Hydrology and
Earth System
Sciences

Discussions

Replication of ecologically relevant hydrological indicators following a covariance approach to hydrological model parameterisation

Annie Visser[1], Lindsay Beevers[1], Sandhya Patidar[1]

[1]Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, EH14 4AS, UK

5 *Correspondence to*: Annie Gallagher Visser (av96@hw.ac.uk)

Abstract. Hydrological models can be used to assess the impact of hydrologic alteration on the river ecosystem. However, there are considerable limitations and uncertainties associated with the replication of the required, ecologically relevant hydrological indicators. Vogel and Sankarasubramanian's covariance approach to model parameterisation represents a shift away from the traditional calibration-validation goodness-of-fit paradigm. Using the covariance structures of the observed

10 input and simulated output time-series, the region of parameter space which best captures (replicates) the characteristics of a hydrological indicator may be identified. Through a case study, a modified covariance approach is applied with a view to replicating a suite of seven ecologically relevant hydrological indicators. Model performance and consistency are assessed relative to four comparative studies. The ability of the approach to address the limitations associated with traditional calibration-validation is further considered. Benefits of the approach include an overall reduction in model uncertainty whilst

15 also reducing overall time-demands. Difficulties in the replication of complex indicators, such as rate of change, are in line with prior work. Nonetheless, the study illustrates that consistency in the replication of hydrological indicators is achievable; additionally, the replication of magnitude indices is markedly improved upon.

1 Introduction

Water is the most essential natural resource (World Water Assessment Programme, 2009; Vörösmarty et al., 2010). The

20 principle source of freshwater, rivers, lakes and groundwater hold only 0.7% of the water on the planet (Shiklomanov, 1993). Rivers support prosperity, health, and well-being through the provision of ecosystem services; examples include: water security, energy production, hydro-hazard regulation, and water purification (Gilvear et al., 2017). The river flow regime is considered the major determinant in the structuring and functioning of the riverine ecosystem and provision of these services (Poff et al., 1997). Despite this, relentless pressures from both societal water demand and climate change raise significant

25 questions over the long-term sustainability of this resource (Gleick 1998, 2016; Klaar et al., 2014).

The need to balance the conflicting demands of both human society and those of the ecosystem has led to recent environmental flows research. This is defined as: *"...the quantity, timing, and quality of water flows required to sustain freshwater and estuarine ecosystems and the human livelihood and well-being that depend on..."* (Brisbane Declaration, 2007). Richter et al. (1996) identified five facets of the flow regime required to support the riverine ecosystem: magnitude, frequency, duration,

timing and rate of change. To date, over 200 ecologically relevant hydrologic indices (HIs) have been proposed (Olden and Poff, 2003; Monk et al., 2006; Thompson et al., 2013). Determined from flow time-series, simulated via hydrological model, these HIs may be used to assess the impact of hydrological change on the river ecosystem in the future, for examples, see: Richter et al. (1996); Carlisle et al. (2010); Poff and Zimmerman (2010); Murphy et al. (2012); You et al. (2014); De Girolamo

5 et al. (2017); Williams (2017). Example applications include understanding the effect of a changed climate, engineering intervention or in the establishment of environmental flow limits.

A hydrological (rainfall-runoff) model is essentially a simplification of the hydrologic system: hydroclimatological variables, such as temperature and precipitation, are used to estimate river flow. The paramount aim of hydrological modelling is to determine a suitable model structure and corresponding parameter set that provides a realistic representation of the hydrological

10 processes in the catchment of interest (Seibert, 2000; Beven, 2012a). Structural deficiencies, as opposed to inadequate calibration (Beven, 2010; Beven, 2012b), should be the principal cause for model rejection (Westerberg et al., 2011).

Traditionally, hydrological models are parameterised following a calibration-validation approach based on Klemeš (1986) split-sample technique (Vogel and Sankarasubramanian, 2003); often with multiple calibration-validation trials. The calibration of the hydrological model focuses on the goodness-of-fit (GOF) between the observed and simulated flow time-

15 series for a defined objective function; the Nash Sutcliffe model efficiency criteria (NSE) is among the most widely used. The sensitive nature of these traditional measures of GOF (objective functions) has been addressed through the consideration of modified NSE criteria and multi-criteria calibration (for example Gupta et al. (1998); Seibert (2000); Efstratiadis and Koutsoyiannis (2010); Pushpalatha et al. (2012)). However, despite improvements, certain problems remain, including (Westerberg et al., 2011): (1) the potential for bias in the model parameterisation as a result of measurement error (due to poor

20 accuracy and/or calibration) and uncertainty in the flow data (Pelletier, 1988; Montanari et al., 2013); (2) the arbitrary nature of the GOF behavioural thresholds; and (3) the problem of equifinality, where similar GOF may be achieved across multiple calibration trials, but with different parameters sets (Beven, 2006).

The use of hydrological models to determine HIs implicitly premises that the underlying hydrological processes of the catchment are sufficiently captured. Where this premise proves false, it directly impacts accuracy in the HIs, leading to high

25 levels of variability (Shrestha et al., 2014, 2016; Vis et al., 2015; Pool et al., 2017). For example, Shrestha et al. (2014) evaluated the ability of the VIC hydrologic model to replicate a number of HIs; HIs relating to annual and peak flows were simulated well whilst minimum flows and flow pulses were not. A focus on the characteristics of the flow regime, or hydrological signatures, has been shown to limit the influence of input uncertainties on the performance and consistency of hydrological models (Westerberg et al., 2011; Euser et al., 2013).

30 Vogel and Sankarasubramanian's, 2003, covariance approach to model parameterisation without calibration addresses many of these problems. The objective is to identify the region of parameter space which captures (replicates) the characteristics of a specified HI. This is achieved by focussing on the ability of the hydrological model to capture the observed covariance structure of the input and output time-series. Presently, the approach is limited by its focus on a single HI, preventing its use for the determination of a suite of ecologically relevant HIs. This paper builds on the covariance approach, adapting the

methodology to consider multiple ecologically relevant HIs. To determine the applicability of this modified covariance approach, the method is applied to a case study using the four-parameter hydrological model GR4J. The modelling objective is the replication of seven ecologically relevant HIs (as part of a larger work determining the impact of climate change on instream hydroecological response; Visser et al. (2018b)). The performance and consistency of the modified covariance

5   approach is evaluated in terms of the replication ability of the hydrological model and with reference to prior studies with similar modelling objectives (Shrestha et al., 2014; Vis et al., 2015; Shrestha et al., 2016; Pool et al., 2017).

Three specific research questions are answered in this paper:

1) Is the modified covariance approach able to satisfactorily replicate a suite of hydrological indicators? (With regards to performance and consistency; definitions below.)

10   2) How do the outcomes (replication of the ecologically relevant hydrological indicators) compare with those studies with similar modelling objectives?

3) Does the covariance approach advance progress towards addressing the limitations inherent in traditional hydrological model calibration (as described above)?
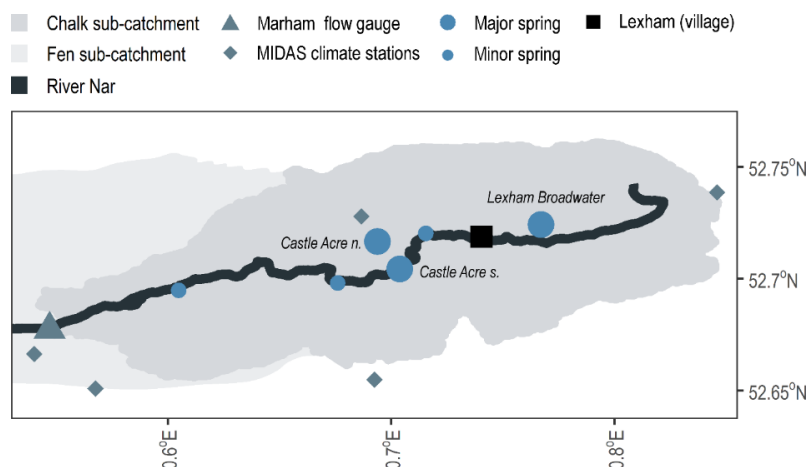
Through this paper we refer to model performance and consistency. After Euser et al. (2013), model performance is defined

15   as the ability to mimic the behaviour of catchment hydrological processes; consistency represents the ability of the hydrological model to reproduce the suite of HIs (i.e. multiple hydrological signatures simultaneously using the same parameter set).

