Year: 2018

Uncertainty assessment of synthetic design hydrographs for gauged and ungauged catchments

Brunner, Manuela I ; Sikorska, Anna E ; Furrer, Reinhard ; Favre, Anne-Catherine

DOI: https://doi.org/10.1002/2017WR021129

Originally published at:
Brunner, Manuela I; Sikorska, Anna E; Furrer, Reinhard; Favre, Anne-Catherine (2018). Uncertainty assessment of synthetic design hydrographs for gauged and ungauged catchments. Water Resources Research, 54(3):1493-1512.
DOI: https://doi.org/10.1002/2017WR021129

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR021129

Key Points:

- Framework for quantifying the uncertainty of synthetic design hydrographs at three levels of complexity
- The major uncertainty sources in design hydrograph construction are the record length and the choice of a flood sampling strategy
- The uncertainty of design hydrographs for ungauged catchments is slightly higher than for gauged catchments

Correspondence to:
M. I. Brunner,
manuela.brunner@geo.uzh.ch

Citation:
Brunner, M. I., Sikorska, A. E., Furrer, R., & Favre, A.-C. (2018). Uncertainty assessment of synthetic design hydrographs for gauged and ungauged catchments. *Water Resources Research*, *54*. https://doi.org/10.1002/2017WR021129

Received 18 MAY 2017
Accepted 9 OCT 2017

Uncertainty Assessment of Synthetic Design Hydrographs for Gauged and Ungauged Catchments

Manuela I. Brunner[1,2], Anna E. Sikorska[1,3], Reinhard Furrer[4], and Anne-Catherine Favre[2]

[1]Department of Geography, University of Zurich, Zurich, Switzerland, [2]University of Grenoble-Alpes, CNRS, IRD, Grenoble INP, IGE, Grenoble, France, [3]Department of Hydraulic Engineering, Warsaw University of Life Sciences SGGW, Warsaw, Poland, [4]Department of Mathematics, University of Zurich, Zurich, Switzerland

Abstract Design hydrographs described by peak discharge, hydrograph volume, and hydrograph shape are essential for engineering tasks involving storage. Such design hydrographs are inherently uncertain as are classical flood estimates focusing on peak discharge only. Various sources of uncertainty contribute to the total uncertainty of synthetic design hydrographs for gauged and ungauged catchments. These comprise model uncertainties, sampling uncertainty, and uncertainty due to the choice of a regionalization method. A quantification of the uncertainties associated with flood estimates is essential for reliable decision making and allows for the identification of important uncertainty sources. We therefore propose an uncertainty assessment framework for the quantification of the uncertainty associated with synthetic design hydrographs. The framework is based on bootstrap simulations and consists of three levels of complexity. On the first level, we assess the uncertainty due to individual uncertainty sources. On the second level, we quantify the total uncertainty of design hydrographs for gauged catchments and the total uncertainty of regionalizing them to ungauged catchments but independently from the construction uncertainty. On the third level, we assess the coupled uncertainty of synthetic design hydrographs in ungauged catchments, jointly considering construction and regionalization uncertainty. We find that the most important sources of uncertainty in design hydrograph construction are the record length and the choice of the flood sampling strategy. The total uncertainty of design hydrographs in ungauged catchments depends on the catchment properties and is not negligible in our case.

1. Introduction

Hydrograph volume and shape, in addition to peak discharge, are important hydrograph characteristics for flood risk management tasks, such as the planning of retention basins and drawing hazard maps (Deutsche Vereinigung für Wasserwirtschaft Abwasser und Abfall, 2012; Klein et al., 2010; Schumann et al., 2010; Tung & Yen, 2005). A complete hydrograph is essential for all designs involving storage (Pilgrim, 1986) where the peak discharge, hydrograph volume, and the hydrograph shape provide complementary information. However, flood frequency analyses often focus on peak discharges without considering their dependence on hydrograph volumes. Brunner et al. (2017) therefore proposed an approach to construct synthetic design hydrographs (SDHs) that provide information on the peak discharge and the corresponding hydrograph volume together with the hydrograph shape. This approach takes into account the dependence between peak discharges and hydrograph volumes and models the hydrograph shape via a probability density function (Yue et al., 2002). In a follow up paper, Brunner et al. (2018) assessed how such SDHs can be transferred from gauged to ungauged catchments and identified the most suitable regionalization model. These previous studies suggested, that the construction and regionalization of SDHs may be linked with nonnegligible uncertainty which should be quantified in a next step.

Studies involving flood estimation entail various sources of uncertainty, such as measurement errors, various assumptions, sample selection, the choice of a suitable distribution function, the choice of a parameter estimation method, and sampling uncertainty (Merz & Thieken, 2005). Measurement errors comprise errors in water level measurements and errors coming from transferring water levels into discharge values via a rating curve (McMillan & Westerberg, 2015; Sikorska et al., 2013). Assumptions include those of stationarity, homogeneity, and independence of the data. Sample selection is associated with the choice of a representative observation period and the choice of a sampling strategy (annual maxima sampling versus peak-over-threshold sampling). Also, several distribution functions have been used to model the distribution of flood samples and the choice of one suitable distribution over another one is linked to uncertainty. The parameters of such a distribution can be estimated using different estimation techniques such as maximum likelihood and the method of moments or L-moments. Sampling uncertainty describes the uncertainty introduced by not knowing the population underlying a data set (Merz & Thieken, 2005). Among these sources of uncertainty, data availability and model choice are said to be the most important (Apel et al., 2004; Botto et al., 2014; Merz & Thieken, 2005). Yet, each step in the modeling process can introduce uncertainty (Kidson & Richards, 2005) and the overall uncertainty of the flood estimates results from the interaction of several uncertainty sources (Beven & Hall, 2014; Merz et al., 2008), which do not have to be additive (Montanari & Koutsoyiannis, 2012). Despite its importance, this uncertainty is often overlooked (Pappenberger & Beven, 2006) even though its consideration has several advantages (Juston et al., 2013). Uncertainty analysis allows the identification of uncertain parameters (Tung & Yen, 2005), a quantitative assessment of model reliability (Merz & Thieken, 2005; Montanari & Koutsoyiannis, 2012; Tung & Yen, 2005), and it provides a means of analyzing the robustness of flood risk management decisions. Furthermore, an analysis of the contribution of individual sources indicates where potential improvements in the method could have the greatest impact (Cullen & Frey, 1999; Hall & Solomatine, 2008; Sikorska et al., 2012) and therefore how uncertainty could be reduced (Qi et al., 2016), which is especially important for ungauged catchments (Sikorska et al., 2012). Although, uncertainty cannot be eliminated, its assessment at least enables its management (Koutsoyiannis, 2014).

Previous studies have dealt mainly with uncertainty analyses for univariate design variable quantiles (Serinaldi, 2009) usually estimated based on a sample of peak discharges. There, rather simple analytical and bootstrap methods allow the exploration of the uncertainty of extreme quantiles. In a univariate framework, the effect of the choice of the marginal distribution (Merz & Thieken, 2005; Qi et al., 2016), parameter uncertainty of the marginal distribution (Qi et al., 2016), data uncertainty from threshold selection (Qi et al., 2016; Xu et al., 2010), and the effect of the choice of annual maxima sampling versus peak-over-threshold sampling (e.g., Madsen et al., 1997; Martins & Stedinger, 2001a, 2001b; Sun et al., 2017) have been considered. In a bivariate framework that allows for the joint consideration of peak discharges and hydrograph volumes, the effect of the choice of annual maxima sampling versus peak-over-threshold sampling has, to our knowledge, not yet been analyzed. Furthermore, the uncertainty of the marginal distributions combines with the uncertainty of their dependence structure and infinite combinations of the studied variables exist that share the same joint probability (Serinaldi, 2013). Recently, Serinaldi (2013) and Dung et al. (2015) proposed several parametric and nonparametric bootstrap algorithms to compute confidence intervals for bivariate quantiles. However, there is still a lack of understanding of combined and interactive contributions of different uncertainty sources in bivariate quantile estimation (Qi et al., 2016). Furthermore, it is not clear how the uncertainty of bivariate design estimates describing the magnitude of an event interacts with the uncertainty related to the hydrograph shape. The goal of this study is therefore threefold:

1. Assessing the effect of the choice of a peak-over-threshold versus an annual maxima sampling strategy on design hydrograph construction in a bivariate framework.
2. Identifying the most important sources of uncertainty in design hydrograph construction and regionalization.
3. Assessing the uncertainty of synthetic design hydrographs for gauged and ungauged catchments.

To answer these questions, we propose an uncertainty assessment framework based on simulations with three levels of complexity. In a first step, the effect of different uncertainty sources on SDH construction is assessed. This allows for the identification of relevant uncertainty sources and therefore enables the refinement of the SDH construction (Brunner et al., 2017) and regionalization procedures (Brunner et al., 2018) in order to reduce uncertainty. Then, we assess the total uncertainty of SDHs for gauged and ungauged catchments that originates from individual steps in the SDH construction (gauged) and in the regionalization approach (ungauged). Finally, we propagate the uncertainty of the constructed SDHs in gauged catchments through SDH regionalization to ungauged catchments. This enables the quantification of the coupled uncertainty of SDHs that is composed of uncertainty from both the SDH construction and regionalization. Such SDHs with corresponding uncertainty bands should be provided to engineers and practitioners as reliable flood estimates (Chowdhury & Stedinger, 1991).

There is a lack of uniform terminology and a general disagreement about appropriate methodologies for uncertainty quantification in hydrological applications (Nearing et al., 2016). We will adopt an uncertainty definition often used in flood frequency analysis, where uncertainty is expressed as the variability of the design value under consideration. Serinaldi (2013) stated that complementing accurate point estimates with realistic confidence intervals (CIs), which clearly highlight the lack of information, is probably the most correct approach to communicate results of hydrological frequency analyses.

The choice of the uncertainty estimation method often depends on the sources of uncertainty considered. The influence of a model choice can be best quantified through comparing a number of models (Kidson & Richards, 2005). On the contrary, sampling uncertainty related to parameter estimation is usually either assessed via the distribution of maximum likelihood (ML) estimators or via resampling approaches (Beven & Hall, 2014). A resampling approach often used is bootstrapping which involves randomly selecting data points, with replacement, from the original sample and then estimating the extreme flow quantile from each of the resampled data sets (Burn, 2003; Efron & Tibshirani, 1993; Hall et al., 2004; Meylan et al., 2012; Wasserman, 2006).

The presence of several uncertainty sources requires their joint consideration. The errors of the various sources are usually not independent from each other and are therefore not necessarily additive (Montanari & Koutsoyiannis, 2012; Sikorska & Renard, 2017). Hence, the total uncertainty is usually not necessarily equal to the sum of its contributing sources. Moreover, most of the models or design procedures used in hydrosystems engineering and analysis are nonlinear and highly complex. This prohibits an analytical derivation of the probability distribution of the model outputs. Engineers therefore frequently resort to methods that yield approximations for the statistical properties of model outputs that are subject to uncertainty (Chang et al., 1994; Tung & Yen, 2005). A method often used to propagate uncertainties through a model chain is the Monte Carlo (MC) approach (Beven et al., 2010). This approach is often used to assess the total uncertainty of a hydrological model output that involves observational, model, and parameter uncertainty. Montanari and Koutsoyiannis (2012) proposed to estimate the distribution of the output of a process-based hydrological model via multiple simulation runs by perturbing input data, parameters, and model output. In this study, we adopt this idea to a flood frequency model. We conduct a bootstrap experiment to assess the distribution of synthetic design hydrographs for a specific catchment while considering different uncertainty sources.

