




## RESEARCH ARTICLE

# When is a hydrological model sufficiently calibrated to depict flow preferences of riverine species?

Jens Kiesel<sup>1,2</sup>  | Karan Kakouei<sup>1,3</sup> | Björn Guse<sup>2,4</sup> | Nicola Fohrer<sup>2</sup>  |  
 Sonja C. Jähnig<sup>1</sup> 

<sup>1</sup>Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

<sup>2</sup>Department of Hydrology and Water Resources Management, Institute for Natural Resource Conservation, Christian-Albrechts University of Kiel, Kiel, Germany

<sup>3</sup>Institute of Biology, Free University of Berlin, Berlin, Germany

<sup>4</sup>Section Hydrology, GFZ German Research Centre for Geosciences, Potsdam, Germany

## Correspondence

Jens Kiesel, Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, Berlin 12587, Germany. Email: kiesel@igb-berlin.de

## Funding information

Bundesministerium für Bildung und Forschung, Grant/Award Number: 01LN1320A; Deutsche Forschungsgemeinschaft, Grant/Award Number: 1466/1-1; Horizon 2020 Framework Programme, Grant/Award Number: 642317

## Abstract

Riverine species have adapted to their environment, particularly to the hydrological regime. Hydrological models and the knowledge of species preferences are used to predict the impact of hydrological changes on species. Inevitably, hydrological model performance impacts how species are simulated. From the example of macroinvertebrates in a lowland and a mountainous catchment, we investigate the impact of hydrological model performance and the choice of the objective function based on a set of 36 performance metrics for predicting species occurrences. Besides species abundance, we use the simulated community structure for an ecological assessment as applied for the Water Framework Directive. We investigate when a hydrological model is sufficiently calibrated to depict species abundance. For this, we postulate that performance is not sufficient when ecological assessments based on the simulated hydrology are significantly different (analysis of variance,  $p < .05$ ) from the ecological assessments based on observations. The investigated range of hydrological model performance leads to considerable variability in species abundance in the two catchments. In the mountainous catchment, links between objective functions and the ecological assessment reveal a stronger dependency of the species on the discharge regime. In the lowland catchment, multiple stressors seem to mask the dependence of the species on discharge. The most suitable objective functions to calibrate the model for species assessments are the ones that incorporate hydrological indicators used for the species prediction.

## KEYWORDS

ecological assessment, hydrological modelling, model optimization, species abundance, species preferences

## 1 | INTRODUCTION

Streamflow is one of the most important abiotic parameters that govern the occurrence and distribution of freshwater biota (Poff et al.,

1997; Wu et al., 2018). In particular, macroinvertebrates have evolved distinct adaptations to flow conditions and hence are affected by their changes (Domisch et al., 2017). These dependencies on flow have been used frequently to assess the occurrence (Pyne & Poff, 2017) and diversity (Poff & Zimmerman, 2010) of riverine species, such as fish (O'Keeffe et al., 2018), benthic invertebrates (Armanini, Horrigan, Monk, Peters, & Baird, 2011), or phytoplankton (Qu, Wu, Guse, &

Jens Kiesel and Karan Kakouei contributed equally to this paper.

Nicola Fohrer and Sonja C. Jähnig share last authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Ecohydrology published by John Wiley & Sons Ltd

Fohrer, 2018), using microcosm experiments (Ceola et al., 2013), statistical models (Kakouei et al., 2018), or process-based models (Mondy & Schuwirth, 2017). In the absence of direct measurement data or for scenario assessments, modelled streamflow is often used as a data basis for such analysis. Although it is important to match the spatial scales on which streamflow, for example, catchment scale, is produced and species are modelled, for example, habitat scale, uncertainties and inaccuracies in simulated streamflow remain and will consequently affect the simulated species.

Hydrological simulations are impacted by the quality as well as the spatio-temporal simplification of the input data (Melsen et al., 2016), which hydrological model type, model algorithms, and depicted processes are chosen (Fenicia, Kavetski, & Savenije, 2011), the type and mathematical formulation of the model algorithms (Clark et al., 2015), the uncertainty and equifinality of the model parameters (Beven, 2007), the quality and type of observations (e.g., time step interval, time period length, and quality of rating curves), to which the model results are compared (Seibert & McDonnell, 2002), and the type (Pfannerstill et al., 2017) and the number of objective functions used to parameterize the model (Shafii & Tolson, 2015). The factor integrating all of these dependencies is the overall model performance, measured by a variety of hydrological metrics that compare simulations to observations (Guse et al., 2017; Reusser, Blume, Schaeffli, & Zehe, 2009) and that are used for model calibration and validation.

Hydrological literature is available that contains guidelines and thresholds for certain metrics that enable an assessment of when model performance is sufficient for hydrological applications; for instance, Ritter and Muñoz-Carpena (2013) list limits for the Nash-Sutcliffe efficiency (NSE) above which hydrological model performance is acceptable, good, and very good. Such recommendations do not exist for ecological applications because, until recently, it was not possible to assess how hydrological model performance impacts species responses because, to our knowledge, no quantitative link between flow and macroinvertebrate abundances existed. For Germany, Kakouei, Kiesel, Kail, Pusch, and Jähnig (2017) established these flow-species linkages for macroinvertebrates. By applying these linkages on simulated streamflow, it can now be tested how modelled species abundance changes for different hydrological model performances.

Kakouei et al. (2017) developed these linkages using the indicators of hydrological alteration (IHAs; Olden & Poff, 2003). Multiple studies showed that a successful representation of IHAs in hydrological models requires a targeted optimization process towards these IHAs (Pool, Vis, Knight, & Seibert, 2017). Kiesel et al. (2017) developed a methodology for a tailored optimization of hydrological models for these IHAs.

However, a key problem in assessing when a hydrological model has sufficient performance to model species occurrences remains, because species abundance alone is not yet a clear indicator for a riverine ecosystem status. A complex assessment considering the ecoregion, stream type, species richness and diversity, as well as its community structure is needed to assess the health of the riverine ecosystem for the European Water Framework Directive. In Europe, the assessment calculations are supported by ASTERICS software

(Hering, Borja, Carvalho, & Feld, 2013), which calculates the ecological status of rivers as different metrics based on benthic invertebrate taxalists. The assessment metrics are defined in classes, and if similar classes arise from different assessments, the results can be considered stable and robust.

We are attempting to answer two research questions: (1) Do different objective functions and does different model performance matter for predicting species occurrences? (2) When does a hydrological model have sufficient performance to simulate species occurrences so that ecological assessments based on this simulation are stable? Both are pertinent research questions because the improvement of hydrological model performance requires significant efforts in minimizing the effects in the above-mentioned dependencies on model performance and may limit the application of species predictions to well-researched and data-rich study regions.

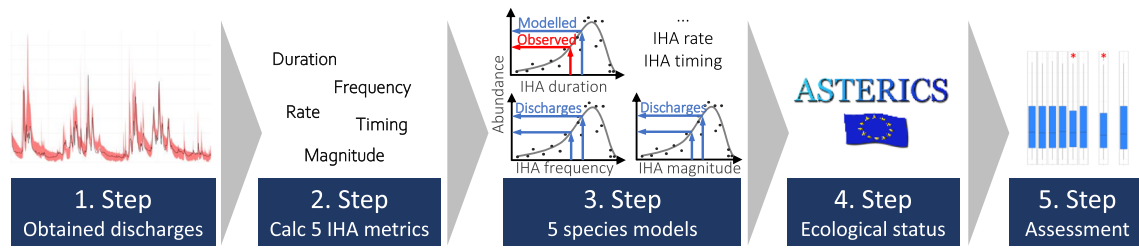
To answer these questions, we will assess the importance of hydrological model performance for simulating macroinvertebrate species in two mesoscale catchments in Germany. Therefore, species predictions are made with hydrological model simulations optimized (a) to the exact species flow preferences (IHAs), (b) to multi-objective functions (MOFs) considering the trade-off between multiple flow preferences, and (c) to standard hydrological performance criteria (HPC) on daily, monthly, and annual time steps. To evaluate the significance of these optimization steps, a comparison is made to species predictions using the observed flow conditions and models without any optimization.

