Technical Note: Design flood under hydrological uncertainty

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)

Abstract. Planning and verification of hydraulic infrastructures demands for a design estimate of hydrologic variables, usually provided by frequency analysis, neglecting hydrologic uncertainty. However, when hydrologic uncertainty is accounted for, the design flood value is no longer a deterministic value, but should be treated as a random variable itself. As a consequence, the design flood is no longer univocally defined, making the design process undetermined.

The Uncertainty Compliant Design Flood Estimation (UNCODE) procedure is a novel approach which allows one to fix the ambiguity in the selection of the design flood under uncertainty, by considering an additional constraint based on a cost-benefit criterion. This paper contributes to disseminate the UNCODE procedure without resorting to numerical computation, but using a correction coefficient that modifies the standard (i.e., uncertainty-free) design value on the basis of sample length and return period only. The procedure is robust and parsimonious, as it does not require additional parameters with respect to the traditional uncertainty-free analysis.

Simple equations to compute the correction term are provided for a number of probability distributions commonly used to represent the flood frequency curve. This new design tool provides a robust way to manage the hydrologic uncertainty and to go beyond the use of traditional safety factors. With all the other parameters being equal, an increase of the sample length reduces the correction factor, and thus the construction costs, still keeping the same safety level.

1 Introduction

The flood frequency curve is commonly used to derive the design flood as the quantile Q_T corresponding to a fixed return period T. For practical reasons, Q_T is commonly expressed only as a single value; however, Q_T can only be expressed in this way if its frequency distribution and its parameters are known perfectly. In practice, one can only estimate the frequency distribution and its parameters using a sample of observed data, thereby inflating the uncertainty in the estimate of Q_T. However, the design of an hydraulic infrastructure demands for a single design value to be selected. A gap therefore exists between theory and practice. Quantitative methods to measure the uncertainty associated to the quantiles of the flood frequency curve (e.g., through their variance or probability distribution) have been proposed (e.g., Cameron et al., 2000; De Michele and Rosso, 2001; Brath et al., 2006; Blazkova and Beven, 2009; Laio et al., 2011; Liang et al., 2012; Viglione et al., 2013), but very few suggestions are provided about how to extract a single design value from the probability distribution of possible design values.

Botto et al. (2014), with the development of the Uncertainty Compliant Design Flood Estimation (UNCODE) procedure, have shown that it is possible to select meaningful flood quantiles from their distribution by considering an additional constraint based on a cost-benefit criterion. Hence, the output is a unique design flood value Q_T^*. Before illustrating the UNCODE approach, it is worth recalling the working principles of the cost-benefit analysis, which is a core element of the procedure. Cost-benefit analysis can be used to estimate the design flood as the flow value which minimizes the total expected cost function, defined as the sum of the actual cost to build a flood protection infrastructure (cost function) and the expected damages caused by a flood event. An illustrative example of this approach is reported in Fig. 1a. The cost function is rather easy to understand, being an increasing function of the design flood. Instead, the expected damage function needs to be computed point-by-point: for any single tentative design flood value (see the inset in Fig. 1a) it equals the integral of the product of the probability density function (pdf) of the flood flow values and a specific damage function. The latter indicates the damage occurring when the flood exceeds the flow value used to design the infrastructure. The damage function depends on a number of parameters such as the exposure and vulnerability of the flooded goods, the flooding dynamics and the topography, to mention a few. For these reasons the damage function turns out to be very site-specific and often unavailable for the lack of information needed to compute it (Menoni et al., 2016).

The cost-benefit method thus appears as an attractive design approach; however, due to the cost and damage functions not easy to be determined and hardly generalizable, it often results inapplicable. To face this problem Botto et al. (2014) made the assumption that costs and damages can be represented by piecewise linear functions, with slope c and d respectively, as illustrated in Fig. 1b. Given this assumption, the total cost C_{TOT}, can be computed as

$$C_{TOT} = c \cdot Q^* + \int_{Q^*}^{\infty} d \cdot (Q - Q^*) \cdot p(Q|\mathbf{\Theta})\,\mathrm{d}Q, \tag{1}$$

where Q^* is the generic design flood value and $p(Q|\mathbf{\Theta})$ is the probability density function of the flood flow with parameters $\mathbf{\Theta}$. The optimal design flood of the (uncertainty-free) cost-benefit framework can be then calculated as the value that minimizes Eq. (1). Examples of cost-benefit analysis in the hydrologic/hydraulic context can be found in the literature (Bao et al., 1987; Ganoulis, 2003; Jonkman et al., 2004; Tung, 2005), with only a few of them accounting for uncertainty (Al-Futaisi and Stedinger, 1999; Su and Tung, 2013).

Botto et al. (2014) further demonstrated that the optimal design flood obtained from the cost-benefit analysis with linear cost and damage functions is equivalent to the design flood Q_T obtained from the standard frequency analysis, provided that uncertainty is not accounted for and the ratio between d and c equals the return period T. This result can be shown by setting to 0 the derivative of C_{TOT} with respect to Q^*, in order to find the minimum of Eq. 1; this leads to the equivalence

$$\frac{d}{c} = \frac{1}{1 - P(Q^*|\mathbf{\Theta})} = T, \tag{2}$$

where $P(\cdot)$ is the cumulative distribution function of the flood values and T is the return period. This is valid provided that the probability distribution used in the cost-benefit framework is the same used in the standard frequency analysis.

The UNCODE approach is founded on these findings through the joint use of the cost-benefit approach of Eq. (1) and the constraint derived in Eq. (2). The rationale behind this approach is that it is possible to apply the cost-benefit framework with standard, but meaningful, cost and damage functions. This is particularly convenient because the cost-benefit framework can be easily extended to include the hydrological uncertainty due to the limited data availability. This allows one to extend also the UNCODE framework to account for this kind of uncertainty. In uncertain conditions, the parameters of the flood frequency distribution, $\boldsymbol{\Theta}$, become a random vector; hence, the uncertainty can be included in the cost benefit analysis by compounding C_{TOT} over all the possible values of $\boldsymbol{\Theta}$. In mathematical terms, the cost-benefit framework with uncertainty is summarized by the equation

$$Q_T^* = \underset{Q^*}{\arg\min} \left[\int_{\boldsymbol{\Theta}} C_{TOT} \left(Q^* | c, d, p(\boldsymbol{\Theta}) \right) \cdot h(\boldsymbol{\Theta}) \, \mathrm{d}\boldsymbol{\Theta} \right], \tag{3}$$

where $h(\boldsymbol{\Theta})$ is the joint pdf of the parameters of the flood frequency curve. Equation (3) represents the full UNCODE model, which adopts linear cost and damage functions and accounts for uncertainty in a cost-benefit framework.

It is worth noting that, as a consequence of the inherent equivalence of Eq. (2), there are no additional parameters in the cost-benefit framework; in fact, c and d are related through the known value of the return period T. The remaining free parameter can be shown to affect only the magnitude of the integral in Equation (3), but not the position of its minimum thus avoiding the need for further parameters in the UNCODE framework with respect to the standard design flood procedure.

To simplify the UNCODE application, which requires the use of numerical computation of Q_T^*, we provide here an approximated, though reliable, method to estimate Q_T^* starting from Q_T. Other than a useful practical tool for design purposes, the analysis reported in this note also provides a method to quantify the "value" of newly available hydrological information or the effect of data scarcity on Q_T^* due to uncertainty.

2 Practical estimation of the UNCODE design flood

The UNCODE design flood, Q_T^*, results systematically larger than its corresponding standard value Q_T. The relative difference between the two values,

$$y = \frac{Q_T^* - Q_T}{Q_T}, \tag{4}$$

has been shown (Botto et al., 2014) to increase with the return period (as the quantile uncertainty increases) as well as, for fixed T, with the standard deviation of the probability distribution of Q_T (i.e., with the uncertainty of Q_T). We propose to calculate the approximated estimate of the UNCODE design flood, hereafter referred to as $\hat{Q}_T^*$, directly by inversion of Eq. (4), without resorting to the numerical solution of Eq. (3). This solution reads:

$$\hat{Q}_T^* = (1 + \hat{y}) \cdot Q_T, \tag{5}$$

where the correction factor $\hat{y}$ (i.e., the approximated estimator of y) needs to be computed separately. Given this background, we propose to model $\hat{y}$ according to the equation

$$\hat{y} = 10^{-2} \cdot \exp\left[a_0 + a_1 \sqrt{n} + a_2 \ln T \right], \tag{6}$$

where T is the return period and n is the sample length which can be considered as a proxy of the standard deviation of Q_T; n can be computed from at-site records or as an equivalent sample length from the regional estimate of Q_T.

