Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Calibration of hydrological models for ecologically-relevant streamflow predictions: a trade-off between performance and consistency

Thibault Hallouin[1], Michael Bruen[1], and Fiachra E. O'Loughlin[1]

[1]UCD Dooge Centre for Water Resources Research, University College Dublin, Ireland.

Correspondence: Thibault Hallouin (th.thibault.hallouin@gmail.com)

Abstract. The ecological integrity of freshwater ecosystems is intimately linked to natural fluctuations in the river flow regime. Anthropogenic alterations in flow regimes threaten water security and freshwater biodiversity in many regions of the world. The impacts of climate change on the hydrological cycle change local flow regimes and thus impact on the ecological systems. In catchments with little human-induced hydro-morphological changes, existing hydrological models can be used to predict changes in local flow regime in order to assess whether its rivers remain a suitable living environment for endemic species. However, hydrological models are traditionally calibrated to give a good general fit between observed and simulated hydrographs, e.g., using an optimisation with an objective function such as the Nash-Sutcliffe, or the Kling-Gupta efficiencies. Much ecological research has shown that aquatic species respond to very specific characteristics of the hydrograph, whether magnitude, frequency, duration, timing, and rate of change of flow events. Since each community in a river may be particularly sensitive to a few very specific streamflow characteristics, alternative hydrological model calibration strategies are needed, focussing on good performance for those specific characteristics. This study investigates the performance of a set of specially developed, bespoke, objective functions made of combinations of specific streamflow characteristics relevant for fish and invertebrate communities. These are compared with the more traditional objective functions on a set of 33 Irish catchments with little human regulation. A split-sample test with a rolling-window procedure is applied to reduce the influence of variations between the calibration/evaluation periods on the conclusions. These bespoke objective functions are shown to be better suited to predict the targetted streamflow characteristics in terms of performance in evaluation; however, traditional objective functions yield more consistent behavioural parameter sets, indicating a trade-off between model performance and model consistency when predicting streamflow characteristics, especially when the number of target streamflow characteristics are low.

1 Introduction

River flow is the cornerstone of freshwater ecosystems, the ecological integrity of which relies on the natural fluctuations in the river flow regime (Poff et al., 1997). A long history of human alterations of river flow regime for water supply, irrigation, flood protection, or hydropower threatens water security and freshwater biodiversity in many regions of the world (Vörösmarty et al., 2010). Richter et al. (1997) raised the overarching research question "How much water does a river need?". In order to quantify these needs and assess the effects of altered flow regime on freshwater ecology, many different hydrological indices

Hydrology and Earth System Sciences
Open Access
Discussions

have been used, whether they are referred to as streamflow characteristics (SFC) (Vis et al., 2015; Pool et al., 2017), ecologically relevant flow statistics (ERFS) (Caldwell et al., 2015), or indicators of hydrological alterations (IHA) (Richter et al., 1996). Each index describes specific aspects of the response of a catchment that can be extracted from the streamflow hydrograph. Olden and Poff (2003) listed a range of such hydrological indices used to characterise river flow regime in relation to ecological species' preferences. The prediction of these streamflow conditions has historically being done using statistical analyses such as regional regression models that relate them to some climatic and physiographic descriptors (e.g., Carlisle et al., 2011; Knight et al., 2014) as well as by hydrological models that produce streamflow hydrographs from which the streamflow characteristics can be computed (e.g., Shrestha et al., 2014; Caldwell et al., 2015).

Most rainfall-runoff models used to predict these SFCs relevant for the stream ecology require calibration to determine their effective model parameter values. The selection of the objective function(s) in the calibration process is of great importance for the quality of the predictions of SFCs (Vis et al., 2015; Kiesel et al., 2017; Pool et al., 2017). For example, Vis et al. (2015) found that even parameter sets performing similarly on the Nash-Sutcliffe efficiency criterion (NSE) (Nash and Sutcliffe, 1970) fitted to flows can yield very different performances when looking at the prediction of SFCs. They found that certain combinations of objective functions fitted to flows each focussing on specific aspects of the streamflow hydrograph tend to be more suitable across a range of streamflow characteristics than using a single objective function fitted to flows, e.g., the NSE criterion alone. Although there does not seem to be any single overall best combination of objective functions fitted to flows for predicting all SFCs at once (Vis et al., 2015). This exposes the limitations of models in representing the entirety of real-world processes in a catchment. Indeed, because of uncertainties in model structure, model forcing, and evaluation data (Beven, 2016), the identification of a single perfect parameter set is usually unachievable (Beven, 2006), and in practice trade-offs are required between modelling different aspects of the hydrograph. The choice of the objective function for model calibration directly influences which trade-offs are made.

The calibration of a rainfall-runoff model using the Nash-Sutcliffe efficiency is known to give higher importance to fitting flow peaks because of its quadratic formulation, and this is reflected in the generally better performance of a model calibrated on such a criterion to predict streamflow characteristics for high flow conditions (Shrestha et al., 2014). Composite objective functions such as the Kling-Gupta efficiency (KGE) are now often preferred, since KGE explicitly considers linear correlation, bias, and variability in a balanced or customisable way (Gupta et al., 2009; Kling et al., 2012). Nonetheless, the quadratic formulation on flows remains in the linear correlation component of KGE. In order to improve the model calibration for the predictions of different ranges of flows, the whole or parts of the flow duration curve have also been found useful in improving the model simulation of the whole hydrograph (e.g., Westerberg et al., 2011; Pfannerstill et al., 2014). However, the flow duration curve does not embed any information about the timing or duration of flow events, which can be essential for ecological species (Arthington et al., 2006). In this situation, a pragmatic approach is to directly use an objective function fitted to the target SFC(s) in the expectation that this will improve predictions of these same SFCs. Such an approach was recently explored by Kiesel et al. (2017) and by Pool et al. (2017) using different combinations of SFCs in different regions (relevant for benthic stream invertebrates in Germany, and for fish communities in the Southeastern United States) and using different hydrological models (SWAT-3S, and HBV-light, respectively). These studies found that a combination of multiple

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

SFCs is generally equivalent or better than a single traditional objective function fitted to flows (KGE, and NSE, respectively) when predicting the chosen SFCs. These results highlight the importance of the choice of the objective function(s) to predict a bundle of SFCs for ecological applications.

Hydrological models are generally less accurate than regional regression models in predicting particular SFCs because separate regression models can be purposely developed for each target SFC individually (Murphy et al., 2013). Similar behaviour has been found for calibrated rainfall-runoff models, where specific calibration focussed on the target SFC is the best performing calibration option for that SFC (Kiesel et al., 2017; Pool et al., 2017). However, when calibrating on a specific SFC, while the model's ability to predict that indicator may improve, the physical representation of the catchment's overall behaviour, captured in the effective parameter values of the model could be compromised, preventing the use of the model for predicting other indicators. For instance, Pool et al. (2017) found that using a combination of SFCs as an objective function does not perform as well as the Nash-Sutcliffe efficiency fitted to flows to predict SFCs not included in the combination. Poff and Zimmerman (2010) showed that each stream species is sensitive to its own combination of SFCs relating to its own preferences for living conditions, which constitutes the ecological flow regime (Knight et al., 2012). When several species are considered simultaneously, the number of SFCs to predict will increase accordingly. In this context, there is a case for traditional objective functions to remain a more parsimonious strategy to guarantee that all the target aspects of the hydrograph for the entire assemblage of species in the river are predicted accurately.

Given the previous research efforts in the field of ecologically-relevant hydrological predictions, specific combinations of the relevant SFCs are strong contenders as calibration criteria when focussing on one ecological community. However, it remains unclear which objective function would be more suited to predict preferences for several communities at once, implying a larger set of target SFCs. Indeed, increasing the number of SFCs exacerbates the number of trade-offs to be made to find suitable parameter sets, which raises questions over model equifinality issues (Beven, 2006). It can be hypothesised that there is an upper limit on the number of SFCs that can be used while maintaining a parsimonious measure of goodness of fit for the calibration of the model. In such situations, a traditional objective function could remain a better compromise. The objective of this study is to explore the capabilities of different objective functions, both traditional and bespoke definitions, to calibrate a rainfall-runoff model intended to simulate ecologically-relevant SFCs. To do so, a fourteen split-sample test using the conceptual SMART rainfall-runoff model (Mockler et al., 2016) for 33 Irish catchments is undertaken. Of particular interest are the model performance stability across split-samples, and the consistency of the behavioural parameter sets identified by the objective functions. The three hypotheses challenged in this study are:

H1: Using a target vector of SFCs as objective function is a better strategy for simulating those SFCs than using a traditional objective function no matter the combination of SFCs targetted.

H2: The use of traditional objective functions for calibration produces more stable predictions across study periods that when using a small number of SFCs as objective function.

H3: Traditional objective functions are more efficient than a combination of SFCs as objective function for the identification of consistent behavioural parameter sets across study periods.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

2 Data and model

2.1 Streamflow characteristics

The streamflow characteristics selected for this study have previously been identified as representative of the habitat preferences of fish communities in the Southeastern US (Knight et al., 2014; Pool et al., 2017), and of invertebrate communities in Germany (Kakouei et al., 2017; Kiesel et al., 2017). The use of the same set of indices allows for straightforward comparisons with the two previous studies on the choice of the objective function for model calibration for ecological applications in a different region of the world with a different model.