2. Methods

2.1. Synthetic Design Hydrographs

Synthetic design hydrographs (SDHs) describe not only the peak discharge of a flood but also its hydrograph volume and shape (Brunner et al., 2017).

2.1.1. Construction of Synthetic Design Hydrographs in Gauged Catchments

Brunner et al. (2017) proposed a method for the construction of SDHs in gauged catchments based on runoff data only. The method models the entire shape of the hydrograph using a probability density function (PDF), and estimates the design variable quantiles peak discharge and hydrograph volume considering their dependence. It consists of eight steps:

1. Flood sampling using a peak-over-threshold (POT) approach.
2. Base flow separation using the recursive digital filter proposed by Eckhardt (2005) whose two parameters need to be estimated for each catchment.
3. Identification of the median hydrograph and its normalization. The median hydrograph is defined using the *h*-mode depth for functional data (Cuevas et al., 2007). In its normalized form, we refer to it as the representative normalized hydrograph (RNH).
4. Fitting of a lognormal probability density function (PDF) (Yue et al., 2002) to the RNH. The parameters of the PDF are computed as a function of the time to peak, the peak discharge, and the time base of the RNH (Nadarajah, 2007; Rai et al., 2009).
5. Determination of marginal distributions of peak discharges and hydrograph volumes. The Generalized Pareto distribution (GPD) is used to model the marginal distribution of peak discharges and the Generalized extreme value (GEV) distribution to model the marginal distribution of the hydrograph volumes.

6. Dependence modeling between peak discharges and hydrograph volumes using the Joe copula (Genest & Favre, 2007; Joe, 2015) independently of the choice of their marginal distributions.
7. Estimation of the design variable quantiles peak discharge (Q_T) and hydrograph volume (V_T) for a chosen return period. Computation of the duration of the design event (D_T).
8. Composition of the design hydrograph using the shape of the hydrograph given by the PDF ($f(t)$), the design variable quantiles (V_T and $D_T = V_T / Q_T$), and the base flow (B) as described by

$$Q_T(t) = f(t) V_T / D_T + B. \tag{1}$$

We refer to the procedure described above as the *standard configuration* for obtaining an SDH. For a further, detailed description of the methodology, the reader is referred to Brunner et al. (2017). The proposed methodology is generally applicable to any data set of interest, however, the model assumptions made in step 4 (choice of PDF), step 5 (choice of marginal distributions), and step 6 (choice of copula family) might need to be refined for a different data set than the one used in this study.

The design flood hydrographs obtained using this method are composed of 10 parameters, which we herein refer to as *SDH parameters*. Specifically, two parameters are needed to model the shape of the hydrograph defined by the lognormal PDF with a location and a scale parameter. Three parameters each (location, scale, and shape) are needed to model the marginal distributions of the design variables peak discharge and hydrograph volume. One parameter defines the dependence between these two variables and one parameter characterizes the proportion of base flow with respect to the direct hydrograph.

2.1.2. Regionalization of Synthetic Design Hydrographs to Ungauged Catchments

The SDH parameters can be transferred from gauged to ungauged catchments using methods based on the relation between catchment characteristics and model parameters, approaches based on spatial proximity, or on homogeneous regions. Brunner et al. (2018) tested regionalization methods belonging to five categories: (1) linear regression models establishing a linear relationship between SDH parameters and catchment characteristics, (2) nonlinear regression models additionally exploiting nonlinear relationships between SDH parameters and catchment characteristics, (3) spatial approaches, (4) regional mean models for fixed regions formed according to catchment size and elevation, and (5) methods forming homogeneous regions to transfer the whole parameter set from similar catchments to the ungauged catchment. The methods of the second category were found to be most suitable for the transfer of SDH parameters to ungauged catchments. Among these methods, the nonlinear regression method, boosted regression trees, performed best. A boosted regression tree model is a linear combination of many regression trees (CART models). It can be thought of as a regression model where each term is a tree (Elith et al., 2008; Hofner et al., 2014). It builds successive trees in a stagewise procedure where new trees depend on previous trees. Only a proportion of the observations is selected at each step to fit the tree model in order to prevent from overfitting (Friedman, 2001; Hastie et al., 2008). For a detailed overview and description of the methods tested, the reader is referred to Brunner et al. (2018). For another application of boosted regression trees in the hydrological context, see Tisseuil et al. (2010).

2.2. Uncertainty Assessment Framework

Our uncertainty analysis takes into account several uncertainty sources inherent in both the construction process and in the regionalization process of SDHs. Therefore, we distinguish between these two processes and three levels of complexity for analyzing uncertainty (for an illustration see Figure 1). On a first level (A), we assess the individual uncertainty sources present in the SDH construction and the SDH regionalization separately (section 2.1). On the second level (B), we assess the total uncertainty of the SDH construction and the total uncertainty of the SDH regionalization resulting from different sources separately (section 2.1). On the third level (C), we propagate the uncertainty of the SDH construction through the regionalization process to couple it with the total uncertainty present in regionalization (section 2.1). All these levels of the uncertainty analysis were based on bootstrap simulations. The basic principle was to construct a set of SDHs using various model configurations and to compare the characteristics of this set to the characteristics of an SDH obtained as a *best estimate* under the standard configuration (i.e., when no uncertainty is considered). The number of simulation runs was determined by a convergence analysis (Beven et al., 2010). We tried to minimize computation time by keeping the number of runs low while ensuring stable estimates. This resulted in different numbers of simulation runs for different levels of the analysis (generally 100 to

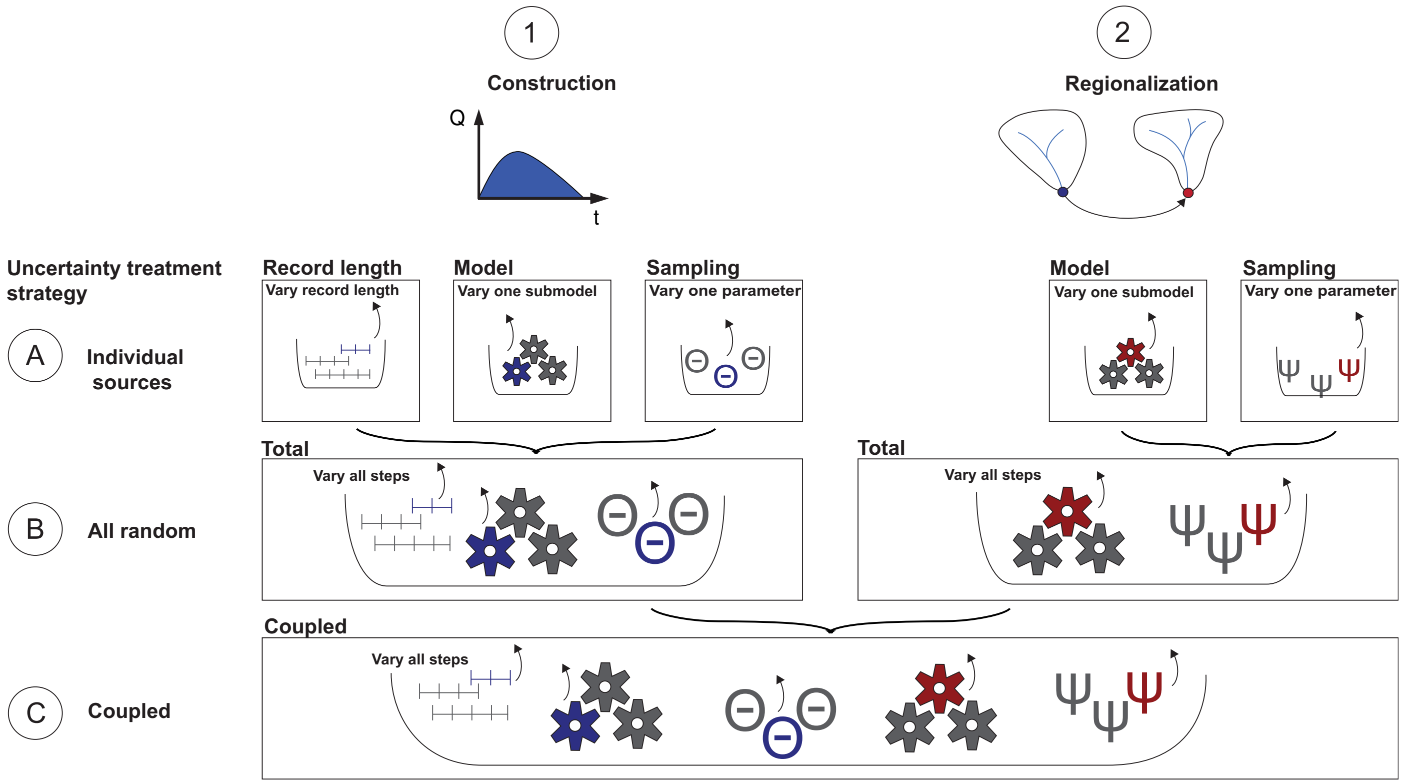

Figure 1. Illustration of the uncertainty framework proposed in this study. The uncertainty of constructed and regionalized SDHs was assessed on three levels of complexity: (A) uncertainty introduced by individual sources, specifically, record length, model choices, and parameter estimation; (B) total uncertainty of the constructed SDH and total uncertainty of the regionalized SDH resulting from the sources described in (A); (C) coupled uncertainty of the SDH when construction uncertainty (steps 1A–1B) is propagated through regionalization (steps 2A–2B) onto the final regionalized SDH.

1,000 runs were sufficient, 2,000 were used for the coupled uncertainty analysis. The exact numbers are indicated in the respective paragraphs). We did not consider observational uncertainty of runoff time series because neither sufficient information on runoff measurement error nor measurement technique was available for all the catchments in the data set. Furthermore, we did not consider the influence of the choice of the parameter estimation method since Dung et al. (2015) found that confidence regions for bivariate quantiles are hardly influenced by the choice of the parameter estimation method.

2.2.1. Uncertainty Due to Individual Sources

An investigation of the impact of individual uncertainty sources on the overall output uncertainty provides important information regarding the relative contribution of uncertainty sources to the overall uncertainty of the model output (here SDH) (Sikorska et al., 2012; Tung & Yen, 2005).

We distinguished between three categories of uncertainty: (1) uncertainty due to a limited record length, (2) model uncertainty resulting from the choice of one model over another feasible model, and (3) sampling uncertainty resulting from estimating the model parameters based on an available flood sample that only approximates the characteristics of the underlying population. The first and the third category are closely related. While the first category explicitly considers the effect of the sample size, the third category keeps the sample size constant but considers that a slightly different sample could have been observed leading to different estimates. The steps in the SDH construction and regionalization procedure concerned with model and sampling uncertainty are listed in Table 1. It indicates which models were found to be most appropriate for each step (see section 2.1) and which models would have also been feasible. Further, it highlights steps of parameter estimation which are subject to sampling uncertainty.