The coefficients a_0, a_1 and a_2 have been evaluated from an extensive simulation study in which the full UNCODE procedure has been systematically applied to many simulated 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, 90% of the synthetic records fall within the ranges: 0.28 $\leq$ L-CV $\leq$ 0.40, 0.14 $\leq$ L-skewness $\leq$ 0.40 and 0.07 $\leq$ L-kurtosis $\leq$ 0.32, which correspond well with values typically encountered in real-world applications. The standard design flood Q_T as well as the (exact) UNCODE estimator Q_T^* have been computed for each record of the simulated dataset. This step has been performed by adopting a suitable fitting distribution $\mathcal{F}$ to the whole synthetic dataset. To make the results more general, $\mathcal{F}$ has been selected 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. The correction factor y (Eq. 4) has been computed for all the available records in the simulated dataset and for different return periods T (respectively equal to 50, 100, 200, 500 and 1000 years). It depends on the fitting distribution $\mathcal{F}$ adopted in the frequency analysis. Finally, the exact y values have been regressed against n and T to obtain their estimate $\hat{y}$ (using an ordinary least squares linear regression on the log-transformed terms of Eq. 6). Different forms of Eq. (6) have also been tested, but are not reported as they provide less accurate results.

Coefficients a_0, a_1 and a_2 are reported in Table 1 for different fitting distributions commonly used in the hydrological practice to compute the design flood (in fact, the fitting distribution is always known, while the parent is not). It can be noticed that, when increasing the sample length n, the difference between Q_T^* and Q_T is reduced, due to the negative value of the coefficient a_1. Table 1 reports also some diagnostics of the regressions used to estimate the coefficients. 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. 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. The MAE and the RMSE take values around 0.02, corresponding to a 2% variation in the design flood estimation, which is negligible in many situations. In general, the PE3 probability distribution is that with the best performances in terms of residuals analysis and R^2_{adj}.

The reliability of the approximated correction factor $\hat{y}$ estimated with the regression model has also been evaluated by comparing the $\hat{Q_T^*}$ value obtained through Eq. (5) and (6) with its exact counterpart calculated with the full UNCODE procedure (Eq. 3). As a reference, time series listed in Botto et al. (2014, Table 1) with at least 30 years of record length have been analyzed, assuming the LN3 and the GEV as possible fitting distributions and different return periods. Results show a very good agreement between the exact (Q_T^*) and the approximated ($\hat{Q}_T^*$) UNCODE design flood values, as reported in Fig. 2, where each panel shows the estimates for all series and all the return periods.

A synthesis of the obtained results is shown in Fig. 3, where the values of $\hat{y}$ have been reported for the studied distributions, based on a set of typical sample length and return period values. As mentioned, a direct comparison of the results between different distributions is not possible, but it is relevant to observe that for all the distributions $\hat{y}$ evolves in the same way for varying n and T values. In general, the correction factor does not exceed 10% of the standard value Q_T for intermediate return periods (e.g., $T = 200$ years) even for small samples, although a significative variability is associated to the distribution type. It is around 10% for $T = 500$ years with sample length values ($n = 50$) commonly available at many gauged stations. On the other hand, the sample length plays an important role: for example, considering $T = 500$ years, the GEV distribution and varying the sample size, the reduction of the y value is about 0.075 between $n = 30$ and $n = 50$, and to 0.040 between $n = 50$ and $n = 70$.

3 Discussion of the application conditions

The UNCODE approach to flood frequency analysis provides an answer to the ambiguity due to the uncertainty in the quantile estimation. Application of the full UNCODE procedure may be cumbersome and computationally demanding. For a quick estimation of the design value an approximated but reliable framework has been proposed here to easily compute the UNCODE flood starting from the standard design value.

The extensive simulation analysis at the base of this study shows that the coefficients relating the UNCODE value $\hat{Q}_T^*$ to the standard value Q_T are distribution-dependent. For the most used distributions they have been computed and provided. The choice of the distribution and the management of its associated uncertainty is a problem of model selection; hence it cannot be solved by the UNCODE procedure, but depends on the preliminary standard flood frequency analysis.

The obtained results demonstrate that an increase in the length of relatively short samples has a noticeable impact in terms of reduction of $\hat{y}$ and of the UNCODE estimate $\hat{Q}_T^*$. This implies that, while the infrastructure keeps the same safety level, additional data reduce construction costs as the actual design value is reduced. The mentioned results agree with findings recently obtained by Ganora and Laio (2016) in a study on the relative role of regional and at-site flood frequency modeling approaches, where the value of at-site data has been highlighted and regarded as a reliable way to improve regional predictions, even with short records. Under this perspective, the correction factor can be used as a metric for uncertainty comparison and quantification, thus providing a further tool to combine different modeling approaches, similarly to the applications of Kjeldsen and Jones (2007) and Ganora et al. (2013) who, with different methodologies, have exploited measures of hydrologic uncertainty to merge regional and at-site information. The coefficient $\hat{y}$ can be considered a measure of the value of data. In fact, with all other parameters being equal, increasing n leads to a reduced $\hat{y}$ value and, consequently, to a reduced UNCODE design flood $\hat{Q}_T^*$. As a consequence, while the design value is still based on the same return period, costs will reduce.

Finally, the correction factor is a new and easy-to-implement design tool which provides a quantitative way to determine the design flood value accounting for hydrologic uncertainty, while keeping the same design hazard level considered in standard uncertainty-free analyses. This is a novel approach when compared to the common engineering practice, which accounts for hydrologic uncertainty by considering, for instance, the hydraulic freeboard. The use of the freeboard is equivalent to increasing the design flood value, but without accounting for the size of the system (e.g., the basin area), nor for the hydrologic information available at the section (i.e., observed of equivalent record length used to compute the standard design flood); this approach is thus not tailored to the specific case study. The correction factor represents an advance with respect to the use of "all-encompassing" safety factors and towards a clearer way to manage the different sources of uncertainty in hydrological and hydraulic design.

Acknowledgements. Funding from the ERC Consolidator Grant 2014 n. 647473 "CWASI - Coping with water scarcity in a globalized world" is acknowledged. D. Ganora also acknowledges the RTD Starting Grant from Politecnico di Torino. The work is based on simulated data.

References

Al-Futaisi, A. and Stedinger, J.: Hydrologic and Economic Uncertainties and Flood-Risk Project Design, Journal of Water Resources Planning and Management, 125, 314–324, doi:10.1061/(ASCE)0733-9496(1999)125:6(314), http://dx.doi.org/10.1061/(ASCE)0733-9496(1999)125:6(314), 1999.

Bao, Y., Tung, Y., and Hasfurther, V.: Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures, Water Resources Research, 23, 2023–2029, doi:10.1029/WR023i011p02023, http://dx.doi.org/10.1029/WR023i011p02023, 1987.

Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resources Research, 45, n/a–n/a, doi:10.1029/2007WR006726, http://dx.doi.org/10.1029/2007WR006726, 2009.

Botto, A., Ganora, D., Laio, F., and Claps, P.: Uncertainty compliant design flood estimation, Water Resources Research, 50, doi:10.1002/2013WR014981, 2014.

Brath, A., Montanari, A., and Moretti, G.: Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty), Journal of Hydrology, 324, 141 – 153, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.10.001, http://www.sciencedirect.com/science/article/pii/S0022169405004816, 2006.

Cameron, D., Beven, K., Tawn, J., and Naden, P.: Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), Hydrology and Earth System Sciences, 4, 23–34, 2000.

De Michele, C. and Rosso, R.: Uncertainty Assessment of Regionalized Flood Frequency Estimates, Journal of Hydrologic Engineering, 6(6), 2001.

Ganora, D. and Laio, F.: A comparison of regional flood frequency analysis approaches in a simulation framework, Water Resources Research, pp. n/a–n/a, doi:10.1002/2016WR018604, http://dx.doi.org/10.1002/2016WR018604, 2016.

Ganora, D., Laio, F., and Claps, P.: An approach to propagate streamflow statistics along the river network, Hydrological Sciences Journal, 58, 41 – 53, 2013.

Ganoulis, J.: Risk-based floodplain management: A case study from Greece, International Journal of River Basin Management, 1, 41–47, 2003.

Hosking, J. and Wallis, J.: Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press, 1997.

Jonkman, S., Brinkhuis-Jak, M., and Kok, M.: Cost benefit analysis and flood damage mitigation in the Netherlands, Heron, 49 (1), 2004.

Kjeldsen, T. and Jones, D.: Estimation of an index flood using data transfer in the UK, Hydrological Sciences Journal-Journal des Sciences Hydrologiques, 52, 86–98, 2007.

Laio, F., Ganora, D., Claps, P., and Galeati, G.: Spatially smooth regional estimation of the flood frequency curve (with uncertainty), Journal of Hydrology, 408, 67–77, doi:10.1016/j.jhydrol.2011.07.022, 2011.

Liang, Z., Chang, W., and Li, B.: Bayesian flood frequency analysis in the light of model and parameter uncertainties, Stochastic Environmental Research and Risk Assessment, 26, 721–730, doi:10.1007/s00477-011-0552-y, http://dx.doi.org/10.1007/s00477-011-0552-y, 2012.

Menoni, S., Molinari, D., Ballio, F., Minucci, G., Mejri, O., Atun, F., Berni, N. and Pandolfo, C.: Flood damage: a model for consistent, complete and multipurpose scenarios, Natural Hazards and Earth System Sciences, 16, 2783–2797, doi:10.5194/nhess-16-2783-2016, http://dx.doi.org/10.5194/nhess-16-2783-2016, 2016.

Su, H. and Tung, Y.: Incorporating uncertainty of distribution parameters due to sampling errors in flood-damage-reduction project evaluation, Water Resources Research, 49, 1680–1692, doi:10.1002/wrcr.20116, http://dx.doi.org/10.1002/wrcr.20116, 2013.