The indices are listed and detailed in Table 2. Only two hydrological indices are common between the two communities' respective streamflow preferences. Each community's preferences feature all aspects of the flow regime, i.e., magnitude, frequency, duration, timing, and rate of change, for high flow, average flow, and/or low flow events. Except for Q85 that is directly derived from the flow duration curve, all streamflow characteristics are defined in Olden and Poff (2003) and their calculation is based on computations in the R-package EflowStats (Henriksen et al., 2006; Archfield et al., 2014). However, all computations for this study were carried out in Python where the calculations were vectorised to allow for reasonable computation time over large parameter sample sets (Hallouin, 2019a).

2.2 Study catchments

This study used discharge records with a minimum of 14 hydrological years with complete daily discharge data in the period from the 1st of October 1986 to the 30th of September 2016. If any daily value was missing, the hydrological year was discarded. The calculation of some streamflow characteristics requires a strictly continuous daily streamflow time series and it can be difficult to find time series with no missing discharge measurement at all. The length of 14 years was set as the minimum requirement in order to have at least seven years for calibration and seven years for evaluation for each catchment. The data availability for the gauges meeting these requirements is presented on Figure 2. In most catchments, these 14 hydrological years were not necessarily consecutive. The daily discharge data used in this study is provided by the Office of Public Works (2019), and Ireland's Environmental Protection Agency (2019).

Catchment selection was also influenced by the quality of the discharge data, including the quality of the rating curve at the gauge as determined by Webster et al. (2017). Heavily regulated rivers were discarded. A total of thirty-three catchments featured sufficient data of good quality to be used in this study, some of which are nested catchments Figure 3. The selected catchments cover 26 % of the Republic of Ireland. They are located throughout the country, hence representing a good sample of the diversity of Irish soils and geology. Their average annual rainfall ranges from 916 to 1660 $\mathrm{mm\,yr^{-1}}$, and the average annual potential evapotranspiration varies from 497 to 578 $\mathrm{mm\,yr^{-1}}$. The size of the catchments varies from 25 to 2462 km^2, while their average elevation ranges from 5 to 910 m above sea level.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

2.3 Rainfall-runoff model

The Soil Moisture Accounting and Routing for Transport (SMART) model used here is an enhancement of the SMARG lumped, conceptual, rainfall-runoff model (Soil Moisture Accounting and Routing with Groundwater) developed in University College Galway (Kachroo, 1992) and based on the soil layers concept (O'Connell et al., 1970; Nash and Sutcliffe, 1970). Separate soil layers were introduced to capture the decline with soil depth in ability of plant roots to extract water for evapotranspiration. SMARG was originally developed for flow modelling and forecasting and was incorporated into the Galway Real-Time River Flow Forecasting System (GFFS) (Goswami et al., 2005). The SMART model reorganised and extended SMARG to provide a basis for water quality modelling by separating explicitly the important flow pathways in a catchment, needed for an EPA funded project "Pathways", and it has been successfully fitted to over 30 % of Irish catchments (Mockler et al., 2016).

The routing component distinguishes between five runoff pathways: overland flow, drain flow, interflow, shallow groundwater flow, and deep groundwater flow (Figure 1). It usually runs at an hourly or daily time-step, requires inputs of measures of precipitation and estimates of potential evapotranspiration, and produces estimates of discharge from the catchment. It normally has ten free parameters (Table 1). During energy-limited periods, the model first estimates effective or excess rainfall by applying a scaling correction T and subtracting any direct evaporation. A threshold parameter H determines how much (if any) of this becomes direct surface runoff through the Horton (infiltration excess) mechanism. Any surplus rainfall is assumed to infiltrate the top layer of the soil. The soil is modelled as six layers with a total soil moisture capacity of Z. As the capacity of a layer is exceeded, surplus moisture moves to a deeper layer if it has capacity or else is intercepted by drains or moves to the shallow or deep groundwater stores. In water-limited periods, the model attempts to meet the evapotranspiration demand by supplying moisture from the soil layers, starting from the top layer but when this is dry from lower layers but with an increasing difficulty expressed by a parameter C. Each of the above pathways is modelled as a single linear reservoir, each with its own parameter (SK for overland and drain flow, FK for interflow, GK for shallow and deep groundwater flow). The outputs of all of these are routed through a single linear reservoir representing river routing (RK). Note, a detailed description of the conceptual model is provided in the Supplement.

3 Method

3.1 Model setup

The SMART model is forced with daily rainfall and daily potential evapotranspiration provided by the national meteorological office Met Éireann (2019). A five-hydrological-year warm-up period is used to determine the initial states of the soil layers and reservoirs in the model. A Python implementation of the SMART model is used to simulate the hydrological response in all study catchments (Hallouin et al., 2019).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

3.2 Model calibration

The calibration of the model is done using six different objective functions. The calibration procedure is illustrated in Figure 4, steps (a) to (d). This methodology is applied for each study catchment individually. First, in step (a), the model parameter space is explored using a Latin Hypercube Sampling (LHS) strategy (McKay et al., 1979) to generate 10^5 random parameter sets well distributed in the parameter space. The feasible parameter ranges used to define the boundaries of the parameter space are based on a previous study by Mockler et al. (2016) providing typical ranges for Irish catchments. The model is then used in step (b) to simulate the catchment response with each of these 10^5 parameter sets, which produces as many hydrographs. Then, the best 1% parameter sets are retained on the basis of their performance on the chosen objective function. This is similar to the GLUE methodology (Beven and Binley, 1992, 2014) without a threshold for acceptability. Instead of defining a threshold of acceptability, here it is preferred to analyse the statistics of equally sized parameter sets with each of the different objective functions.

In step (c) (Figure 4), six different objective functions are used to calculate the model performance by comparing the simulated and observed catchment responses. Three variants of the Kling-Gupta efficiency (Gupta et al., 2009) are tested. First, the KGE criterion is computed on the untransformed discharge series, that is E_{hi} (Equation 1); since the linear correlation coefficient included in KGE is more sensitive to errors on flow peaks (Krause et al., 2005), it is considered to put more emphasis on high flow conditions. Second, the KGE criterion is computed on the inverted discharge series, that is E_{lo} (Equation 2); this objective function on transformed flows is used to put more emphasis on low flow conditions. Inverted flows are preferred over log-transformed flows in order to retain the dimensionless character of the objective function allowing for comparison across catchments, however, any transformation of flows before computing KGE leads to the loss of the physical interpretability of the three components of KGE (Santos et al., 2018), but it is not required here. Third, an average of both objective functions is used, that is E_{av} (Equation 3); this objective function is used in order to equally consider high flow and low flow conditions in calibration (Garcia et al., 2017). These variants of KGE, referred to as traditional objective functions hereafter, assess the suitability of the model parameters by comparing the entirety of the observed and simulated hydrographs. In addition, three combinations (vectors) of streamflow characteristics (SFCs) are used as bespoke objective functions, and referred as such hereafter. They assess the model performance by comparing the observed and simulated values of the SFCs extracted from the observed and simulated hydrographs, respectively. For each vector of SFCs, the Euclidean distance (Equation 4) separating the observed and simulated points in the multi-dimensional space formed by each dimension in the vector of SFCs is calculated. One distance is calculated for the invertebrate community D_{inv} (vector of 7 SFCs), one is calculated for the fish community D_{fsh} (vector of 13 SFCs), and one is calculated for both communities at once D_{all} (vector of 18 SFCs). Similar to Kiesel et al. (2017), each SFC is normalised prior the calculation of the Euclidean distance so that its value is bounded between 0 and 1, effectively giving all SFCs the same weight in the computation of the Euclidean distance. Eventually, in step (d), the calibration with each of these six objective functions yields a set of 10^3 best performing behavioural parameter sets on the given function.

$$E_{hi} = E_{KG}(q_{obs}, q_{sim}) = 1 - \sqrt{\left(\frac{cov(q_{obs}, q_{sim})}{\sigma_{q_{obs}} \cdot \sigma_{q_{sim}}} - 1\right)^2 + \left(\frac{\sigma_{q_{sim}}}{\sigma_{q_{obs}}} - 1\right)^2 + \left(\frac{\mu_{q_{sim}}}{\mu_{q_{obs}}} - 1\right)^2} \tag{1}$$

$$E_{lo} = E_{KG}\left(\frac{1}{q_{obs} + 0.01 \cdot \mu_{q_{obs}}}, \frac{1}{q_{sim} + 0.01 \cdot \mu_{q_{sim}}}\right) \tag{2}$$

$$E_{av} = \frac{E_{hi} + E_{lo}}{2} \tag{3}$$

where cov, σ, and µ correspond to the covariance, the standard deviation, and the arithmetic mean, respectively; q_{obs}, and q_{sim} correspond to the time series of observed discharge, and simulated discharge, respectively. Noteworthy, a constant is added to the inverted discharge values in Equation 2 in order to avoid zero flows issues, and a hundredth of the arithmetic mean of the corresponding discharge series is used as recommended by Pushpalatha et al. (2012).

$$D_{target} = \sqrt{\sum_{i=1}^{N_{target}} \left(C_{obs,i} - C_{sim,i}\right)^2} \tag{4}$$

where N_{target} corresponds to the number of SFCs contained in the targetted combination of SFCs (the specific SFCs contained in each targetted combination can be found in Table 2), and where $C_{obs,\,i}$, $C_{sim,\,i}$ correspond to the i^{th} observed SFC value in the combination, and the i^{th} simulated SFC value in the combination, respectively.