We focused on one uncertainty source at a time to assess the impact of the individual uncertainty sources on the estimated SDH. To do so, we constructed various SDHs using the standard model configuration while we either varied one model choice or considered the uncertainty of one parameter at a time. All the other model choices and parameters were fixed to the standard configuration (see Table 1: Model used). The set of SDHs

Table 1
List of Individual Uncertainty Sources Inherent in the Steps of the Construction and Regionalization of Synthetic Design Hydrographs (Section 2.1)

Step	Model used (standard configuration)	Other feasible models	Estimated parameters
1	Peak-over-threshold (POT) approach sampling four events per year on average	POT approach sampling two events per year on average Annual maxima in terms of peak discharges Annual maxima in terms of hydrograph volumes	
2	Two parameter recursive digital filter (Collischonn & Fan, 2013; Eckhardt, 2005)		α: recession coefficient estimated using linear regression; B_{max}: maximum base flow index which follows from α and discharge
3	Depth notion *h*-mode	Fraiman-Muniz depth Band depth Random projection depth (Cuevas et al., 2007)	
4	Lognormal PDF	Normal PDF Fréchet PDF Weibull PDF Logistic PDF Gamma PDF Inverse gamma PDF Beta PDF	
5	GPD for Q_p and GEV for V	GEV for Q_p GPD for V	Location, scale, and shape parameters of the GPD and the GEV estimated using the ML estimation
6	Use of Joe copula to model the dependence between Q_p and V	Gumbel Survival Clayton Tawn	θ: dependence parameter estimated based on pseudo-observations using maximum pseudolikelihood estimation
7	Choice of the event with ML on the isoline in the bivariate space	Choice of other event on isoline	
8	Mean event base flow		Estimation of mean event base flow
9	Nonlinear regression model boosting	Linear regression: lasso Spatial: universal kriging Regional mean: elevation zones Formation of homogeneous regions: median parameter set from most similar catchment.	Parameters of the boosting model estimated by minimizing a loss function

Note. The methods used in each step are listed together with other feasible models and information on the parameters estimated.

obtained by these simulations gives an idea of the variability introduced by each source of uncertainty considered. The uncertainty assessment framework proposed is applicable to any return period sensible in the light of data availability. In our case study, we focused on a return period of 100 years because it is frequently used as protection goals for agricultural areas and settlements in Switzerland (Camezind-Wildi, 2005). A preliminary analysis has shown that the relative importance of different uncertainty sources is independent of the return period. In this work, we considered the following uncertainty sources (see also Table 1):

1. *Record Length*. We computed SDHs to assess the effect of the sample size on the SDH estimates using subsamples of the original time series (Botto et al., 2014; Burn, 2003). The samples were drawn without replacement from the original sample to exclude a potential effect of the chronological order of the events. We started with drawing a time series of 20 years and increased the length of the time series by 5 years at a time until the maximum available record length was reached. The minimum length was set to 20 years because a flood frequency analysis based on a shorter time series would provide unreliable

flood estimates (DVWK, 1999). The stations with only 20 years of observations (6% of all stations) were excluded from this part of the analysis.

2. *Flood Sampling*. We assessed model uncertainty due to flood sampling by constructing SDHs using flood events based on four flood sampling strategies: (1) peak-over-threshold approach (Lang et al., 1999) sampling four events on average per year (POT_4); (2) peak-over-threshold approach sampling two events on average per year (POT_2); (3) annual peak maxima sampling (AM_Q); (4) annual volume maxima sampling (AM_V). To sample annual volume maxima, we computed the runoff volume over a running window of 72 h, the maximum length of a frontal storm in Switzerland, and identified the window with the maximum volume per year.
3. *Base Flow Separation*. Model uncertainty due to the base flow separation method was not assessed since the two-parameter recursive digital filter proposed by Eckhardt (2005) was previously found to outperform alternative models for the application under consideration (Brunner et al., 2017). The recursive digital filter used requires the determination of two parameters: a recession coefficient α and a maximum base flow coefficient B_{max}. The uncertainty due to these two parameters was assessed via a parametric bootstrap by sampling the two parameters $N_B=100$ times from their distributions. α was sampled from a Weibull distribution fitted to the α's overall study catchments while B_{max} was sampled from a normal distribution fitted to the B_{max}'s over all catchments. The Weibull and normal distributions were found to fit the data well based on the Kolmogorov-Smirnov goodness of fit test (level of significance of 0.05).
4. *Normalization and Identification of a Representative Normalized Hydrograph*. The identification of the representative normalized hydrograph (RNH) does, on the contrary to the normalization of the hydrograph, introduce uncertainty. The RNH was defined as the median hydrograph of the catchment under consideration. The *h*-mode depth was chosen to define the median hydrograph within a catchment among a set of four suitable definitions of depth functions for functional data. We considered the model uncertainty coming from the choice of one depth function over the others. For this, we constructed four SDHs using each of the suitable depth functions: *h*-mode depth, Fraiman-Muniz depth, random projection depth (Cuevas et al., 2007), and band depth (López-Pintado & Romo, 2009). The definition of the RNH via the median does not involve parameter estimation, therefore, sampling uncertainty did not need to be considered.
5. *Fitting of a Probability Density Function to the RNH*. The fitting of a PDF to the RNH to model the shape of the SDH introduces model uncertainty. One could fit another PDF to the RNH instead of the lognormal PDF, since the best PDF depends on the catchment. We chose the lognormal PDF because it showed the best average fit in terms of the Nash-Sutcliffe and Kling-Gupta efficiencies (Gupta et al., 2009) when considering all catchments in the data set. This model uncertainty was assessed by constructing SDHs using eight different PDFs: lognormal, normal, Fréchet, Weibull, logisitic, gamma, inverse gamma, and beta PDF. Sampling uncertainty due to the fitting of a PDF to the RNH was not considered since the estimation method chosen is based on an analytical expression (Nadarajah, 2007; Rai et al., 2009). It expresses the parameters of the PDF in terms of the time to peak and the time base of the RNH instead of inferring the parameters via classical estimation approaches such as maximum likelihood or method of moments.
6. *Determination of Marginal Distributions of Peak Discharges and Hydrograph Volumes*. The determination of the marginal distributions for the two design variables peak discharge (Q_p) and hydrograph volume (V) introduced both model and sampling uncertainty. Goodness of fit tests (Kolmogorov-Smirnov, Anderson-Darling, and upper tail Anderson-Darling test, *Chernobai et al*., 2015) of different model alternatives (Generalized Pareto (GPD), Generalized Extreme Value (GEV), Gumbel, generalized logistic, Pearson type III, normal, lognormal, exponential, and Weibull) showed that only the hypothesis of GPD and GEV distributions were not rejected in most of the catchments at $\alpha=0.05$. The results of the goodness of fit tests suggested that the GPD and GEV models could both be used to model Q_p and V and it was difficult to decide which distribution was a better predictor over another (Beven & Hall, 2014). We therefore assessed model uncertainty for the marginal distributions of peak discharges and hydrograph volumes by constructing two SDHs based on each of the two admissible marginal distributions. The parameters of the marginal distributions were estimated using ML estimation (Held & Sabanés Bové, 2014), which introduced sampling uncertainty. The ML approach assumes that the parameter estimates follow a parametric distribution. This distribution can be used to assess sampling uncertainty. We

constructed $N_B=500$ SDHs by sampling parameters from a multivariate Normal distribution defined by the means and covariance matrices of the location, scale, and shape parameters of the marginal distributions. We only allowed positive location and scale parameters to be sampled because the location and scale parameters must be positive.

7. *Dependence Modeling Between Q_p and V*. The modeling of the dependence between Q_p and V introduced both model and sampling uncertainty. Model uncertainty was introduced because several copula models were nonrejected in most catchments (Brunner et al., 2017). A statistical bootstrap procedure was applied to compute a p value for the Cramér-von Mises statistic (Genest et al., 2009) of 12 copulas (Independence, Gumbel, Clayton, Joe, Frank, AMH, normal, Student, Tawn, t-EV, Plackett, and Survival Clayton) (Joe, 2015). Only four copulas were not rejected at a level of significance of $\alpha=0.05$ in most of the catchments: Joe, Gumbel, Survival Clayton, and Tawn. We used each of the nonrejected copulas to construct four SDHs. Sampling uncertainty was introduced because the copula parameter θ could only be estimated based on a sample as the underlying population was unknown. We assessed this uncertainty using a parametric bootstrap experiment in which we constructed $N_B=500$ SDHs sampling the θ from a lognormal distribution fitted to the θs of all the study catchments. The lognormal distribution was found to fit the θs well based on the Kolmogorov-Smirnov goodness of fit test. The θ of the Joe copula must be positive. Therefore , we only allowed the sampling of positive values.
8. *Estimation of the Design Variable Quantiles for Peak Discharge and Hydrograph Volume*. The choice of the definition of the return period is defined by the problem at hand (Brunner et al., 2016; Serinaldi, 2015) and the return period usually defined in national guidelines (Requena et al., 2013). However, several design events have the "same" probability of occurrence in bivariate frequency analysis and thus lie on an isoline. Often, the most likely of these pairs of design variables is chosen for practical application (Brunner et al., 2016). This choice, however, introduces uncertainty because one could choose another pair of design variables on the isoline instead of the one with the highest likelihood. This source of uncertainty was assessed by a bootstrap experiment with $N_B=1{,}000$ SDHs sampling one point on the isoline according to the probability density function of the points on the isoline. Sampling according to the probability density function ensures that more points close to the most likely point and very few extreme points are sampled.
9. *Composition of the Design Hydrograph*. The SDH can be composed using equation (1) based on the shape of the hydrograph given by the PDF and the magnitude of the event given by the design variable quantiles and the base flow component. The uncertainty of the shape and of the magnitude has been considered above. Additional uncertainty may come from adding the base flow component. We assessed these sources of uncertainty using a bootstrap experiment with $N_B=100$ SDHs sampling a mean event base flow index from a normal distribution defined by a mean and a standard deviation derived from the mean base flow indices over all catchments. The normal distribution was used because it was found to fit the base flow indices over all catchments well based on the Kolmogorov-Smirnov goodness of fit test.
10. *Regionalization*. The regionalization of the SDHs from gauged to ungauged catchments introduced both model and sampling uncertainty. Brunner et al. (2018) found that the nonlinear regression model boosting was most suitable to regionalize the magnitude of the SDH. However, they tested models of different categories among which several could be found that performed as well as the boosting model. Thus, the choice of one regionalization model may have introduced model uncertainty. This source of uncertainty was assessed by regionalizing the SDH parameters (best estimates obtained using the standard configuration) using five regionalization methods. The five models considered were found to have an acceptable performance in the regionalization of SDHs in the previous study (Brunner et al., 2018). These were:

    1. Linear regression: lasso.
    2. Nonlinear regression: boosted regression trees (boosting).
    3. Kriging: universal kriging with catchment area as explanatory variable.
    4. Regional mean: elevation zones.
    5. Formation of homogeneous regions: median parameter set from the five most similar catchments determined using hydrological reasoning.

The sampling uncertainty was assessed using a bootstrap experiment which focused on the most suitable regionalization method boosting (Brunner et al., 2018). We sampled with replacement $n=163$ (original sample size) catchments from the catchment set. The boosting model was fitted for each SDH parameter using the data from the resampled catchment set. The 10 SDH parameters were predicted for the original catchment set using the respective regionalization model. This procedure was repeated $N_B = 500$ times resulting in a separate distribution for each SDH parameter in each catchment. These catchment specific distributions were fitted by a normal distribution since parametric bootstrap is to be preferred over non-parametric bootstrap (Kysely, 2008). However, the normal distribution did not provide a good fit for the location and scale parameters of the marginal distributions of peak discharges and hydrograph volumes. These distributions were found to be bimodal or skewed in most catchments (roughly 90%) and were therefore difficult to represent by a theoretical distribution. For consistency reasons, we still used the normal distribution to represent the distribution of all SDH parameters in all catchments. Using a theoretical distribution allowed the dependency between the location and scale parameter of the marginal distributions for Q_p and V to be considered by sampling from a multivariate normal distribution. This would not have been possible when using the empirical distributions. The use of a normal distribution instead of the empirical distribution, however, implied that central values were overrepresented in samples supposed to show skewed or bimodal rather than centered distributions. We constructed $N_B = 500$ SDHs for each catchment randomly sampling the 10 SDH parameters from the catchment specific parameter distributions.