2 Methods

2.1 Study area

The River Nar is a chalk stream located in Norfolk, south-east England. With two distinct river units, the River Nar has been

20   designated a Site of Special Scientific Interest (SSSI); the upper Nar overlies a chalk scarp to Marham, whilst the lower alluvial reach forms a fen basin. Despite its high conversation value, the River Nar is subject to significant pressures, inhibiting the ecological potential of the river (NRT, 2012). The river has been subject to continuing research into its flow regime, and their governing factors (Garbe et al., 2016; Visser et al., 2017, 2018a).

The focus of this paper is on the 153.3 km^2 upper catchment (chalk reach) only. Flow is primarily sustained by springs at West

25   Lexham and near Castle Acre (Fig. 1), and upstream of Lexham, through groundwater seepage and surface water runoff; these upstream reaches are considered particularly vulnerable to low-flows (Sear et al., 2005). With a highly seasonal flow regime, the hydrology of the River Nar is characteristic of pure chalk streams (Sear et al., 2005); aquifer recharge occurs in the autumn months, with a progressive rise in flow March/April. Flow is relatively low, over the available 1961-2015 record the median flow is 1.11 m^{3}s^{-1}, whilst Q10 and Q90 flows are 1.96 and 0.47 m^{3}s^{-1}. As of September 2017 (the most recent data currently

30   available), the year 1991 saw the most extreme hydrological drought recorded at the Marham gauge (Garbe et al., 2016), with flow falling below Q95 values for 178 consecutive days.

Figure 1. River Nar case study catchment focussed on the chalk reach; the locations of the flow gauge and MIDAS climate stations are detailed. The three major springs sustaining flow in the reach are also indicated.

2.2 Hydrological model

5   The principle of parsimony, known as Occam's razor, posits that a solution should be no more complex than necessary. In the context of hydrological modelling, model simplicity relative to performance is thus made key (Kokkonen and Jakeman, 2002; Perrin et al., 2003; Beven, 2012a). To this end, GR4J, a four-parameter model from the GR-J series of hydrological models was selected (Perrin et al., 2003). The GR-J series of models have been applied in a variety of hydrological contexts, examples include: Le Moine et al. (2008); Perrin et al. (2008); Coron et al. (2012); Smith et al. (2012); Coron et al. (2017).

10   The model GR4J is a lumped model based on soil moisture accounting (Fig. 2). The model inputs, P, the catchment rainfall depth, and E, the average depth of (potential) evapotranspiration), fill the production store with a capacity of $x1$ mm. The routed depth of water, Pr, is determined by the rate of percolation, $F(S, x1)$, as well as water in excess of the storage capacity. To simulate the time difference between rainfall event and flow peak, Pr is divided into two flow components and routed through unit hydrographs, time base $F(x4)$ days. Finally, the groundwater exchange term gw, $F(x2)$, acts on the routed, Qr,

15   and direct flow, Qd, components; a positive value indicates inflow from groundwater whilst a negative represents water export. The total flow, Q, is determined by summing the routed and direct flows. The model is applied using the R package `airGR` (Coron et al., 2017; Coron et al., 2018). Parameter limits are summarised in Table 1.

2.3 Modelling application

2.3.1 Data

20   Continuous (daily) time-series of mean flow, precipitation and potential evapotranspiration for the period 1961-2015 serve as model input. Flow data from the Marham gauge (Fig. 1) was provided by the National River Flow Archive (CEH (2018)). The required climate data was computed from daily average rainfall and hourly temperature recordings at 5 MIDAS stations (Fig. 1; Met Office (2016)); potential evapotranspiration was estimated using a temperature-based PE model (Oudin et al., 2005).
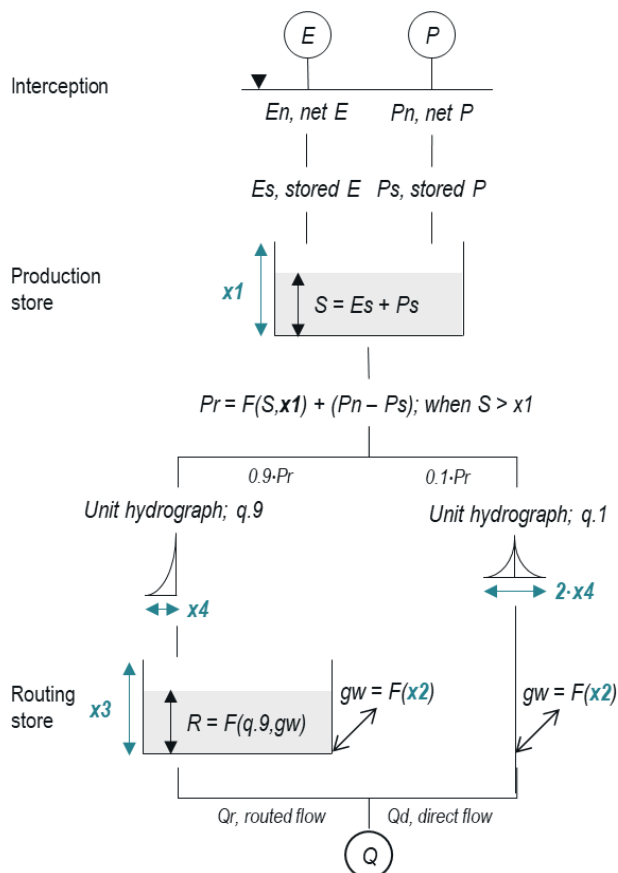
Figure 2. Structure of the GR4J hydrological model; based on Perrin et al. (2003).

5 **Table 1. Parameter limits specified for the hydrological model GR4J.**

	Description	Limits
x1	Capacity of production store (mm)	(100,1200}
x2	Groundwater transfer (mm/day; positive indicates flow *from* aquifer)	(-5,25}
x3	Capacity of routing store (mm)	(20,1000}
x4	Time lag between rainfall event and flow (days)	(0.5,30}

The ecologically relevant HIs were determined in Visser et al. (2018b) as part of the development of a hydroecological model following an Information Theory (IT) approach (Visser et al., 2018a). The selected HIs are summarised in Table 2 along with their relative importance (according to IT). These seven were selected from a set of 63 ecologically relevant HIs, based on

10 Olden and Poff (2003), Monk et al. (2006) and Thompson et al. (2013). To reflect seasonality in the flow regime, the indices are differentiated by season (Table 2): winter (October-March) and summer (April-September).

Hydrology and
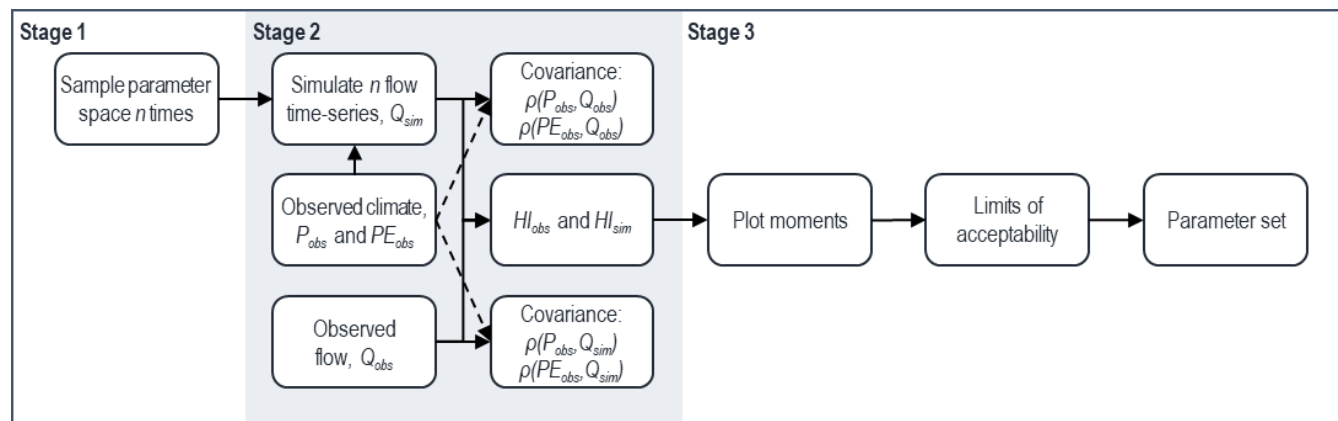Earth System
Sciences
Discussions

2.3.2 Covariance approach

The covariance approach was first developed by Vogel and Sankarasubramanian (2003), where the aim was to replicate a HI rather than the flow time-series. Here, modification of the covariance approach allows for the consideration of a suite of ecologically relevant HIs. The modified covariance approach is implemented over three stages (Fig. 3).