3.3 Model evaluation

The method used to evaluate the performance of the predictions with a model calibrated with each of the six different objective functions is described in steps (e) to (h) of Figure 4. Again, this methodology is applied for each study catchment individually. First, in step (e), the model is run separately with each of the behavioural (10^3) model parameter sets to simulate its catchment response in the evaluation period, which produces 10^3 hydrographs. From each hydrograph, in step (f) the vectors of SFCs are calculated for each of the three combinations of SFCs (fish, invertebrates, and both), which yields 10^3 vectors of SFCs for each target community. The model prediction is then evaluated in step (g) by calculating the Euclidean distance between the observed and simulated vectors of SFCs, producing 10^3 Euclidean distances (Equation 4) for each community (D_{inv}, D_{fsh}, and D_{all}). Finally, in order to compare the predictive performance in each catchment, in step (h) a measure of central tendency, the median, is used to summarise the performance of the best behavioural parameter sets identified with each of the six objective functions.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

3.4 Split-sample tests

Split-sample tests are commonly used in the evaluation of hydrological predictions (Klemeš, 1986). Coron et al. (2012) proposed a generalised split-sample test using a sliding window for calibration, and evaluating the model performance on all other independent windows in the period. de Lavenne et al. (2016) adapted this strategy to calibrate a catchment model with a sliding window, and to evaluate the simulations on all other years (before and/or after the sliding window) in the study period. These approaches have the advantage of reducing the influence of the calibration/evaluation periods, compared with a single split-sample test that divides the study period into fixed calibration and validation periods.

The split-sampling strategy adopted for this study is adapted from the original approach by de Lavenne et al. (2016) in that it uses each hydrological year the same number of times in each of the 14 split-sample tests. For each catchment, the 14-hydrological-year series of discharge measurements is split into two seven-hydrological-year periods, and the split is repeated 14 times (Figure 5). It is implicitly assumed that any combination of hydrological years can be used, even if they are not consecutive years. Given this assumption, there are theoretically 3432 different combinations of seven-year periods in a 14-year study period. These combinations would represent all possible climatic combinations represented in the data for the study period; however, given the large dataset it would represent, it was decided to work only on 14 combinations.

It is to be noted that, in order to be able to assess the model consistency, a single Latin Hypercube sampling of the 10^5 parameter sets per catchment is used. That is to say that the Latin Hypercube is generated once and it is used on the 14 different calibration-evaluation periods in order to be able to determine whether a behavioural parameter set identified as behavioural remains behavioural on a different test.

4 Results

The analysis of the median (and inter-quartile range) performance of the behavioural parameter sets constitutes the first level of analysis, i.e., at the catchment level for each split-sample test taken individually. The split-sampling strategy is then used as a second level of analysis, in order to reach conclusions in each catchment that can be generalised for different calibration periods. To do so, the mean of the medians is used as a summary statistic across the split-sample tests. Finally, the performances of the six objective functions is assessed by comparing the average and standard deviation of these mean values across the 33 catchments, in order to reach conclusions that can be generalised for different catchments (Figure 4).

4.1 Are the candidate objective functions capable of reproducing the catchment hydrograph?

Prior to the comparison of the different objective functions for calibration in view to predict SFCs, the performance of each objective function for calibration in view to predict the overall hydrograph is analysed (Figure 6). Because a hydrological model is used to make the streamflow predictions, it is important to check whether the different objective functions are capable of finding parameter values that are able to reproduce the catchment hydrological response relatively well. Moreover, this gives more confidence in the model structure as being a plausible approximation of the hydrological processes in the study catchments.

To do so, the original definition of the Kling-Gupta efficiency E_{hi} is used, and the performances are compared in calibration and in evaluation. On average, all six objective functions score highly in calibration, with scores ranging from 0.69 to 0.83 (Figure 6a); they also score highly in model evaluation, with only a marginal drop compared to calibration with, on average, scores from 0.69 to 0.81 (Figure 6b). Note that D_{fsh} and D_{all} perform almost as well as E_{hi} itself, indicating that these combinations of SFCs are good candidates for general purpose hydrological studies.

4.2 Which objective function provides the most accurate predictions?

The performances of the six objective functions under scrutiny to predict SFCs is first analysed on average across the 14 split-sample tests, and across the 33 study catchments. The comparison reveals that the differences in performance between most objective functions are small (Figure 7). Nonetheless, the best performance in predicting all SFCs, i.e., the shortest Euclidean distance, for each of the three set of SFCs targetted in this study is always obtained using this same combination of SFCs as the objective function. Furthermore, the largest combination featuring 18 SFCs D_{all} is a competitive option, even when the focus is on smaller subsets of SFCs, and it outperforms any of the three formulation of the KGE criterion.

The formulation of KGE on low flows E_{lo} is the worst objective function to predict any of the three combinations of SFCs. E_{hi} is the best traditional objective function for the two largest combinations of SFCs, while E_{av} performs better for the smallest combination D_{inv}. This can be explained by the fact that a majority of SFCs in the D_{inv} set focus on low flow conditions, while a majority of SFCs in the D_{fsh} and and D_{all} sets focus on high flow conditions (Table 2). Thus, it is surprising to find that E_{av} clearly outperforms E_{lo} even on D_{inv}, however, Garcia et al. (2017) also found E_{av} to be better suited than E_{lo} to predict low-flow indices.

In addition, the dispersion of the performance across the 33 study catchments, measured by means of the standard deviation (Figure 7), is smaller for the better performing objective functions, which indicates that in addition to be predict well on average, they produce less variability in performance across the different study catchments.

4.3 Which objective function provides the most stable predictions?

The 14 split-sample tests strategy offers the opportunity to explore the stability of the performance obtained with the various objective functions, and hence the independence of the performance from the specificities of the period used in calibration. First, examining the average performance across the split-sample tests shows that the differences in performance between the objective functions are small (Figure 8). The best bespoke objective function on the targetted combination of SFCs is compared with the best traditional objective function on the same combination in order to investigate whether the more holistic definition that traditional objective functions represent brings an advantage when it comes to performance stability, while yielding poorer performances, as found above.

The comparison of E_{av} and D_{inv} to predict the SFCs contained in D_{inv} reveals that some catchments, e.g., 24003 and 16009 (Figure 8a), showcase the hypothesised behaviour with more stable predictions with the traditional objective function. However, this appears as a marginal behaviour amongst the set of 33 study catchments, and stability for most of the catchments are comparable between the traditional and bespoke objective functions, excluding any underlying bias in performance. As a matter of fact, the opposite behaviour is found when targetting the larger combinations of SFCs, e.g., 16003 (Figure 8b,c) or 27002 (Figure 8c). In these catchments, the bespoke objective functions yield more stable performances than the traditional objective function E_{hi}, in both cases. These are again outliers compared to the rest of the study catchments that do not exhibit significant differences in performance stability. Nonetheless, these results suggest that, if not equivalent, the stability in performance can be better with bespoke objective functions when they feature larger numbers of SFCs.

4.4 Which objective function yields the most consistent behavioural sets?

The concept of consistency has been previously used as a guide in the selection from competing model structures (Euser et al., 2013). Originally used as the capacity of a model structure to predict a range of hydrological signatures with the same parameter set, the idea of consistency is applied to objective functions in this study. It is used to compare different objective functions according to their ability to identify the same parameter sets as behavioural across the 14 split-sample tests, described above. This objective function consistency is defined as the ratio of parameter sets identified as behavioural in all 14 split-sample tests, with an optimal value of 1. Consistency establishes whether similar performance results were obtained with largely different parameter sets.

Unlike the measures of average model performance and performance stability, the objective function consistency measure reveals more significant differences between the six candidate objective functions (Figure 9). Indeed, on average, E_{hi} clearly outperforms all other five objective functions with a consistency of 51 %. This means that more than half of the behavioural parameter sets remain selected as behavioural across all 14 split-sample tests. The largest combination of SFCs D_{all} comes second best in terms of consistency with 33 %, closely followed by E_{av}, third with 32 %. The two least consistent objective functions are E_{lo} and D_{fsh}, with consistency ratios of 19 %, and 13 %, respectively.

Remarkably, the consistency ratios for the bespoke objective functions are related to the number of SFCs they contain. For example, D_{inv} containing only seven SFCs comes last, while D_{all} containing all 18 SFCs is third and very close to the ratio for the traditional objective function E_{av}. However, this may be related to the range of flow events the SFCs represent, and their relative difficulty in predicting them, since D_{inv} contains a majority of SFCs for low flow conditions, while the two other bespoke objective functions contain a majority of SFCs for high flow conditions.

Some catchments show close to zero consistency no matter what the objective function used, e.g., 15005 or 34024. This suggests that the model structure is not adequate for these catchments. Overall, the ranking of the objective functions is in accordance with the average behaviour described above.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

5 Discussion

5.1 On the definition of SFC-based objective functions for ecologically-relevant streamflow predictions

The choice of the objective function for ecological applications is known to influence the predictive performance of the hydrological model for specific streamflow characteristics (Vis et al., 2015; Kiesel et al., 2017; Pool et al., 2017). In particular, specially chosen composite objective functions containing the target SFCs have improved the prediction of these SFCs (e.g., Kiesel et al., 2017). This study confirmed these separate findings on three different set of SFCs using the same catchment model for all three sets. However, the consistency analysis revealed that the sample of parameter sets found suitable in calibration are less consistent across different split-sample tests with this type of objective function. The difference between traditional and bespoke objective functions in this regard is amplified by the reduction in the number of SFCs used to define the objective function (cf. section 4.4).