2.2.2. Total Uncertainty

On the previous level of complexity of the uncertainty analysis, we considered the effect of each uncertainty source independently of the other uncertainty sources. Here, we combine all these sources and assess the resulting total uncertainty of the (1) SDH construction and of the (2) SDH regionalization. However, we did not yet consider here that the constructed SDHs used to fit a regionalization model are uncertain themselves. That means we considered the regionalization uncertainty independently of the construction uncertainty at this level of the analysis.

1. The total uncertainty coming from the SDH construction process was assessed for a return period of $T = 100$ years by an all "random" strategy. We use the term "all random" strategy for a bootstrap simulation in which model choices were not fixed (instead models were also sampled) and where sampling uncertainty was taken into account. At each step of the construction procedure, we randomly sampled one option from the model and/or parameter space (see Table 1 for a list of admissible models). The uncertainty sources considered in the "all random" strategy were all the same as already considered in the assessment of the uncertainty coming from the individual uncertainty sources (see section 2.1). The procedures applied were also the same except that the sampling uncertainty due to the dependence modeling (step 7) was extended from the Joe copula to the other three suitable copulas. The θs of the Gumbel and Survival Clayton copulas, as those of the Joe copula, could be described by a lognormal distribution. However, no theoretical distribution could be found for the θs of the Tawn copula. Therefore, we used the empirical distribution of the Tawn θs over all catchments. This strategy was repeated $N_B = 1{,}000$ times to construct 1000 SDHs.
2. The total uncertainty coming from the SDH regionalization was also assessed by a bootstrap simulation. In each iteration, a regionalization model was first sampled from the model space. Then, the model was fitted for each SDH parameter using the data from a resampled catchment data set (with replacement) to also consider sampling uncertainty. The sample size was kept the same as the original sample size (i.e., number of catchments $n = 163$) to not confuse uncertainty coming from a limited sample size with sampling uncertainty due to not knowing the true population. However, the resampling with replacement introduced numerical problems (noninvertible covariance matrix) when using the regionalization model universal kriging. To solve this problem, we removed redundant stations (approximately one third of the samples) from the resampled catchment set when the regionalization model universal kriging was sampled in the first step. The fitted models were then used to make predictions for the SDH parameters of the catchments in the original catchment set. The bootstrap experiment was repeated $N_B = 2{,}000$ times (determined by a convergence analysis) to construct 2,000 regionalized SDHs.

2.2.3. Coupled Uncertainty

In the previous step, we assumed that the regionalization uncertainty was independent from the construction uncertainty. On the third level of complexity, we combined both uncertainty sources, i.e., the uncertainty in the construction of SDHs with the uncertainty in the regionalization of SDHs. To this end, the uncertainty of the constructed SDHs of a return period of $T = 100$ years was propagated through the regionalization process. We refer to this uncertainty as the coupled uncertainty of SDHs. To propagate the construction uncertainty, catchment specific distributions of the 10 SDH parameters had to be defined. This was done by running the "all random" strategy already used for the assessment for the total construction uncertainty $N_B=1{,}000$ times for each catchment. This provided 1,000 values for each of the 10 SDH parameters which were assumed as the empirical distribution of these parameters within a catchment. For the uncertainty propagation, we then sampled one value from each of the 10 empirical parameter distributions for each catchment and fitted a randomly sampled regionalization model to the data of a resampled (with replacement) set of catchments. We then used the fitted model to predict the SDH parameters for the original catchment set. This procedure was repeated $N_B=2{,}000$ times. This corresponds to the procedure already applied when assessing the total regionalization uncertainty. The only difference between the assessment of the total regionalization uncertainty and the coupled uncertainty assessment was that in the former the 10 parameters in each catchment were fixed whereas in the latter the 10 SDH parameters for each catchment were sampled from catchment specific empirical distributions.

3. Case Study

This uncertainty assessment study was performed using 163 Swiss catchments (for a map and a complete list of the catchments and their catchment characteristics see Brunner et al., 2018) with a wide range of catchment characteristics and flood behaviors. The catchments selected have hourly flow series of at least 20 years duration ranging up to 53 years. The catchments' runoff is neither altered by regulated lakes upstream or inland canals nor by highly urbanized areas. The catchments are small to medium-sized (6–1,800 km^2), situated between 400 and 2,600 m.a.s.l. with respect to mean elevation, and have no or only a small percentage of areas with glaciers.

4. Results

4.1. Quantification of Uncertainty Via Median Absolute Relative Error

Each step of the uncertainty analysis provided a set of SDHs that were obtained by varying one or several model choices and/or considering sampling uncertainty. These sets were compared to the *best estimate SDH* obtained when using the standard configuration (i.e., without considering any uncertainty). To this end, we computed the absolute relative error of each SDH in the set when compared to the best estimate SDH. We summarized this information for each catchment by taking the median of the absolute relative errors of all SDHs in the set considered. We refer to this as the median absolute relative error (E_{MAR}). The median was taken instead of the mean because it puts relatively little weight on very extreme errors compared to the mean and is therefore more robust. The E_{MAR}s of all catchments together provided information on the variability of the uncertainty introduced by a certain uncertainty source across catchments. The E_{MAR} was used on all three levels of complexity of the uncertainty assessment framework. It was computed for four important hydrograph characteristics: the peak discharge (Q_p), the hydrograph volume (V), the time to peak (t_p), and the half-recession time (t_{p05}). The half-recession time was defined as the time from the peak to where the hydrograph falls back to 0.5 times the peak discharge. The characteristics Q_p and V describe the magnitude of the event while the characteristics t_p and t_{p05} describe the shape of the hydrograph.

The results from the three levels of complexity are summarized in Table 2 and presented in the next few sections.

4.2. Uncertainty Due to Individual Sources

The E_{MAR} due to the individual uncertainty sources belonging to the categories record length, model, and sampling uncertainty are displayed in Figure 2 for a return period of $T = 100$ years. The median uncertainty of the individual uncertainty sources over all catchments ranged between 0% and roughly 30%. We look at each hydrograph characteristic in turn. The E_{MAR} of Q_p was influenced by most sources of uncertainty considered. Among them, the choice of the marginal distribution of Q_p (median across sites $E_{MAR}=20\%$), the sampling uncertainty of this distribution (10%), the choice of the sampling strategy (8%), base flow addition (8%), and the record length (3%) had the most prominent influence. Other sources of uncertainty, in particular, the ones related to the shape of the hydrograph, only slightly affected the E_{MAR} of Q_p. The E_{MAR} of V was also affected by most sources of uncertainty. The largest E_{MAR} came from the sampling uncertainty of the marginal distribution of V (across sites $E_{MAR}=12\%$), the choice of the sampling strategy (12%), the

Table 2
Summary of Uncertainties Across the Three Levels of Complexity for the Four Hydrograph Characteristics Q_p, V, t_p, and Q_{p05}

	Q_p			V			t_p			t_{p05}		
Uncertainty sources	First	Second	Third	First	Second	Third	First	Second	Third	First	Second	Third
Sample size	0	0.03	0.06	0.02	0.09	0.19	0	0.19	0.5	0	0.17	0.45
Sampling strategy	0.05	0.08	0.17	0.08	0.12	0.2	0.11	0.2	0.27	0.11	0.17	0.24
RNH definition	0	0	0	0.01	0.03	0.04	0.07	0.16	0.23	0.01	0.02	0.03
PDF choice	0	0	0	0.03	0.04	0.06	0.24	0.27	0.32	0.04	0.06	0.09
Copula choice	0	0.01	0.01	0.01	0.02	0.03	0.01	0.01	0.02	0.01	0.01	0.02
Choice on isoline	0.02	0.03	0.04	0.04	0.05	0.07	0.07	0.09	0.11	0.07	0.09	0.11
Margin Q_p	0.12	0.2	0.34	0	0	0	0.09	0.14	0.2	0.09	0.14	0.2
Margin V	0	0	0	0.08	0.12	0.16	0.08	0.12	0.16	0.08	0.12	0.16
Base flow separation	0.03	0.04	0.06	0.07	0.11	0.16	0.09	0.13	0.21	0.1	0.15	0.2
Base flow addition	0.06	0.08	0.1	0.06	0.08	0.1	0	0	0	0	0	0
Copula parameter	0.01	0.01	0.01	0.01	0.02	0.03	0.01	0.01	0.01	0.01	0.01	0.01
Margin Q_p parameter	0.08	0.1	0.14	0	0	0	0.07	0.1	0.14	0.07	0.1	0.14
Margin V parameter	0	0	0	0.13	0.15	0.19	0.13	0.15	0.19	0.13	0.15	0.19
Total construction	0.21	0.26	0.35	0.28	0.34	0.4	0.41	0.45	0.52	0.37	0.44	0.5
Regionalization model	0.23	0.36	0.63	0.21	0.34	0.76	0.12	0.21	0.34	0.19	0.28	0.41
Regionalization sampling	0.11	0.15	0.2	0.12	0.16	0.22	0.18	0.22	0.3	0.18	0.22	0.3
Total regionalization	0.26	0.4	0.67	0.29	0.43	0.68	0.14	0.25	0.41	0.2	0.3	0.44
Coupled	0.47	0.59	0.94	0.45	0.55	0.66	0.43	0.48	0.54	0.44	0.48	0.53

Note. The uncertainties are provided in the form of the first, second, and third quartile of the E_{MAR}s over all catchments. The numbers were rounded to two decimal places.

choice of the marginal distribution of V (12%), the base flow separation (11%), and record length (median 9%). The E_{MAR} of t_p was, as the other characteristics, also influenced by most uncertainty sources. Among them, the choice of the PDF (median across sites $E_{MAR}=27\%$), the choice of the sampling strategy (20%), the record length (19%), the RNH definition (16%) had the largest influence. Similarly, the E_{MAR} of t_{p05} was most strongly influenced by the record length (median across sites $E_{MAR}=17\%$) and the choice of the sampling strategy (17%).

4.2.1. Influence of Sampling Strategy

As shown above, the sampling strategy was one of the uncertainty sources most strongly affecting the E_{MAR} of all four hydrograph characteristics. We therefore have a closer look at this source of uncertainty. Each of the four sampling strategies defined its own flood sample, which formed the basis of the subsequent analysis. The four flood samples were quite different. On the one hand, they had different sample sizes ($POT_4 > POT_2 > AM_Q = AM_V$), on the other hand, they differed in terms of their dependence structure. The sample of the POT_4 strategy was largest and did not reject only four copulas: Joe, Gumbel, Survival Clayton, and Tawn. All these models were able to model tail dependence which was found to be present in the data according to the upper tail dependence coefficient (Poulin et al., 2007). The sample of the POT_2 strategy was half as large as the one obtained by the POT_4 strategy. The nonrejected copulas for this strategy were, in most of the catchments: Survival Clayton, Placket, Student, Independence, Joe, Gumbel, and Tawn. The sampling strategy AM_Q led to quite a small sample size and almost all of the copulas tested, including the independence copula, were nonrejected. The sampling strategy AM_V also led to a small sample and nonrejected the Survival Clayton, t-EV, normal, Frank, and Gumbel copulas in most of the catchments. The dependence between Q_p and V assessed via the three correlation coefficients Pearson, Spearman, and Kendall was highest for the strategy AM_V and was comparable for the other three sampling strategies. The characteristics of the flood sample influenced the estimated SDHs, which is shown in Figure 3 for an example catchment. It shows that design hydrograph estimates derived based on the AM_V sample had generally a smaller magnitude than the estimates derived based on the flood samples obtained by the other strategies. Design flood estimates with very high peaks are mainly related to the sampling strategies POT_2 and POT_4.

