Technical Note: Design flood under hydrological uncertainty (Supplementary Material)

Anna Botto[1*], Daniele Ganora[1], Pierluigi Claps[1], and Francesco Laio[1]

[1]Department of Environment, Land and Infrastructure Engineering - Politecnico di Torino - Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy

[*]Now at the Department of Civil, Environmental and Architectural engineering - Università di Padova - Via Marzolo, 9 - Padova, Italy

Correspondence to: Daniele Ganora (daniele.ganora@polito.it)

The simulation analysis is based on a dataset of generated records, created by combining the following criteria:

1. The parent distribution $\mathcal{P}$ selected from the list: log-Normal (LN3), Generalized Extreme Value (GEV), Generalized Logistic (GLO), Pearson type III (PE3) and log-Pearson type III (LP3). For details on the probability distribution equation and on the relationship between parameters and L-moments the reader is referred to Hosking and Wallis (1997). The LP3 corresponds to the PE3 with log-transformed variate.

2. The sample length n of annual maxima selected from the list: 30, 40, 50, 60, 70 80, 90, 100.

We generated 100 records for each combination of $\mathcal{P}$ and n. Looking at the L-moments space, it results that 90% of the synthetic records have L-moments falling within the ranges: $0.28 \leq$ L-CV ≤ 0.40, $0.14 \leq$ L-skewness ≤ 0.40 and $0.07 \leq$ L-kurtosis ≤ 0.32.

On each record of the simulated dataset the standard design flood Q_T as well as the (exact) UNCODE estimator Q_T^* have been computed. This step has been performed by applying a suitable fitting distribution $\mathcal{F}$ to the whole synthetic dataset. To make the results more general, $\mathcal{F}$ has been selected in turn from the list: LN3, GEV, GLO, PE3, LP3. Note that any $\mathcal{F}$ is used to fit records from any parent $\mathcal{P}$ as in real cases the exact parent distribution is not known a priori. In this way, the error due to the misspecification of the fitting distribution is included in the results.

Finally, the correction factor y (i.e., the relative difference between the standard and the UNCODE estimates) has been computed for all the records available in the simulated dataset and for different return periods T (respectively equal to 50, 100, 200, 500 and 1000 years). The correction factor reads:

$$y_{T,\mathcal{F}} = \frac{Q_T^* - Q_T}{Q_T}, \tag{S1}$$

and depends on the fitting distribution $\mathcal{F}$ adopted in the frequency analysis.

Multiple linear regression models have been used to fit the exact $y_{T,\mathcal{F}}$ values from the prediction variables n and T. Different transformations of n and T have been tried to improve the regression; the most accurate results are obtained with Equation (5), which produces non-negative values of the final value y. All the regression models have been tested with the t-Student test, with significance level 10^{-3}.

The global performance of the regressions has been evaluated using the coefficient of determination and residuals analysis (through the mean absolute error, MAE, and root mean squared error, RMSE) for each fitting distribution $\mathcal{F}$. The value of the coefficient of determination ranges from 0.96 in case of the PE3 and 0.94 for the LN3, to 0.85 for the GEV and GLO. As for the residuals, apart from the LP3 which reports a mean absolute error equal to 0.0235, GEV, LN3, GLO and PE3 are associated with an error smaller than 0.02; the distinction is even more evident in case of the RMSE. In general, the PE3 probability distribution is that with the best performances in terms of residuals analysis and R^2_{adj}.

To have a more detailed landscape of the performance of the regression model, some diagnostic plots are provided in Figure S1 (linear axis in the top-row panels; logarithmic scale in the bottom-row panels) for each fitting distribution.
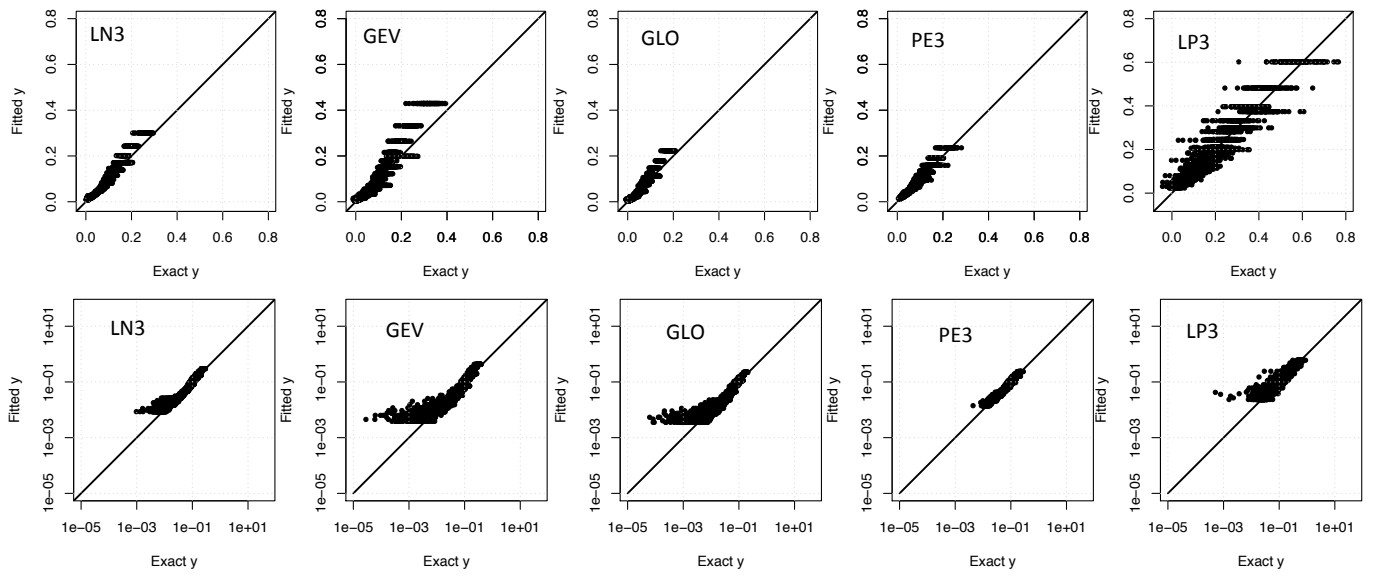
Figure S1. Comparison between exact (full UNCODE procedure) and approximated (regression) correction factor y based on Equation (5) for the 5 fitting distributions (LN3= log-Normal, GEV=Generalized Extreme Value, GLO=General Logistic, PE3=Pearson type III, LP3= log-Pearson type III)