This result highlights that, because streamflow characteristics are found to be ecologically-relevant, this does not imply that they are necessarily hydrologically-relevant. Indeed, while some indicators originally used as ecologically-relevant SFCs (Olden and Poff, 2003) are also used as hydrological signatures (e.g., Yadav et al., 2007; Zhang et al., 2008), their selection as a relevant characteristic for the catchment of interest is driven by different needs in terms of the indicator skills. These indicators are selected as ecologically-relevant SFCs according to their influence on the stream ecology (Poff and Zimmerman, 2010), while they are selected as hydrological signatures to represent the hydrological behaviour of catchments (McMillan et al., 2017), i.e., they are "hydrologically-relevant" SFCs that can be used for catchment classification, or the regionalisation of hydrological information, for example. Hence, ecologically-relevant SFCs are not necessarily very informative when it comes to eliciting suitable parameter values in calibration, because they may not be key descriptors of the emergent hydrological processes at the catchment scale. For example, Pool et al. (2017) defined a composite objective function made of the most informative SFCs at hand, and yet, they were unable to accurately predict SFCs not included in the objective function. Also, their bespoke objective function was not able to capture the catchment hydrological behaviour accurately. The use of a consistency analysis in this study confirms that the bespoke objective functions tested are not skilled in selecting parameter values stable across split-sample tests. Nonetheless, some SFCs can be found useful in calibration. Yadav et al. (2007) suggest that a carefully selected subset of SFCs has the potential to constrain well a model parameter space. Kiesel et al. (2017) even found that the use of single SFCs may be almost as powerful as their complete set of seven SFCs to predict all seven SFCs, suggesting that these individual ecologically-relevant SFCs also have potential to be hydrologically-relevant SFCs in their German study catchment.

In this context, the definition of a good objective function for ecologically-relevant streamflow predictions must be based on SFCs that are key descriptors of the ecological response, while also key descriptors of the hydrological behaviour in the catchment. Otherwise, model consistency will be compromised and the model predictions might not be robust outside its calibration conditions. Nevertheless, a composite traditional objective function such as the Kling-Gupta efficiency remains a strong contender for the prediction of these SFCs. In particular, a non-parametric version of the KGE criterion could prove useful, namely to reduce the emphasis on high flow conditions and to provide a more balanced criterion across various flow conditions (Pool et al., 2018). Alternatively, segments of the flow duration curve have been used to calibrate hydrological models, which also offers opportunities to balance low, average, and high flow conditions (e.g. Yilmaz et al., 2008; Pfannerstill et al., 2014). However, the flow duration curve does not contain information on the timing (or duration) of individual flow events, which is important for aquatic species (Arthington et al., 2006). A combination of different objective functions fitted to flows (Vis et al., 2015), or a combination of objective functions fitted to flows and objective functions fitted to SFCs (Pool et al., 2017) can also be competitive options. In particular, the latter has the potential to overcome the consistency issue found with bespoke objective functions by including traditional objective functions.

5.2 Implications for the study of the impacts of climate change on the stream ecology

Hydrological models are usually preferred over statistical regression models when the impacts of a changing climate on the flow regime and the associated ecologically-relevant SFCs is of interest. Even though regression models may fit historical data better (Murphy et al., 2013), hydrological models have the potential to be run with alternative climate data in order to predict future changes in the catchment hydrograph. The identification of the most suitable objective function is therefore valuable for climate change scenario analysis. Here, we have already established the superiority of bespoke objective functions over a range of fourteen different split-sample tests in which the ranking between the objective functions is relatively stable. However a limitation of the study is that the flow data period from 1986 to 2016 is relatively short in climatological terms and does not contain a severe drought period, although some have been identified in long-term (250-year) precipitation records (Noone et al., 2017), but a corresponding flow record does not exist.

Assuming a suitable set of SFCs has been found, as described in the previous section, the use of a composite definition for the objective function based on absolute error between observed and simulated SFCs, such as in this study, or the recent studies by Kiesel et al. (2017), and by Pool et al. (2017) may not be realistic for practical applications. Indeed, the effects of a deviation in a given SFC value due to climate change might not be as critical for the stream ecology whether it is a positive or a negative one. By defining the objective function based on absolute error, this is not taken into account. Moreover, while SFCs are often normalised to avoid artificially weighing them based on their amplitude, they are not weighed according to the impact a given percentage deviation has on the stream ecology. The use of such weighted Euclidean distance may therefore prove useful to include such consideration in the calibration procedure.

5.3 Implications for ecologically-relevant streamflow predictions in ungauged basins

Understanding the ecological response to altered flow regimes is hindered by the lack of hydrological data where ecological data is available (Poff et al., 2010) because hydrometric gauges may not be in locations where ecological surveys have been carried out. As a result, the usual calibration of a hydrological model is not possible, and a direct method of predicting streamflow characteristics in ungauged locations is required.

One approach to regionalisation is the transfer of optimised parameter values from gauged to ungauged locations (Parajka et al., 2005). Given their higher consistency demonstrated in this study, the original KGE-based criteria appear better suited for regionalisation, rather than the bespoke objective functions tested in this study. Indeed, the optimised parameter values need to be strongly related to catchment behaviour in order for hydrological knowledge to be related to physical features and thus transferred to ungauged locations. While consistency could be improved through the change in model structure (Euser et al., 2013), Caldwell et al. (2015) and Garcia et al. (2017) found the choice of the calibration procedure more decisive than the model used for the prediction of SFCs.

Alternatively, ecologically-relevant streamflow characteristics can be directly transferred from gauged to ungauged locations (e.g., Yadav et al., 2007; Westerberg et al., 2014) and used as calibration information in the ungauged catchment. However, these SFCs are used as hydrological signatures to constrain the model parameter space, and as a result, their potential was assessed in order to predict the hydrograph in ungauged catchments. It remains to be explored whether these regionalised ensemble predictions can prove useful in predicting other SFCs relevant for ecological communities in ungauged catchments.

6 Conclusions

Desirable qualities for a useful objective function are that it performs well in evaluation, i.e., outside calibration, that its performance is independent of the calibration period, and that it consistently identifies the same parameter sets regardless of the study period, i.e., that it describes a consistent catchment hydrological behaviour. This study explored all three aspects for six different objective functions intended to predict three combinations of streamflow characteristics relevant for stream ecology. In relation to the three hypotheses formulated in the introduction, the study reveals that: bespoke objective functions perform marginally better than traditional objective functions to predict all three combinations of SFCs in evaluation periods on average, which provides evidence to confirm the first hypothesis; there was no evidence of any difference in performance stability in simulating the SFCs between the two types of objective functions across the split-sample tests, so that the second hypothesis cannot be confirmed; traditional objectives functions select more consistently the same parameter sets as behavioural across the split-sample tests, which provides evidence to support the third hypothesis.

This study unveils that a gain in fitting performance on the SFCs may hide a loss in consistency in the behavioural parameter sets across the split-sample tests. This highlights that ecologically-relevant SFCs are not necessarily hydrologically-relevant SFCs, and a combination of ecologically-relevant streamflow characteristics is not guaranteed to be a good definition of the key hydrological processes defining the catchment response. Unless, streamflow characteristics are proven to be both at once, carefully selected traditional objective functions fitted to flows are likely to remain preferable to predict ecologically-relevant streamflow predictions.

Code and data availability. The rainfall and potential evapotranspiration daily datasets are available online from Met Éireann (2019). The streamflow observations are available online from Ireland's Environmental Protection Agency (2019), and from the Office of Public Works (2019). The source code of the SMART model is open source and accessible online (Hallouin et al., 2019). The source code for the tools used to calculate the streamflow characteristics and the traditional objective functions are also open source and accessible online (Hallouin, 2019a, b).

Author contributions. This work is part of the PhD research of TH at the UCD Dooge Centre for Water Resources Research, under the supervision of MB and FOL. TH developed the idea. TH collected the data and performed the model simulations. TH wrote the original draft and the final version of this manuscript. MB and FOL reviewed and edited the different drafts of this manuscript.

Competing interests. The authors declare no conflict of interest.

Acknowledgements. The authors would like to thank Ireland's Environmental Protection Agency (EPA) for their financial support to the project ESManage (2014-W-LS-5).

References

Archfield, S. A., Kennen, J. G., Carlisle, D. M., and Wolock, D. M.: An Objective and Parsimonious Approach for Classifying Natural Flow Regimes at a Continental Scale, River Research and Applications, 30, 1166–1183, https://doi.org/10.1002/rra.2710, 2014.

Arthington, A. H., Bunn, S. E., Poff, N. L., and Naiman, R. J.: The challenge of providing environmental flow rules to sustain river ecosystems, Ecological Applications, 16, 1311–1318, https://doi.org/10.1890/1051-0761(2006)016[1311:TCOPEF]2.0.CO;2, 2006.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18–36, https://doi.org/10.1016/J.JHYDROL.2005.07.007, 2006.

Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, Hydrological Sciences Journal, 61, 1652–1665, https://doi.org/10.1080/02626667.2015.1031761, 2016.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, https://doi.org/10.1002/HYP.3360060305, 1992.

Beven, K. and Binley, A.: GLUE: 20 years on, Hydrological Processes, 28, 5897–5918, https://doi.org/10.1002/hyp.10082, 2014.