## 2 | MATERIALS AND METHODS

To test the impact of different model performances, we need to generate different hydrological model parameterizations. These models provide different discharge time series, to which we add the observed discharge to complete the set of discharges that is used for the analysis (Figure 1, Step 1). In Step 2, these different discharges are translated into five IHA metrics related to the duration, frequency, magnitude, rate of change, and timing of the discharge. In Step 3, these five IHA metrics are then used to predict species abundance for each catchment, species, and discharge time series separately; this is the basis to answer Research Question 1. In Step 4, based on the resulting species lists of Step 3, metrics are calculated that define the ecological status originating from the different discharges. In Step 5, the distribution of these ecological metrics is assessed according to their similarity, which is the basis to answer Research Question 2.

### 2.1 | Study areas

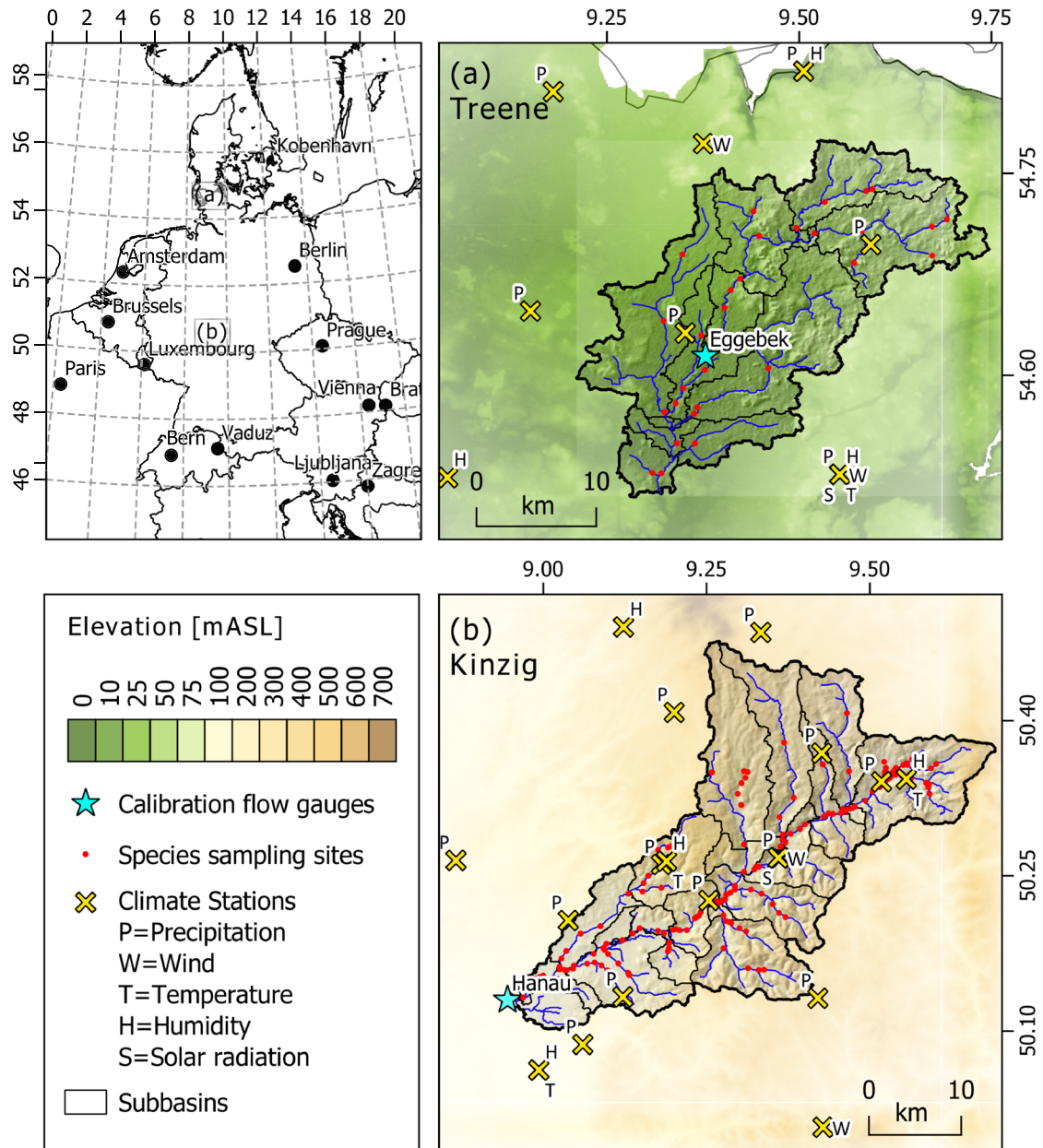
The methodology is applied in two mesoscale catchments in Germany (Figure 2 and Table 1). The Treene is a northern German lowland catchment where hydrological processes are governed by low hydraulic gradients, high groundwater influence, and agricultural land use, which led to artificial tile drainage of approximately one third of



**FIGURE 1** Flow chart of the applied methodology. Abbreviation: IHA, indicator of hydrological alteration

the catchment (Fohrer, Schmalz, Tavares, & Golon, 2007). The Treene contains the catchment of the Kielstau, Germany's first UNESCO eco-hydrological demonstration site (Fohrer & Schmalz, 2012). The Kinzig,

located in the mid-mountain range of Germany, is part of the Rhine-Main-Observatory and is a long-term ecological research (Haase, Frenzel, Klotz, Musche, & Stoll, 2016) site. At this site, different



**FIGURE 2** Catchment locations within Germany with latitude and longitude; subbasin delineation, flow gauges at which the models are calibrated, species sampling sites, climate stations, and elevations of the (a) Treene and (b) Kinzig catchments

**TABLE 1** Main physical, climatic, and hydrological characteristics and information about the macroinvertebrate species of the Treene and Kinzig catchments

Catchment characteristic	Unit	Treene	Kinzig
Longitude/latitude	deg	9.5/54.7	9.3/50.3
Catchment area	km <sup>2</sup>	477 (non-tidal)	921
Elevation gradient	m a.s.l.	4–80	104–624
Major land use classes	—	Agriculture (48%) Pasture (32%)	Forest (45%) Pasture (22%)
Annual precipitation gradient <sup>a</sup>	mm	830–944	623–1094
Temperature daily average spatial gradient in JJA <sup>a</sup>	°C	+16.4 to +16.6	+17.6 to +19.2
Temperature daily average spatial gradient in DJF <sup>a</sup>	°C	+1.6 to +1.6	+1.3 to +2.6
Mean runoff rate <sup>a</sup>	l s <sup>-1</sup> km <sup>-2</sup>	13.2	10.7
q2 runoff rate <sup>a</sup>	l s <sup>-1</sup> km <sup>-2</sup>	3.3	2.8
q98 runoff rate <sup>a</sup>	l s <sup>-1</sup> km <sup>-2</sup>	43.6	45.3
Number of sampling sites (surveys)	—	30 (67)	176 (223)
Number of macroinvertebrate taxa	—	60	134

Abbreviations: JJA, summer (June, July, and August); DJF, winter (December, January, and February).

<sup>a</sup>Data from 1995 to 2015.

taxonomic groups and numerous abiotic variables associated with freshwater and floodplain ecosystems are continuously monitored. The hydrology of the catchment is more complex and influenced by snow, surface runoff, interflow, and groundwater processes (Guse et al., 2019). Stream macroinvertebrates were sampled from 2004 to 2015 in the Treene catchment and from 2005 to 2012 in the Kinzig catchment. Samples were taken based on the multi-habitat sampling protocol (Haase, Pauls, Sundermann, & Zenker, 2004) required for monitoring according to the European Water Framework Directive.