5   Table 2. Overview of the seven ecologically relevant HIs considered in the application of the case study river catchment. Seasons are denoted as W (winter) and summer (S); the facets of the flow regime are denoted as M (magnitude) and R (rate of change).

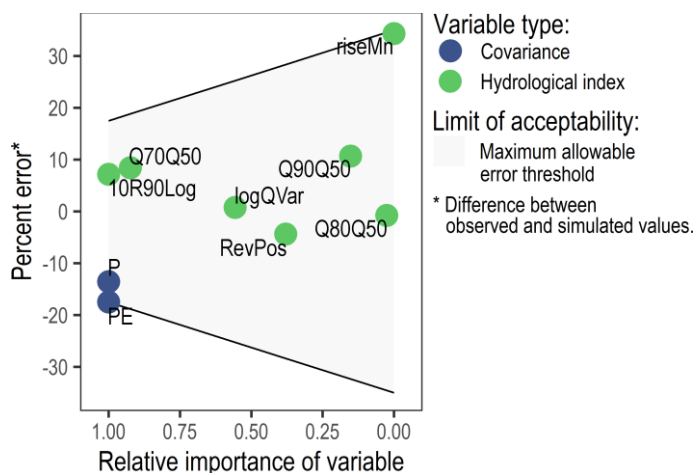| No.	Season	Facet	Index	Streamflow characteristic	Unit	Importance
1	W	M	10R90Log	Log ratio $10^{th}/90^{th}$ percentile flows.	-	0.86
2	W	R	riseMn	Mean rise rate in flow.	m^3s^{-1}	0.07
3	S	M	Q80Q50	Q80 flows relative to median.	-	0.51
4	S	M	logQVar	Variance in log flows.	m^3s^{-1}	0.37
5	S	M	Q90Q50	Q90 flows relative to median.	-	0.19
6	S	M	Q70Q50	Q70 flows relative to median.	-	0.09
7	S	R	RevPos	No. days when flow increases (positive reversals).	days	0.8

Figure 3. Overview of the three stages of the modified covariance approach to model parameterisation.

10   Stage 1: The complete parameter space of the hydrological model was sampled; the number of parameter sets considered is dependent upon the number of free parameters in the hydrological model and the accepted level of uncertainty. To address the issue of parameter sensitivity (Tong and Graziani, 2008; Wu et al., 2017), the parameter space was sampled uniformly based on Sobol quasi-random sequences (a Quasi-Monte Carlo method). Here, 100,000 independent parameter sets were selected.

Stage 2: For each parameter set, flow time-series were simulated based on the observed climate data. For each of these flow
15   time-series, a corresponding set of covariances (between observed climate and simulated flow) and HIs was computed. The observed covariance and HIs are also determined.

Stage 3: Before a parameter set was selected and evaluated, it was necessary to determine if the observed moments lie within the of the simulated moments (sampled parameter space). This was facilitated through plots (Fig. A2) of the observed and

simulated relationship between the (a) covariance between precipitation and flow, $\rho(P,Q)$, and HIs; and (b) covariance between potential evapotranspiration and flow, $\rho(PE,Q)$, and HIs. If the moments do not agree, the model is invalidated. Selection of a model parameter set was based on a specified limit of acceptability, i.e. the ability to replicate or minimise the error, between the observed and simulated covariance structures and HIs. In Vogel and Sankarasubramanian (2003) the focus

5 was on the replication of a single index. Here, the objective was the replication of multiple indices; to account for this, a limit of acceptability was specified per index, with the indices assigned maximum error thresholds based on their relative importance (Fig. 4). The index importance (Table 2) was normalised (rescaled to a range from zero to one) allowing the covariances to be assigned a relative importance of one (equal to the most important index). The limits of acceptability were determined through the linear relationship between the relative importance and a user-specified allowable error range (minimum and maximum;

10 see Fig. 4). Parameter sets which fall below this limit of acceptability were rejected. Here, the minimum and maximum error were specified as 17.5% and 35% ($2 \cdot error_{min}$) respectively; see also Table A1 in the appendix.

Figure 4. Boundaries of the limit of acceptability (shaded) for the case study selected parameter set. The lines indicate the relationship
15 **between the allowable error thresholds and relative importance.**

2.4 Model evaluation

The ability of the performance and consistency of the modified covariance approach was made with reference to the seven HIs (Table 2). The seven HIs were calculated annually for both the observed and simulated flow datasets over the 54-year period. The model was evaluated with reference to prior studies with similar modelling objectives (the replication of ecologically

20 relevant HIs): Shrestha et al. (2014); Vis et al. (2015); Shrestha et al. (2016); Pool et al. (2017). To permit such comparison, this study applied the same (non-parametric) evaluation metrics (Table 3). Two additional measures, Cramér-von Mises and the mean arctangent absolute percentage error (MAAPE), were considered to address limitations associated with certain metrics; see Table 3 for discussion. The modified covariance approach is considered relative to the above-mentioned studies; an overview of these works is provided in Tables B1 and B2.

**Hydrology and
Earth System
Sciences**

Open Access

EGU

Discussions

Table 3. Evaluation metrics. Unless indicated otherwise, measures are based upon Shrestha et al. (2014), 2016; Vis et al. (2015); Pool et al. (2017). In the definition of the indices, *I* is the index value and *n* the number of observations.

	Metric	Description	Definition (or R-function)	Optimal value
Statistical test, correlation	Spearman correlation (r_s)	A correlation coefficient. The strength of the linear relationship between two variables. Non-parametric, applicable when data is not normally distributed.	stats::cor(…, method = "spearman")	1; -1
	Welch's t-test	Variation on correlation where the two samples have unequal variances. Hypothesis is that two populations have equal means.	stats::t.test(…)	$p > 0.05$
Statistical test, distribution	Kolmogorov-Smirnov test (*KS*)	Tests whether samples come from the same population, i.e. follow the same distribution. Non-parametric, applicable when data is not normally distributed.	stats::ks.test(…)	$p > 0.05$
	Cramér-von Mises (*CvM*) (Cramér, 1928; Anderson, 1962)	Addresses limitations of KS test: (1) less focused on the central distribution; (2) more equal weighting on the tails of the distribution.	cramer::cramer(…) (Franz, 2014)	$p > 0.05$
Error	Model efficiency (*NSE*)	Nash Sutcliffe efficiency. A measure of the goodness of fit of the HI to the 1:1 line.	$1 - \dfrac{\sum(I_{obs} - I_{sim})^2}{\sum(I_{obs} - \overline{I_{obs}})^2}$	1
	Mean absolute relative error (*MARE*)	Mean absolute error relative to index measured.	$1 - \dfrac{1}{n}\sum\dfrac{|I_{obs} - I_{sim}|}{I_{obs}}$	1
	Mean arctangent absolute percentage error (MAAPE) (Kim and Kim, 2016)	A modification of MARE. Considers the relative error as an angle rather than a slope, reducing the bias of large errors.	$\dfrac{1}{n}\sum \arctan\left(\dfrac{I_{obs} - I_{sim}}{I_{obs}}\right)$	0
Additional descriptive stats	Normalised error (NSFC in Pool et al. (2017))	Index error relative to the range of values the index can take.	$\dfrac{I_{obs} - I_{sim}}{\Delta I_{obs}}$	0
	Hydrologic alteration factor (*HAF*)	A factor developed as part of the Indicators of Hydrologic Alteration (Mathews and Richter, 2007). Tests the replicability of sections of the probability distribution (lower-tail, IQR and upper-tail) for a given index.	$\dfrac{F_{sim} - F_{obs}}{F_{obs}}$ Where F is frequency, no. values lying within the probability distribution.	0

3 Results

3.1 Model parameters

5   The relationship between the covariance of the input and output time-series and each HI, for all 100,000 simulations, are summarised in Fig. A2. The observed moments can be seen to lie within the simulated moments, validating the use of the GR4J hydrological model. The parameters of the production (x1) and routing (x3) store capacities were estimated as 511 and 311 mm respectively; time elapsed for the routing of the flow is 1.17 days (x2). A positive groundwater exchange coefficient (x4) of 2.84 mm per day represents the inflow from the chalk aquifer. The percentage error for the covariances and HIs are

depicted in Fig. 4 previously. A covariance error range of -10-20% indicates a possible underestimation in the simulation of the observed flow. Error in the HIs is below ±10% in all instances, with the exception of the least important index, *riseMn*, where the error is equal to the upper threshold of +35%.