Caldwell, P. V., Kennen, J. G., Sun, G., Kiang, J. E., Butcher, J. B., Eddy, M. C., Hay, L. E., LaFontaine, J. H., Hain, E. F., Nelson, S. A. C., and McNulty, S. G.: A comparison of hydrologic models for ecological flows and water availability, Ecohydrology, 8, 1525–1546, https://doi.org/10.1002/eco.1602, 2015.

Carlisle, D. M., Wolock, D. M., and Meador, M. R.: Alteration of streamflow magnitudes and potential ecological consequences: a multiregional assessment, Frontiers in Ecology and the Environment, 9, 264–270, https://doi.org/10.1890/100053, 2011.

Colwell, R. K.: Predictability, Constancy, and Contingency of Periodic Phenomena, Ecology, 55, 1148–1153, https://doi.org/10.2307/1940366, 1974.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, https://doi.org/10.1029/2011WR011721, 2012.

de Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., and Ramos, M.-H.: Spatial variability of the parameters of a semi-distributed hydrological model, Proceedings of the International Association of Hydrological Sciences, 373, 87–94, https://doi.org/10.5194/piahs-373-87-2016, 2016.

Environmental Protection Agency: Daily Discharge Data, https://www.epa.ie/hydronet/#Flow, 2019.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci, 17, 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Garcia, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, Hydrological Sciences Journal, pp. 1–18, https://doi.org/10.1080/02626667.2017.1308511, 2017.

Goswami, M., O'Connor, K. M., Bhattarai, K. P., and Shamseldin, A. Y.: Assessing the performance of eight real-time updating models and procedures for the Brosna River, Hydrology and Earth System Sciences, 9, 394–411, https://doi.org/10.5194/hess-9-394-2005, 2005.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: 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, 2009.

Hallouin, T.: EFlowCalc: Ecological Streamflow Characteristics Calculator (Version 0.0.2), https://doi.org/10.5281/zenodo.2566757, 2019a.

Hallouin, T.: HydroEval: Evaluator for Streamflow Simulations (Version 0.0.2), https://doi.org/10.5281/zenodo.2591218, 2019b.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Hallouin, T., Mockler, E., and Bruen, M.: SMARTpy: Top-Down Rainfall-Runoff Model (Version 0.2.0), https://doi.org/10.5281/zenodo.2564042, 2019.

Henriksen, J. A., Heasley, J., Kennen, J. G., and Nieswand, S.: Users' Manual for the Hydroecological Integrity Assessment Process Software (Including the New Jersey Assessment Tools), Tech. rep., https://doi.org/10.3133/ofr20061093, 2006.

Kachroo, R.: River flow forecasting. Part 5. Applications of a conceptual model, Journal of Hydrology, 133, 141–178, https://doi.org/10.1016/0022-1694(92)90150-T, 1992.

Kakouei, K., Kiesel, J., Kail, J., Pusch, M., and Jähnig, S. C.: Quantitative hydrological preferences of benthic stream invertebrates in Germany, Ecological Indicators, 79, 163–172, https://doi.org/10.1016/J.ECOLIND.2017.04.029, 2017.

Kiesel, J., Guse, B., Pfannerstill, M., Kakouei, K., Jähnig, S. C., and Fohrer, N.: Improving hydrological model optimization for riverine species, Ecological Indicators, 80, 376–385, https://doi.org/10.1016/J.ECOLIND.2017.04.032, 2017.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, https://doi.org/10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424-425, 264–277, https://doi.org/10.1016/J.JHYDROL.2012.01.011, 2012.

Knight, R. R., Gain, W. S., and Wolfe, W. J.: Modelling ecological flow regime: an example from the Tennessee and Cumberland River basins, Ecohydrology, 5, 613–627, https://doi.org/10.1002/eco.246, 2012.

Knight, R. R., Murphy, J. C., Wolfe, W. J., Saylor, C. F., and Wales, A. K.: Ecological limit functions relating fish community response to hydrologic departures of the ecological flow regime in the Tennessee River basin, United States, Ecohydrology, 7, https://doi.org/10.1002/eco.1460, 2014.

Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89–97, https://doi.org/10.5194/adgeo-5-89-2005, 2005.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21, 239, https://doi.org/10.2307/1268522, 1979.

McMillan, H., Westerberg, I., and Branger, F.: Five guidelines for selecting hydrological signatures, Hydrological Processes, 31, 4757–4761, https://doi.org/10.1002/hyp.11300, 2017.

Met Éireann: Daily Meteorological Data, https://www.met.ie/climate/available-data/daily-data, 2019.

Mockler, E. M., O'Loughlin, F. E., and Bruen, M.: Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis, Computers & Geosciences, 90, 66–77, https://doi.org/10.1016/j.cageo.2015.08.015, 2016.

Murphy, J. C., Knight, R. R., Wolfe, W. J., and S. Gain, W.: Predicting Ecological Flow Regime at Ungaged Sites: A Comparison of Methods, River Research and Applications, 29, 660–669, https://doi.org/10.1002/rra.2570, 2013.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Noone, S., Broderick, C., Duffy, C., Matthews, T., Wilby, R., and Murphy, C.: A 250-year drought catalogue for the island of Ireland (1765-2015), International Journal of Climatology, 37, 239–254, https://doi.org/10.1002/joc.4999, 2017.

O'Connell, P., Nash, J., and Farrell, J.: River flow forecasting through conceptual models part II - The Brosna catchment at Ferbane, Journal of Hydrology, 10, 317–329, https://doi.org/10.1016/0022-1694(70)90221-0, 1970.

Office of Public Works: Daily Discharge Data, https://waterlevel.ie/hydro-data/, 2019.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Discussions

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Research and Applications, 19, 101–121, https://doi.org/10.1002/rra.700, 2003.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrology and Earth System Sciences, 9, 157–171, https://doi.org/10.5194/hess-9-157-2005, 2005.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, Journal of Hydrology, 510, 447–458, https://doi.org/10.1016/J.JHYDROL.2013.12.044, 2014.

Poff, N. L. and Zimmerman, J. K. H.: Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows, Freshwater Biology, 55, 194–205, https://doi.org/10.1111/j.1365-2427.2009.02272.x, 2010.

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C.: The Natural Flow Regime, BioScience, 47, 769–784, https://doi.org/10.2307/1313099, 1997.

Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C., Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., O'Keeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E., and Warner, A.: The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, Freshwater Biology, 55, 147–170, https://doi.org/10.1111/j.1365-2427.2009.02204.x, 2010.

Pool, S., Vis, M. J. P., Knight, R. R., and Seibert, J.: Streamflow characteristics from modeled runoff time series – importance of calibration criteria selection, Hydrology and Earth System Sciences, 21, 5443–5457, https://doi.org/10.5194/hess-21-5443-2017, 2017.

Pool, S., Vis, M., and Seibert, J.: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, Hydrological Sciences Journal, 63, 1941–1953, https://doi.org/10.1080/02626667.2018.1552002, 2018.

Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420-421, 171–182, https://doi.org/10.1016/J.JHYDROL.2011.11.055, 2012.

Richter, B., Baumgartner, J., Wigington, R., and Braun, D.: How much water does a river need?, Freshwater Biology, 37, 231–249, https://doi.org/10.1046/j.1365-2427.1997.00153.x, 1997.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A Method for Assessing Hydrologic Alteration within Ecosystems, Conservation Biology, 10, 1163–1174, https://doi.org/10.1046/j.1523-1739.1996.10041163.x, 1996.

Santos, L., Thirel, G., and Perrin, C.: Technical note: Pitfalls in using log-transformed flows within the KGE criterion, Hydrology and Earth System Sciences Discussions, pp. 1–14, https://doi.org/10.5194/hess-2018-298, 2018.

Shrestha, R. R., Peters, D. L., and Schnorbus, M. A.: Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators, Hydrological Processes, 28, 4294–4310, https://doi.org/10.1002/hyp.9997, 2014.

Vis, M., Knight, R., Pool, S., Wolfe, W., and Seibert, J.: Model Calibration Criteria for Estimating Ecological Flow Characteristics, Water, 7, 2358–2381, https://doi.org/10.3390/w7052358, 2015.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, Nature, 467, 555–561, https://doi.org/10.1038/nature09440, 2010.

Webster, K. E., Tedd, K., Coxon, C., and Donohoe, I.: Environmental Flow Assessment for Irish Rivers, Tech. rep., Ireland's Environmental Protection Agency, Dublin, http://www.epa.ie/pubs/reports/research/water/EPARR203finalweb-3.pdf, 2017.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves, Hydrology and Earth System Sciences, 15, 2205–2227, https://doi.org/10.5194/hess-15-2205-2011, 2011.

Hydrology and Earth System Sciences
Open Access
Discussions

Westerberg, I. K., Gong, L., Beven, K. J., Seibert, J., Semedo, A., Xu, C. Y., and Halldin, S.: Regional water balance modelling using flow-duration curves with observational uncertainties, Hydrology and Earth System Sciences, 18, 2993–3013, https://doi.org/10.5194/hess-18-2993-2014, 2014.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756–1774, https://doi.org/10.1016/J.ADVWATRES.2007.01.005, 2007.

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, https://doi.org/10.1029/2007WR006716, 2008.

Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, Water Resources Research, 44, https://doi.org/10.1029/2008WR006833, 2008.