## 2.2 | Hydrological model

The Soil and Water Assessment Tool (SWAT) model (Arnold, Srinivasan, Muttiah, & Williams, 1998) in the version SWAT3S (Pfanterstill, Guse, & Fohrer, 2014a) was used to simulate the hydrological processes in the catchments. In contrast to the original Soil and Water Assessment Tool model, SWAT3S uses two groundwater storages that can be independently controlled for groundwater flow into the stream and a third storage that may be used to account for percolation into geologic formations that are not connected to the stream. The catchment is divided into subbasins (Figure 2; black lines within catchment boundary), which contain a stream channel and are further divided into hydrological response units (HRUs), a spatial entity of unique soil, land use, and slope. For each HRU, the processes of plant growth, evaporation, surface runoff, infiltration, lateral flow, soil moisture, groundwater flow of two aquifers, and potential percolation losses are simulated on a daily time step. Water leaving the HRUs via surface runoff, lateral flow, and groundwater discharge is received in the stream channel where the water is routed to the catchment outlet.

The model was parameterized using 25-m-resolution digital elevation models (Hessian Administration for Soil Management and Geoinformation, 2011; LVA, 1992), vector-based land use maps

(GeoBasis-DE/BKG, 2013), and 1:200,000 soil maps (BGR, 1995). Climate data were derived from precipitation, temperature, wind speed, solar radiation, and humidity stations (DWD, 2016; Figure 2). Channel geometry was taken from satellite images (Google Earth, 2016) and field observations. Sowing, fertilization, harvest, and tillage data followed standard German agricultural practices (KTBL, 2009). Tile drains were implemented according to the methodology described by Guse, Reusser, and Fohrer (2014), where HRUs with slopes smaller than 1.25% and agricultural land use patterns and soils prone to water logging were classified as “drained.” This parameterization is designated as the “default” run.

## 2.3 | Obtain discharges with different model performances (Step 1)

To obtain different simulated discharges, the hydrological models were run 20,000 times for a 6-year calibration period from 2010 to 2015. Parameter combinations were identified by Latin hypercube sampling of the parameter space presented in Table 2. The analysis was performed according to the methodology described by Pfanterstill, Guse, and Fohrer (2014b) using the R-package Calibration, Sensitivity and Monte Carlo Analysis in R (FME) (Soetaert & Petzoldt, 2010). These parameters influence the major hydrological processes of snow accumulation and snowmelt, surface runoff, soil moisture, and groundwater. Thirty-six metrics (Table 3) were calculated for all simulations to assess the model performances gained from the 20,000 parameterizations. The selected metrics can be categorized into three groups: nine IHAs, three MOFs, and 24 standard HPC.

The IHAs were selected as optimization criteria to ensure that the hydrological model depicts the individual IHAs and, therefore, the species preferences as well as possible. This is necessary because

**TABLE 2** Parameter description and ranges used for optimization

Process	SWAT parameter	Description	Treene		Kinzig	
			Value min	Value max	Value min	Value max
SNOW	SNOCVMX	Minimum snow water content that corresponds to 100% snow cover [mm]	1	50	1	50
	SFTMP	Snow fall temperature [°C]	-1	0.99	-1	0.99
	SMTMP	Snow melt temperature [°C]	1	4	1	4
SURFACE RUNOFF	CN2	Curve number value [-] <sup>a</sup>	0.85	1.15	0.9	1.1
	SURLAG	Surface runoff lag coefficient, lower value—higher surface retention [-]	0.01	1	0.01	1
SOIL	ESCO	Soil evaporation compensation factor, lower value—higher soil evaporation [-]	0.01	0.5	0.01	0.5
	SOL_AWC	Soil available water capacity [-] <sup>a</sup>	—	—	0.9	1.5
	SOL_K	Soil saturated hydraulic conductivity [mm h <sup>-1</sup> ] <sup>a</sup>	—	—	0.4	0.6
GROUNDWATER	RCHRGssh	Ratio of percolated water entering the slow shallow aquifer [-]	0.2	0.8	0.2	0.8
	GWDELAYfsh	Groundwater delay time to fast shallow aquifer [d]	5	30	5	30
	ALPHA_BFfsh	Controls the response time of the fast shallow aquifer to recharge [1/d]	0.1	1	0.1	1
	ALPHA_BFssh	Controls the response time of the slow shallow aquifer to recharge [1/d]	0.001	0.05	0.001	0.05

Abbreviation: SWAT, Soil and Water Assessment Tool.

<sup>a</sup>Value is multiplied; all others are substituted.

hydrological models perform weakly in predicting IHAs if they are not included in the optimization process (Kiesel et al., 2017; Pool et al., 2017; Vigiak et al., 2018). The IHAs were selected based on Kakouei et al. (2018), who investigated the most important and not cross-correlated IHAs for the occurring species in the Treene and Kinzig. Kakouei et al. (2018) selected the most important variables for the taxa of each catchment by applying a boosted regression tree analysis. A stepwise process was then applied to exclude an IHA variable with the lower relative influence on the taxa, once cross-correlation between two variables reached the sensitivity threshold of 0.7.

The MOFs include multiple IHAs and minimize the error in depicting all the IHAs combined by optimizing the Euclidean distance (ED) between the normalized IHA errors (Kiesel et al., 2017). The MOFs were included to find the minimum trade-off between simulating all the IHAs important for the species. Three MOFs were selected, which include (a) the most important IHAs for species occurring in the Treene (ED<sub>Tr</sub>) and the Kinzig (ED<sub>Ki</sub>) separately, (b) the most important indicators for the species occurring in the Treene and the Kinzig combined (ED<sub>TrKi</sub>), and (c) 14 IHAs that focus on hydrological extremes (ED<sub>Ext</sub>) according to Richter, Baumgartner, Powell, and Braun (1996) because hydrological extremes significantly impact species occurrence (Stubbington et al., 2009).

The HPCs were selected to evaluate the impact of applying the optimization methodology commonly used in hydrological modelling. Therefore, standard performance metrics were selected that were optimized on daily (subscript D), monthly (subscript M), and yearly

(subscript Y) time steps. These different time steps were used to show the impact of using a broad range of optimized hydrological simulations and, therefore, a wide range of model performance for species predictions.

The simulation runs that performed best for each of the 36 metrics were selected for further analysis. In case multiple simulations performed similarly well in depicting a single metric, the run with the best Kling–Gupta efficiency (KGE<sub>D</sub>) value was chosen. The KGE was selected because it is based on a balanced optimization of model bias, variability, and temporal fit. It is superior to the widely used NSE because models optimized to the NSE tend to underestimate the variability in the simulations (Gupta, Kling, Yilmaz, & Martinez, 2009). Because some of the optimized metrics yielded the same model run, 28 and 27 runs were identified in the Treene and Kinzig catchments, respectively. In addition to these runs, the observed flow time series and the default model run (no optimization) were also included in this selection. The default (DEF) run represents the lower boundary (worst case), and the observed (OBS) flow time series represents the upper boundary (best case; Seibert, Vis, Lewis, & van Meerveld, 2018). Finally, the hydrographs for all the selected runs are compared, and the values of all the metrics are compared for each run to evaluate the trade-off when optimizing individual metrics. This is important because, for instance, the run that shows the highest performance for the correlation coefficient *r* may not show low model bias, or the best run for peak flow indicators may yield unsatisfactory low-flow indicator values.