3.2 Model evaluation

The model is evaluated with reference to the distribution of each HI as well as the evaluation metrics used in Shrestha et al., 2014, 2016; Vis et al., 2015 and Pool et al., 2017 (Table 3). The ability of the model to replicate the seven HI is considered in terms of performance and consistency.

The distribution of the HIs is presented in Fig. 5. In the case of the empirical cumulative distribution functions (ECDF; top row), the level of agreement between the observed and simulated HIs is indicative of good overall performance. However, the probability density functions (PDF; bottom row) indicate a lack of consistency across the indicator *RevPos,* and the low flow indicators *(Q70Q50, Q80Q50* and *Q90Q50)*. With the tails being well replicated, issues lie principally within the central distribution. For *RevPos* this is confirmed by a lack of correlation (Spearman; Table A2) and the statistical tests, where the null hypothesis was not rejected (observed-simulated HI do not agree in terms of mean and distribution).

A summary of the NSE values and measures of error is provided in Fig. 6, optimal values are indicated by the dashed line; for the numerical values see Table A2. With a maximum of 0.54, the NSE values are suggestive of relatively poor model performance; again, highlighting difficulties in replicating the index *RevPos*. The metrics MARE and MAAPE suggest a more positive outcome, however consistency across the HIs is lacking. In contrast to the prior results, MARE indicates that the index *RevPos* is well replicated, with *riseMn* exhibiting the poorest performance. Conversely, MAAPE, which is intended to reduce the bias of large error values (see Table 3), identifies *riseMn* as the best performing index, whilst *10R90Log* is deemed worst.

The normalised error associated with each HI, a measure of the difference between observed-simulated values relative to the observed range, is précised in Fig. 7. The range of values that the indices *10R90Log* and *RevPos* can take are low (±0.25 and ±0.5 respectively), consequently, no inference can be made. In contrast to the ECDFs and PDFs, a high level of consistency is observed for the index *logQVar* and low flow indicators (across mean, range, distribution and bias of the normalised error). A strong negative bias, that was not clear in the Fig. 5distributions, is in evidence for the least important index (Table 2), *riseMn*.

The hydrologic alteration factor (HAF) is adapted from the IHA approach (Table 3). It is a measure of the simulated and observed frequencies of values within three target percentile ranges: 0-25th, 25-75th, and 75-100th. As a measure of distribution, HAF is essentially a simplification of the distribution functions in Fig. 5. The acceptable range of HAF values is defined as ±0.33 (Mathews and Richter, 2007). HAF values for each HI are presented in Fig. 8.

The central distribution (25-75th percentiles) of all HIs lie within this acceptable range, indicating good performance and consistency. Performance across the tails of the distributions is generally good (lying within or on the bounds of the acceptable range), though there is a lack of consistency in the direction of bias. With a HAF value of 1.82 (not pictured to preserve figure resolution), the positive bias in the index *RevPos* exhibits the largest such deviation. With bias at both tails, inconsistencies in the replication of the index *riseMn* are again highlighted.
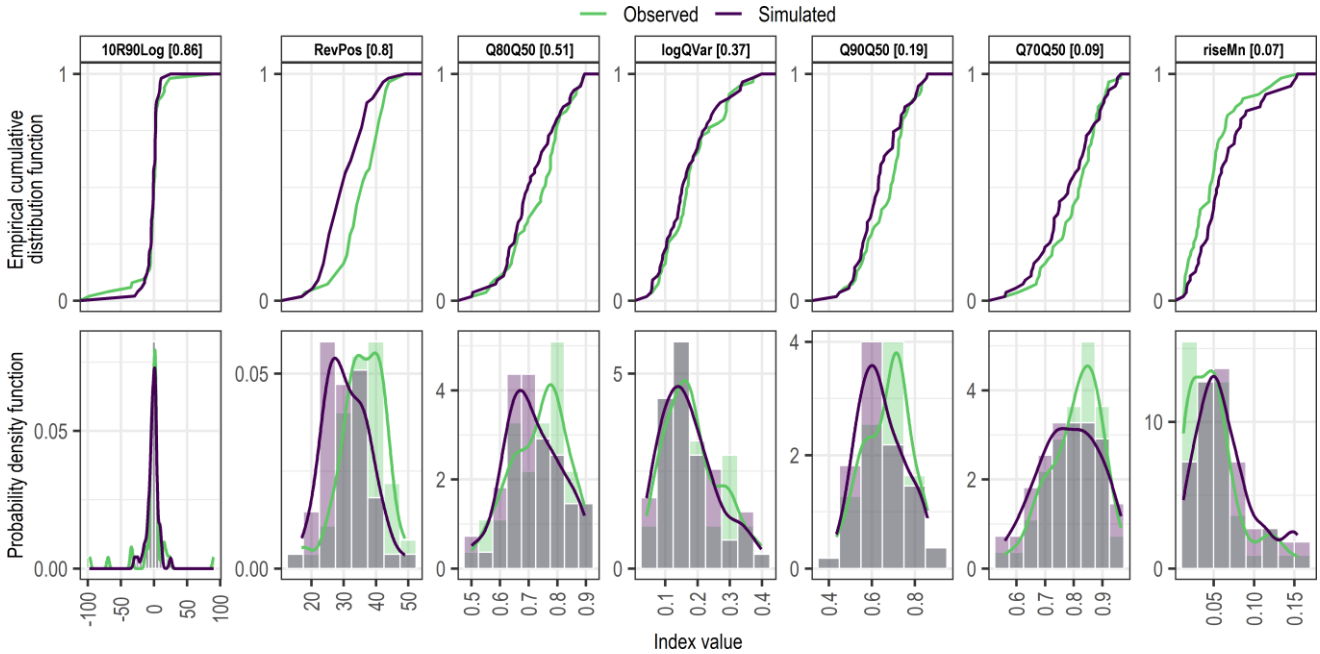
Figure 5. Empirical cumulative distribution functions (top) and probability density functions (bottom) for observed-simulated HIs.
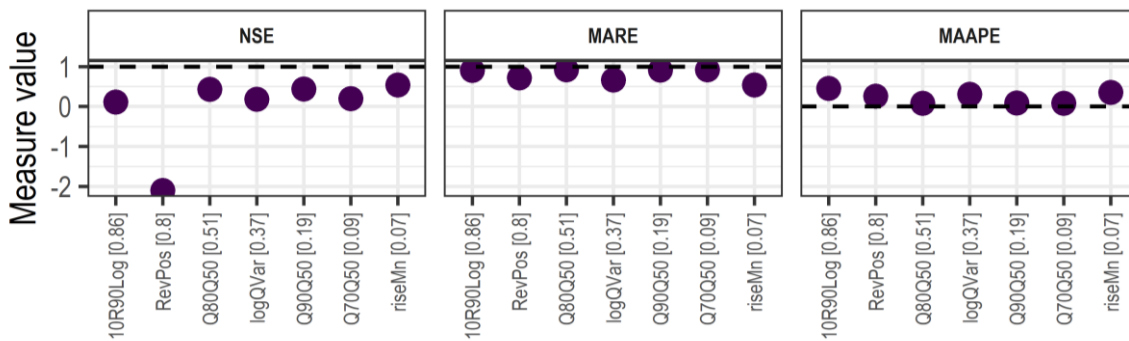