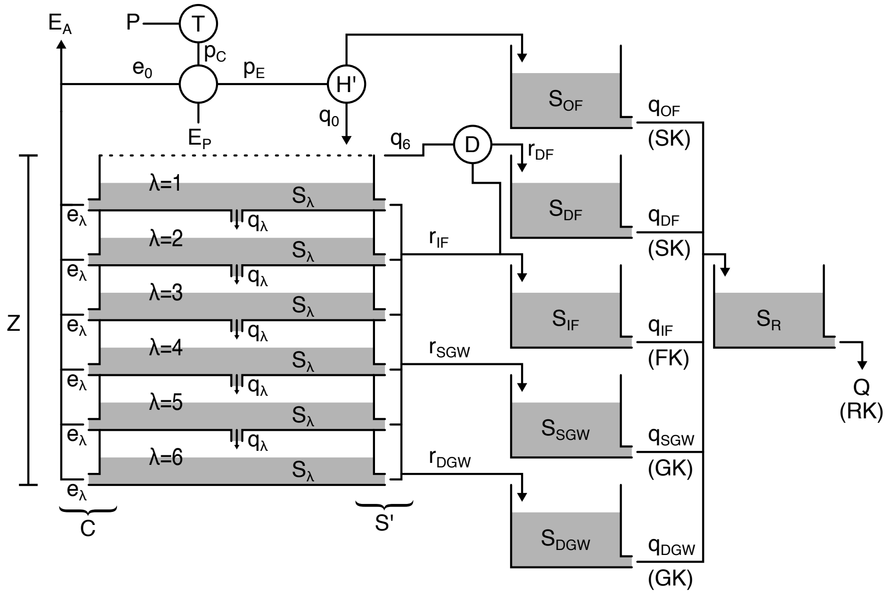

Figure 1. Conceptual representation of the SMART model structure. P and E_P, precipitation and potential evapotransporation, respectively, are the model inputs; Q and E_A, discharge and actual evapotranspiration, respectively, are the model outputs. For full description of the parameters, states, and fluxes presented on the figure, as well as the conceptual model equations, the reader is referred to the documentation provided in the Supplement.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

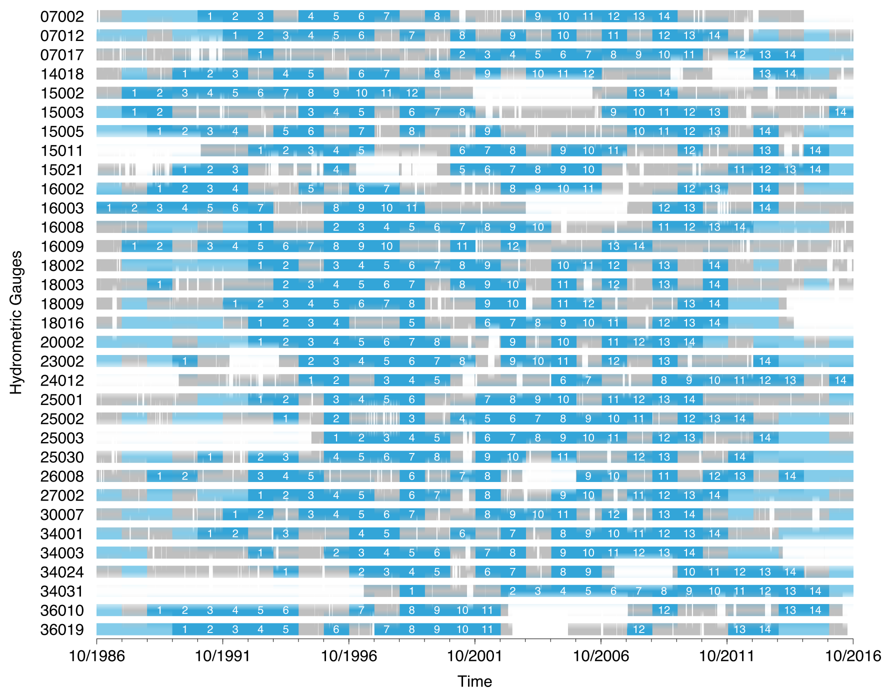

Figure 2. Discharge data availability for the 33 study catchments. The 14 complete hydrological years selected are represented in dark blue and annotated from 1 to 14. Years in light blue are other complete hydrological years not retained. Discontinuous grey years contain missing data represented as a discontinuity in the bar.

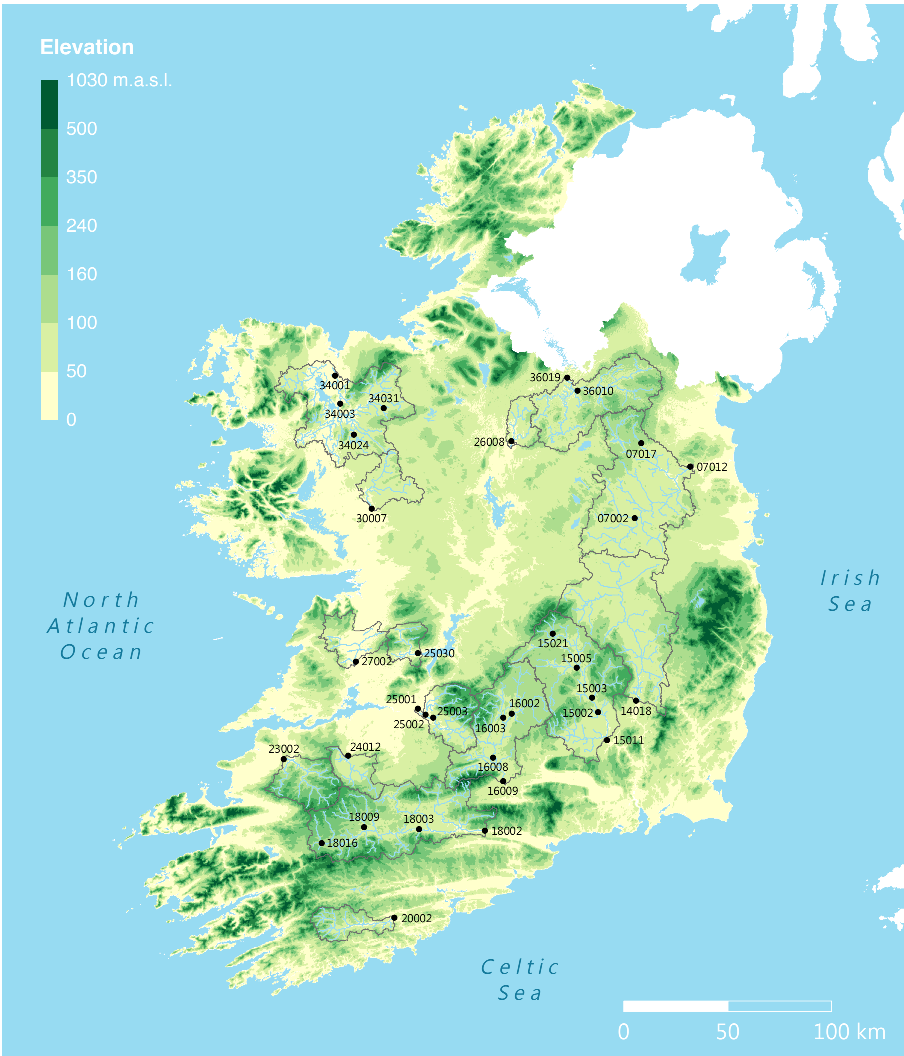

Figure 3. Location of the 33 hydrometric gauges forming the 33 nested study catchments on the elevation map of the Republic of Ireland (source: Ireland's EPA). Each number corresponds to the code of a hydrometric gauge.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

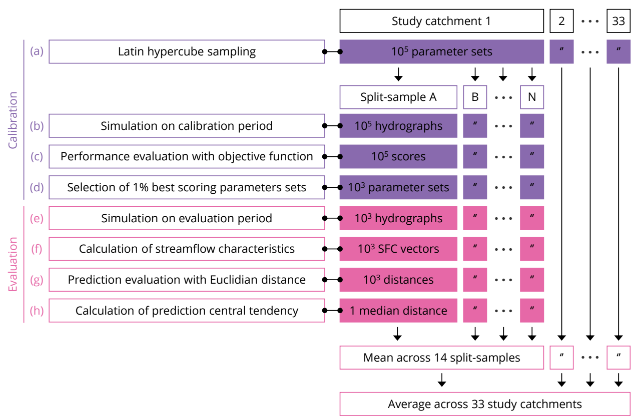

Figure 4. Model calibration and evaluation strategy for the prediction of SFCs with different objective functions. Steps (a) to (d) correspond to model calibration, while steps (e) to (h) correspond to model evaluation. These steps are replicated for each study catchment, and for each split-sample test.

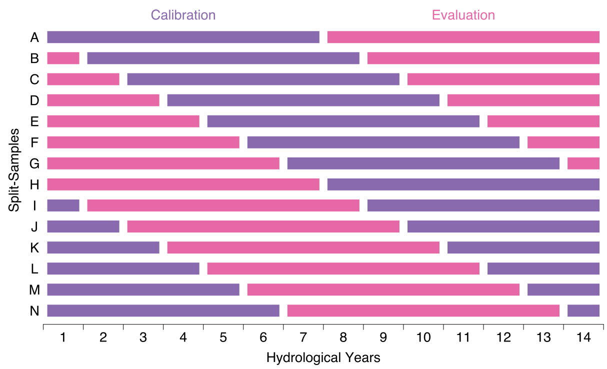

Figure 5. Split-sampling strategy using a seven-year rolling window, adapted from de Lavenne et al. (2016). Each period of 14 hydrological years enumerated in Figure 2 are represented on the x-axis and split into two seven-year periods, one for model calibration (in purple), and one for model evaluation (in pink).