4.2.2. Regionalization Uncertainty

Model and parameter choices both introduced uncertainty to the regionalized SDHs (see Figure 4). All four hydrograph characteristics were more affected by model uncertainty than by sampling uncertainty. The median E_{MAR} across all catchments related to model uncertainty lay around 30% for all hydrograph

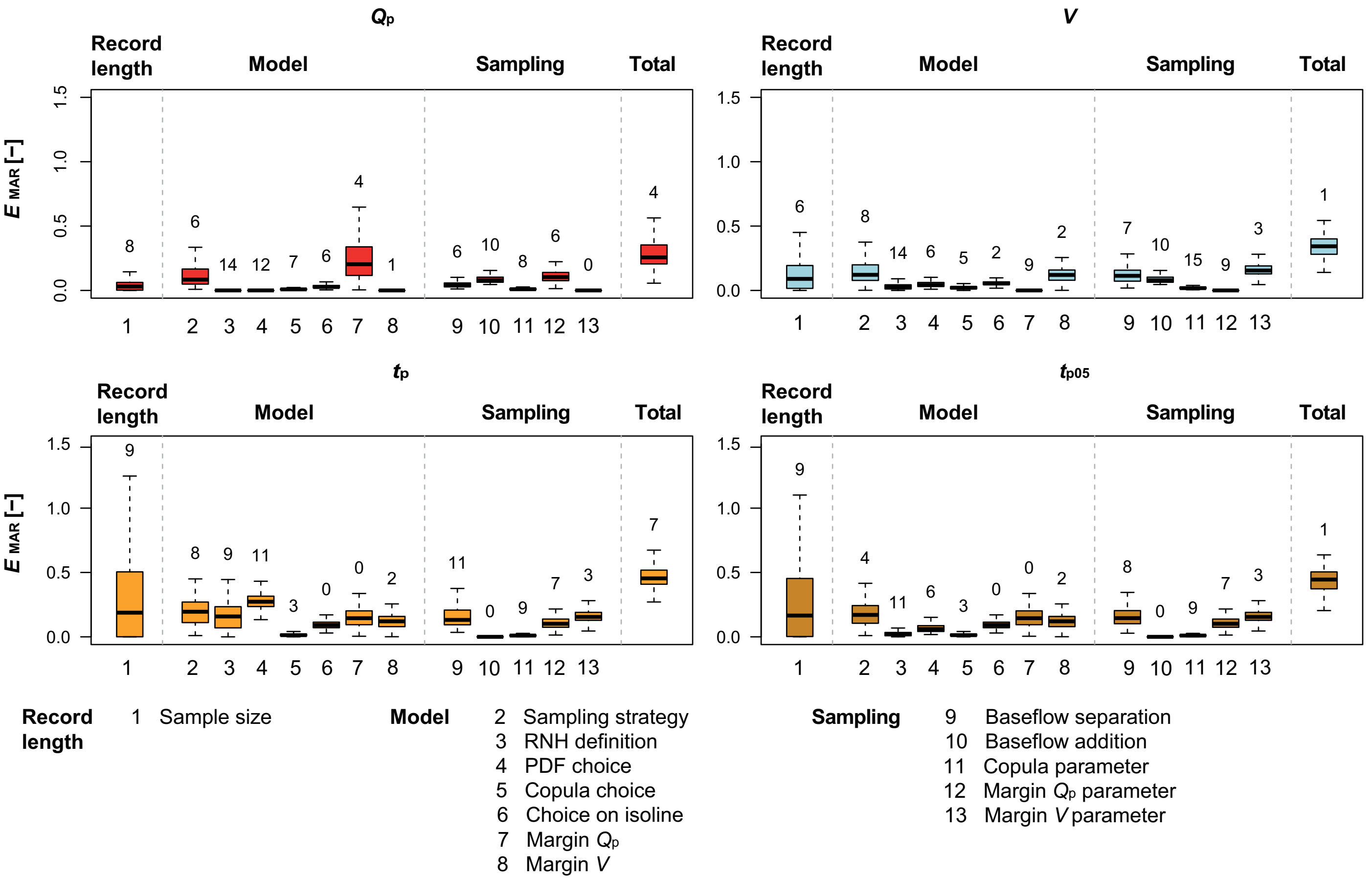

Figure 2. Median relative errors within the catchments of four hydrograph characteristics for an SDH of $T=100$ years due to different uncertainty sources. The hydrograph characteristics assessed were peak discharge (Q_p), hydrograph volume (V), time to peak (t_p), and half-recession time (t_{p05}). The whiskers extend to 1.5 times the interquartile range. Outliers are not displayed, however, their number among the 163 study catchments is given above the upper whisker of the boxplot.

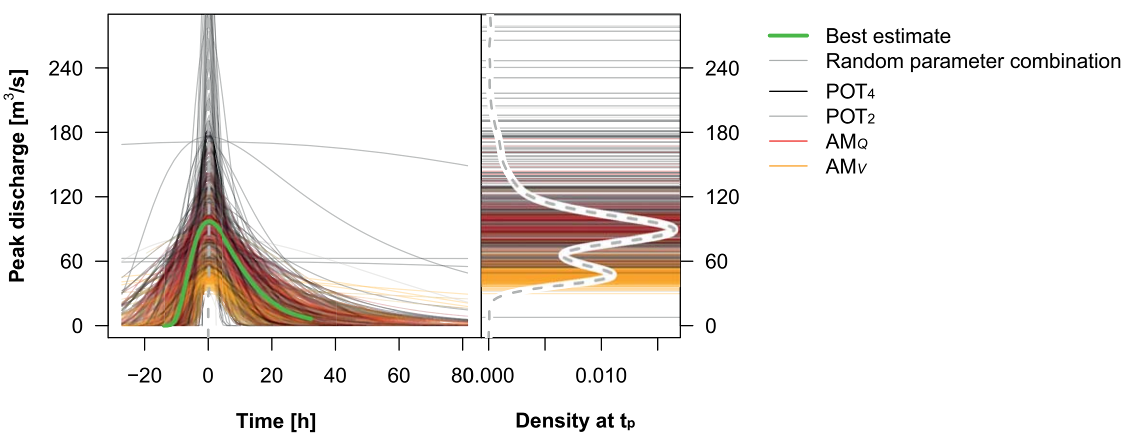

Figure 3. (left plot)$N_B=1,000$ SDHs obtained using the all random strategy for different sampling strategies for the Birse-Moutier catchment. SDHs obtained using a peak-over-threshold approach sampling four events per year on average (POT_4) are colored in black, and two events per year on average (POT_2) are colored in gray, SDHs obtained by annual maxima peak sampling (AM_Q) are depicted in red and those obtained by annual maxima volume sampling (AM_V) in orange. The best estimate SDH constructed using observed runoff data is given in green. The density of peak discharges is given in the right plot. Peaks are colored as in the left plot.

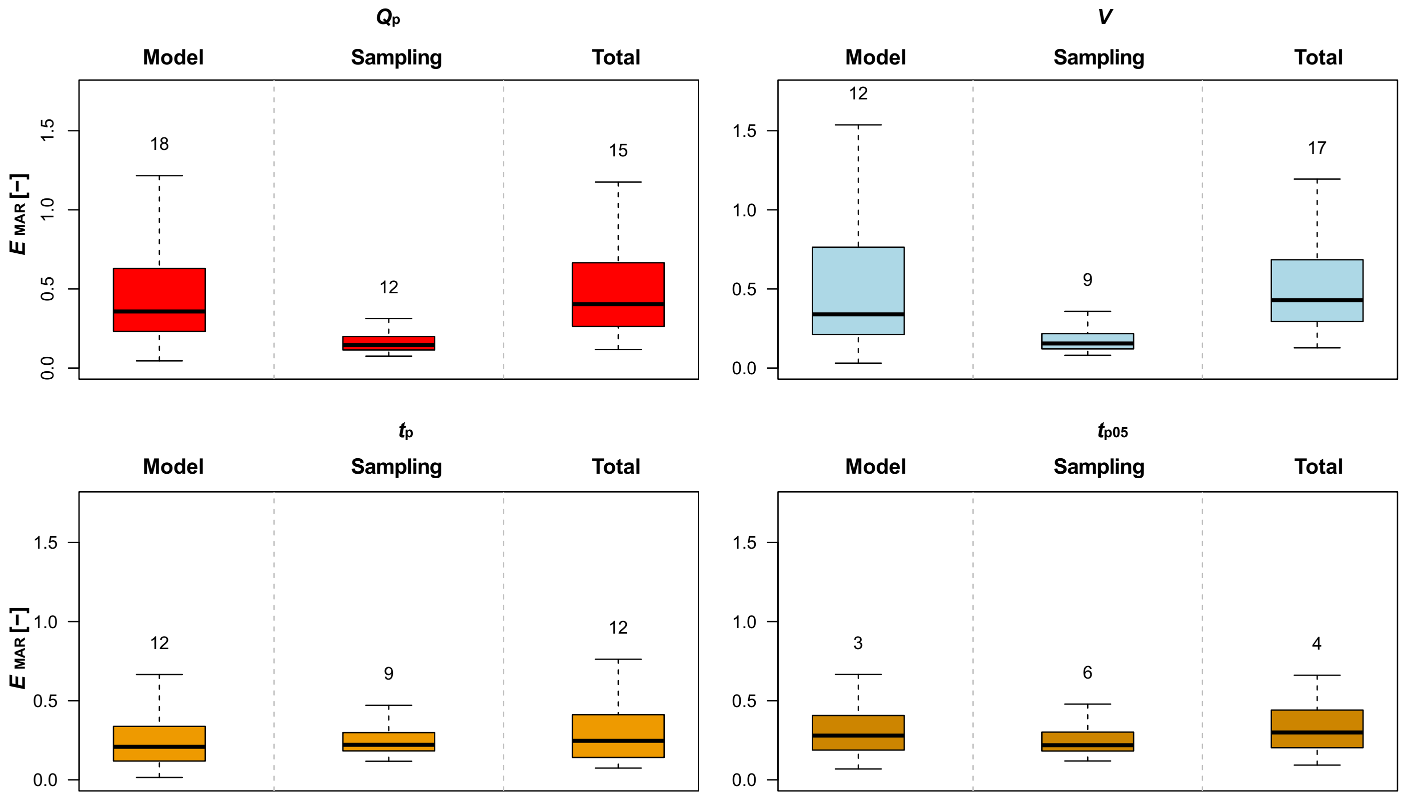

Figure 4. Regionalization uncertainty expressed as the median relative errors within the catchments of the four hydrograph characteristics: peak discharge (Q_p), hydrograph volume (V), time to peak (t_p), and half-recession time (t_{p05}). Model and sampling uncertainty are given separately and combined (total). The whiskers extend to 1.5 times the interquartile range. Outliers are not displayed, however, their number among the 163 study catchments is given above the upper whisker of the boxplot.

characteristics, while the median E_{MAR} related to sampling uncertainty was roughly 15% for peak discharges and hydrograph volumes and 20% for the characteristics related to the shape of the hydrograph. The characteristics describing the hydrograph shape (t_p and t_{p05}) were more affected by sampling uncertainty than the characteristics describing the hydrograph magnitude (Q_p and V).