Tung, Y.: Flood Defense Systems Design by Risk-Based Approaches, Water International, 30, 50–57, doi:10.1080/02508060508691836, 2005.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resources Research, 49, 675 – 692, 2013.

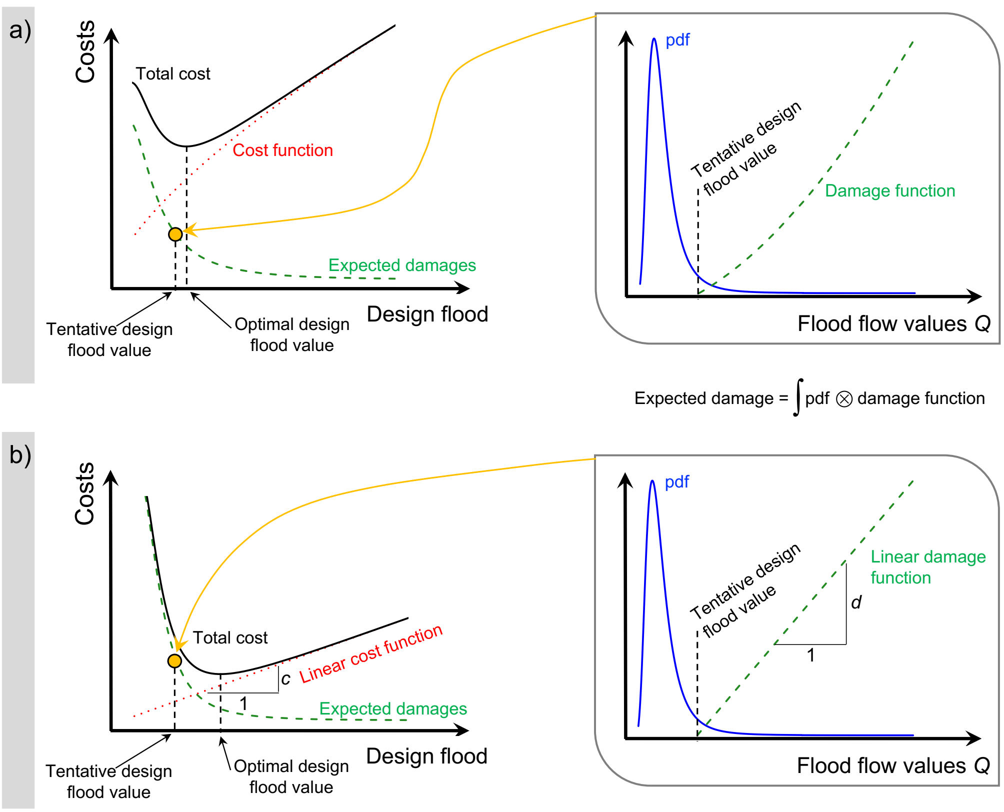

Figure 1. Illustrative example (without uncertainty) of the application of the cost-benefit framework to compute the design flood. Two generic cost and damage functions are reported in panel a, while panel b shows the linear functions adopted in the UNCODE framework.

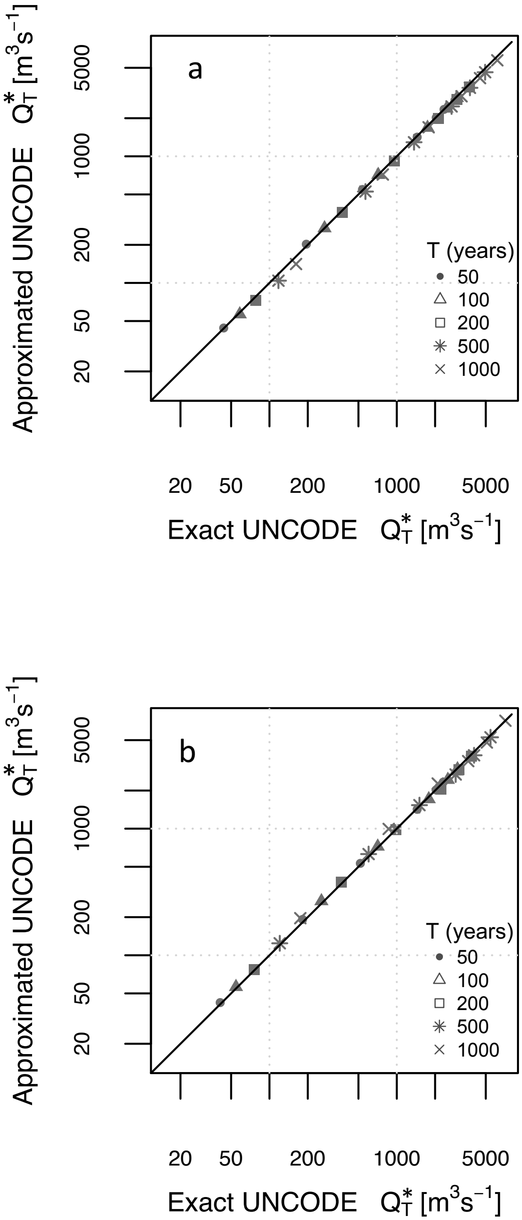

Figure 2. Comparison between the exact, Q_T^*, and the approximated, $\hat{Q}_T^*$, UNCODE estimators of the design flood for a pool of 6 flood records considered in Botto et al. (2014, Table 1) with at least 30 years of data. Different return periods are listed in the legend. The reference distribution used for this flood frequency analysis is the 3-parameter log-Normal (LN3) in panel “a” and the generalized extreme value (GEV) in panel “b”.

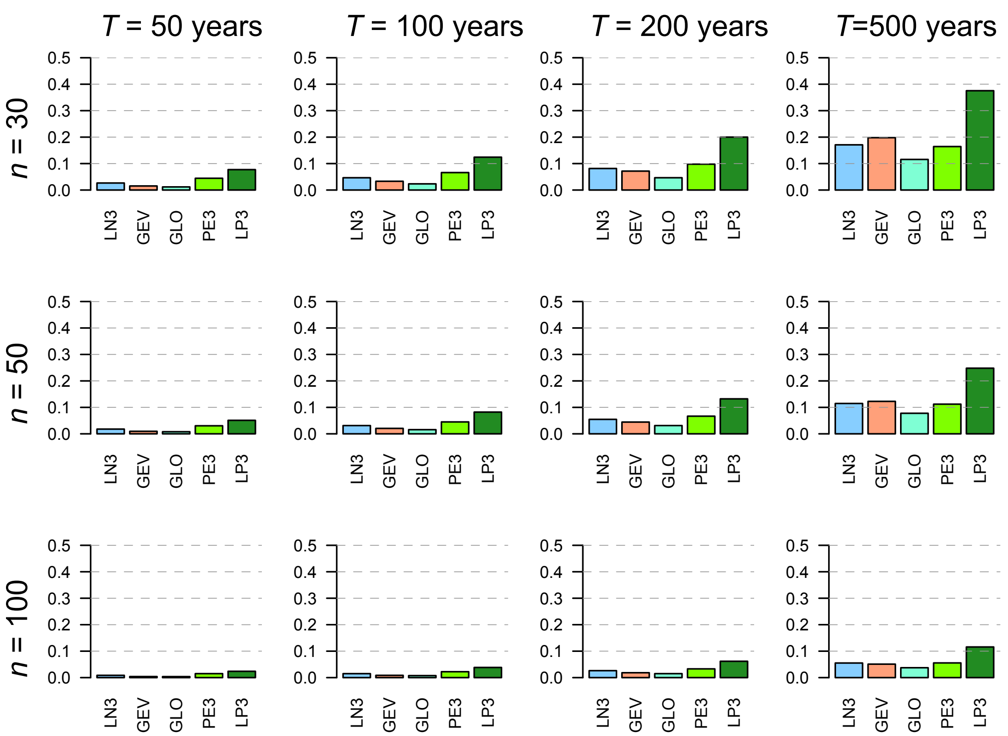

Figure 3. Values of the correction factor $\hat{y}$ from Eq. (6) for some values of the sample length n and return period T and for different three-parameter fitting distributions (LN3 = log-Normal; GEV= generalized extreme value; GLO = generalized logistic; PE3 = Pearson type III; LP3 = log Pearson type III). The LP3 corresponds to the PE3 with log-transformed variate.

Table 1. Coefficients to be used to estimate $\hat{y}$ based on the sample length n and the return period T (eq. 6) and corresponding regression diagnostics, for different 3-parameter fitting distributions (LN3 = log-Normal; GEV= generalized extreme value; GLO = generalized logistic; PE3 = Pearson type III; LP3 = log Pearson type III). The LP3 corresponds to the PE3 with log-transformed variate.

	a_0	a_1	a_2	R^2_{adj}	MAE	$RMSE$
LN3	-0.82	-0.25	0.809	0.94	0.0107	0.0160
GEV	-2.27	-0.3	1.110	0.85	0.0190	0.0321
GLO	-2.36	-0.25	0.994	0.85	0.0096	0.0145
PE3	0.59	-0.24	0.567	0.96	0.0080	0.0115
LP3	0.78	-0.26	0.687	0.89	0.0235	0.0363

We thank the Referees and the Editor for their detailed analysis of the manuscript and useful comments. We report below the Referees' comments (italic font) along with our response (regular font).