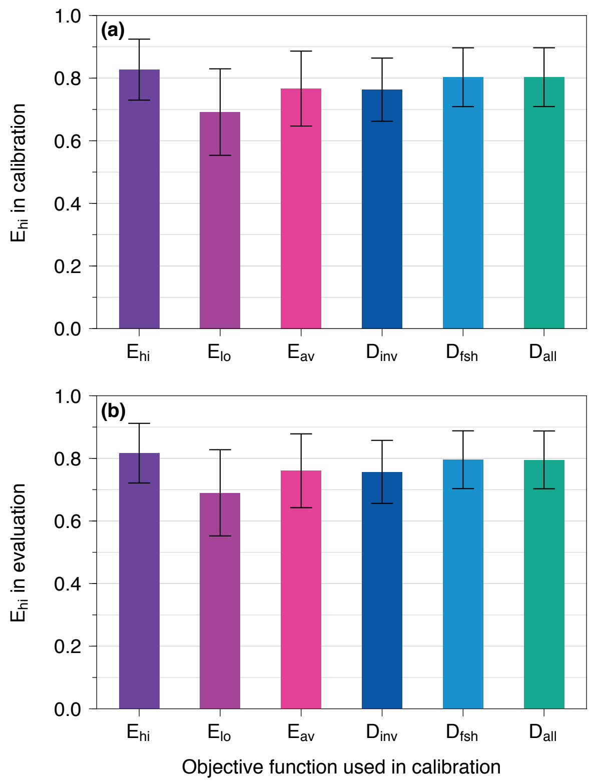

Figure 6. Comparison of the performance for general purpose hydrological predictions (a) in calibration and (b) in evaluation using six different objective functions for calibration. The top and bottom panels correspond to the performance on E_{hi} in calibration, and in evaluation, respectively. Each bar in a panel corresponds to the average performance across the 33 catchments as detailed in Figure 4. The standard deviation across the 33 catchments is represented by error bars on each bar.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

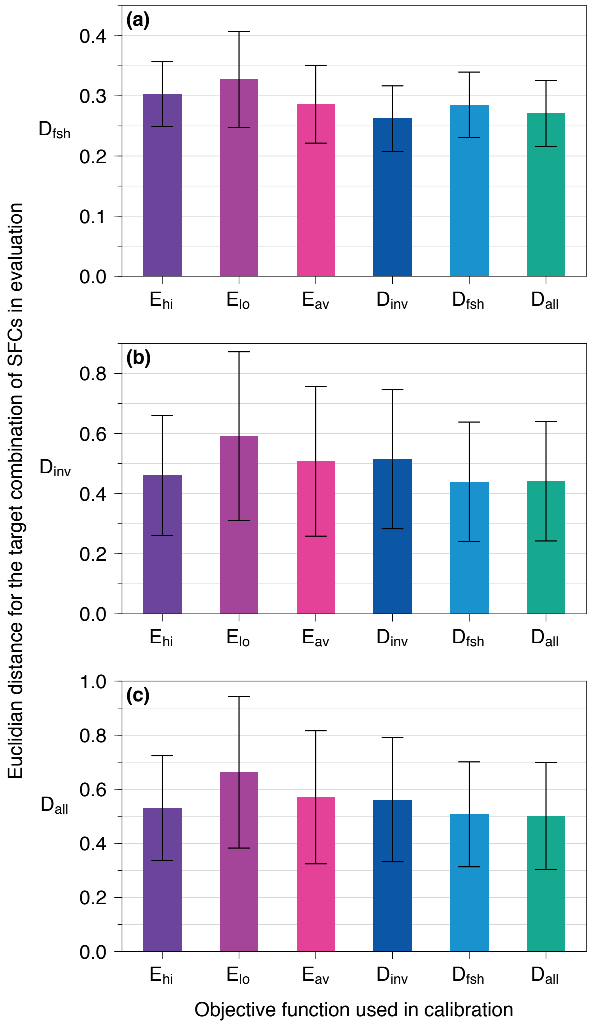

Figure 7. Comparison of the predictive performance in evaluation of three combinations of streamflow characteristics using six different objective functions for calibration. Each panel corresponds to a different combination specified on the y-axis ((a) D_{inv}, (b) D_{fsh}, and (c) D_{all}), while each bar in the panel corresponds to the average performance across the 33 catchments as detailed in Figure 4. The standard deviation across the 33 catchments is represented by error bars on each bar.

Hydrology and Earth System Sciences
Open Access
Discussions

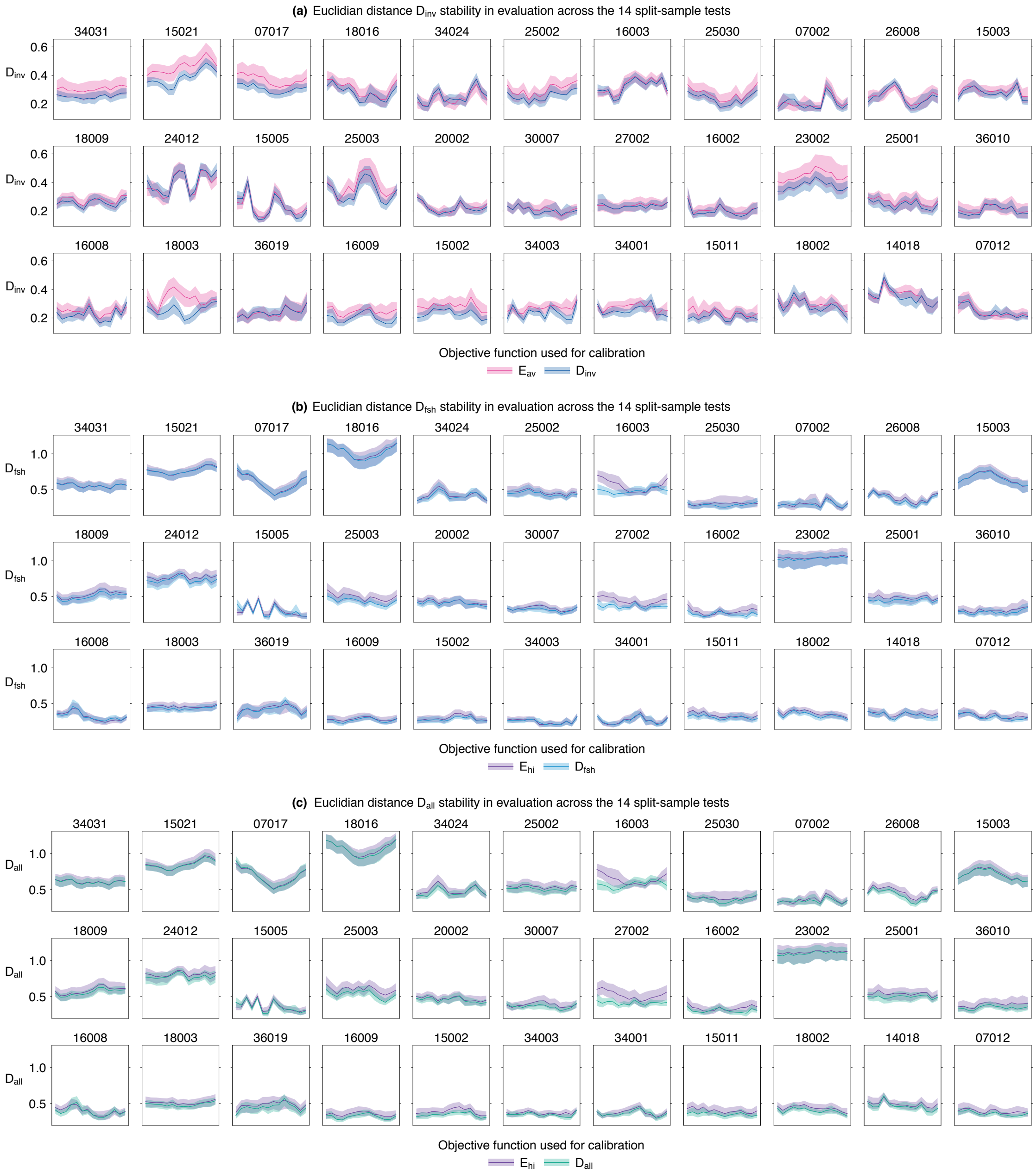

Figure 8. Performance stability in predicting the three combinations of SFCs (a) D_{inv}, (b) D_{fsh}, and (c) D_{all}. For each, two objective functions are compared across the 14 split-sample tests for each study catchment taken individually. On each panel, the 14 split-sample tests are displayed along the x-axis, while the y-axis corresponds to the performance expressed as a Euclidean distance. For each objective function, the bold line corresponds to the median as detailed in Figure 4 while the shaded band around it corresponds to its associated inter-quartile range.