**TABLE 3** Description of metrics used for optimization

Group	Metric	Description	Reference
IHAs	<b>dh4</b>	Annual maximum of 30-day moving average flow	Olden and Poff (2003)
	<i>fl1</i>	Low-flow (<25th percentile) pulse count	Olden and Poff (2003)
	<b>fl2</b>	Variability in low-flow pulse count	Olden and Poff (2003)
	<b>ml16</b>	Median of annual minimum flows	Olden and Poff (2003)
	<i>ml18</i>	Variability in baseflow index	Olden and Poff (2003)
	<i>ra4</i>	Variability in fall rate	Olden and Poff (2003)
	<b>ra7</b>	Negative change in flow	Olden and Poff (2003)
	<b>ta3</b>	Seasonal predictability of flooding	Olden and Poff (2003)
	<i>th3</i>	Seasonal predictability of non-flooding	Olden and Poff (2003)
Multi-objective function	ED	Euclidean distance for Treene (including <b>dh4</b> , <b>fl2</b> , <b>ml16</b> , <b>ra7</b> , and <b>ta3</b> ) Euclidean distance for Kinzig (including <i>dh4</i> , <i>fl1</i> , <i>ml18</i> , <i>ra4</i> , and <i>th3</i> )	Kakouei et al. (2018)
	ED <sub>Extr</sub>	Euclidean distance for 14 IHA of extreme events	Richter et al. (1996)
	ED <sub>TrKi</sub>	Euclidean distance of all above-listed IHAs	Kakouei et al. (2018)
Hydrological performance criteria	PBIAS <sub>D</sub> , PBIAS <sub>M</sub> , PBIAS <sub>Y</sub>	Percent bias	Moriasi et al. (2007)
	NSE <sub>D</sub> , NSE <sub>M</sub> , NSE <sub>Y</sub>	Nash–Sutcliffe efficiency	Nash and Sutcliffe (1970)
	RMSE <sub>D</sub> , RMSE <sub>M</sub> , RMSE <sub>Y</sub>	Root mean square error	Moriasi et al. (2007)
	RSR <sub>D</sub> , RSR <sub>M</sub> , RSR <sub>Y</sub>	RMSE divided by observed standard deviation	Moriasi et al. (2007)
	KGE <sub>D</sub> , KGE <sub>M</sub> , KGE <sub>Y</sub>	Kling–Gupta efficiency	Kling, Fuchs, and Paulin (2012)
	<i>r<sub>D</sub></i> , <i>r<sub>M</sub></i> , <i>r<sub>Y</sub></i>	KGE <i>r</i> (correlation coefficient)	Kling et al. (2012)
	<i>beta<sub>D</sub></i> , <i>beta<sub>M</sub></i> , <i>beta<sub>Y</sub></i>	KGE beta (bias)	Kling et al. (2012)
	<i>gamma<sub>D</sub></i> , <i>gamma<sub>M</sub></i> , <i>gamma<sub>Y</sub></i>	KGE gamma (variability)	Kling et al. (2012)

Note. IHAs used in the Treene are in bold, and IHAs used in the Kinzig are in italics.

Abbreviations: D, daily time step; IHAs, indicators of hydrological alteration; M, monthly time step; Y, yearly time step.

## 2.4 | Calculation of IHA metrics for species models (Step 2)

As described in Table 3, we selected **dh4**, **fl2**, **ml16**, **ra7**, and **ta3** in the Treene and *dh4*, *fl1*, *ml18*, *ra4*, and *th3* in the Kinzig for the simulation of species abundance, because these were found to be most important for the communities of stream macroinvertebrates in each catchment (Kakouei et al., 2018). This was done based on the observations as well as the 28 and 27 model parameterizations that were selected in Step 1 for the Treene and Kinzig, respectively. Using the flow accumulation approach (Kakouei et al., 2018), these daily discharge time series ( $\text{m}^3 \text{s}^{-1}$ ) were interpolated to the sampling sites of each catchment. The time-series discharge data of each sampling site were then used to calculate the five IHA metrics per catchment at each site over a 6-year period from 2010 to 2015 (i.e., the same period as the calibration period).

## 2.5 | Set-up and application of the species models (Step 3)

The statistical relationships between the IHA values and species abundance were set up by Kakouei et al. (2018) for these five IHA metrics

for 60 and 134 stream macroinvertebrate species inhabiting 30 (67 surveys) and 176 (223 surveys) sites in the Treene and Kinzig catchments, respectively. These statistical relationships were used to simulate the abundance of individual species in single sites over the entire 6-year period for the observation and all hydrological models in each catchment.

Simulating the abundance of species over the period similar to the calibration period guaranteed that the model depicts the species preferences ideally. These data, representing taxa lists, were used in the subsequent ecological assessments. In addition, to assess and visualize the variability in the abundance of species in each catchment over all the hydrological models, the abundance of each species was rounded to the closest integer and averaged over all the sites in each catchment, resulting in 29 and 30 mean abundance values for each species in the Kinzig and Treene catchments, respectively, according to the observations and all the hydrological models.

## 2.6 | Calculate ecological status (Step 4)

Biological diversity has widely been used to assess ecosystem health (Hering et al., 2013). To estimate whether a change in species abundance would result in an ecological effect, we computed different

**TABLE 4** Hydrological performance statistics (daily time step) for the best run of each individual metric

Metric	Treene						Kinzig						
	$\beta_D$	$r_D$	$\gamma_D$	PBIAS <sub>D</sub>	KGE <sub>D</sub>	NSE <sub>D</sub>	$\beta_D$	$r_D$	$\gamma_D$	PBIAS <sub>D</sub>	KGE <sub>D</sub>	NSE <sub>D</sub>	
IHA	dh4	1.00	0.93	0.95	-0.40	0.91	0.87	1.05	0.92	1.00	4.80	0.91	0.83
	fl1	1.00	0.90	1.02	0.10	0.90	0.80	1.37	0.44	0.86	37.00	0.31	-0.46
	fl2	1.03	0.92	1.02	3.20	0.91	0.83	1.07	0.78	0.90	7.40	0.75	0.57
	ml16	0.99	0.94	1.00	-0.60	0.94	0.87	1.02	0.92	0.96	2.20	0.90	0.83
	ml18	0.93	0.93	1.00	-6.60	0.91	0.86	1.03	0.93	1.02	2.80	0.92	0.85
	ra4	1.02	0.84	0.72	1.70	0.68	0.70	1.10	0.89	0.98	9.80	0.85	0.76
	ra7	1.00	0.94	1.01	0.40	0.94	0.87	1.01	0.93	1.02	1.30	0.93	0.86
	ta3	1.02	0.90	1.05	1.70	0.89	0.79	1.03	0.94	1.01	2.90	0.93	0.87
	th3	1.02	0.92	1.00	1.60	0.92	0.84	1.09	0.90	0.95	8.70	0.86	0.79
MOF	ED	1.02	0.89	0.91	1.90	0.86	0.79	1.02	0.90	0.95	1.80	0.88	0.80
	ED <sub>Extr</sub>	1.04	0.83	1.04	3.50	0.82	0.63	1.03	0.93	0.95	3.40	0.91	0.85
	ED <sub>TrKi</sub>	1.06	0.89	0.68	6.50	0.66	0.76	1.01	0.91	0.92	0.60	0.88	0.83
Hydrological Performance Criteria	PBIAS <sub>D</sub>	1.00	0.93	0.98	0.00	0.93	0.87	1.00	0.91	0.97	0.00	0.91	0.83
	NSE <sub>D</sub>	1.00	0.94	0.90	-0.50	0.88	0.88	1.06	0.94	0.95	6.40	0.90	0.88
	RMSE <sub>D</sub>	1.00	0.94	0.90	-0.50	0.88	0.88	1.06	0.94	0.95	6.40	0.90	0.88
	RSR <sub>D</sub>	1.00	0.94	0.90	-0.50	0.88	0.88	1.06	0.94	0.95	6.40	0.90	0.88
	$r_D$	0.99	0.94	1.07	-0.80	0.91	0.87	1.10	0.94	0.92	10.10	0.86	0.87
	$\beta_aD$	1.00	0.92	1.43	0.00	0.56	0.59	1.00	0.93	1.08	0.00	0.89	0.84
	$\gamma_aD$	1.03	0.88	1.00	2.80	0.88	0.76	1.08	0.92	1.00	8.10	0.88	0.81
	KGE <sub>D</sub>	1.00	0.94	1.01	0.40	0.94	0.87	1.03	0.94	1.01	2.90	0.93	0.87
	PBIAS <sub>M</sub>	1.00	0.93	0.98	0.00	0.93	0.87	1.00	0.93	0.99	-0.10	0.93	0.86
	NSE <sub>M</sub>	1.01	0.93	0.89	1.30	0.87	0.87	1.05	0.94	0.93	5.30	0.89	0.88
	RMSE <sub>M</sub>	1.01	0.93	0.89	1.30	0.87	0.87	1.05	0.94	0.93	5.30	0.89	0.88
	RSR <sub>M</sub>	1.01	0.93	0.89	1.30	0.87	0.87	1.05	0.94	0.93	5.30	0.89	0.88
	$r_M$	1.00	0.93	1.28	-0.40	0.71	0.75	1.08	0.94	1.04	7.70	0.89	0.84
	$\beta_aM$	1.00	0.87	1.27	-0.10	0.70	0.59	1.00	0.92	0.84	0.00	0.82	0.84
	$\gamma_aM$	1.09	0.86	0.90	8.70	0.81	0.72	1.10	0.92	1.01	9.60	0.87	0.80
	KGE <sub>M</sub>	1.00	0.93	0.92	-0.10	0.89	0.87	1.00	0.93	0.99	-0.10	0.93	0.86
	PBIAS <sub>Y</sub>	1.00	0.93	0.99	-0.10	0.93	0.86	1.00	0.92	0.96	-0.20	0.92	0.85
	NSE <sub>Y</sub>	0.99	0.88	0.95	-0.60	0.87	0.77	1.00	0.90	1.01	0.30	0.90	0.79
	RMSE <sub>Y</sub>	0.99	0.88	0.95	-0.60	0.87	0.77	1.00	0.90	1.01	0.30	0.90	0.79
	RSR <sub>Y</sub>	0.99	0.88	0.95	-0.60	0.87	0.77	1.00	0.90	1.01	0.30	0.90	0.79
	$r_Y$	1.05	0.83	0.87	4.60	0.79	0.68	1.07	0.93	0.88	7.10	0.84	0.86
	$\beta_aY$	1.00	0.90	1.34	0.20	0.65	0.62	1.00	0.93	1.06	-0.20	0.91	0.85
	$\gamma_aY$	1.03	0.85	0.98	3.20	0.84	0.69	1.10	0.74	0.68	10.40	0.57	0.53
	KGE <sub>Y</sub>	0.97	0.86	0.96	-3.20	0.85	0.73	1.11	0.89	0.88	11.50	0.80	0.77
DEF	1.29	0.70	0.93	28.90	0.58	0.13	1.37	0.44	0.86	37.00	0.31	-0.46	