The Referees will find the "Introduction" section of the manuscript deeply revised. While the reported concepts are the same as in the original manuscript, the new version is intended to give a concise, but more complete and stand-alone description of the UNCODE methodology, which was one of the major issues rised by both the Referees. Section 2 has been also reformulated to be clearer and to fix some misleading notations; moreover, details about the simulation study previously reported in the "Supplementary Material" are now included in the main text (the Supplementary Material is no longer provided with the current version of the manuscript being no longer necessary).

All the other general and specific issues are discussed below. Some questions, regarding the same topic, have been addressed together in the following response to provide a more comprehensive explanation.

Interactive comment on "Technical Note: Design flood under hydrological uncertainty" by Anna Botto et al.
Anonymous Referee #1

I find the topic and scope of the manuscript appropriate as a technical note in HESS. Simple yet technically sufficient approaches to include uncertainty in hydrologic design are of great significant in practice. However, the technical note is lacking critical details and requires clarification in some areas, which makes the manuscript difficult to fully assess at this time. The manuscript itself does not stand-alone and assumes that the reader is familiar with the details of the Botto et al. (2014) study. I would recommend major revision with additional review before publication in HESS. I have provided specific areas which need to be addressed with additional details or clarification.

We thank the Referee for this comment and appreciate the opportunity to better clarify our research objectives and results. As mentioned above, the Introduction section has been completely revised to be stand alone, although references to Botto et al. (2014) are still used to refer to specific details. To better explain the UNCODE procedure, a graphical definition of the cost – benefit approach in its general form is provided by including a description in the text and a new figure (Fig. 1 in the revised manuscript). In particular, the new Fig. 1 specifies the cost and damage functions of a general cost-benefit framework (panel a) versus the functions used in the UNCODE context (panel b). This direct comparison is intended to better exemplify the hypotheses made in the model.

More technical details are needed in the following areas:

- *Eqn 15: C_TOT is not defined.*
- *p. 2, line 18: The authors state that setting C_TOT to zero gives the "optimal design flood value." Please help the reader connect why this is the case. Is this because this condition gives the local minima for C_TOT? Defining C_TOT will help with this point.*

The total cost function, C_TOT, is now defined and discussed in a more coherent way in the Introduction section. The computation of the optimal design value is also discussed in more detail. The new Fig. 1 helps the reader visualize the "optimal" value and its meaning (i.e., the minimum of the total cost function, which is the sum of costs and damages).

- Parameters c and d are not well explained in the manuscript (p. 2, lines 17-18). The manuscript only indicates that they are site-specific and "influenced by topography and land use among others. . .." A reader should have enough information in order to understand these parameters and how they are estimated without having to refer back to Botto et al. (2014).

-p. 2, line 30: Please explain how c and d are related to the known value of the return period T

These issues have been now explicitly discussed in the revised manuscript. We clarified that in general, in a cost-benefit framework, the cost and especially the damage functions are difficult to be estimated, the latter depending on a number of specific variables such as the exposure and vulnerability of the flooded goods, the topography, the land use, the flooding dynamics to mention but a few (page 2 lines from 10 to 14). This function turns out to be very site-specific and often unavailable for the lack of information needed to compute it. A new reference (Menoni et al, 2016) has been added to support this statement.

Moreover, the manuscript now reports that the UNCODE procedure makes use of simplified cost and damage functions. These have been reported in Fig.1b, and discussed throughout section 1. More details about the relationships between the classical flood frequency analysis and the cost – benefit approach are now provided, being a key concept in the UNCODE approach. A concise demonstration of the equivalence d/c=T is also provided.

- The derivation of equation 5 and its relation to the pdf is not immediately obvious. Please provide a more detailed explanation of the origin of this equation and its conceptual meaning.

- The variable y needs to be better explained and defined as it serves as the focus of the contribution of the note, in my opinion. In the manuscript, y is only described as a non-negative number depending only on the pdf used to fit the flood frequency data (p. 3, line 12-13), which seems too simplistic. Please provide additional explanation. Later in line 29, the text states that y is estimated from regression, further confusing the reader as to how to interpret its meaning. The reader should not have to refer to the supplementary information to understand this.

Section 2 has been also revised to facilitate the interpretation of the terms. Firstly, the notation has been changed to easily identify the full-UNCODE design flood (QT*) and the approximated-UNCODE design flood (hatQT*). According to this notation, also the correction factor "y" has an exact (y) and an approximated version (hat y).

The new notation, along with some more details moved from the Supplementary Material to the main text, helps to understand the origin of Eq. (5) (Eq. 6 in the revised version). Exact "y" values have been computed for a number of random samples and then a simplified equation (regression) has been used to compute its estimator (hat y) with minimal information: the sample length n (always know for at-site analysis, but equivalent n can be also evaluated at un-gauged sites) and the return period T.

- More description is needed in the main text to discuss the computation of the a_j's and y values. Also, include in the main text the regressors used in the regression model.

The relationship between y and the variables n and T has been obtained by means of linear regressions. Equation 5 shows the general formula to estimate y. There, the coefficients a_0, a_1, a_2 depend on the specific probability distribution function chosen for the inference procedure, which are reported in Table 1. Supposing we want to calculate the correction y for the return period T and sample length n with two different pdf: this can be pursued just choosing the appropriate coefficients from Table 1 and the same T and n in both cases. Details of the simulation and regression procedure have been now fully reported in the main text. The Supplementary Material is thus no longer necessary.

- p. 3, line 20: It is unclear why the reference to Hosking and Wallis is placed at the end of this sentence. The reference placement implies that Hosking and Wallis (1997) have some comment as to

the values of the a_j's, which is not the case. Remove the reference here but keep the reference in Table 1.

We completely agree with the referee. The correction has been implemented.

- Clarification is needed in the following areas:

- There needs to be a clarification in the introduction regarding the understanding of where uncertainty in the design flood arises. I agree that the design flood can be known as a single value when, as the authors state (p. 1, lines 18-19), "the frequency distribution and its parameters are known without uncertainty." My understanding is that the motivation for this work is because the frequency distribution and its parameters can only be estimated from a sample of flood data. It is for this reason, uncertainty arises and must be considered in practice. I recommend reworking lines 18-22 to something such as this:

". . .period T. For practical reasons, Q_T is commonly expressed only as a single value; however, Q_T can only be expressed in this way if its frequency distribution and its parameters are known perfectly. In practice, one can only estimate the frequency distribution and its parameters using a sample of observed data, thereby creating uncertainty in the estimate of Q_T. However, the design of a hydraulic. . ."

We thank the referee for the suggestion which makes the point clearer; we included the correction in the Introduction section of the manuscript.

- *My understanding is that the authors only consider parameter uncertainty and not uncertainty in the choice of the pdf. Clarify this in the text.*

It is correct. The point, originally discussed in Sect. 3, has been clarified in page 5 lines 17-20; the procedure does not include the issue of uncertainty in model selection, but provides a way to compute the UNCODE estimator tailored for different distributions commonly used for this kind of analysis.

Minor editorial comments:

p. 1, line 5: Avoid using a reference in the abstract.

Amended

p. 3, line 4: Change to read: "Botto et al. (2014) shows that the UNCODE. . ."

Amended

p. 5, line 2: Change to "increasing"

Amended

Interactive comment on "Technical Note: Design flood under hydrological uncertainty" by Anna Botto et al.
A. Pugliese (Referee)
alessio.pugliese3@unibo.it

The paper "Technical Note: Design flood under hydrological uncertainty" by Botto et al., shows how to quickly find an estimate of hydrologic uncertainty to be added to classical statistical inference for the design flood. The paper is well written and is rather complete in all its sections, and it has the potential to be extremely useful for practitioners and engineers. I believe it is suitable for the publication in HESS as a technical note after some minor improvements, essentially due to miscommunicated reasonings, which, in my view, the authors might consider to take into account.

Finally, I think there is enough material for another paper here, so I encourage the authors to consider deepening the analysis in the future, in order to find operational ranges and domains with real world data too, since the presented practical method needs to be as much robust as possible. For instance, an idea could be to extended the presented method using regional flood frequency analysis, in order to overcome possible unsuitability of the presented procedure in data scarce regions.

We would like to thank the reviewer for the useful comments to improve the paper and for the kind encouragement. We have addressed all the comments as explained below.

Minor comments

The authors report that the parameters c and d are site-specific and controlled by topography and land use among others. These two parameters control the magnitude of the total cost function, and, ultimately, the design flood, but in most cases collecting cost data can be a cumbersome process, as they are usually unavailable. As far as I understood, the calibration of the empirical law (Eq. 5) does not need the knowledge of either c or d, so that the resulting coefficients a_0*,* a_1*,* a_2*, are rather general and independent from a specific site. Is it? Can the coefficients in Table 1 be used without any restriction on the location? I think the authors should address better the following thoughts:*

The Referee has brought up some good points and we appreciate the opportunity to clarify better our research objectives and results.

- *what is the role of the two parameters c and d, how do they transfer their information about site-specific cost rates to the parameters* a_0*,* a_1*,* a_2*, if any;*

A detailed discussion of the role of the parameters c and d has been included in the revised version of the manuscript (Section 1) to make the theoretical description of the UNCODE procedure concise, but completely stand-alone. More details about the meaning and the estimation of a0, a1 and a2 are now included in section 2.

