

Regional simulation in package `lmomRFA`

Prepared for `lmomRFA` version 3.5, February 2023

Introduction

Hosking and Wallis (2005) (HW for short) Section 6.4 describes a method, based on Monte Carlo simulation, for computing root mean square errors (RMSEs) and constructing error bounds for quantile estimates obtained using the regional L -moment algorithm. We now recommend some modifications to this method. These modifications are implemented in full in R package `lmomRFA`, version 3.5 and later, and affect results returned by functions `regsimq`, `regquantbounds`, and `sitequantbounds` in that package.

HW equation (6.19) shows how to calculate error bounds for at-site quantiles from simulation results, but HW recommends (p. 93, last paragraph) that the region that we simulate from should not have the same population L -moments as the sample L -moments of the actual region. So, as in the example in HW Section 6.5, to simulate from North Cascades region we assign the sites L -CVs that increase linearly from site 1 to site 19, and the error bounds that we obtain for site 1 reflect this — in the lower tail the quantile estimates are negatively biased, and the lower error bound is much closer to the estimate than the upper error bound is (e.g., for $F = 0.01$ the bounds are $\hat{Q}(F)/1.008$ and $\hat{Q}(F)/0.830$). But the actual site 1 has higher L -CV than the regional average, so the results for simulated site 1 can't apply to it. We now describe a modified simulation procedure that avoids this and similar issues.

A modified simulation procedure

A site in the simulations is specified by its sample size n_i and its at-site frequency distribution $Q_i(F)$ (implied by the site's population L -moments and distribution family). Writing the site- i frequency distribution as $Q_i(F) = \mu_i q_i(F)$ (as in HW, eq. (6.17)), it is helpful to consider the index flood μ_i and the at-site growth curve $q_i(F)$ separately. The quantities n_i and μ_i directly correspond to those of the actual site i . But the at-site growth curves of sites in the simulations are intended merely to be representative of a region as heterogeneous as the actual region, with no direct correspondence between simulated and real sites. Indeed, for estimating accuracy at actual site i it is reasonable to use a set of simulations in which the site- i growth curve is, at each simulation run, equally likely to be any of the specified set of at-site growth curves. When generating successive simulations, it is therefore appropriate to permute the at-site growth curves, using a different permutation at each simulation run in such a way that across the entire set of runs each simulated site is assigned each at-site

growth curve in approximately equal numbers. One way to achieve this is to randomly permute the at-site growth curves at each simulation run.

We also recommend some modifications to the accuracy computations — RMSEs and error bounds — in HW Section 6.4. In particular, HW eq. (6.15) needs modification because permutation of the at-site growth curves in the simulations means that the site- i quantile function $Q_i(F)$ is not the same at each repetition. Let $Q_i^{[m]}(F)$ denote the quantile function used for site i at the m th repetition: then (6.15) becomes

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}^{[m]}(F) - Q_i^{[m]}(F)}{Q_i^{[m]}(F)} \right\}^2 \right]^{1/2}. \quad (1)$$

We consider three cases where accuracy computations may be needed: for the regional growth curve, for quantile estimates at sites used in the regional L -moment algorithm, and for quantile estimates at “ungaged” sites, i.e., sites that have no direct observations but where the site-specific scaling factor (index flood) has been estimated by some other means.

Accuracy computations for the regional growth curve

The simulations provide estimates $\hat{q}^{[m]}(F)$, $m = 1, \dots, M$, of the regional growth curve. The m th repetition enables comparison of $\hat{q}^{[m]}(F)$ with each true at-site growth curve $q_i(F)$, $i = 1, \dots, N$, so the set of simulations provides a total of MN comparisons. As noted above, the at-site growth curve for a specified site in the simulated region is regarded as equally likely to be any of the specified set of at-site growth curves. Each of the MN comparisons is therefore equally valid for any of the simulated sites, and the natural estimator of the relative RMSE of the estimated regional growth curve is the relative RMSE across all MN possible comparisons:¹

$$R^R(F) = \left[(MN)^{-1} \sum_{i=1}^N \sum_{m=1}^M \left\{ \frac{\hat{q}^{[m]}(F) - q_i(F)}{q_i(F)} \right\}^2 \right]^{1/2}. \quad (2)$$

The absolute RMSE of the regional growth curve is then given by $R^R(F)$ scaled by the estimated regional growth curve $\hat{q}(F)$ obtained from the regional L -moment algorithm:

$$R^{\text{abs,R}}(F) = \hat{q}(F) R^R(F). \quad (3)$$

¹Note that $R^R(F)$ in (2) is not the same as in HW eq. (6.16): it is the RMSE rather than the arithmetic mean of the relative RMSEs in (6.15). But for simulations in which the at-site growth curves are permuted, the sites’ relative RMSEs are approximately equal, and their RMSE and arithmetic mean are also approximately equal.