4.3. Total Uncertainty

4.3.1. Construction Uncertainty

The E_{MAR} of the total construction uncertainty for the four hydrograph characteristics (for $T = 100$ years) was larger than each of the individual uncertainty sources (Figure 2), as we had expected. The largest median E_{MAR} (45%) over all catchments was found for the hydrograph characteristics related to the shape of the hydrograph (t_p and t_{p05}) with an interquantile range of 40% to 50%. The smallest median E_{MAR} was found for Q_p (25%) with an interquartile range spanning from 20% to 35%. The median E_{MAR} of V over all catchments was 35% and the interquartile range spanned from 30% to 40%.

The SDHs obtained by the "all random" strategy are displayed in Figure 5 for three example catchments of different size: Langete-Huttwil (60 km^2), Mentue-Yvonand (105 km^2), and Birs-Münchenstein (911 km^2). The 90% confidence interval includes the best estimate SDH.

4.3.2. Regionalization Uncertainty

The E_{MAR} of the total regionalization uncertainty for the four hydrograph characteristics assessed for $T = 100$ years was as expected larger than the separate sampling and model uncertainties (Figure 4) even though the interquartile ranges are quite similar as those of the model uncertainties. The largest median E_{MAR} (40%) over all catchments was found for the hydrograph characteristics related to the magnitude of the hydrograph (Q_p and V) with an interquantile range of 25–65% and 30–70%, respectively. The median E_{MAR} was lowest for t_p (25%) with an interquartile range spanning from 15% to 40%. The median E_{MAR} of t_{p05} was 30% and the interquartile range spanned from 20% to 45%.

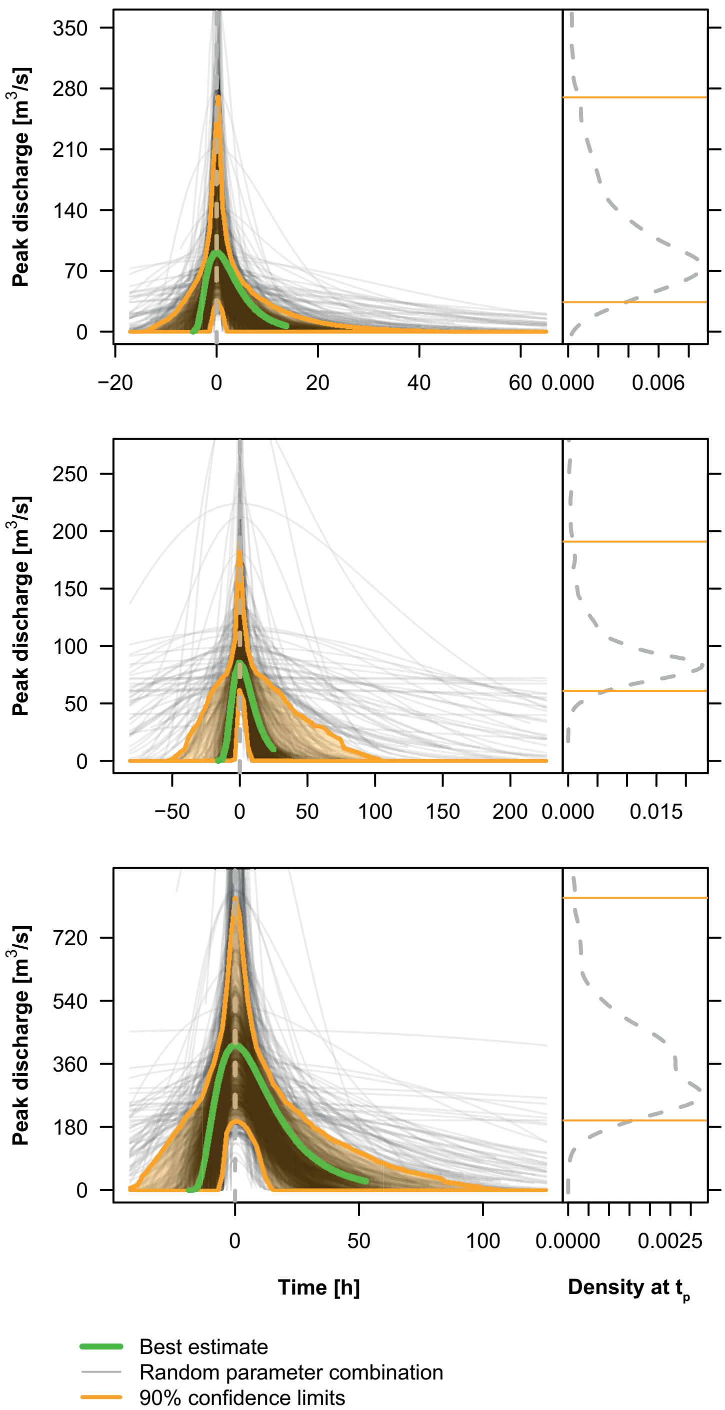

Figure 5. Total uncertainty of SDH construction assessed via the all random strategy for three catchments of different size: Langete-Huttwil (60 km^2), Mentue-Yvonand (105 km^2), and Birs-Münchenstein (911 km^2) (from top to bottom). The best estimate obtained based on observed data (green) is shown together with the 90% confidence limits (orange) and the 1,000 SDHs obtained by randomly picking model choices and parameters (gray). The orange shaded areas represent the 90% confidence regions. The right plot shows the density of peak discharges and the corresponding 90% confidence limits.

4.4. Coupled Uncertainty

The median E_{MAR} of the four catchment characteristics Q_p, V, t_p, and t_{p05} assessed via the coupled uncertainty strategy lay around 50% when looking at the median over all catchments (Figure 6). It was slightly

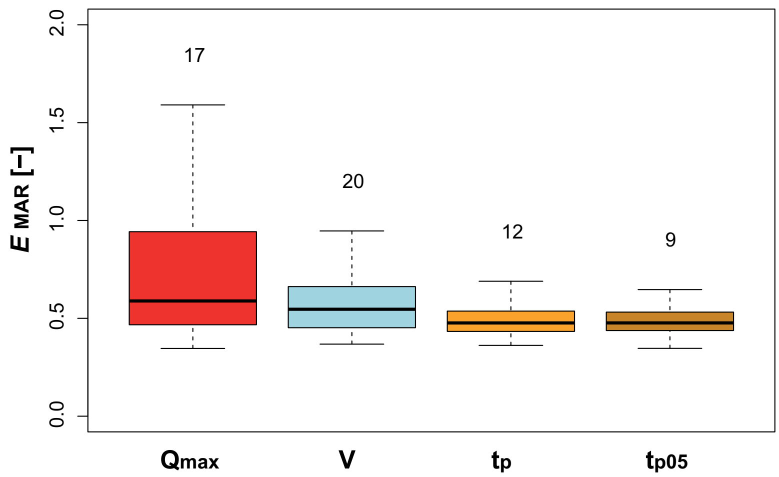

Figure 6. Uncertainty obtained by propagating construction uncertainty through regionalization. Uncertainty is expressed as the median relative errors within catchments of the four hydrograph characteristics: peak discharge (Q_p), hydrograph volume (V), time to peak (t_p), and half-recession time (t_{p05}). The whiskers extend to 1.5 times the interquartile range. Outliers are not displayed, however, their number among the 163 study catchments is given above the upper whisker of the boxplot.

lower for the characteristics related to the shape of the hydrograph than for those related to the magnitude. The E_{MAR} of the characteristics related to the magnitude showed a higher variability across the catchments. This variability was most pronounced for Q_p.

5. Discussion

5.1. Uncertainty Due to Individual Uncertainty Sources

5.1.1. Record Length

We showed that the record length is a major source of uncertainty in design hydrograph construction for all four hydrograph characteristics considered. Serinaldi (2013) stated that many joint distributions and copulas can be fitted to small samples because goodness of fit tests cannot discriminate between alternative models due to the lack of power. An increase in the length of relatively short samples could noticeably reduce the uncertainty of design variable quantiles (Botto et al., 2016) and therefore, as shown in our study, the uncertainty of the whole design hydrograph. The problem of a limited sample size could for practical applications be partially overcome by temporal, spatial, or causal information expansion. Historical floods could be introduced to expand information in time, data from neighboring or similar stations could be used to expand data in space, and runoff processes could be considered for causal information expansion (Blöschl et al., 2013).

5.1.2. Sampling Strategy

Another major uncertainty source for all four studied hydrograph characteristics was the choice of the sampling strategy. It directly influences the sample size and the characteristics of the flood sample. In most catchments, the flood sample chosen by the strategy AM_V was characterized by lower peak discharges than the flood samples determined by the strategies POT_4 and AM_Q. This is the case because a high volume is not always related to a high peak discharge for all floods. Brunner et al. (2017) have shown that the dependence between Q_p and V is rather small for specific short duration flood-types such as flash floods and short-rain floods compared to long-rain floods (Merz & Blöschl, 2003; Sikorska et al., 2015). From a safety point of view, the choice of the strategy POT_4 seems to be judicious since the magnitude of the SDHs is not underestimated compared to the other strategies. The choice of the sampling strategy also influences the choice of a suitable copula. The sampling strategies leading to a small sample size (AM_Q and AM_V) make the choice of a suitable copula inconclusive. Almost all copulas tested were not rejected because of the small power of the goodness of fit test in the case of small sample sizes (Cullen & Frey, 1999; Genest et al., 2009). On the contrary, the choice of a sampling strategy with a larger sample size (POT_4) is quite conclusive since only a few copulas are nonrejected. The choice of the sampling strategy POT_4 therefore seems to be a suitable choice for the estimation of SDHs.

5.1.3. Model Uncertainty

Besides record length and sampling strategy, various other uncertainty sources influenced the four analyzed characteristics of a design hydrograph. The magnitude of the event was largely influenced by the choice of the marginal distributions for peak discharges and hydrograph volumes and their corresponding sampling uncertainty. Still, modelers are faced with the problem of determining one model to be applied to a catchment for a particular modeling task (Ajami et al., 2007; Kuczera et al., 2010; Marshall et al., 2005). Beven and Hall (2014) and Kite (1975) stated that the choice of one out of several models that are not rejected by statistical goodness of fit tests might not be straightforward but does not matter in the range of the observed data. However, the choice of one model over another might matter a lot when extrapolating beyond the range of the data (Hosking & Wallis, 1997). Sampling uncertainty related to the estimation of the parameters of the marginal distributions has been found to be important in univariate (Hailegeorgis & Alfredsen, 2017; Lamb & Kay, 2004) and in bivariate quantile estimation (Dung et al., 2015; Serinaldi, 2013). Unlike Serinaldi (2009), we did not find that model uncertainty is smaller than sampling uncertainty. The relative contribution of error due to model choice and sampling error most likely depends again on the record length. For large samples, the standard deviation of the estimate becomes small in comparison to the bias caused by the wrong distribution choice (Strupczewski et al., 2002).

The magnitude of the flood event is only slightly affected by the choice of one of the nonrejected copulas. This is in accordance with findings by Xu et al. (2010) who found that the uncertainty originating from method selection was largest and the uncertainty caused by ignoring the dependence was smallest. However, this might be different if one is interested in estimates for higher return periods, where the ability of a copula to correctly describe the tail dependence in the data might be crucial (Poulin et al., 2007). Our results highlight the importance of a sufficiently large sample as obtained by POT_4 compared to the annual maxima sampling strategies. The choice of a larger sample narrows down the number of admissible copulas. It prevents from having to choose a dependence structure based on a goodness of fit test with limited power when several probabilistic models cannot be rejected despite their poor fit to the data with the limited size of an annual maxima sample (Cullen & Frey, 1999). The uncertainty related to the choice of a design event on the isoline of equally likely events is also rather small compared to the other uncertainty sources considered. This is important since this uncertainty source depends on model selection and not on the sample size and is therefore irreducible.