In general, in a cost-benefit framework the cost and, especially, the damage functions are difficult to be estimated, the latter depending on a number of specific variables such as the exposure and vulnerability of the flooded goods, the topography, the land use, the flooding dynamics to mention a few (page 2 lines from 10 to 14). This function turns out to be very site-specific and often unavailable for the lack of information needed to compute it. A new reference (Menoni et al, 2016) has been added to support this statement. This point has been clarified in the revised manuscript including a new figure (Fig. 1) that sketches the cost-benefit procedure and help visualizing the involved functions. The main text has been integrated accordingly.

With respect to a general cost-benefit analysis, the UNCODE procedure adopts specific cost and damage functions (linear with slope c and d) that are also included in the new Fig.1 for an easy comparison with the general cost and damage functions. The revised version of the Introduction section also reports a more detailed description of the derivation of the equivalence d/c=T that, in practice, makes the estimation of c and d values unnecessary. This point has been reported on page 3 lines 10-14 …, highlighting that these parameters influence only the magnitude of the integral in Eq. (3) (i.e., the total expected costs under uncertainty), while its minimum (i.e., the UNCODE design flood) is unaffected.

Moreover, it is true that both sets of parameters (c, d) and (a_0, a_1, a_2) involve T; however, the two sets of parameters derive from different concepts. In the first case the use of the return period is a consequence of the application of the cost-benefit framework under some simplifying assumption and it can be considered as an element of the UNCODE model (as already derived by Botto et al., 2014). In the second case, the return period is a proxy (together with n) of the quantile uncertainty (i.e., it considers the typical spread of the confidence bands of the flood frequency curve with the return period).

- *or, is this transferring perhaps delivered by the parameters of the flood frequency curve only?*
- *are the parameters a_0, a_1, a_2 general (independent by site, computed once and for all) or, perhaps, does the end-user have to fit the empirical law when needed on a specific site? If so, it would be useful to have a general point- by-point procedure, like an algorithm, to let the end-user implement it on a specific dataset or location;*

The estimation of the UNCODE design flood is a non-linear process that depends on the sample uncertainty and how it propagates to the distribution quantiles. The proposed method aims at simplifying this complex process by requiring only the estimate of the standard quantile QT and the multiplying factor y that includes all the elements that drive the "propagation" of the uncertainty to the distribution quantile. The developed approach is thus general so that the regression parameters (a_0, a_1, a_2) are computed once and for all, although they can be considered accurate only within a subset of the T-n-Lmoments space, as specified in the main text of the revised manuscript.

The application still depends on the choice of the distribution as the UNCODE method does not account for the model uncertainty. However, this does not limit the application of the procedure as different distributions can be tested as in the classic flood frequency analysis.

I think the same sentence L20-23 P1 of supplementary material can be included into the manuscript, or at least the authors should mention exhaustively about the choice of such empirical expression for y.

The description of the regression has been detailed in the revised manuscript (sec. 2), including the information originally reported in the Supplementary Material (the Supplementary Material is no longer provided with the current version of the manuscript).

The authors should consider to report the accuracy of the fitted empirical law in the body of the text too, at least for LN3 and GEV distributions.

This suggestion has been implemented in the revised manuscript, reporting a more complete analysis of fitting residuals.

In my view there is ambiguity in the mathematical notation between exact UNCODE solution QT and predicted (approximated) UNCODE, which is reported with the same variable QT . The authors might consider to change notation on one of the two, indeed the approximated UNCODE introduces one more source of error brought by the selected empirical law y.

The notation has been revised to properly identify the exact and the approximated values of both $\hat{y}$ and $\widehat{Q_T^*}$. This is now highlighted in the first part of section 2.

L19 P3. The sentence in parentheses is put aside the main sentence, but I think it is rather important for the reader to know that regional analyses can be used where there is lack of data. The author should consider to expand the reasoning here, without parentheses.

We thank the Referee for the note, the sentence has been modified.

Notes and misspellings

I agree with reviewer 1, I would remove the citation Botto et al., 2014 from the abstract to let it be more general.

Amended

L5 P3. Replace "methods" with "method".

Amended

L27 P3. I would add the range of variation of the index j, so "The coefficients aj" will be "The coefficients aj with j = 0, 1, 2".

Amended, index j has been removed and coefficients are now explicitly defined, a_0, a_1, a_2

Technical Note: Design flood under hydrological uncertainty

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)

Abstract. Planning and verification of hydraulic infrastructures demands for a design estimate of hydrologic variables, usually provided by frequency analysis, neglecting hydrologic uncertainty. However, when hydrologic uncertainty is accounted for, the design flood value is no longer a deterministic value, but should be treated as a random variable itself. As a consequence, the design flood is no longer univocally defined, making the design process undetermined.

~~Botto et al. (2014), with the development of the~~ The Uncertainty Compliant Design Flood Estimation (UNCODE) procedure ~~, have shown that it is possible~~ is a novel approach which allows one to fix the ambiguity in the selection of the design flood under uncertainty, by considering an additional constraint based on a cost-benefit criterion. This paper contributes ~~with an easy-to-use framework to implement~~ to disseminate the UNCODE procedure without resorting to numerical computation, but using a correction coefficient that modifies the standard (i.e., uncertainty-free) design value on the basis of sample length and return period only. The procedure is robust and parsimonious, as it does not require additional parameters with respect to the traditional uncertainty-free analysis.

Simple equations to compute the correction term ~~to the standard estimate~~ are provided for a number of probability distributions commonly used to represent the flood frequency curve. This new design tool provides a robust way to manage the hydrologic uncertainty and to go beyond the use of traditional safety factors. With all the other parameters being equal, an increase of the sample length reduces the correction factor, and thus the construction costs, still keeping the same safety level. ~~This improvement is shown to be more effective when short samples are extended.~~

1 Introduction

The flood frequency curve is commonly used to derive the design flood as the quantile Q_T corresponding to a fixed return period T. ~~The design value~~ For practical reasons, Q_T is ~~defined~~ commonly expressed only as a single value~~when the~~; however, Q_T can only be expressed in this way if its frequency distribution and its parameters are known ~~without uncertainty. When uncertainty in the parameters or in the probabilistic model is accounted for, this propagates to the quantile; this means that for the same return period T the quantile is no longer a single value, but should be treated as a random variable itself. As a consequence, the design flood is no longer univocally defined.~~ perfectly. In practice, one can only estimate the frequency distribution and its parameters using a sample of observed data, thereby inflating the uncertainty in the estimate of Q_T. However, the design of an hydraulic infrastructure demands for a single design value to be selected. A gap therefore exists between theory and practice. Quantitative methods to measure the uncertainty associated to the quantiles of the flood frequency curve (e.g., through their variance or probability distribution) have been proposed (e.g., Cameron et al., 2000; De Michele and Rosso, 2001; Brath et al., 2006; Blazkova and Beven, 2009; Laio et al., 2011; Liang et al., 2012; Viglione et al., 2013), but very few suggestions are provided about how to ~~obtain~~ extract a single design value from the probability distribution of possible design values.

Botto et al. (2014), with the development of the Uncertainty Compliant Design Flood Estimation (UNCODE) procedure, have shown that it is possible to select meaningful flood quantiles from their distribution by considering an additional constraint based on a cost-benefit criterion. Hence, the output is a unique design flood value Q_T^*. ~~For theoretical and practical aspects of the procedure the reader is referred to the original paper, whereas here we recall only the core concepts of the UNCODE approach.~~

~~Botto et al. (2014) primarily demonstrated that~~ Before illustrating the UNCODE approach, it is worth recalling the working principles of the cost-benefit analysis, which is a core element of the procedure. Cost-benefit analysis can be used to estimate the design flood ~~Q_T obtained with the standard flood frequency analysis (without uncertainty) is equivalent to~~ as the flow value which minimizes the total expected cost function, defined as the sum of the actual cost to build a flood protection infrastructure (cost function) and the expected damages caused by a flood event. An illustrative example of this approach is reported in Fig. 1a. The cost function is rather easy to understand, being an increasing function of the design flood~~obtained with a~~. Instead, the expected damage function needs to be computed point-by-point: for any single tentative design flood value (see the inset in Fig. 1a) it equals the integral of the product of the probability density function (pdf) of the flood flow values and a specific damage function. The latter indicates the damage occurring when the flood exceeds the flow value used to design the infrastructure. The damage function depends on a number of parameters such as the exposure and vulnerability of the flooded goods, the flooding dynamics and the topography, to mention a few. For these reasons the damage function turns out to be very site-specific and often unavailable for the lack of information needed to compute it (Menoni et al., 2016).