Note. Dark grey cells show worse performance (excluding the default run); white cells show best performance per column, which is zero for PBIAS<sub>D</sub> and unity for all other performance criteria.

Abbreviations: DEF, default; ED, Euclidean distance; IHA, indicator of hydrological alteration; KGE, Kling–Gupta efficiency; MOF, multi-objective function; NSE, Nash–Sutcliffe efficiency; PBIAS, per cent bias; RMSE, root mean square error; RSR, RMSE divided by observed standard deviation.

ecological assessment metrics. These are a selection from the metrics available in the PERLODES assessment protocol implemented in ASTERICS software, being the official German assessment method compliant with the European Water Framework Directive (<http://www.fliessgewaesserbewertung.de/en/>, Hering et al., 2013). The selected metrics are (a) ecological status, (b) general degradation (GD), (c) German saprobic index (GSI), and (d) German fauna index. The generated taxa lists (available for observed flow and all simulations) are prepared in a matrix of all the sampling sites (columns), including the information on the abundance of each species for each IHA metric (rows) occurring at each site. The ecological status considers a variety of individual sub-metrics, such as saprobic pollution, acidification, and GD, to make a final assessment of the ecological state of a sampling site according to the proportion and abundance of taxa occurring there. The ecological state of each site takes values ranging from 1 (*high status*) to 5 (*bad status*). The module for GD reflects the impact of various stressors, such as hydromorphological degradation and changes in stream hydrology, with values ranging from 0 (*high degradation*) to 1 (*no degradation*; Rolauffs, Hering, Sommerhäuser, Rödiger, & Jähnig, 2003). The GSI ranges from 1 to 4, with higher values indicating higher tolerance of macroinvertebrates of a sampling site to organic pollution, that is, higher saprobic pollution (Rolauffs et al., 2003). The German fauna index, which indicates species response to morphological degradation, ranges between  $-2$  and  $2$ , where higher values indicate the existence of species that prefer to occur in near-natural sites and lower values for species with preferences for hydromorphologically degraded sites (Lorenz, Hering, Feld, & Rolauffs, 2004).

## 2.7 | Assessment (Step 5)

Finally, we compared the values of the four selected ecological assessment (ASTERICS) metrics resulting from the observations and the hydrological simulations over all the sampling sites. As long as no significant differences ( $p > .05$ ) are detected between ecological status classes, it can be argued that the respective model simulations have no significant ecological effect and can therefore be accepted as suitable.

## 3 | RESULTS AND DISCUSSION

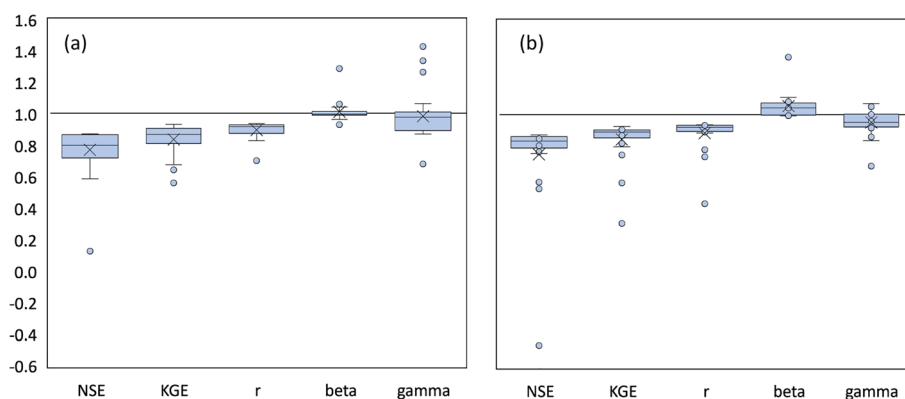
### 3.1 | Hydrological model optimization

The detailed statistical performance of the selected simulations is shown in Table 4 and in the Supporting Information. Table 4 shows the commonly used HPCs per cent bias (PBIAS), NSE, KGE, and the three components of the KGE, beta ( $\beta$ ), correlation coefficient ( $r$ ), and gamma ( $\gamma$ ) on a daily time step for each run that performs best for the metric given in each row of the column "Metric." This means that, for instance, the simulation that performs best for the row  $r_D$  achieves the highest value of  $0.94 r_D$  in both the Kinzig and the Treene. As can be seen in the other columns for row  $r_D$ , this best  $r_D$  run does not yield the best values for the remaining HPCs, for example, only a KGE of 0.91 or 0.86 for Treene and Kinzig, respectively. Similarly, when optimizing for the single IHA metrics, low performance can occur in the daily HPCs. For instance, for the rows *low-flow pulse count* (fl1) in the Kinzig or the *variability in fall rate* (ra4) in the Treene, performance in all columns is low, indicating that matching these indicators compromises the general hydrological processes. To some extent, this is also the case for the *seasonal predictability of flooding* (ta3) and the *variability in baseflow index* (ml18) in the Treene and the *seasonal predictability of non-flooding* (th3) in the Kinzig. Optimizing the hydrological model for the IHAs generally leads to errors close to 0 in depicting these IHAs (Tables S1 and S2). However, this compromises the representation of the remaining IHAs and can lead to large deviations up to the range of the default model run.

The MOF simulations lead to overall good HPCs in the Kinzig but only acceptable HPCs in the Treene. The MOF simulations show a balanced compromise in depicting all the IHAs (Tables S1 and S2).

Optimizing to the HPCs shows a difference between the optimization time steps and the chosen metric. The  $KGE_D$  and  $KGE_M$  metrics yield a balanced result for the other HPCs, whereas the  $NSE_D$  and  $NSE_M$  as well as  $r$  and gamma lead to higher model biases. Not surprisingly, the performance in the daily HPCs declines with optimizing for increasing time steps.