Similarly the MN ratios $\hat{q}^{[m]}(F)/q_i(F)$, $m = 1, \dots, M$, $i = 1, \dots, N$, provide the basis for computing quantiles for use in error bounds: for example, the 5th and 95th percentiles of these ratios are values $L_{.05}(F)$ and $U_{.05}(F)$ that can be used, as in HW eq. (6.19), to obtain error bounds for the regional growth curve:

$$\frac{\hat{q}(F)}{U_{.05}(F)} \leq q(F) \leq \frac{\hat{q}(F)}{L_{.05}(F)}. \quad (4)$$

Accuracy computations for a site used in the regional L-moment algorithm

For a site used in the regional L -moment algorithm, eq. (1) can be used directly to give the relative RMSE $R_i(F)$ of estimated quantiles. For site i , with estimated quantiles $\hat{Q}_i(F)$ obtained from the regional L -moment algorithm, the absolute RMSE of estimated quantiles is correspondingly given by

$$R_i^{\text{abs}}(F) = \hat{Q}_i(F) R_i(F). \quad (5)$$

The M ratios $\hat{Q}_i^{[m]}(F)/Q_i^{[m]}(F)$, $m = 1, \dots, M$, provide the basis for computing quantiles for use in error bounds: for example, the 5th and 95th percentiles of these ratios are values $L_{.05}(F)$ and $U_{.05}(F)$ that can be used, as in HW eq. (6.19), to obtain error bounds for the regional growth curve:

$$\frac{\hat{Q}_i(F)}{U_{.05}(F)} \leq Q_i(F) \leq \frac{\hat{Q}_i(F)}{L_{.05}(F)}. \quad (6)$$

Accuracy computations for an ungaged site

At an ungaged site we assume that there is an estimate μ_0 of the site-specific scaling factor (index flood), computed by a method independent of the estimated regional growth curve $\hat{q}(F)$ obtained from the regional L -moment algorithm. This method might, for example, involve estimation of a regression relationship between the index flood and site characteristics, as described in HW Section 8.4. Quantile estimates for the ungaged site are computed as $\hat{Q}_0(F) = \hat{\mu}_0 \hat{q}(F)$.

In regional simulations we specify a true value μ_0 for the index flood at the ungaged site and a distribution of the index-flood estimate. Typically we set $\mu_0 = \hat{\mu}_0$, so that the distributions of simulated index-flood estimates is centred on the index-flood estimate from the actual ungaged site. We can then generate simulated values of the estimated index flood, and relative RMSEs and error bounds can be computed for quantile estimates at the ungaged site. Let $\hat{\mu}_0^{[m]}$, $m = 1, \dots, M$, be simulated values of $\hat{\mu}_0$. These can be combined with simulations of the regional growth curve from the regional L -moment algorithm to give simulated quantile estimates at the ungaged site: $\hat{Q}_0^{[m]}(F) = \hat{\mu}_0^{[m]} \hat{q}^{[m]}(F)$, $m = 1, \dots, M$. Concordantly with our overall simulation

approach, the at-site growth curve for the unaged site is regarded as equally likely to be any of the specified set of at-site growth curves, so the true quantile at the unaged site is equally likely to be any of the quantities $\mu_0 q_i(F)$, $i = 1, \dots, N$. As with accuracy computations for the regional growth curve, there are MN available comparisons of estimated and true quantiles, and the natural estimator of the relative RMSE of a quantile estimate at the unaged site is the relative RMSE across all MN possible comparisons:

$$R_0(F) = \left[(MN)^{-1} \sum_{i=1}^N \sum_{m=1}^M \left\{ \frac{\hat{Q}_0^{[m]}(F) - \mu_0 q_i(F)}{\mu_0 q_i(F)} \right\}^2 \right]^{1/2}. \quad (7)$$