Hydrology and Earth System Sciences
Open Access
Discussions

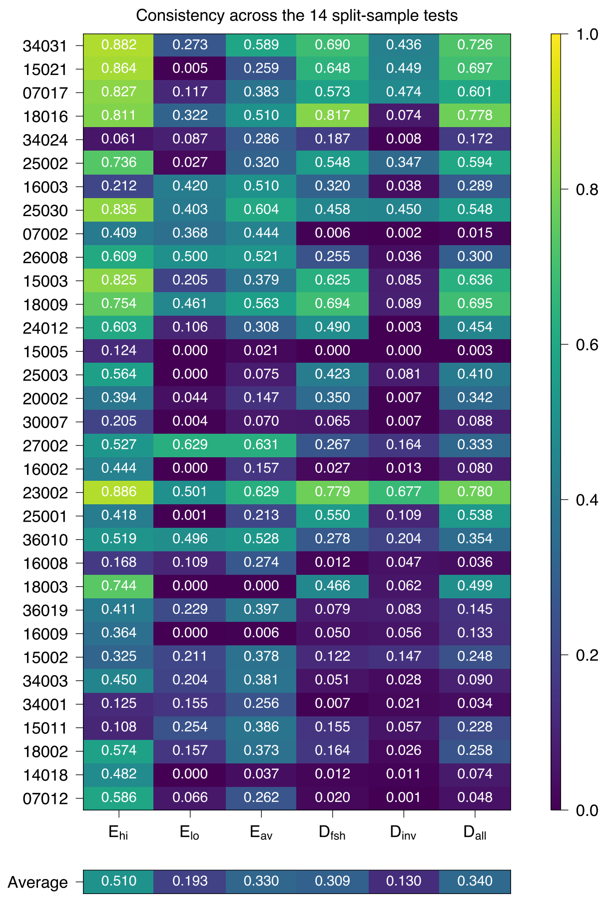

Figure 9. Heatmap of the consistency of the behavioural sets identified in calibration with the six different objective functions. Catchments are displayed on the y-axis while the six objective functions used for calibration are displayed on the x-axis. Each cell of the heatmap contains the mean ratio of behavioural parameter sets that remain behavioural across the 14 split-sample tests. An average across the 33 catchments is also provided at the bottom of the figure.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Table 1. List and Description of the ten parameters of the SMART model.

Parameter	Description	Unit
T	Rainfall aerial correction factor	–
C	Evaporation decay coefficient	–
H	Quick runoff ratio	–
D	Drain flow ratio	–
S	Soil outflow coefficient	–
Z	Effective soil depth	mm
SK	Surface reservoir residence time	time step
FK	Interflow reservoir residence time	time step
GK	Groundwater reservoir residence time	time step
RK	Channel reservoir residence time	time step

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Table 2. List and description of the 18 selected streamflow characteristics. Detailed calculations for each SFC in available in Table A1. The three last columns indicate whether a given SFC is included (∈) or not included (∉) in Equation 4 for the definition of each of the three Euclidean distances used as bespoke objective functions.

Category	Indicator	Description	Unit	Target species	D_{inv}	D_{fsh}	D_{all}
Magnitude							
Average flows	MA26	Variability in March mean flow	%	Fish	∉	∈	∈
	MA41	Annual mean daily flow	$m^3\ s^{-1}$	Fish	∉	∈	∈
Low flows	ML17	Base flow ratio 1	–	Invertebrates	∈	∉	∈
	ML20	Base flow ratio 3	–	Fish	∉	∈	∈
	Q85	Flow exceeded 85% of the time	$m^3\ s^{-1}$	Fish	∉	∈	∈
High flows	MH10	Mean October highest flood	$m^3\ s^{-1}$	Fish	∉	∈	∈
Frequency							
Low flows	FL2	Variability in low flow pulse count	%	Both	∈	∈	∈
High flows	FH6	Frequency of moderate floods	yr^{-1}	Fish	∉	∈	∈
	FH7	Frequency of large floods 1	yr^{-1}	Fish	∉	∈	∈
	FH9	Frequency of large floods 2	yr^{-1}	Invertebrates	∈	∉	∈
Duration							
Low flows	DL9	Variability in annual minimum 30-day mean flow	%	Invertebrates	∈	∉	∈
High flows	DH4	Annual maximum of 30-day moving mean flow	$m^3\ s^{-1}$	Invertebrates	∈	∉	∈
	DH13	Variability in annual maximum 30-day mean flow	–	Fish	∉	∈	∈
	DH16	Variability in high flow pulse duration	%	Fish	∉	∈	∈
Timing							
Average flows	TA1	Flow constancy	–	Both	∈	∈	∈
Low flows	TL1	Timing of annual minimum flow	Julian day	Fish	∉	∈	∈
Rate of change							
All flows	RA2	Variability in flow rise rate	%	Invertebrates	∈	∉	∈
	RA7	Flow recession rate	$m^3\ s^{-1}$	Fish	∉	∈	∈

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences Discussions

Table A1. Detailed computations for the 18 selected streamflow characteristics.

SFC	**Description** Detailed calculations
MA26	**Variability in March mean flow** Compute the mean and standard deviation in daily flows in March for each hydrological year. Divide the standard deviations by the means. Calculate the mean of these ratios to get MA26.
MA41	**Annual mean daily flow** Compute the mean daily flow for each hydrological year. Divide the means by the drainage area in square kilometers. Calculate the mean of these ratios to get MA41.
ML17	**Base flow ratio 1** Compute the 7-day rolling mean for each hydrological year. Calculate the minimum rolling mean and divide by the mean daily flow for each hydrological year. Calculate the mean of these ratios to get ML17.
ML20	**Base flow ratio 3** Break down the entire record of daily flows into 5-day blocks. Calculate the minimum flow in each block. This minimum is set as the baseflow for the block if 90% of its value is less than the minimum flow of its preceding and following blocks. Otherwise baseflow for this block is unassigned. Replace all unassigned baseflow values using linear interpolation on the already assigned baseflow values. Calculate the total baseflow by summing up the baseflow values in each 5-day block, and the total flow for the entire record. Calculate the ratio of these two totals to get ML20.
Q85	**Flow exceeded 85% of the time** Calculate the 15^{th} percentile for the entire record to get Q85.
MH10	**Mean October highest flood** Compute the maximum daily flow In October for each hydrological year. Calculate the mean of these values to get MH10.
FL2	**Variability in low flow pulse count** Calculate the 25^{th} percentile for the entire record. Calculate the number of flow events that are below this percentile for each hydrological year. Calculate the coefficient of variation (i.e., standard deviation divided by mean) of these values and multiply by 100 to get FL2.
FH6	**Frequency of moderate floods** Calculate the median for the entire record. Calculate the number of flow events that are above 3 times this median for each hydrological year. Calculate the mean of these values to get FH6.
FH7	**Frequency of large floods 1** Calculate the median for the entire record. Calculate the number of flow events that are above 7 times this median for each hydrological year. Calculate the mean of these values to get FH7.
FH9	**Frequency of large floods 2** Calculate the 25^{th} percentile for the entire record. Calculate the number of flow events that are above this percentile for each hydrological year. Calculate the mean of these values to get FH9.

continued on next page

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-161
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 6 May 2019

Hydrology and Earth System Sciences
Open Access
Discussions

continued from previous page

SFC	**Description** Detailed calculations
DL9	**Variability in annual minimum 30-day mean flow** Compute the 30-day rolling mean for the entire record. Calculate the minimum of this rolling mean for each hydrological year. Calculate the coefficient of variation (i.e., standard deviation divided by mean) of these values and multiply by 100 to get DL9.
DH4	**Annual maximum of 30-day moving mean flow** Compute the 30-day rolling mean for the entire record. Calculate the maximum of this rolling mean for each hydrological year. Calculate the mean of these values to get DH4.
DH13	**Variability in annual maximum 30-day mean flow** Compute the 30-day rolling mean for the entire record. Calculate the maximum of this rolling mean for each hydrological year. Calculate the mean of these values and divide by the median daily flow for the entire record to get DH13.
DH16	**Variability in high flow pulse duration** Calculate the 75th percentile for the entire record. Calculate the average duration of flow events above this percentile for each hydrological year. Calculate the coefficient of variation of these values and multiply by 100 to get DH16.
TA1	**Flow constancy** Decimal log-transform the entire record of daily flows. Calculate the decimal log of the mean daily flow for the entire record. Compute the Colwell (1974) matrix featuring 365 rows for 365 days in a year (ignoring last day of February for leap years) and 11 columns for 11 flow states (break points are 0.10, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, and 2.25 times the log mean daily flow calculated previously) for each hydrological year, incrementally adding to the tally in each cell from year to year. Calculate the sum of each column Y (vector), and the sum of the whole matrix Z (scalar). Divide the elements of vector Y by scalar Z. Multiply the elements of the new vector by their respective decimal log-transformed value, sum the elements of the vectors to obtain a scalar and multiply by minus one to obtain the uncertainty with respect to the states H(Y). Divide H(Y) by the decimal log of the number of states (11), and subtract this ratio from one to get TA1.
TL1	**Timing of annual minimum flow Julian day** Determine the date of the annual minimum daily flow in the Julian calendar for each hydrological year. Convert these values into an angle in the unit circle. Compute their coordinates (i.e., cosine and sine). Calculate the mean of these two values separately. Calculate the ratio of this mean sine divided by this mean cosine. Calculate the arc tangent of this ratio to get the angle corresponding to these mean coordinates. Convert this angle back to a Julian date to get TL1.
RA2	**Variability in flow rise rate** Compute the difference in daily flows between each consecutive days for the entire record. Calculate the coefficient of variation (i.e., standard deviation divided by mean) for the positive differences (i.e., rising limbs) and multiply by 100 to get RA2.
RA7	**Flow recession rate** Natural log-transform the entire record of daily flows. Compute the difference in this log-transformed daily flows between each consecutive days for the entire record. Calculate the median of the negative differences (i.e., recession limbs) to get RA7.