Although the KGE is superior to the NSE for the general optimization of the hydrological models, thresholds for assessing model performance have been published only for the NSE (Ritter & Muñoz-Carpena, 2013). The  $NSE_D$  for the runs selected from the 36 metrics



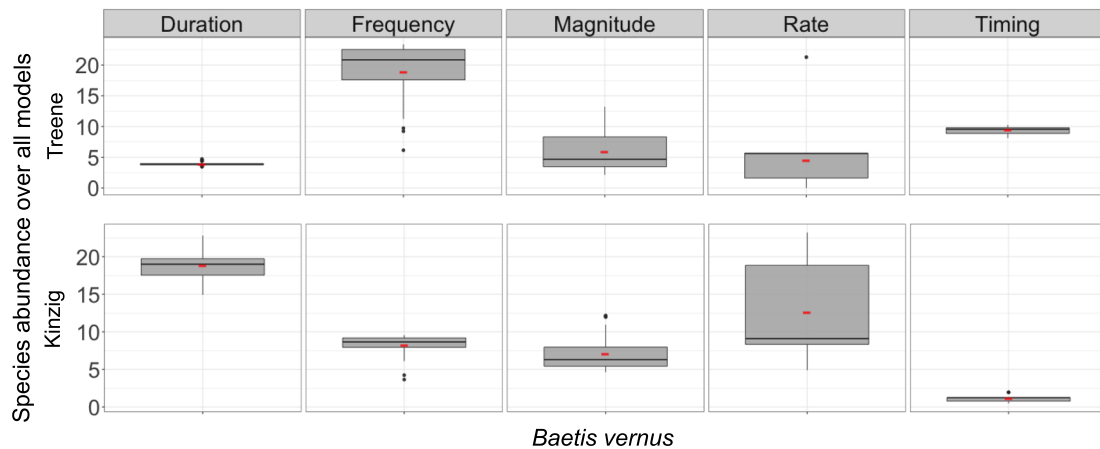
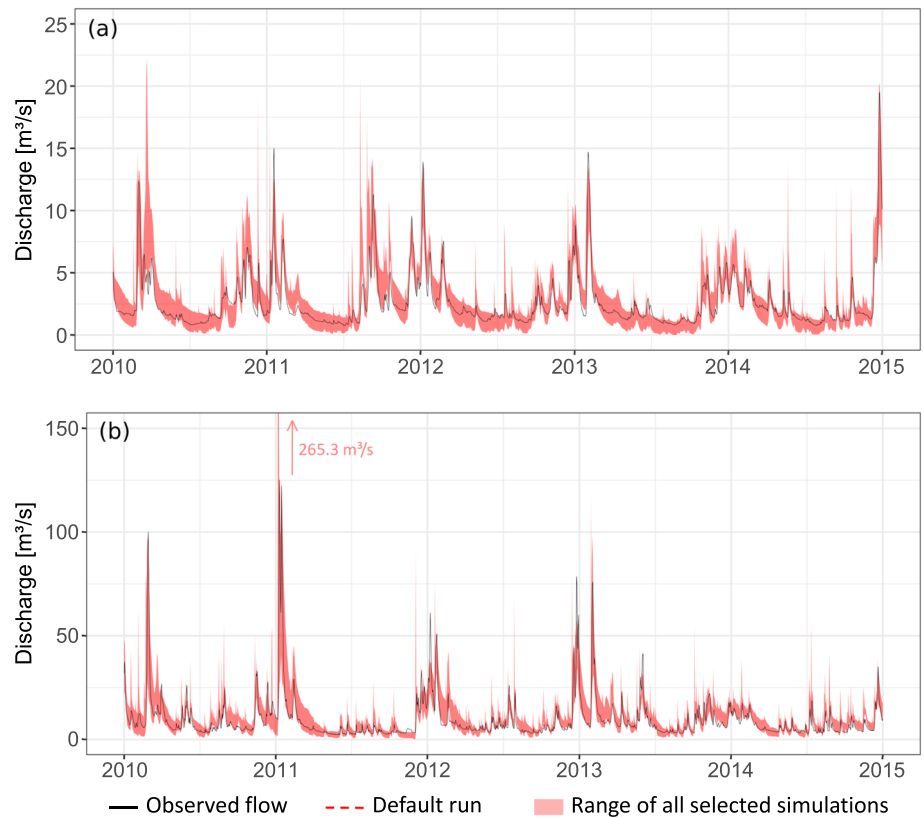
**FIGURE 3** Range in performance criteria shown in Table 3 for all the chosen runs for (a) Treene and (b) Kinzig; horizontal line at unity marks the best performance. Abbreviations: KGE, Kling-Gupta efficiency; NSE, Nash-Sutcliffe efficiency



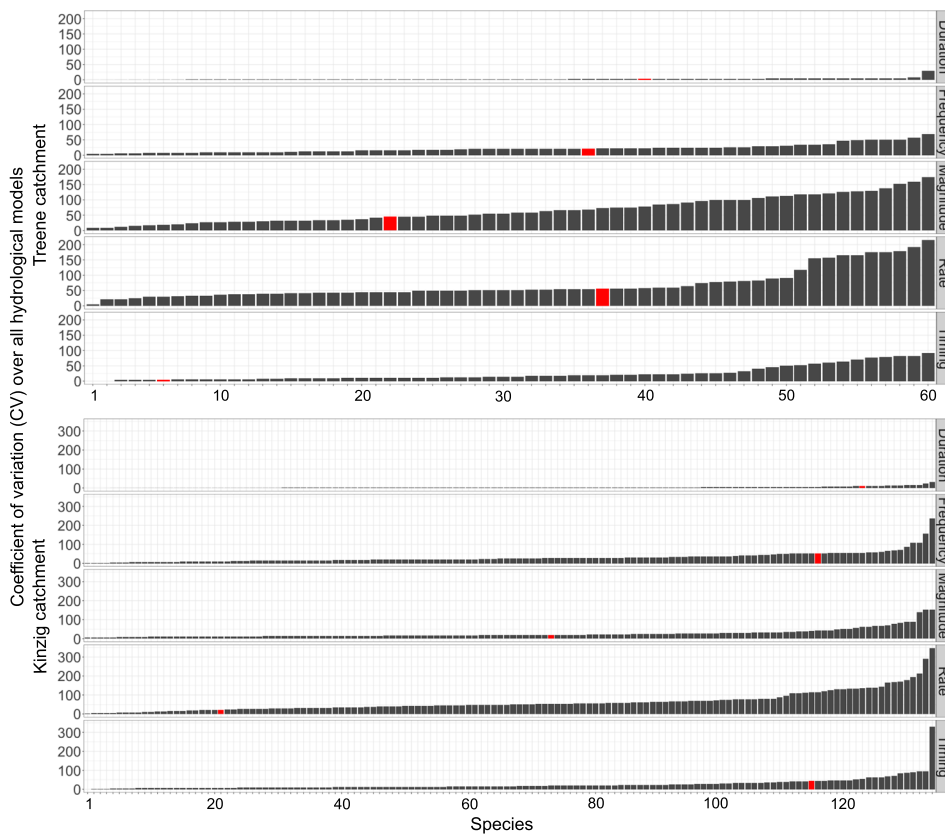
lie between 0.59 and 0.88 for the Treene and between 0.53 and 0.88 for the Kinzig, which range from *unsatisfactory* to the upper end of *good* model performance (Figure 3). The default model run with  $NSE_D$  of  $-0.26$  and  $-0.46$  for the Treene and Kinzig, respectively, perform worse than the mean of the observations ( $NSE = 0$ ).  $NSE = 1$  would represent an ideal model where the simulation is equal to the observation. As seen from the three components of the KGE (Figure 3), beta (bias ratio) and gamma (variability ratio) reach ideal values of unity,

whereas  $r$  (correlation coefficient) does not. The highest KGE performance in the two catchments is 0.94 and 0.93 for the Treene and Kinzig, respectively. Using the observed time series together with all the selected simulations for the species prediction yields a wide performance range. From the performance range, we can also deduce that, when optimizing for the IHAs only, it is important to keep the HPC in mind because good results in depicting the IHAs may mean sacrificing hydrological consistency.

