstationary data.” *Proceedings of the International Association of Hydrological Sciences*, 371(1), 95–98.

O’Brien, N. L., and Burn, D. H. (2014). “A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow.” *Journal of Hydrology*, Elsevier B.V., 519(PB), 2040–2048.

Obeysekera, J., and Salas, J. D. (2014). “Quantifying the uncertainty of design floods under non-stationary conditions.” *Journal of Hydrologic Engineering*, 19(7), 1438–1446.

Sung, J. H., Kim, Y. O., and Jeon, J. J. (2018). “Application of distribution-free nonstationary regional frequency analysis based on L-moments.” *Theoretical and Applied Climatology*, Theoretical and Applied Climatology, 133(3–4), 1219–1233.