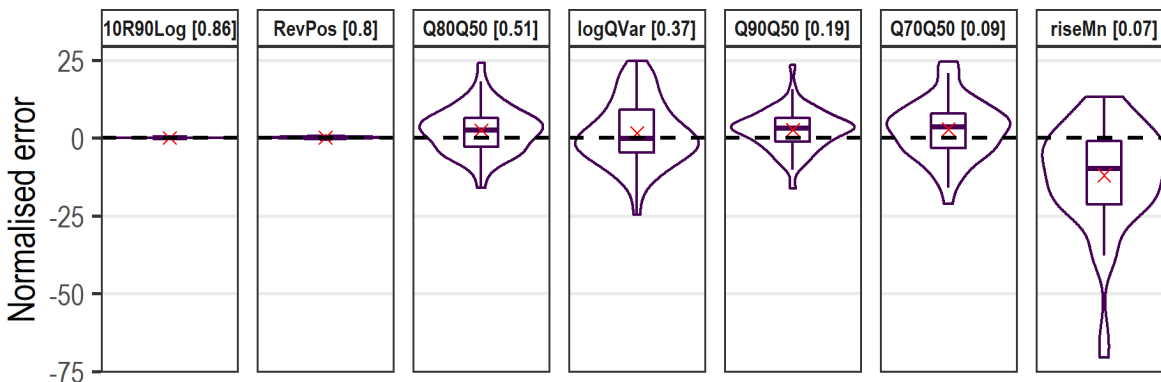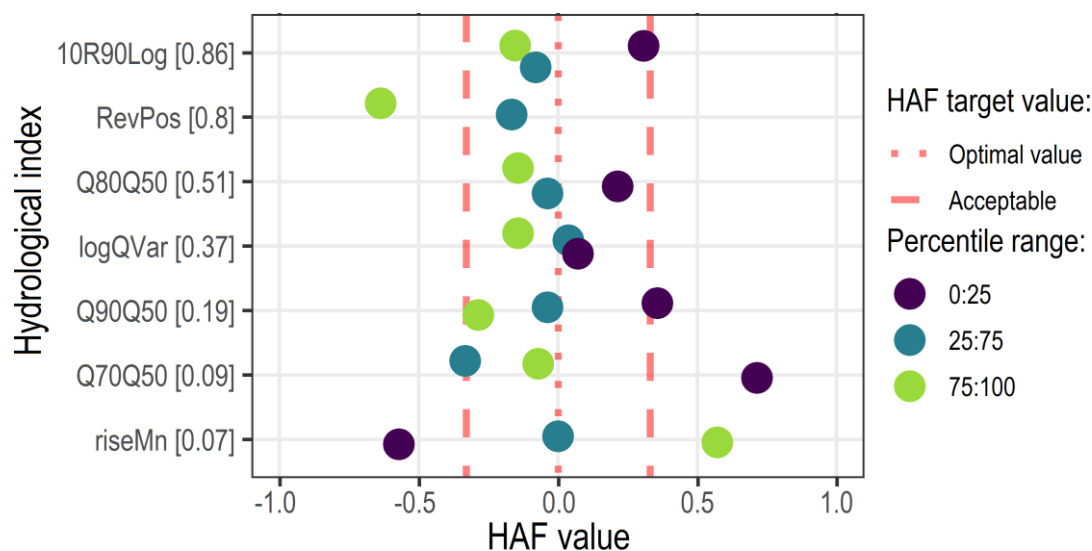**Figure 6. Evaluation metric summary by HI; optimal values are indicated by the dashed line.**

5

Figure 7. Normalised error (percentage) for each HI. A value of zero represents the optimum (no error; dashed line).

Figure 8. Hydrologic alteration factor (HAF) values for the three percentile ranges for each HI. The acceptable range of HAF values
is defined as ±0.33; HAF > 0 represents an increas in frequenecy relative to the observed whilst HAF < 0 represents a decrease.

4 Discussion

4.1 Modified covariance approach

There is a clear need to understand the impact of hydrologic change on the river ecosystem. To assess this, hydrological models are used to simulate flow time-series from which HI of ecological relevance are derived. In this study, a modification of Vogel and Sankarasubramanian's (2003) covariance approach was considered, with a focus on the replication of a suite of seven ecologically relevant HIs.

The first aim of this study was to determine whether this modified covariance approach is able to satisfactorily replicate the suite of HIs. The hydrological model was successfully parameterised with observed moments lying within the bounds of the simulated moments for all HIs. Overall, replication of the HIs was good. Indices related to magnitude where best replicated, whilst difficulties were observed in replicating rate of change and integer indices.

In terms of distribution (Fig. 5) and evaluation metrics (Table 3), the best performing and most consistent HIs are the measures of magnitude: *logQVar*, *Q80Q50* and *Q90Q50*. This is a clear indication that the model can successfully replicate the variation in flow and quantiles (specifically low-flows, Q80 and Q90). Measures of model performance associated with the remaining four indices present a conflicting picture, leading to a lack of clarity as to the full capacities of the hydrological model and relative success of the covariance approach. The two most important indices, *10R90Log* and *RevPos*, are among the less well-replicated; however, this may be due to their sheer inherent complexity (discussed further below).

4.2 Comparison

A number of studies have investigated the ability of hydrological models to replicate ecologically relevant HI. Subjecting the outcomes to a comparison is the second objective of this study. The comparative studies, Shrestha et al., 2014, 2016, Vis et al., 2015 and Pool et al., 2017, follow a traditional calibration-validation approach, considering an array of objective functions

5   and performance metrics (Table 3); see Appendix B for details. Comparison is made with reference to the facets of the flow regime, specifically magnitude and rate of change (Table 2).

Both Shrestha et al. (2014) and Vis et al. (2015) observed poor model performance in the replication of low flow HIs. Consistent with the literature (Westerberg et al., 2011; Pushpalatha et al., 2012), Shrestha et al. (2014) attribute this to the use of objective functions tuned to high-flow periods (i.e. NSE and volume error; Table B1). As shown in this study and Pool et

10   al., 2017, this is largely redressed through explicit consideration of low-flow HIs in parameterisation of the hydrological model. Mean flows are similarly accounted for in Pool et al., 2017. Performance and consistency in the replication of high flow indicators was consistent across all studies, with the exception of Shrestha et al., 2016, where summer (June-September) high flows exhibit a distinct negative bias. No studies, this work inclusive, observed difficulties in replicating indicators related to flow variability directly.

15   Whilst inconsistency in the replication of the rate of change indicator *RevPos* is clear, a lack of agreement in the evaluation metrics leads to difficulties in assessing the performance of *riseMn*. Such observations are found consistently across three out of the four studies: Shrestha et al., 2016 excluded frequency and rate of change indicators due to large negative bias observed in Shrestha et al., 2014, whereas Vis et al., 2015 saw inconsistencies across the calibration-validation and performance metrics (NSE and Spearman). Performance improvements were, however, seen in Pool et al., 2017 when the HI is considered as the

20   objective function.

Of the studies considered, only Vis et al., 2015 and Pool et al., 2017 are directly linked to the outcomes of hydroecological modelling, replicating a suite of ecologically relevant HIs. In Vis et al. (2015) performance was inconsistent, varying considerably across HIs and objective function; model evaluation was limited due to the focus on model efficiency (NSE) as an evaluation metric. Pool et al. (2017) concluded that the choice of objective function strongly influenced the accuracy in

25   replication, with the best results achieved when the models were calibrated on the HI of interest.

It is clear that no approach has been able to achieve adequate performance and consistency in the replication of more complex HIs, specifically those related to rate of change. Whilst Pool et al., 2017 saw improvements, the need to calibrate the model to each HI in question would strongly call into question the reliability of the hydrological model (due to the inability of the hydrological model to simulate catchment hydrological processes simultaneously). The consistency with which (the majority

30   of the) HIs are replicated here illustrates that this is not a necessary limitation of hydrological models.

4.3 Evaluation metrics

This work, and the comparative studies, highlight a number of shortcomings in the evaluation metrics used in the evaluation of hydrological models. One consequence being difficulty in evaluating the most important indices. Considerable conflicts have arisen between and across the measures; the following problems are highlighted: (1) the statistical tests of agreement

5 (Table 3) are generally limited to the mean or central distribution; (2) the error measures NSE and MARE exhibit known bias (Table 3; Vis et al. (2015); Kim and Kim (2016)); (3) the HAF index is not well-suited for the evaluation of HIs which are integer counts or dimensionless. Number 3 is best illustrated through consideration of the index *RevPos*, an integer count of the number of positive reversals in the summer season. When assessing *RevPos* based on percentile ranges (as per HAF), integers equal to the percentile boundary are not considered (see Fig. A1). This suggests that HAF may not be applicable for

10 those indices which take integer values (i.e. counts, days or Julian day time of year). Indeed, the difficulties observed here and in the comparative studies suggest that certain HIs, such as rate of change indicators like *RevPos*, may simply not be practically replicable by a hydrological model. Given redundancy in many HIs, it may be possible to identify another more suitable index capable of providing the same information. Such efforts would not arbitrarily improve the replication of the HIs, but rather, confidence in the performance of the outputs.