The cost-benefit ~~analysis with specific damage and cost curves. Examples of cost-benefit analysis in the hydrologic/hydraulic context can be found in the literature (Bao et al., 1987; Ganoulis, 2003; Jonkman et al., 2004; Tung, 2005), but only a fewof them include hydrologic uncertainty (Al-Futaisi and Stedinger, 1999; Su and Tung, 2013). Cost and damage curves obtained by Botto et al. (2014) are~~ method thus appears as an attractive design approach; however, due to the cost and damage functions not easy to be determined and hardly generalizable, it often results inapplicable. To face this problem Botto et al. (2014) made the assumption that costs and damages can be represented by piecewise linear functions, with slope c and d respectively~~that, combined with the probability density function p of the flood values Q, give~~, as illustrated in Fig. 1b. Given this assumption, the total cost ~~function (i.e., actual costs plus damages):~~ C_{TOT}, can be computed as

$$C_{TOT} = c \cdot Q^* + \int_{Q^*}^{\infty} d \cdot (Q - Q^*) \cdot p(Q|\boldsymbol{\Theta})\,\mathrm{d}Q, \qquad (1)$$

where Q^* is the generic design flood value and ~~$\boldsymbol{\Theta}$ is the vector of parameters of the probability distribution, which depends on the hydrologic characteristic of the site. Parameters c and~~ $p(Q|\boldsymbol{\Theta})$ is the probability density function of the flood flow with parameters $\boldsymbol{\Theta}$. The optimal design flood of the (uncertainty-free) cost-benefit framework can be then calculated as the value that minimizes Eq. (1). Examples of cost-benefit analysis in the hydrologic/hydraulic context can be found in the literature (Bao et al., 1987; Ganoulis, 2003; Jonkman et al., 2004; Tung, 2005), with only a few of them accounting for uncertainty (Al-Futaisi and Stedinger, 1999; Su and Tung, 2013).

Botto et al. (2014) further demonstrated that the optimal design flood obtained from the cost-benefit analysis with linear cost and damage functions is equivalent to the design flood Q_T obtained from the standard frequency analysis, provided that uncertainty is not accounted for and the ratio between d ~~are also generally site-specific, being influenced by topography and land use among others. Taking~~ and c equals the return period T. This result can be shown by setting to 0 the derivative of C_{TOT} with respect to Q^* ~~and setting it to 0 gives the optimal design flood value of the (uncertainty-free) cost-benefit framework, and leads also~~ , in order to find the minimum of Eq. 1; this leads to the equivalence

$$\frac{d}{c} = \frac{1}{1 - P\left(Q^* | \boldsymbol{\Theta}\right)} = T, \tag{2}$$

where $P(\cdot)$ is the cumulative distribution function of the flood values and T is the return period. This is valid provided that the probability distribution used in the cost-benefit framework is the same used in the standard frequency analysis. ~~Equation (2) links the standard flood frequency analysisto the~~

The UNCODE approach is founded on these findings through the joint use of the cost-benefit approach ~~.~~

~~In uncertain condition~~ of Eq. (1) and the constraint derived in Eq. (2). The rationale behind this approach is that it is possible to apply the cost-benefit framework with standard, but meaningful, cost and damage functions. This is particularly convenient because the cost-benefit framework can be easily extended to include the hydrological uncertainty due to the limited data availability. This allows one to extend also the UNCODE framework to account for this kind of uncertainty. In uncertain conditions, the parameters of the flood frequency distribution, $\boldsymbol{\Theta}$~~becomes~~ , become a random vector; hence, ~~hydrologic uncertainty should~~ the uncertainty can be included in the cost benefit analysis by compounding C_{TOT} over all the possible values of $\boldsymbol{\Theta}$. In mathematical terms, the cost-benefit framework with uncertainty is summarized by the equation

$$Q_T^* = \underset{Q^*}{\arg\min} \left[\int_{\boldsymbol{\Theta}} C_{TOT}\left(Q^* | c, d, p\left(\boldsymbol{\Theta}\right)\right) \cdot h\left(\boldsymbol{\Theta}\right) \mathrm{d}\boldsymbol{\Theta} \right], \tag{3}$$

where $h\left(\boldsymbol{\Theta}\right)$ is the joint pdf of the parameters of the flood frequency curve.

Equation (3) ~~is valid in general upon specification of c and d, which are usually unavailable. However,~~ represents the full UNCODE model, which adopts linear cost and damage functions and accounts for uncertainty in a cost-benefit framework.

It is worth noting that, as a consequence of the inherent equivalence ~~between the cost-benefit and the quantile-based approaches defined by~~ of Eq. (2)~~reduces the degrees of freedom of the~~ , there are no additional parameters in the cost-benefit framework~~, as~~ ; in fact, c and d are ~~not independent, but are~~ related through the known value of the return period T. The remaining free parameter can be shown to affect only the magnitude of the integral in ~~Eq.~~ Equation (3), but not the position of its minimum ~~. As a consequence, the UNCODE framework does not add any further parameter~~ thus avoiding the need for further parameters in the UNCODE framework with respect to the standard design flood procedure~~, but it allows one to frame the uncertainty analysis into a cost-benefit framework~~.

~~We have shown in Botto et al. (2014) that the UNCODE design flood,~~ To simplify the UNCODE application, which requires the use of numerical computation of Q_T^*~~, is always larger than its corresponding standard value Q_T. However, computation of Q_T^* requires application of a numerical methods. To simplify the UNCODE application~~, we provide here an approximated, though reliable, method to estimate Q_T^* starting from Q_T. Other than a useful practical tool for design purposes, the analysis reported in this note also provides a method to quantify the "value" of newly available hydrological information or the effect of data scarcity on Q_T^* due to uncertainty.

2 Practical estimation of the UNCODE design flood

~~To compute the~~

The UNCODE design flood, Q_T^*, ~~we consider the equation:~~

$$\cancel{Q_T^* = (1+y)\cdot Q_T}$$

results systematically larger than its corresponding standard value Q_T. The relative difference between the two values,

$$y = \frac{Q_T^* - Q_T}{Q_T}, \tag{4}$$

~~where y is a-non negative coefficient which depends on the probability distribution used to fit the flood frequency curve, and Q_T is the standard design flood.~~

~~As noted by Botto et al. (2014), the relative distance between Q_T^* and Q_T increases~~ has been shown (Botto et al., 2014) to increase with the return period (as the quantile uncertainty increases) as well as, for fixed T, with the standard deviation of the probability distribution of Q_T (i.e., with the uncertainty of Q_T). We propose ~~here to model~~ to calculate the approximated estimate of the UNCODE design flood, hereafter referred to as $\hat{Q}_T^*$, directly by inversion of Eq. (4), without resorting to the numerical solution of Eq. (3). This solution reads:

$$\hat{Q}_T^* = (1+\hat{y})\cdot Q_T, \tag{5}$$

where the correction factor $\hat{y}$ (i.e., the approximated estimator of y) needs to be computed separately. Given this background, we propose to model $\hat{y}$ according to the equation

$$\cancel{y}\hat{y} = 10^{-2}\cdot \exp\left[a_0 + a_1\sqrt{n} + a_2 \ln T\right], \tag{6}$$

where T is the return period and n is the sample length which can be considered as a proxy of the standard deviation of Q_T~~(~~; n can be computed from at-site records or as an equivalent sample length from the regional estimate of Q_T~~). Coefficients a_0, a_1 and a_2 are reported in Table 1 for different fitting distributions commonly used in the hydrological practice to compute the design flood (Hosking and Wallis, 1997). Table 1 clearly shows the effect of increasing the sample length n, which reduces the difference between Q_T^* and Q_T due to the negative value of the coefficient a_1~~.

The coefficients a_0, a_1 and a_2 have been evaluated ~~through~~ from an extensive simulation study ~~(details are reported in the Supplementary Material) in~~ in which the full UNCODE procedure has been systematically applied to many simulated records, ~~considering a number of different combinations of parent and fitting distribution , sample length (with~~ 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 ~~from~~ of annual maxima selected from the list: 30~~to 100)~~, 40, 50, 60, 70 80, 90, 100.

We generated 100 records for each combination of $\mathcal{P}$ and ~~return periods (with T from~~ n. Looking at the L-moments space, 90% of the synthetic records fall within the ranges: 0.28 $\leq$ L-CV $\leq$ 0.40, 0.14 $\leq$ L-skewness $\leq$ 0.40 and 0.07 $\leq$ L-kurtosis $\leq$ 0.32, which correspond well with values typically encountered in real-world applications. The standard design flood Q_T as well as the (exact) UNCODE estimator Q_T^* have been computed for each record of the simulated dataset. This step has been performed by adopting a suitable fitting distribution $\mathcal{F}$ to the whole synthetic dataset. To make the results more general, $\mathcal{F}$ has been selected 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. The correction factor y (Eq. 4) has been computed for all the available records in the simulated dataset and for different return periods T (respectively equal to 50~~to~~, 100, 200, 500 and 1000 ~~). Different forms of Eq. (6) have been also tested. For each run, the empirical~~ years). It depends on the fitting distribution $\mathcal{F}$ adopted in the frequency analysis. Finally, the exact y ~~value has been recorded. The coefficients a_j have been finally estimated through~~ values have been regressed against n and T to obtain their estimate $\hat{y}$ (using an ordinary least squares linear regression on the log-transformed terms of Eq. 6). Different forms of Eq. (6) ~~. Table 1 reports~~ have also been tested, but are not reported as they provide less accurate results.