Contrary to the magnitude, the characteristics related to the hydrograph shape were more influenced by the definition of the representative normalized hydrograph and the choice of a probability density function to model the shape of the hydrograph. The importance of the different uncertainty sources differed for the four hydrograph characteristics. Consequently, it is important to consider both uncertainty sources related to the magnitude of the event and sources related to the shape of the event when assessing the uncertainty of a design hydrograph.

5.1.4. Regionalization Method

The relative importance of model and sampling uncertainty due to regionalization differed for the hydrograph characteristics related to the flood magnitude (Q_p and V) and those related to the hydrograph shape (t_p and t_{p05}). Model uncertainty was higher than sampling uncertainty for most hydrograph characteristics. However, sampling uncertainty seemed to be quite important in the regionalization of the hydrograph shape.

5.2. Total Uncertainty

The total uncertainty of constructed hydrographs was higher than each of the individual uncertainty sources considered. However, it was lower than the sum of all the individual uncertainty sources. In the case of regionalization, the spread of the E_{MAR}s referring to model uncertainty were comparable to those referring to total uncertainty. This implies that model uncertainty alone introduces a lot of uncertainty to SDH regionalization and the effect of sampling uncertainty is negligible. The nonadditive property of individual uncertainty sources highlights the need of considering all sources jointly when the total uncertainty is of interest. A simple adding of individual sources would result in highly overestimated and unrealistic uncertainties. The assessment of individual sources may be thus only used to compare their relative contributions and to indicate which uncertainty part could be reduced with the highest benefit.

5.3. Coupled Uncertainty

The coupled SDH uncertainty considering construction and regionalization uncertainty amounted to a bit more than 50% depending on the hydrograph characteristic considered. It was slightly higher for the hydrograph characteristics related to the magnitude of the event when looking at a return period of $T = 100$ years. For lower return periods, it would most likely be the other way around since it would be in the interpolation range of the bivariate extreme value distribution rather than the extrapolation range. The higher uncertainty for magnitude-related hydrograph characteristics compared well with the total uncertainties obtained for the regionalization. On the contrary, the total uncertainty of the constructed SDHs was higher for the hydrograph characteristics related to the hydrograph shape than for those related to the hydrograph magnitude. The coupled uncertainty was slightly higher than the total uncertainty in gauged catchments. This suggests that already SDH construction contributes a big part of the uncertainty and regionalization just adds a little uncertainty. Yet, these two uncertainty sources are not necessarily additive. The coupled uncertainty was lower than the sum of the total construction and the total regionalization uncertainty which implies that these have to be considered jointly, as it was done here, to get realistic uncertainty bands.

5.4. Value of the Proposed Uncertainty Framework

The three level uncertainty framework proposed here allows the quantification of the uncertainty of synthetic design hydrographs as they would be provided to engineers or practitioners. It is therefore the first step toward communicating the uncertainty of design estimates to practitioners. SDHs complemented with uncertainty bands allow more informed decisions than using SDHs without uncertainty information (Hall, 2003). These simulation results obtained could be provided to the practitioner as a set of values comprising the best estimate, the first and third quantile of the simulation results and the maximum simulated SDH. Depending on the question of interest, the practitioner could choose the value to work with. If safety is crucial, a more conservative estimate could be used. If cost-efficiency is crucial, a smaller estimate could be chosen. Ideally, an ensemble-based approach could be adopted. An uncertainty assessment based on three levels of complexity further points out which individual uncertainty sources and which steps in the analysis offer the largest potential for improvement. This information can be used to decide which avenues to take for future research or in depth analysis. Our results suggest that uncertainty could be successively reduced if more data became available or by using alternative data.

We have also shown that the choice of a flood sampling strategy is quite an important source of uncertainty. We showed that using an annual maxima sampling strategy might neglect important information. Singh et al. (2005) also found that floods within a year may not be adequately represented by their maximum and suggested the use of two-independent subpopulations: snowfall and rainfall generated floods. Such a conscious selection of more than one flood event per year might be an alternative to the selection of flood events using a peak-over-threshold approach without considering flood generating mechanisms. A subdivision into floods generated by different processes, as proposed by Sikorska et al. (2015) and Merz and Blöschl (2003) would further allow for a representation of different phenomena by different statistical distributions (Shu & Ouarda, 2008). Working with subpopulations would also allow the representation of different hydrograph shapes within a catchment. This would give a better impression of hydrograph shape variability within a catchment due to different processes and to different rainfall inputs than working with only one catchment specific hydrograph shape. The construction of flood type specific hydrographs as proposed by Brunner et al. (2017) goes into this direction.

5.5. Limitations and Perspectives

The uncertainty assessment framework proposed here is generally applicable to any return period of interest. However, several choices need to be made dependent on the data set: the choice of a set of distribution functions to fit the peak discharges and hydrograph volumes, the choice of a set of suitable copula models to model the dependence between these two variables, and the choice of a set of PDFs to model the hydrograph shapes to be sampled from the simulations. Conducting a simulation study on several catchments is computationally quite expensive, however, such an analysis is usually only done once.

The uncertainty analysis presented here considered most uncertainty sources affecting synthetic design hydrographs. One possibly important uncertainty source was not considered here: observational uncertainty due to measurement errors and due to the conversion of water level to discharge via a rating curve approach. The findings on the importance of this source of uncertainty differ quite a bit between various authors. Apel et al. (2004) showed that the stage-discharge relationships are relatively less important in the uncertainty analysis while Kundzewicz (2002) claimed that uncertainty resulting from the need to extrapolate the rating curve to extreme values, where no direct runoff measurements exist, was considerable. We did not consider this uncertainty source because the uncertainty of the rating curve was not known for the whole set of catchments. Therefore, it could not be incorporated into the uncertainty assessment framework proposed here. Still, rating curve uncertainty might be substantial especially when looking at extreme events (Di Baldassarre & Montanari, 2009) as it will propagate through the whole SDH construction process. In addition, the uncertainty due to the rating curve is catchment dependent and information on this uncertainty is often limited and case-specific (Di Baldassarre & Montanari, 2009; Sikorska et al., 2013), which makes the generalization of the rating curve uncertainty difficult. The effect of the rating curve uncertainty could therefore be potentially assessed only for a few catchments in our data set where enough information on the rating curve and its uncertainty is available.

6. Conclusions

Synthetic design hydrographs are inherently uncertain. They are affected by various uncertainty sources comprising a limited record length, model, and sampling uncertainties. This uncertainty needs to be assessed and communicated to the practitioner to ensure reliable flood estimates. In this work, we proposed a framework for assessing the uncertainty of such synthetic design hydrographs which consists of three levels of complexity using an extensive set of 163 Swiss catchments. First, we assessed the influence of individual sources of uncertainty on the estimation of synthetic design hydrographs. Second, we quantified the total uncertainty of constructed and regionalized synthetic design hydrographs considered separately. Third, we quantified the coupled uncertainty of a regionalized hydrograph when considering that already constructed hydrographs are uncertain. We identified the record length and the choice of the sampling strategy as having the strongest effect on the uncertainty of a synthetic design hydrograph characterized by peak discharge, hydrograph volume, and hydrograph shape. The magnitude of the design hydrograph was further strongly affected by the choice of the marginal distributions for peak discharges and hydrograph volumes. The shape of the hydrograph, however, was more affected by the definition of a representative hydrograph shape, the choice of the probability density function used to model this shape, and base flow separation. The total uncertainty of design hydrographs constructed based on observed runoff data differed between the catchments analyzed and had a median of roughly 25% for Q_p, 35% for V, 45% for t_p and t_{p05}. The median E_{MAR}s for hydrographs obtained by regionalization assuming that the data used for regionalization (constructed hydrographs) is known lie around 40% for Q_p and V, around 25% for t_p and around 30% for t_{p05}. The coupled uncertainty of synthetic design hydrographs for ungauged catchments lay around 50% but differed quite a bit between catchments especially in terms of peak discharges. We also demonstrated that the uncertainty framework provides insights into promising avenues for future research. The uncertainty of synthetic design hydrographs could be most effectively reduced by enlarging the sample size by successively collecting more data or by considering alternative information such as historical flood data or regional data. Alternatively, we could aim at more adequately describing the variability of hydrograph types. Flood estimates complemented with uncertainty bands computed via the uncertainty assessment framework proposed in this study should allow the engineer to make profound decisions based on a cost-benefit analysis.

Acknowledgments
We thank the Federal Office for the Environment (FOEN) for funding the project (contract 13.0028.KP/M285–0623) and for providing runoff measurement data. We also thank MeteoSwiss for providing precipitation data. The data used in this study is available upon order from the FOEN and MeteoSwiss. For the hydrological data of the federal stations, the order form under http://www.bafu.admin.ch/wasser/13462/13494/15076/index.html. can be used. The hydrological data of the cantonal stations can be ordered from the respective cantons. The meteorological data can be ordered via https://shop.meteoswiss.ch/index.html. We thank the three anonymous reviewers for their detailed and constructive feedback which helped us to improve the previous version of the manuscript.

References

Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, *43*, W01403. https://doi.org/10.1029/2005WR004745

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Science*, *4*(2), 295–308. https://doi.org/10.5194/nhess-4-295-2004

Beven, K., & Hall, J. (2014). *Applied uncertainty analysis for flood risk management* (672 p.). London: Imperial College Press.

Beven, K., Leedal, D., & Alcock, R. (2010). Uncertainty and good practice in hydrological prediction. *VATTEN*, *66*(3–4), 159–163.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff prediction in ungauged basins* (465 p.). Cambridge, UK: Cambridge University Press.

Botto, A., Ganora, D., Claps, P., & Laio, F. (2016). Technical note: Design flood under hydrological uncertainty. *Hydrology and Earth System Sciences Discussions*, *21*, 3353–3358. https://doi.org/10.5194/hess-2016-637

Botto, A., Ganora, D., Laio, F., & Claps, P. (2014). Uncertainty compliant design flood estimation. *Water Resources Research*, *50*, 4242–4253. https://doi.org/10.1002/2013WR014981

Brunner, M. I., Furrer, R., Sikorska, A., Viviroli, E., Seibert, D. J., & Favre, A.-C. (2018). Synthetic design hydrographs for ungauged catchments: A comparison of regionalization methods. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-018-1523-3

Brunner, M. I., Seibert, J., & Favre, A.-C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. *Wire's Water*, *3*, 819–833. https://doi.org/10.1002/wat2.1173

Brunner, M. I., Viviroli, D., Sikorska, A. E., Vannier, O., Favre, A.-C., & Seibert, J. (2017). Flood type specific construction of synthetic design hydrographs. *Water Resources Research*, *53*, 1390–1406. https://doi.org/10.1002/2016WR019535

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. https://doi.org/10.1623/hysj.48.1.25.43485

Camezind-Wildi, R. (2005). *Empfehlung Raumplanung und Naturgefahren* (technical reportArt.-Nr.: 812.046.d). Bern: Bundesamt für Raumentwicklung, Bundesamt für Wasser und Geologie, Bundesamt für Umwelt, Wald und Landschaft.

Chang, C. H., Tung, Y. K., & Yang, J. C. (1994). Monte Carlo simulation for correlated variables with marginal distributions. *Journal of Hydraulic Engineering*, *120*(3), 313–331. https://doi.org/10.1061/(ASCE)0733-9429(1994)120:3(313)

Chernobai, A., Rachev, S. T., & Fabozzi, F. J. (2015). Composite goodness-of-fit tests for left-truncated loss samples. In C.-F. Lee & J. Lee (Eds.), *Handbook of financial econometrics and statistics* (Chapter 20, pp. 575–596). New York: Springer.