15 ### 4.4 Addressing the limitations of traditional hydrological model calibration

As discussed at the outset of this paper, a number of limitations of a traditional approach to hydrological model calibration have been identified. These include: (1) bias and uncertainty as a result of measurement error, i.e. disinformative data; (2) the arbitrary nature of GOF behaviour thresholds; and (3) equifinality. Determining whether the modified covariance approach serves to address any of these limitations represents the final aim of this study.

20 #### 4.4.1 Disinformative data

Models calibrated following a traditional approach are particularly sensitive to measurement error (Westerberg et al., 2011). Lack of agreement in the observed-simulated time-series, even for a single event, may bias the objective function, leading to rejection of an otherwise well-performing parameter set (Beven, 2010; Westerberg et al., 2011). Methods which do not focus on the replication of time-series, such as the modified covariance approach, limit the influence of input uncertainty (Westerberg

25 et al., 2011; Euser et al., 2013). Additionally, length of the time-series is a significant factor, with shorter time-series featuring greater bias (Westerberg et al., 2011); depending on the area of application, long-term climate variability (e.g. El Niño/La Niña) may exacerbate this. It is worth noting that, amongst the comparative studies, none feature data in excess of 29 years; it is possible that, an alternative approach to parameterisation that does not focus on the time-series, such as the covariance approach, may reduce input uncertainty, leading to improvements in the replication of indicators.

4.4.2 Behaviour thresholds

The observed and simulated moments (Fig. A2) clearly illustrate whether a given hydrological model structure is able to capture the hydrological processes in the catchment. In this way, the modified covariance approach is not dependent on an arbitrary behavioural threshold to validate the use of the hydrological model. However, in the absence of a numerical measure

5 of the relative importance of each HI, an element of subjectivity is necessarily introduced into the parameterisation of the model. An approach such as the Generalised Likelihood Uncertainty Evaluation (GLUE) framework (Beven and Binley, 2014) may represent a viable alternative where HI importance is unspecified or irrelevant.

4.4.3 Equifinality and parameter space

Equifinality, reaching the same outcome by different means, is a major challenge of hydrological modelling. In the modified
10 covariance approach the entire parameter space is considered. The range of possible solutions, i.e. parameter sets, is reduced by focussing on the region which is best able to replicate the characteristics of the HIs, thereby reducing the uncertainty associated with equifinality (Wu et al., 2017). Additionally, the approach ensures the selection of the global optimum.
In hydrological uncertainty analyses, the size of the parameter space is highly variable; for example, Wilby (2005) considered 10,000 simulations, whilst Ballio and Guadagnini (2004) looked at 200,000. Here, a total of 100,000 simulations were
15 considered in order to verify the method of investigation; upscaling this for the 16 parameter HBV model in Vis et al., 2015 and Pool et al. would necessitate 400,000 simulations. Whilst the large number of simulations may seem prohibitive, this demand may be offset. Unlike the calibration-validation paradigm, where selection algorithms may introduce issues of speed and accuracy (Seibert, 2000), finite time is needed to apply the covariance approach. All simulations of the hydrological model are performed at the outset; once the full suite of parameter sets have been simulated the hydrological model need not be run
20 again. In traditional calibration-validation, where the HI serves as the objective function (e.g. Pool et al., 2017), the HIs must be specified at the outset. This is not the case in the modified covariance approach; the n Monte Carlo simulations can be performed in advance of HI selection. Further, multiple sets of HIs may be considered at a time (e.g. all rate of change or magnitude indicators), or at a later date, with limited additional time outlay.

4.4.4 Additional limitations

25 The outcomes of this, and comparative studies, highlight the present inability of hydrological to simulate a wide range of HIs concurrently (e.g. rate of change plus other facets of the flow regime). Any attempted improvements may, f come at the cost of parsimony and equifinality. Additionally, as a simplification of the hydrologic system, it is impossible, by definition, for a hydrological model to accurately replicate all aspects of the flow regime simultaneously (Beven, 2012b). For this reason, it may be that a focus on the replication of HIs, leads to a poor representation of the flow hydrograph (Seibert, 2000), limiting
30 the use of such models to the initial modelling objective (replication of HIs).

4.5 Wider applicability and further work

The applicability of the modified covariance approach is not limited to hydroecological studies and the simulation of ecologically relevant HI, being suited to the simulation of any HIs or hydrological signatures. Indeed, a focus on hydrological signatures may serve to improve the simulation of underlying hydrological processes (Seibert, 2000; Euser et al., 2013). In

5   this context, example applications include the replication of water resource management indicators (monthly, seasonal and annual flows). Additionally, the clarity with which model structures are accepted or rejected makes the approach ideally suited for use in combination with model selection frameworks such as the Framework for Assessing the Realism of Model Structures (FARM; Euser et al. (2013)).

There remains a need to explore the capability of the modified covariance approach in replicating a wide range of HIs,

10   specifically the indicators of hydrologic alteration (IHA). Given the inherent limitations on the number of indices which may be considered at a given time, this may be explored through application to a database of catchments of different flow regimes and/or hydrological models.

5 Concluding remarks

The performance and consistency of a modified covariance approach, relative to traditional calibration-validation, was

15   explored. The focus of the study was the replication of ecologically relevant HIs. Compared to similar studies, the modified covariance approach shows an ability to maintain both consistency and performance when replicating hydrological indicators. A major advantage of the approach lies in the identification of the region of parameter space which best captures (replicates) the characteristics of the HIs, thereby providing a greater understanding of the suitability, limitations and uncertainties of the hydrological model. In order to understand the range of HIs which may be considered at a given time, further work across a

20   database of catchments and the range of IHAs remains necessary.

Data availability: The data used in this study is freely available from the Met Office (2016) and the National River Flow Archive (CEH, 2018). See section 2.3.1 Data for additional details.

25   *Author contributions:* AV developed the code and performed the data analysis. AV prepared the manuscript whilst LB provided review and edits. Both LB and SP provided supervision.

Competing interests. The authors declare that they have no conflict of interest.

Hydrology and
Earth System
Sciences

Discussions

Appendix A – Supplementary data and results

Table A1. Linearly determined limits of acceptability for each HI and covariance; see also Fig. 4.

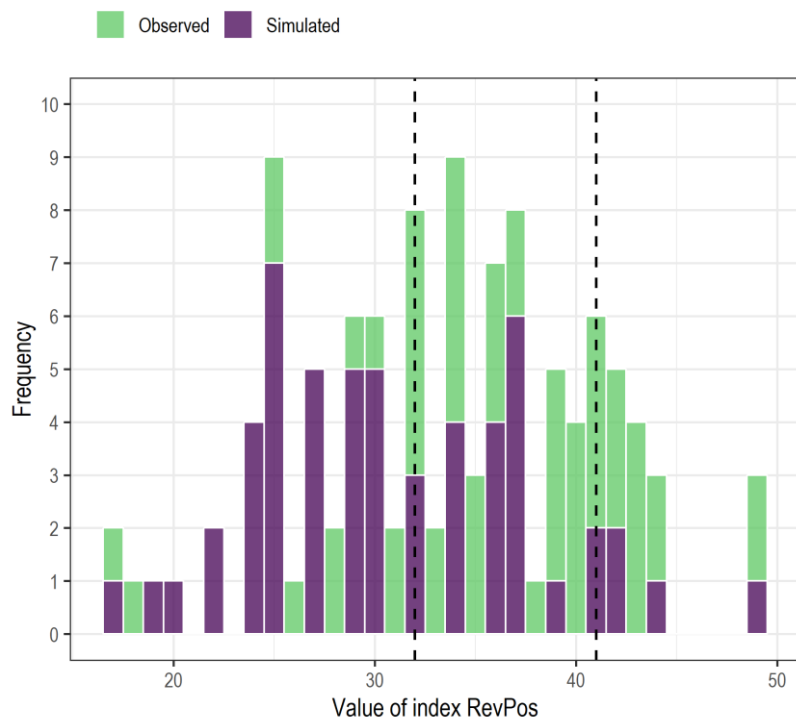| Variable type	Variable	Importance	Relative importance	Limit of acceptability – absolute error (%)	Error (%)
Covariance	P	-	1.00	17.50	-13.54
Covariance	PE	-	1.00	17.50	-17.48
HI	10R90Log	0.86	1.00	17.50	7.16
HI	RevPos	0.80	0.92	18.83	8.45
HI	Q80Q50	0.51	0.56	25.25	0.80
HI	logQVar	0.37	0.38	28.35	-4.36
HI	Q90Q50	0.19	0.15	32.34	10.69
HI	Q70Q50	0.09	0.03	34.56	-0.74
HI	riseMn	0.07	0.00	35.00	34.36

Table A2. Model evaluation summary by HI; see also Fig. 6. NA denotes where an approach is not applicable for a given measure. Optimal values are indicated in Table 3.