**FIGURE 4** Observed and simulated flow of all metrics for the (a) Treene and (b) Kinzig catchments



**FIGURE 5** Range of abundances for *Baetis vernus* in the Treene and Kinzig catchments for each indicator of hydrological alteration metric. The number of data points in each box plot is equal to the number of discharge time series used (36 metrics + 1 default run + 1 observed discharge)



**FIGURE 6** Relative variability (coefficient of variation) in species abundance over all hydrological models for each indicator of hydrological alteration (IHA) category in the Treene and Kinzig catchments, ordered according to increasing coefficient of variation for the average over all the IHA groups. The red-coloured bar marks species *Baetis vernus* shown in Figure 5. The bars are ordered according to the values of the coefficient of variation for each IHA metric

The resulting hydrographs selected from the 36 metrics (Table 3) and the default run are compared with the observed flow in Figure 4 to give a visual impression of the calculated performance statistics. Analysing the daily flow values shows that 83% and 76% of the observed values are within the range of simulations in the Treene and Kinzig, respectively. It can be seen that the simulated low-flow periods show a high range in the Treene, which is due to the strong groundwater influence. In the Kinzig, the recession phases show a high range in the simulations. The default model setting causes a single high peak flow in January 2011.

### 3.2 | Species abundance

Figure 5 shows the simulated species abundance for each IHA group on the example of *Baetis vernus*. Species abundances are based on species responses along the gradient of each IHA group (Figure S2); thus, each IHA metric leads to different abundances. The box plots show the range in species abundance over all the hydrological models and observations for each IHA category. The wider the box plot is, the higher the influence of the model calibration for the respective metric on the species. For the selected species, rate is an important indicator class in both catchments, whereas timing in both catchments and duration in the Treene and frequency in the Kinzig are less sensitive. The results for all the species are shown in Figure 6, which shows the variability directly. In both catchments, a strong gradient can be observed between the species that are sensitive to flow changes or

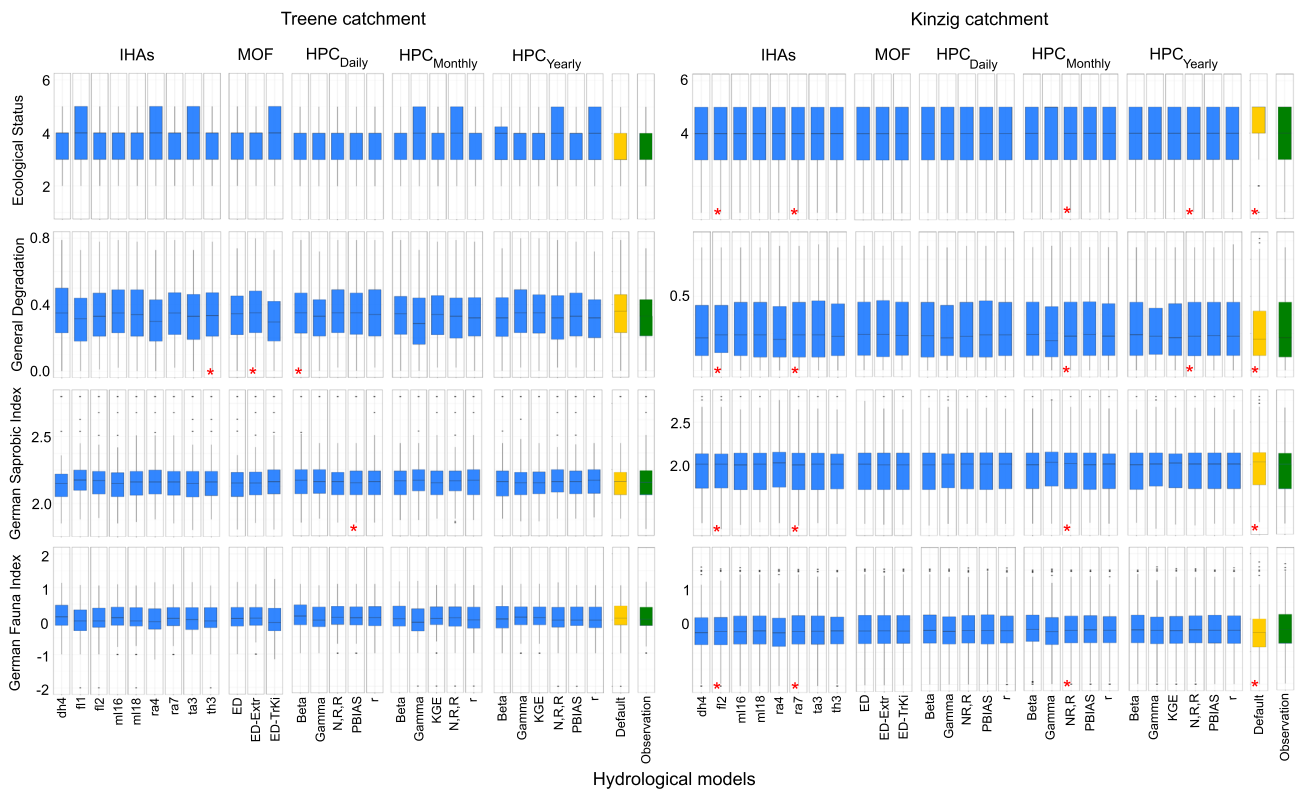
not sensitive (Figure 6). In the Treene catchment, the strongest variability was detected according to magnitude and rate of change in flow events, whereas changes in frequency and rate caused the strongest variability in the abundance of species in the Kinzig catchment. The full range of species abundance is supplied as box plots for each species in Figure S1a,b.

The bar plots enable the assessment of whether species are generally more susceptible to flow changes in all the IHA categories or whether they are “specialists” for certain IHA variables only. *B. vernus* is among the most sensitive species in the Kinzig for duration and frequency, but its sensitivity varies across metrics and catchments (Figure 6, red bars). These results suggest that for a holistic assessment of overall species abundance, it is necessary to optimize the hydrological model to sufficiently depict all the indicators simultaneously.

### 3.3 | Ecological assessment

The response of communities to the different species abundances (Figure 6) may result in significantly different ecological assessment metrics (Figure 7). The number of significant differences is higher in the Kinzig than in the Treene catchment.

In the Kinzig, five hydrological model runs, which are the default model run, the annual and monthly NSE ( $NSE_Y$  and  $NSE_M$ ), and calibration on the two IHAs *negative change of flow (ra7)* and *variability in low-flow pulse count (fl2)*, show significantly different results in all



**FIGURE 7** The four ASTERICS metrics describing community responses across all the sampling sites and indicator groups of the (a) Treene and (b) Kinzig catchments according to the gauging data (observations shown in the green box plot) and all the hydrological model realizations (blue; yellow = default). The number of significant differences in the ecological status classes is counted for each discharge time series (each column). Significantly (analysis of variance, Tukey's honest significant difference test,  $p < .05$ ) different distributions of ASTERICS metrics from the observation are shown by red asterisks (\*) below the box plots. N,R,R on the x-axis represents the same model run for Nash–Sutcliffe efficiency, root mean square error, and root mean square error divided by observed standard deviation. Abbreviations: ED, Euclidean distance; IHAs, indicators of hydrological alteration; HPC, hydrological performance criteria; KGE, Kling–Gupta efficiency; MOF, multi-objective function; PBIAS, per cent bias