Chowdhury, J. U., & Stedinger, J. R. (1991). Confidence interval for design floods with estimated skew coefficient. *Journal of Hydraulic Engineering*, *117*(7), 811–831. https://doi.org/10.1061/(ASCE)0733-9429(1991)117:7(811)

Collischonn, W., & Fan, F. M. (2013). Defining parameters for Eckhardt's digital baseflow filter. *Hydrological Processes*, *27*, 2614–2622.

Cuevas, A., Febrero, M., & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, *22*(3), 481–496. https://doi.org/10.1007/s00180-007-0053-0

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment* (335 p.). New York: Society for Risk Analysis.

Deutsche Vereinigung für Wasserwirtschaft Abwasser und Abfall (2012). *Merkblatt DWA-M 552*. Hennef, Germany: DWA.

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*, 913–921. https://doi.org/10.5194/hessd-6-39-2009

Dung, N. V., Merz, B., Bárdossy, A., & Apel, H. (2015). Handling uncertainty in bivariate quantile estimation - An application to flood hazard analysis in the Mekong Delta. *Journal of Hydrology*, *527*, 704–717. https://doi.org/10.1016/j.jhydrol.2015.05.033

DVWK (1999). Statistische Analyse von Hochwasserabflüssen (technical report). Bonn: Deutscher Verband für Wasserwirtschaft.

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrological Processes*, *19*, 507–515. https://doi.org/10.1002/hyp.5675

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap* (436 p.). Dordrecht, Netherlands: Chapman & Hall/CRC.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, *12*(4), 347–367. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, *44*, 199–213. https://doi.org/10.1016/j.insmatheco.2007.10.005

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies*, *9*, 104–126. https://doi.org/10.1016/j.ejrh.2016.11.004

Hall, J., & Solomatine, D. (2008). A framework for uncertainty analysis in flood risk management decisions. *International Journal of River Basin Management*, *6*(2), 85–98. https://doi.org/10.1080/15715124.2008.9635339

Hall, J. W. (2003). Handling uncertainty in the hydroinformatic process. *Journal of Hydroinformatics*, *5*(4), 215–232.

Hall, M. J., van den Boogaard, H. F. P., Fernando, R. C., & Mynett, A. E. (2004). The construction of confidence intervals for frequency analysis using resampling techniques. *Hydrology and Earth System Sciences*, *8*(2), 235-246. https://doi.org/10.5194/hess-8-235-2004

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning*, Springer series in statistics (745 p.). Stanford, CA: Springer.

Held, L., & Sabanés Bové, D. (2014). *Applied statistical inference. Likelihood and Bayes* (376 p.). Berlin: Springer.

Hofner, B., Mayr, A., Robinzonov, N., & Schmid, M. (2014). Model-based boosting in R-A hands-on tutorial using the R Package mboost. *Computational Statistics*, *29*, 3–35. https://doi.org/10.1007/s00180-012-0382-5

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis* (238 p.). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511529443

Joe, H. (2015). *Dependence modeling with copulas* (479 p.), Boca Raton, FL: CRC Press.

Juston, J. M., Kauffeldt, A., Montano, B. Q., Seibert, J., Beven, K. J., & Westerberg, I. K. (2013). Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling. *Hydrological Processes*, *27*(7), 1117–1122. https://doi.org/10.1002/hyp.9625

Kidson, R., & Richards, K. S. (2005). Flood frequency analysis: Assumptions and alternatives. *Progress in Physical Geography*, *29*(3), 392–410. https://doi.org/10.1191/0309133305pp454ra

Kite, G. W. (1975). Confidence limits for design events. *Water Resources Research*, *11*(1), 48–53. https://doi.org/10.1029/WR011i001p00048

Klein, B., Pahlow, M., Hundecha, Y., & Schumann, A. (2010). Probability analysis of hydrological loads for the design of flood control systems using copulas. *Journal of Hydrologic Engineering*, *15*(5), https://doi.org/10.1061/(ASCE)HE.1943-5584.0000204

Koutsoyiannis, D. (2014). Reconciling hydrology with engineering. *Hydrology Research*, *45*, 2–22. https://doi.org/10.2166/nh.2013.092

Kuczera, G., Renard, B., Thyer, M., & Kavetski, D. (2010). There are no hydrological monsters, just models and observations with large uncertainties!. *Hydrological Sciences Journal*, *55*(6), 980–991. https://doi.org/10.1080/02626667.2010.504677

Kundzewicz, Z. W. (2002). Floods in the context of climate change and variability. In M. Beniston (Ed.), *Climatic change: Implications for the hydrological cycle and for water management* (pp. 225–247). Netherlands: Kluwer Academic Publishers.

Kysely, J. (2008). A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *Journal of Applied Meteorology and Climatology*, *47*, 3236–3251. https://doi.org/10.1175/2008JAMC1763.1

Lamb, R., & Kay, A. L. (2004). Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resources Research*, *40*, W07501. https://doi.org/10.1029/2003WR002428

Lang, M., Ouarda, T., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*, 103–117.

López-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, *104*, 718–734. https://doi.org/10.1198/jasa.2009.0108

Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1: At-site modeling. *Water Resources Research*, *33*(4), 747–757.

Marshall, L., Nott, D., & Sharma, A. (2005). Hydrological model selection: A Bayesian alternative. *Water Resources Research*, *41*, W10422. https://doi.org/10.1029/2004WR003719

Martins, E. S., & Stedinger, J. R. (2001a). Generalized Maximum Likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research*, *37*(10), 2551–2557. https://doi.org/10.1029/2001WR000367

Martins, E. S., & Stedinger, J. R. (2001b). Historical information in a generalized Maximum Likelihood Framework with partial duration and annual maximum series. *Water Resources Research*, *37*(10), 2559–2567. https://doi.org/10.1029/2000WR000009

McMillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, *29*(7), 1873–1882. https://doi.org/10.1002/hyp.10419

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resource Research*, *39*(12), 1340. https://doi.org/10.1029/2002WR001952

Merz, B., Kreibich, H., & Apel, H. (2008). Flood risk analysis: Uncertainties and validation. *Österreichische Wasser- Und Abfallwirtschaft*, *60*(5–6), 89–94. https://doi.org/10.1007/s00506-008-0001-4

Merz, B., & Thieken, A. H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, *309*(1–4), 114–132. https://doi.org/10.1016/j.jhydrol.2004.11.015

Meylan, P., Favre, A.-C., & Musy, A. (2012). *Predictive hydrology. A frequency analysis approach* (212 p.). St. Helier, Jersey, British Channel Islands: Science Publishers.

Montanari, A., & Koutsoyiannis, D. (2012). A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, *48*, W09555. https://doi.org/10.1029/2011WR011412

Nadarajah, S. (2007). Probability models for unit hydrograph derivation. *Journal of Hydrology*, *344*, 185–189. https://doi.org/10.1016/j.jhydrol.2007.07.004

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrologic uncertainty. *Hydrological Sciences Journal*, *61*, 1666–1678. https://doi.org/10.1080/02626667.2016.1183009

Pappenberger, F., & Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research*, *42*, W05302. https://doi.org/10.1029/2005WR004820

Pilgrim, D. H. (1986). Bridging the gap between flood research and design practice. *Water Resources Research*, *22*(9), 165–176.

Poulin, A., Huard, D., Favre, A.-C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, *12*(4), 394–403. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(394)

Qi, W., Zhang, C., Fu, G., & Zhou, H. (2016). Imprecise probabilistic estimation of design floods with epistemic uncertainties. *Water Resources Research*, *52*, 4823–4844. https://doi.org/10.1002/2015WR017663

Rai, R. K., Sarkar, S., & Singh, V. P. (2009). Evaluation of the adequacy of statistical distribution functions for deriving unit hydrograph. *Water Resources Management*, *23*, 899–929. https://doi.org/10.1007/s11269-008-9306-0

Requena, A. I., Mediero, L., & Garrote, L. (2013). A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrological Earth System Sciences*, *17*, 3023–3038. https://doi.org/10.5194/hess-17-3023-2013

Schumann, A. H., Nijssen, D., & Pahlow, M. (2010). Handling uncertainties of hydrological loads in flood retention planning. *International Journal of River Basin Management*, *8*(3–4), 281–294. https://doi.org/10.1080/15715124.2010.512561

Serinaldi, F. (2009). Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles. *Journal of Hydrology*, *376*, 528–541. https://doi.org/10.1016/j.jhydrol.2009.07.065

Serinaldi, F. (2013). An uncertain journey around the tails of multivariate hydrological distributions. *Water Resources Research*, *49*, 6527–6547. https://doi.org/10.1002/wrcr.20531

Serinaldi, F. (2015). Dismissing return periods!. *Stochastic Environmental Research and Risk Assessment*, *29*, 1179–1189. https://doi.org/10.1007/s00477-014-0916-1

Shu, C., & Ouarda, T. (2008). Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology*, *349*, 31–43. https://doi.org/10.1016/j.jhydrol.2007.10.050

Sikorska, A., & Renard, B. (2017). Calibrating a hydrological model in stage space to account for rating curve uncertainties: General framework and key challenges. *Advances in Water Resources*, *105*, 51–66. https://doi.org/10.1016/j.advwatres.2017.04.011

Sikorska, A. E., Scheidegger, A., Banasik, K., & Rieckermann, J. (2012). Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. *Hydrology and Earth System Sciences*, *16*(4), 1221–1236. https://doi.org/10.5194/hess-16-1221-2012

Sikorska, A. E., Scheidegger, A., Banasik, K., & Rieckermann, J. (2013). Considering rating curve uncertainty in water level predictions. *Hydrology and Earth System Sciences*, *17*, 4415–4427. https://doi.org/10.5194/hess-17-4415-2013

Sikorska, A. E., Viviroli, D., & Seibert, J. (2015). Flood type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, *51*, 7959–7976. https://doi.org/10.1002/2015WR017326

Singh, V. P., Wang, S. X., & Zhang, L. (2005). Frequency analysis of nonidentically distributed hydrologic flood data. *Journal of Hydrology*, *307*(1–4), 175–195. https://doi.org/10.1016/j.jhydrol.2004.10.029

Strupczewski, W., Singh, V., & Węglarczyk, S. (2002). Asymptotic bias of estimation methods caused by the assumption of false probability distribution. *Journal of Hydrology*, *258*, 122–148.

Sun, H., Jiang, T., Jing, C., Su, B., & Wang, G. (2017). Uncertainty analysis of hydrological return period estimation, taking the upper Yangtze River as an example. *Hydrology and Earth System Sciences Discussions*, 1–26. https://doi.org/10.5194/hess-2016-566

Tisseuil, C., Vrac, M., Lek, S., & Wade, A. J. (2010). Statistical downscaling of river flows. *Journal of Hydrology*, *385*(1–4), 279–291. https://doi.org/10.1016/j.jhydrol.2010.02.030

Tung, Y.-K., & Yen, B.-C. (2005). *Hydrosystems engineering uncertainty analysis* (285 p.). New York: McGraw-Hill Book Company.

Wasserman, L. (2006). *All of nonparametric statistics* (269 p.). New York: Springer.

Xu, Y.-P., Booij, M. J., & Tong, Y.-B. (2010). Uncertainty analysis in statistical modeling of extreme hydrological events. *Stochastic Environmental Research and Risk Assessment*, *24*, 567–578. https://doi.org/10.1007/s00477-009-0337-8

Yue, S., Ouarda, T., Bobée, B., Legendre, P., & Bruneau, P. (2002). Approach for describing statistical properties of flood hydrograph. *Journal of Hydrologic Engineering*, *7*(2), 147–153. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:2(147)