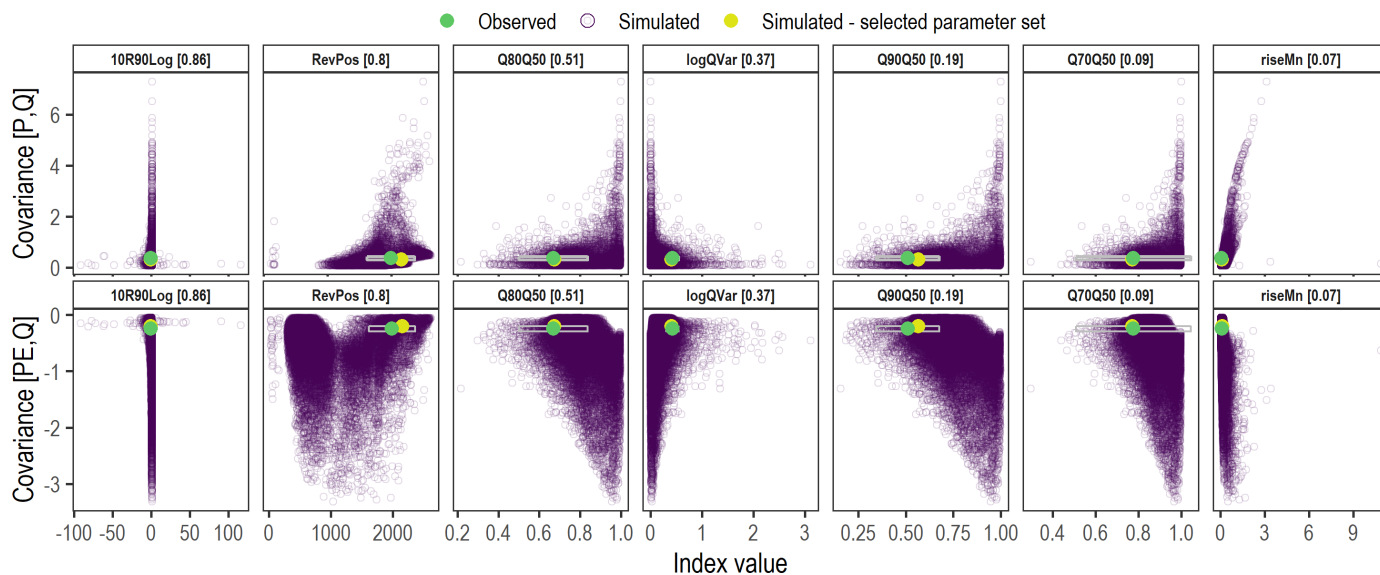| Index	Importance	Spearman corr.	NSE	MARE	MAAPE
10R90Log	0.86	NA	0.11	0.90	0.45
RevPos	0.8	-0.31	-2.10	0.72	0.26
Q80Q50	0.51	0.73	0.43	0.92	0.08
logQVar	0.37	0.62	0.18	0.66	0.31
Q90Q50	0.19	0.76	0.44	0.91	0.09
Q70Q50	0.09	NA	0.20	0.92	0.08
riseMn	0.07	0.81	0.54	0.54	0.35

Figure A1. Histogram of the observed (green) and simulated (purple) values for the index *RevPos*. The dashed lines indicate the lower and upper boundaries for the HAF.

Figure A2. Relationships between (a) the covariance between precipitation and flow, $\rho(P, Q)$, and each HI; and (b) the covariance between potential evapotranspiration and flow, $\rho(PE, Q)$, and each HI. The shaded points depict the moments of the observed data and selected parameter sets; the hollow points depict the 100,000 Monte Carlo simulations. Grey boxes depict the boundaries of the limits of acceptability for each index as defined in Fig. 4.

Appendix B – Comparative studies

The first two studies, Shrestha et al. (2014) and 2016, are focussed on the replication of the full suite of IHA's rather than case study specific ecologically relevant HIs. Informed by the findings in the first study, Shrestha et al. (2016) excluded rate of

5 change and frequency indicators. The studies conducted by the second research group in Table B1 explicitly focus on replication of ecologically HIs with selection informed by previous hydroecological studies. Uncertainty is reduced through the consideration of longer time-series and multiple (calibration) iterations. A total of four HIs can be considered as analogous with four considered here: RA7, E85, MH10 and MA41; summarised in Table B2. Unfortunately, direct comparison with Pool et al. (2017) in terms of HAF was restricted due to the focus on the central distribution.

10

Table B1. Overview of the studies considered in section Error! Reference source not found. **Comparison.**

Research group	Authors	Case study	Model	Free parameters	No. years	Repetitions	Objective function	Ecologically relevant	No.
							Hydrological modelling	**HIs**	
1	Shrestha, Peters, and Schnorbus (2014)	Fraser catchment, Canada; 2 of 66 sub-catchments	VIC; semi-distributed	5	14	5	(1) Nash Sutcliffe Efficiency (NSE; also known as model efficiency, see Table 3); (2) NSE_{log}, (Pushpalatha et al., 2012); (3) Volume error.	No	38
	Shrestha, Schnorbus and Peters (2016)	Fraser catchment, Canada; 1 of 4 sub-catchments			23	1		No	24
2	Vis, Knight, Pool, Wolfe and Seibert (2015)	Tennessee-Cumberland River catchment, USA; 27 sub-catchments	HBV; semi-distributed	16	29	100	(1) NSE; (2) NSE_{log}; (3) Lindstrom, a combination of NSE and volume error; (4) MARE (Table 3); (5) Spearman rank correlation; (6) Volume error; (7) three objective functions combining the above statistical criteria.	Yes	12
	Pool, Vis, Knight and Seibert (2017)	Tennessee River catchment, USA; 25 sub-catchments					Five functions based on combinations of NSE and error; single and multiple indices.	Yes	16

Table B2. Hydrological indicators considered in Vis et al. (2015) and Pool et al. (2017). The relative comparability with indices in this study are indicated.

Index	Facet	Streamflow characteristic	Comparability	Analogous indices
RA7	Rate of change	Rate of runoff recession, i.e. log-measure of fall rate.	Direct	*riseMn* (rise rate)
			Indirect	Replication of log indices generally: *10R90Log* and *logQVar*
E85	Magnitude	The lowest 15% of daily flow, i.e. Q85.	Direct	Low-flow indices: *Q70Q50*, *Q80Q50* and *Q90Q50*
MH10	Magnitude	Maximum October flow.	Indirect	Magnitude facet.
MA41	Magnitude	Mean annual flow.	Indirect	

References

Anderson, T. W.: On the Distribution of the Two-Sample Cramer-von Mises Criterion, Ann. Math. Statist., 33, 1148-1159, http://dx.doi.org/10.1214/aoms/1177704477, 1962.

Ballio, F., and Guadagnini, A.: Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology, Water Resources Research, 40, W04603, http://dx.doi.org/10.1029/2003WR002876, 2004.

Beven, K.: A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18-36, http://dx.doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

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

Beven, K. J.: Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models, Hydrological Processes, 24, 1537-1547, http://dx.doi.org/10.1002/hyp.7718, 2010.

Beven, K. J.: 1 Down to Basics: Runoff Processes and the Modelling Process, in: Rainfall-runoff modelling: The Primer, 2nd ed., Wiley-Blackwell, Chichester, 2012a.

Beven, K. J.: 7 Parameter Estimation and Predictive Uncertainty, in: Rainfall-runoff modelling: The Primer, 2nd ed., Wiley-Blackwell, Chichester, 2012b.

The Brisbane Declaration. Declaration of the 10th International Environmental Flows Conference, Brisbane, Australia, 3–6 September, 2007., 2007.

Carlisle, M., Falcone, J., Wolock, M., Meador, R., and Norris, H.: Predicting the natural flow regime: models for assessing hydrological alteration in streams, River Research and Applications, 26, 118-136, http://dx.doi.org/10.1002/rra.1247, 2010.

33007 - Nar at Marham - Gauged Daily Flow: http://nrfa.ceh.ac.uk/data/station/meanflow/33007, access: 2018-05-07, 2018.