The absolute RMSE of quantile estimates at the unaged site is then given by $R_0(F)$ scaled by the estimated regional growth curve obtained from the regional L -moment algorithm:

$$R_0^{\text{abs}}(F) = \hat{Q}_0(F) R_0(F) = \hat{\mu}_0 \hat{q}(F) R_0(F). \quad (8)$$

Similarly the MN ratios $\hat{Q}_0^{[m]}(F)/(\mu_0 q_i(F))$, $m = 1, \dots, M$, $i = 1, \dots, N$, provide the basis for computing quantiles for use in error bounds: for example, the 5th and 95th percentiles of these ratios are values $L_{.05}(F)$ and $U_{.05}(F)$ that can be used, as in HW eq. (6.19), to obtain error bounds for the regional growth curve:

$$\frac{\hat{Q}_0(F)}{U_{.05}(F)} \leq Q_0(F) \leq \frac{\hat{Q}_0(F)}{L_{.05}(F)}. \quad (9)$$

Implementation in package `lmomRFA`

The methods described above are implemented in R package `lmomRFA` (Hosking, 2023). In that package, version 2.0 and later, function `regsimq` runs simulations that include permutation of at-site growth curves. RMSEs and error bounds are computed by functions `regquantbounds` and `sitequantbounds`, using intermediate results returned by `regsimq`. In `lmomRFA` version 3.5 and later, RMSEs and error bounds use the expressions in eqs. (3)–(9). In earlier versions the computational methods were somewhat different and gave results that were similar but not identical.

Specifically, the results returned by functions `regsimq`, `regquantbounds`, and `sitequantbounds` are related to quantities described in this note as follows. Note that error bounds can be obtained for arbitrary probabilities β , specified in argument `boundprob=c(0.05,0.95)` of function `regsimq`. The default value `boundprob=c(0.05,0.95)` gives results for upper and lower probabilities .05, as in HW eqs. (6.18)–(6.19).

Function `regsimq` has a return value that includes the following quantities:

`relbounds.rgc` A data frame whose columns include:

`rel.RMSE` $R^R(F)$ in (2).

`rel.bound.beta` $L_\beta(F)$ (if $\beta \leq 0.5$) or $U_{1-\beta}(F)$ (if $\beta > 0.5$) in (4).

`relbounds.by.site` A list of data frames whose i th element includes columns:

`rel.RMSE` $R_i(F)$ in (1).

`rel.bound.beta` $L_\beta(F)$ or $U_{1-\beta}(F)$ in (6).

`true.asgc` True at-site growth curves $q_i(F)$, for F values in argument `fval`.

`sim.rgc` Simulated regional growth curve values $q^{[m]}(F)$, for F values in argument `fval`.

Function `regquantbounds` returns a data frame whose columns include:

`RMSE` $R^{\text{abs},R}(F)$ in (3).

`bound.beta` Bounds $\hat{q}(F)/U_{1-\beta}(F)$ or $\hat{q}(F)/L_\beta(F)$ in (4).

Function `sitequantbounds` with `index` not specified returns results for sites in the region to which the regional L -moment algorithm was applied. The return value includes, for each requested site i , a data frame whose columns include:

`RMSE` $R_i^{\text{abs}}(F)$ in (5).

`bound.beta` Bounds $\hat{Q}_i(F)/U_{1-\beta}(F)$ or $\hat{Q}_i(F)/L_\beta(F)$ in (6).

Function `sitequantbounds` with `index` specified returns results for unged sites. The return value includes, for each requested site, a data frame whose columns include:

`RMSE` $R_0^{\text{abs}}(F)$ in (8).

`bound.beta` Bounds $\hat{Q}_0(F)/U_{1-\beta}(F)$ or $\hat{Q}_0(F)/L_\beta(F)$ in (9).

References

- Hosking, J. R. M. (2023). Regional Frequency Analysis using L-Moments. R package, version 3.5. URL: <https://CRAN.R-project.org/package=lmomRFA>.
- Hosking, J. R. M., and Wallis, J. R. (2005). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.