Coefficients a_0, a_1 and a_2 are reported in Table 1 for different fitting distributions commonly used in the hydrological practice to compute the design flood (in fact, the fitting distribution is always known, while the parent is not). It can be noticed that, when increasing the sample length n, the difference between Q_T^* and Q_T is reduced, due to the negative value of the coefficient a_1. Table 1 reports also some diagnostics of the regressions used to estimate the coefficients. 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. 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. The MAE and the RMSE take values around 0.02, corresponding to a 2% variation in the design flood estimation, which is negligible in many situations. In general, the PE3 probability distribution is that with the best performances in terms of residuals analysis and R^2_{adj}.

The reliability of the approximated correction factor ~~y~~ $\hat{y}$ estimated with the regression model has also been evaluated by comparing the ~~Q_T^*~~ $\hat{Q}_T^*$ value obtained through Eq. (5) and (6) with its exact counterpart calculated with the full UNCODE procedure ~~.~~ (Eq. 3). As a reference, time series listed in Botto et al. (2014, Table 1) with at least 30 years of record length have been analyzed, assuming the LN3 and the GEV as possible fitting distributions and different return periods. Results show a very good agreement between the exact (Q_T^*) and the approximated $\cancel{Q_T^*}$ $(\hat{Q}_T^*)$ UNCODE design flood values, as reported in Fig. 2, where each panel shows the estimates for all series and all the return periods. ~~Panel a) refers to the LN3, while panel b) to the GEV distribution.~~

A synthesis of ~~quantitative~~ the obtained results is shown in Fig. 3, where the values of ~~y from Eq. (6)~~ $\hat{y}$ have been reported for the studied distributions, based on a set of typical sample length and return period values. As mentioned, a direct comparison of the results between different distributions is not possible, but it is relevant to observe that for all the distributions $\cancel{y}$ $\hat{y}$ evolves in the same way for varying n and T values. In general, the correction factor does not exceed 10% of the standard value Q_T for intermediate return periods (e.g., $T = 200$ years) even for small samples, although a significative variability is associated to the distribution type. It is ~~also~~ around 10% ~~around~~ for $T = 500$ years with sample length values ($n = 50$) commonly available at many gauged stations. On the other hand, the sample length plays an important role: for example, considering $T = 500$ years, the GEV distribution and varying the sample size, the reduction of the y value is about 0.075 between $n = 30$ and $n = 50$, ~~while it drops~~ and to 0.040 between $n = 50$ and $n = 70$.

3 Discussion of the application conditions

The UNCODE approach to flood frequency analysis provides an answer to the ambiguity due to the uncertainty in the quantile estimation. Application of the full UNCODE procedure may be cumbersome and computationally demanding. For a quick estimation of the design value an approximated but reliable framework has been proposed here to easily compute the UNCODE flood starting from the standard design value.

The extensive simulation analysis at the base of this study shows that the coefficients $\cancel{a_T}$ relating the UNCODE value $\cancel{Q_T^*}$ $\hat{Q}_T^*$ to the standard value Q_T are distribution-dependent. For the most used distributions they have been computed and provided. The choice of the distribution and the management of its associated uncertainty is a problem of model selection~~and~~; hence it cannot be solved by the UNCODE procedure, but depends on the preliminary standard flood frequency analysis.

The obtained results demonstrate that an increase in the length of relatively short samples has a noticeable impact in terms of reduction of $\cancel{y}$ $\hat{y}$ and of the UNCODE estimate $\cancel{Q_T^*}\hat{Q}_T^*$. This implies that, while the infrastructure keeps the same safety level, additional data reduce construction costs as the actual design value is reduced. The mentioned results agree with findings recently obtained by Ganora and Laio (2016) in a study on the relative role of regional and at-site flood frequency modeling approaches, where the value of at-site data has been highlighted and regarded as a reliable way to improve regional predictions, even with short records. Under this perspective, the correction factor can be used as a metric for uncertainty comparison and quantification, thus providing a further tool to combine different modeling approaches, similarly to the applications of Kjeldsen and Jones (2007) and Ganora et al. (2013) who, with different methodologies, have exploited measures of hydrologic uncertainty to merge regional and at-site information. The coefficient $\cancel{y}$ $\hat{y}$ can be considered a measure of the value of data. In fact, with all other parameters being equal, increasing n leads to a reduced $\cancel{y}$ $\hat{y}$ value and, consequently, to a reduced UNCODE design flood $\cancel{Q_T^*}\hat{Q}_T^*$. As a consequence, while the design value is still based on the same return period, costs will reduce.

Finally, the correction factor is a new and easy-to-implement design tool which provides a quantitative way to determine the design flood value accounting for hydrologic uncertainty, while keeping the same design hazard level considered in standard uncertainty-free analyses. This is a novel approach when compared to the common engineering practice, which accounts for hydrologic uncertainty by considering, for instance, the hydraulic freeboard. The use of the freeboard is equivalent to ~~increase~~ increasing the design flood value, but without accounting for the size of the system (e.g., the basin area), nor for the hydrologic information available at the section (i.e., observed of equivalent record length used to compute the standard design flood); this approach is thus not tailored to the specific case study. The correction factor represents an advance with respect to the use of "all-encompassing" safety factors and towards a clearer way to manage the different sources of uncertainty in hydrological and hydraulic design.

Acknowledgements. Funding from the ERC Consolidator Grant 2014 n. 647473 "CWASI - Coping with water scarcity in a globalized world" is acknowledged. D. Ganora also acknowledges the RTD Starting Grant from Politecnico di Torino. The work is based on simulated data.

References

Al-Futaisi, A. and Stedinger, J.: Hydrologic and Economic Uncertainties and Flood-Risk Project Design, Journal of Water Resources Planning and Management, 125, 314–324, doi:10.1061/(ASCE)0733-9496(1999)125:6(314), http://dx.doi.org/10.1061/(ASCE)0733-9496(1999)125:6(314), 1999.

Bao, Y., Tung, Y., and Hasfurther, V.: Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures, Water Resources Research, 23, 2023–2029, doi:10.1029/WR023i011p02023, http://dx.doi.org/10.1029/WR023i011p02023, 1987.

Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, Water Resources Research, 45, n/a–n/a, doi:10.1029/2007WR006726, http://dx.doi.org/10.1029/2007WR006726, 2009.

Botto, A., Ganora, D., Laio, F., and Claps, P.: Uncertainty compliant design flood estimation, Water Resources Research, 50, doi:10.1002/2013WR014981, 2014.

Brath, A., Montanari, A., and Moretti, G.: Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty), Journal of Hydrology, 324, 141 – 153, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.10.001, http://www.sciencedirect.com/science/article/pii/S0022169405004816, 2006.

Cameron, D., Beven, K., Tawn, J., and Naden, P.: Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), ~~HYDROLOGY AND EARTH SYSTEM SCIENCES~~Hydrology and Earth System Sciences, 4, 23–34, 2000.

De Michele, C. and Rosso, R.: Uncertainty Assessment of Regionalized Flood Frequency Estimates, Journal of Hydrologic Engineering, 6(6), 2001.

Ganora, D. and Laio, F.: A comparison of regional flood frequency analysis approaches in a simulation framework, Water Resources Research, pp. n/a–n/a, doi:10.1002/2016WR018604, http://dx.doi.org/10.1002/2016WR018604, 2016.

Ganora, D., Laio, F., and Claps, P.: An approach to propagate streamflow statistics along the river network, Hydrological Sciences Journal, 58, 41 – 53, 2013.

Ganoulis, J.: Risk-based floodplain management: A case study from Greece, International Journal of River Basin Management, 1, 41–47, 2003.

Hosking, J. and Wallis, J.: Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press, 1997.

Jonkman, S., Brinkhuis-Jak, M., and Kok, M.: Cost benefit analysis and flood damage mitigation in the Netherlands, Heron, 49 (1), 2004.

Kjeldsen, T. and Jones, D.: Estimation of an index flood using data transfer in the UK, Hydrological Sciences Journal-Journal des Sciences Hydrologiques, 52, 86–98, 2007.

Laio, F., Ganora, D., Claps, P., and Galeati, G.: Spatially smooth regional estimation of the flood frequency curve (with uncertainty), Journal of Hydrology, 408, 67–77, doi:10.1016/j.jhydrol.2011.07.022, 2011.

Liang, Z., Chang, W., and Li, B.: Bayesian flood frequency analysis in the light of model and parameter uncertainties, Stochastic Environmental Research and Risk Assessment, 26, 721–730, doi:10.1007/s00477-011-0552-y, http://dx.doi.org/10.1007/s00477-011-0552-y, 2012.

Menoni, S., Molinari, D., Ballio, F., Minucci, G., Mejri, O., Atun, F., Berni, N. and Pandolfo, C.: Flood damage: a model for consistent, complete and multipurpose scenarios, Natural Hazards and Earth System Sciences, 16, 2783–2797, doi:10.5194/nhess-16-2783-2016, http://dx.doi.org/10.5194/nhess-16-2783-2016, 2016.

Su, H. and Tung, Y.: Incorporating uncertainty of distribution parameters due to sampling errors in flood-damage-reduction project evaluation, Water Resources Research, 49, 1680–1692, doi:10.1002/wrcr.20116, http://dx.doi.org/10.1002/wrcr.20116, 2013.

Tung, Y.: Flood Defense Systems Design by Risk-Based Approaches, Water International, 30, 50–57, doi:10.1080/02508060508691836, 2005.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resources Research, 49, 675 – 692, 2013.