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, W05552, http://dx.doi.org/10.1029/2011WR011721, 2012.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling & Software, 94, 166-171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

airGR: Suite of GR hydrological models for precipitation-runoff modelling. R package version 1.0.10.11: http://webgr.irstea.fr/modeles/?lang=en, 2018.

Cramér, H.: On the composition of elementary errors, Scandinavian Actuarial Journal, 1928, 13-74, https://doi.org/10.1080/03461238.1928.10416862, 1928.

5   De Girolamo, A. M., Barca, E., Pappagallo, G., and Lo Porto, A.: Simulating ecologically relevant hydrological indicators in a temporary river system, Agricultural Water Management, 180, 194-204, https://doi.org/10.1016/j.agwat.2016.05.034, 2017.

Efstratiadis, A., and Koutsoyiannis, D.: One decade of multi-objective calibration approaches in hydrological modelling: a review, Hydrological Sciences Journal, 55, 58-78, https://doi.org/10.1080/02626660903526292, 2010.

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

cramer: Multivariate nonparametric Cramer-Test for the two-sample-problem. R package version 0.9-1. https://cran.r-project.org/web/packages/cramer/index.html, 2014.

Garbe, J., Beevers, L., and Pender, G.: The interaction of low flow conditions and spawning brown trout (Salmo trutta) habitat availability, Ecological Engineering, 88, 53-63, https://doi.org/10.1016/j.ecoleng.2015.12.011, 2016.

15 Gilvear, D. J., Beevers, L. C., O'Keeffe, J., and Acreman, M.: Chapter 8 - Environmental Water Regimes and Natural Capital: Free-Flowing Ecosystem Services, in: Water for the Environment, Academic Press, 151-171, 2017.

Gleick, P. H.: WATER IN CRISIS: PATHS TO SUSTAINABLE WATER USE, Ecological Applications, 8, 571-579, http://dx.doi.org/10.1890/1051-0761(1998)008[0571:WICPTS]2.0.CO;2, 1998.

Gleick, P. H.: Water strategies for the next administration., Science, 354, 555-556, 2016.

20 Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resources Research, 34, 751-763, https://doi.org/10.1029/97WR03495, 1998.

Kim, S., and Kim, H.: A new metric of absolute percentage error for intermittent demand forecasts, International Journal of Forecasting, 32, 669-679, https://doi.org/10.1016/j.ijforecast.2015.12.003, 2016.

Klaar, M. J., Dunbar, M. J., Warren, M., and Soley, R.: Developing hydroecological models to inform environmental flow standards: a case
25 study from England, Wiley Interdisciplinary Reviews: Water, 1, 207-217, https://doi.org/10.1002/wat2.1012, 2014.

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

Kokkonen, T. S., and Jakeman, A. J.: Chapter 14 Structural effects of landscape and land use on streamflow response, in: Developments in Environmental Modelling, edited by: Beck, M. B., Elsevier, 303-321, 2002.

30 Le Moine, N., Andréassian, V., and Mathevet, T.: Confronting surface- and groundwater balances on the La Rochefoucauld-Touvre karstic system (Charente, France), Water Resources Research, 44, W03403, http://dx.doi.org/10.1029/2007WR005984, 2008.

Mathews, R., and Richter, B.: Application of the Indicators of Hydrologic Alteration Software in Environmental Flow Setting1, JAWRA Journal of the American Water Resources Association, 43, 1400-1413, https://doi.org/10.1111/j.1752-1688.2007.00099.x, 2007.

MIDAS   Record   listing:
35 http://catalogue.ceda.ac.uk/list/?return_obj=ob&id=1184,1251,1256,1225,1259,1228,1231,1234,1267,1204,1195,1207,1241,1263,1244,1214,1247, access: 2016-10-09, 2016.

Monk, W. A., Wood, P. J., Hannah, D. M., Wilson, D. A., Extence, C. A., and Chadd, R. P.: Flow variability and macroinvertebrate community response within riverine systems, River Research and Applications, 22, 595-615, http://dx.doi.org/10.1002/rra.933, 2006.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre,
5   G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022, Hydrological Sciences Journal, 58, 1256-1275, http://dx.doi.org/10.1080/02626667.2013.809088, 2013.

Murphy, J. C., Knight, R. R., Wolfe, W. J., and W, S. G.: Predicting Ecological Flow Regime at Ungauged Sites: A Comparison of Methods, River Research and Applications, 29, 660-669, https://doi.org/10.1002/rra.2570, 2012.

10   NRT: The River Nar - A Water Framework Directive Local Catchment Plan, Norfolk Rivers Trust, Norfolk, 2012.

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, http://dx.doi.org/10.1002/rra.700, 2003.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling,
15   Journal of Hydrology, 303, 290-306, http://dx.doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Pelletier, P. M.: Uncertainties in the single determination of river discharge: a literature review, Canadian Journal of Civil Engineering, 15, 834-850, https://doi.org/10.1139/l88-109, 1988.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275-289, http://dx.doi.org/10.1016/S0022-1694(03)00225-7, 2003.

20   Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., and Le Moine, N.: Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments, Water Resources Research, 44, W08447, http://dx.doi.org/10.1029/2007WR006579, 2008.

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.

25   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.

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

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

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, http://dx.doi.org/10.1046/j.1523-1739.1996.10041163.x, 1996.

Sear, D. A., Newson, M., Old, J. C., and Hill, C.: Geomorphological appraisal of the River Nar Site of Special Scientific Interest. (N684), English Nature, Peterborough, 2005.

35   Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215-224, https://doi.org/10.5194/hess-4-215-2000, 2000.

Shiklomanov, I.: World Fresh Water Resources, in: Water in Crisis: A Guide to the World's Fresh Water Resources, edited by: P, G., Oxford University Press, New York, 13-24, 1993.

Shrestha, R., Schnorbus, M., and Peters, D.: Assessment of a hydrologic model's reliability in simulating flow regime alterations in a changing climate, Hydrological Processes, 30, 2628-2643, https://doi.org/10.1002/hyp.10812, 2016.

5   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.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A., and Cosgrove, B. A.: The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments, Journal of Hydrology, 418-419, 3-16, https://doi.org/10.1016/j.jhydrol.2011.08.055, 2012.

10 EflowStats: An R package to compute ecologically-relevant streamflow statistics: https://github.com/USGS-R/EflowStats, 2013.

Tong, C., and Graziani, F.: A Practical Global Sensitivity Analysis Methodology for Multi-Physics Applications, in: Computational Methods in Transport: Verification and Validation, Berlin, Heidelberg, 2008, 277-299,

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.

15 Visser, A., Beevers, L., and Patidar, S.: Macro-invertebrate Community Response to Multi-annual Hydrological Indicators, River Research and Applications, 33, 707-717, http://dx.doi.org/10.1002/rra.3125, 2017.

Visser, A., Beevers, L., and Patidar, S.: Complexity in hydroecological modelling, a comparison of stepwise selection and information theory, River Research and Applications, https://doi.org/10.1002/rra.3328, 2018a.

Visser, A., Beevers, L., and Patidar, S.: A coupled modelling framework to assess the hydroecological impact of climate change [In review],
20 2018b.

Vogel, R. M., and Sankarasubramanian, A.: Validation of a watershed model without calibration, Water Resources Research, 39, http://dx.doi.org/10.1029/2002WR001940, 2003.

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,
25 http://dx.doi.org/10.1038/nature09440, 2010.

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, Hydrol. Earth Syst. Sci., 15, 2205-2227, https://doi.org/10.5194/hess-15-2205-2011, 2011.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrological Processes, 19, 3201-3219, http://dx.doi.org/10.1002/hyp.5819, 2005.

30 Williams, J.: Building Hydrologic Foundations for Applications of ELOHA: How Long A Record Should You Have?, River Research and Applications, 34, 93-98, https://doi.org/10.1002/rra.3143, 2017.

World Water Assessment Programme: The United Nations World Water Development Report 3: Water in a Changing World, Paris, UNESCO and London, Earthscan, 2009.

Wu, Q., Liu, S., Cai, Y., Li, X., and Jiang, Y.: Improvement of hydrological model calibration by selecting multiple parameter ranges,
35 Hydrol. Earth Syst. Sci., 21, 393-407, https://doi.org/10.5194/hess-21-393-2017, 2017.

You, G. J.-Y., Thum, B.-H., and Lin, F.-H.: The examination of reproducibility in hydro-ecological characteristics by daily synthetic flow models, Journal of Hydrology, 511, 904-919, https://doi.org/10.1016/j.jhydrol.2014.02.047, 2014.