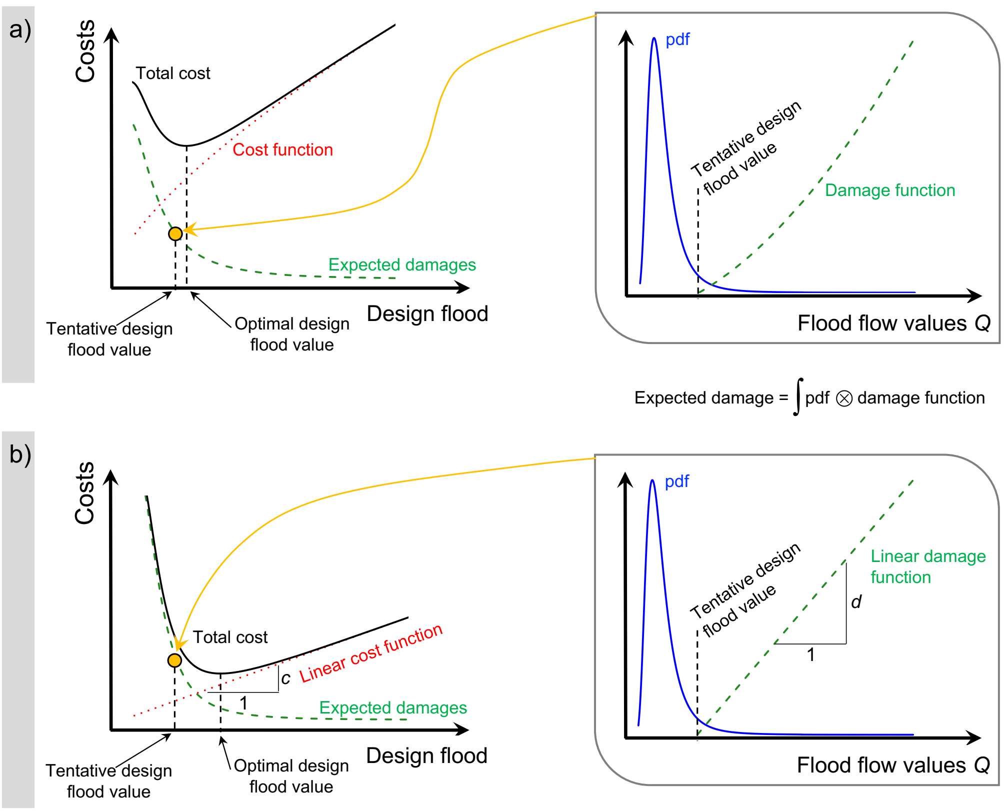

Figure 1. Illustrative example (without uncertainty) of the application of the cost-benefit framework to compute the design flood. Two generic cost and damage functions are reported in panel a, while panel b shows the linear functions adopted in the UNCODE framework.

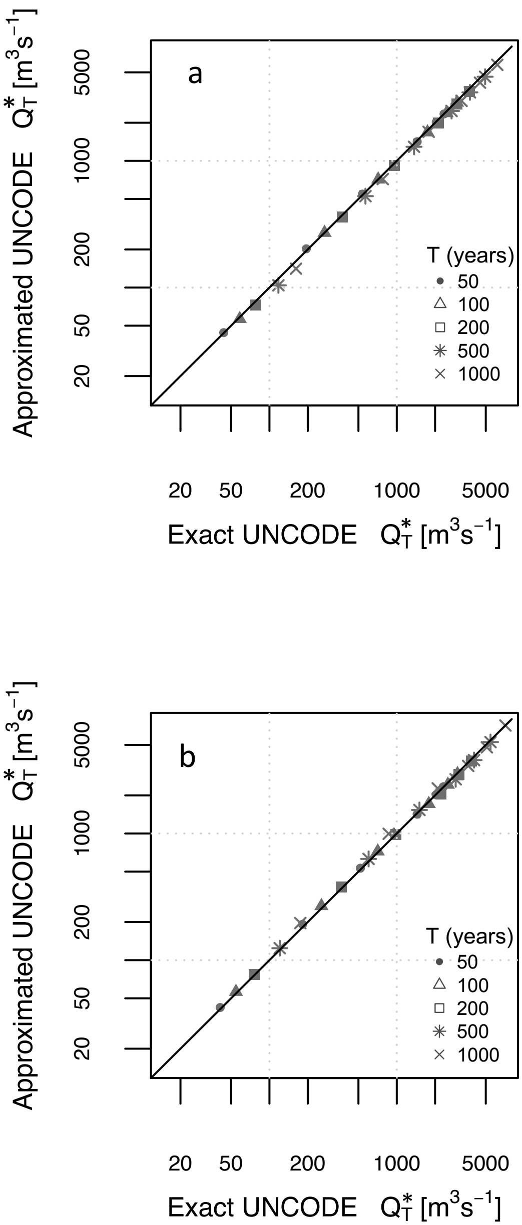

Figure 2. Comparison between the exact, Q^*_T, and the approximated, $\hat{Q}^*_T$, UNCODE estimators of the design flood ~~, Q^*_T,~~ for a pool of 6 flood ~~series~~ records considered in Botto et al. (2014, Table 1) with at least 30 years of data. Different return periods are listed in the legend. The reference distribution used for this flood frequency analysis is the 3-parameter log-Normal (LN3) in panel “a~~)~~” and the generalized extreme value (GEV) in panel “b~~)~~”.

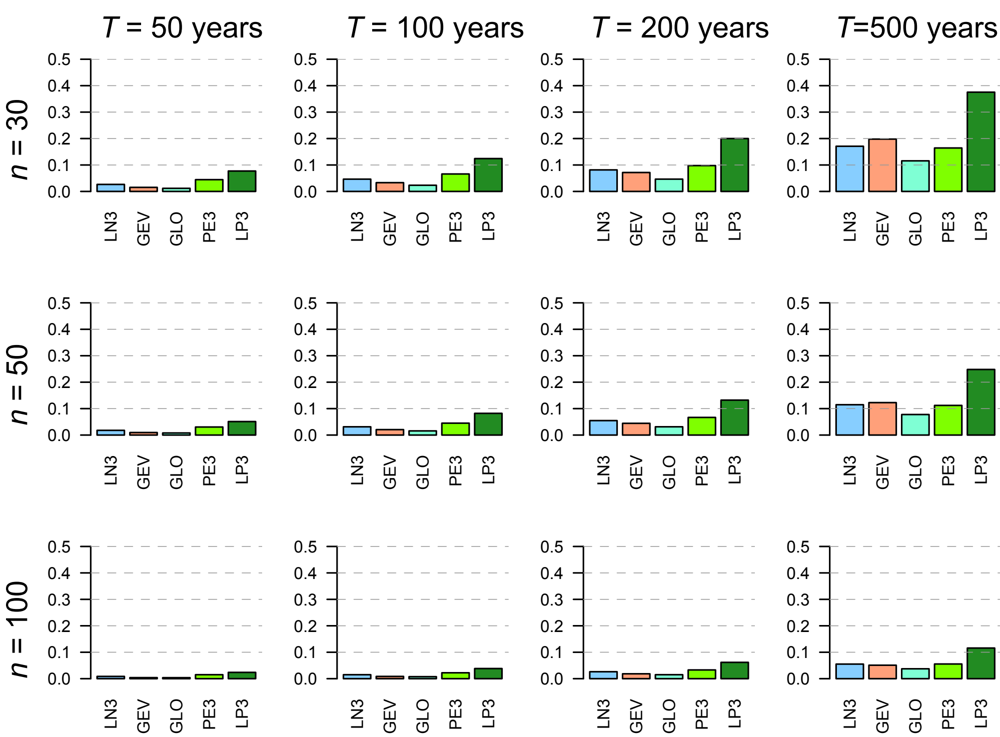

Figure 3. Values of the correction factor ~~y~~ $\hat{y}$ from Eq. (6) for some values of the sample length n and return period T and for different three-parameter fitting distributions (LN3 = log-Normal; GEV= generalized extreme value; GLO = generalized logistic; PE3 = Pearson type III; LP3 = log Pearson type III). ~~Details on the distributions can be found in Hosking and Wallis (1997); the~~ The LP3 corresponds to the PE3 with log-transformed variate.

Table 1. Coefficients to be used to estimate ~~y~~ $\hat{y}$ based on the sample length n and the return period T (eq. 6) and corresponding regression diagnostics, for different 3-parameter fitting distributions (LN3 = log-Normal; GEV= generalized extreme value; GLO = generalized logistic; PE3 = Pearson type III; LP3 = log Pearson type III). ~~Details on the distributions can be found in Hosking and Wallis (1997); the~~ The LP3 corresponds to the PE3 with log-transformed variate.

	a_0	a_1	a_2	R^2_{adj}	MAE	$RMSE$
LN3	-0.82	-0.25	0.809	0.94	0.0107	0.0160
GEV	-2.27	-0.3	1.110	0.85	0.0190	0.0321
GLO	-2.36	-0.25	0.994	0.85	0.0096	0.0145
PE3	0.59	-0.24	0.567	0.96	0.0080	0.0115
LP3	0.78	-0.26	0.687	0.89	0.0235	0.0363