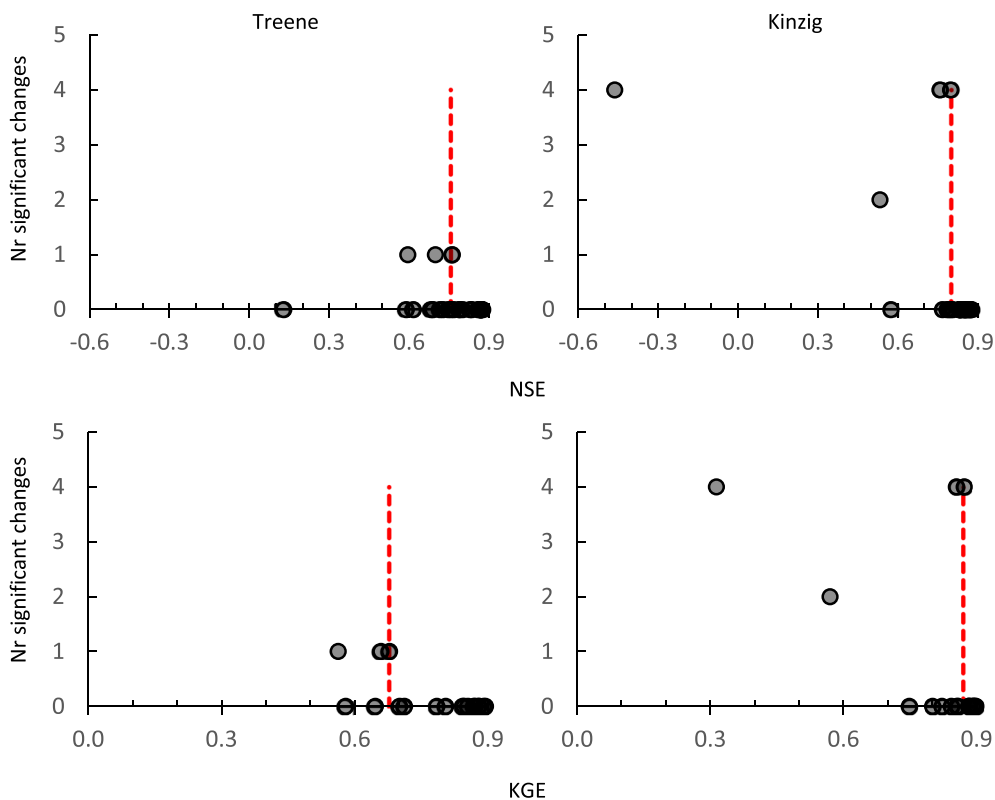
ASTERICS metrics compared with those of the observations. Only the MOF and HPC<sub>D</sub> model groups show no significant deviation from the observations. This is reasonable because those groups also lead to the highest agreement between the simulated and observed discharge patterns.

In the lowland catchment of the Treene, only the models calibrated towards daily PBIAS (PBIAS<sub>D</sub>), the ED between extremely low- and high-flow IHAs (ED<sub>Extr</sub>), and the IHA *seasonal predictability of non-flooding* (th3) lead to significantly different ASTERICS metrics of the GD and GSIs.

If the results are grouped according to each IHA group within the four assessment metrics, it can be seen that the magnitude group in the Kinzig and the rate group in the Treene are subject to significant changes (Figure S3a,b). Similar to Figure 7, over all the IHA groups, more pronounced changes are found in the Kinzig. A possible explanation for the smaller differences in the Treene compared with those in the Kinzig is that the species in the lowland show a stronger dependency on water quality and river morphology and less on the discharge pattern (Kiesel et al., 2015; Schröder et al., 2013). Discharge in the lowlands is generally less

erratic and smoother compared with that of more mountainous catchments due to higher groundwater influence (Guse et al., 2019), whereas water quality and morphological degradation are more of a concern due to the high agricultural impact (Wagner, Hörmann, Schmalz, & Fohrer, 2018). Overall, this may lead to species appearing to be less sensitive to changes in the discharge pattern.

Comparing the hydrological model performance against the number of significant changes in the ecological status classes shows that as soon as model performance reaches values above an NSE of 0.76 (KGE 0.68) for the Treene and 0.8 (KGE 0.87) for the Kinzig, no significant difference in the ecological status to the observation exists (Figure 8). The number of catchments is too small to consider this as a general rule, but it shows that a certain hydrological model fit is likely required to sufficiently simulate species responses. However, please note that KGE and NSE scores cannot be directly compared and that it is suggested to follow a more purpose-based assessment of hydrological model performance (Knoben, Freer, & Woods, 2019), as we have presented it here for the two catchments regarding the impact on ecological status.



**FIGURE 8** Hydrological model performance (Nash–Sutcliffe efficiency [NSE] above; Kling–Gupta efficiency [KGE] below) of all the calibration runs and the corresponding significant changes in the ecological status classes (sum of red asterisks in Figure 7 for each column); red line shows the cut-off model performance above, which no significant changes occur

## 4 | CONCLUSIONS AND OUTLOOK

Our results show that the objective function and model performance influence the prediction of species occurrences and that different calibration efforts lead to different simulated species abundances (Research Question 1). As expected, these results are species dependent, where specialists that accept specific ranges of streamflows are more sensitive than generalists that are distributed over larger flow ranges (Kakouei et al., 2018). Hence, the species response to different calibration stages depends on the sensitivity of the species to the particular IHA (Kakouei et al., 2017) and how well the model is able to replicate this IHA (Kiesel et al., 2017). These results are different for the two catchments, indicating that different stressors in the catchments lead to different species sensitivities to flow changes. To deduce more generalized results from the proposed method, the application of the method to a higher number of heterogeneous catchments is needed. This could potentially reveal the spatial differences between species sensitivity to flow changes.

Research Question 2 (sufficient performance to simulate species occurrences so that ecological assessments are stable) was answered through calculating the ecological assessments from all the simulated species lists and statistically evaluating their similarity. In the Kinzig, plausible results were found, where, generally, hydrologically poor performing models versus observed flows led to significantly different ecological assessments. In the Treene, no clear pattern between hydrological model performance and significantly different ecological assessments could be found; for instance, even the default model setting led to no significant differences. However, a direct

comparison between the hydrological model performance and the number of significant changes in the ecological status classes in both catchments revealed that skilled hydrological models are sufficient to depict species responses, that is, lead to no significant differences between the status classes. This may provide a first careful threshold, but due to our small sample size of two catchments, we argue that studies assessing the impact of hydrological change on species should not evaluate the calibration performance on HPC alone. Until larger catchment sample studies lead to more robust results, it is necessary to first assess the performance of the model to predict the metric used for the species prediction and second assess the sensitivity of the species to this metric. Additionally, although stream discharge is a significant descriptor of macroinvertebrate abundance, our study shows that in catchments where multiple stressors, such as lower water quality and morphological degradation, occur, multiple stressors should be considered in the species simulation.

## ACKNOWLEDGEMENTS

This study was funded through the “GLANCE” project (Global change effects in river ecosystems; 01LN1320A) supported by the BMBF (Bundesministerium für Bildung und Forschung). SCJ and KK acknowledge funding through the “AQUACROSS” project (Knowledge, Assessment, and Management for AQUATIC Biodiversity and Ecosystem Services aCROSS EU policies; 642317) supported by the European Union’s Horizon 2020 Framework Programme for Research, Technological Development and Demonstration. BG acknowledges funding by DFG (Deutsche Forschungsgemeinschaft) via project GU

1466/1-1 (Hydrological Consistency in Modelling). We thank the Hessian Administration for Soil Management and Geoinformation (HVBG) and the Land Survey Office Kiel (LVA) for the 25-m digital elevation models; the German Working Group on Water Issues of the Federal States and the Federal Government (LAWA), the State Agency for Agriculture, Environment and Rural Areas Schleswig-Holstein (LLUR), and the German federal state environmental agencies for the discharge data; and Peter Haase from the Senckenberg Research Institute and Natural History Museum with the Rhine-Main-Observatory (RMO) Long-Term Ecosystem Research (LTER) site Kinzig who contributed the biomonitoring data and supported the field work.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AVAILABILITY OF DATA

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Jens Kiesel  <https://orcid.org/0000-0002-4371-6434>

Nicola Fohrer  <https://orcid.org/0000-0002-7456-6301>

Sonja C. Jähnig  <https://orcid.org/0000-0002-6349-9561>

## REFERENCES

- Armanini, D. G., Horrigan, N., Monk, W. A., Peters, D. L., & Baird, D. J. (2011). Development of a benthic macroinvertebrate flow sensitivity index for Canadian rivers. *River Research and Applications*, 27(6), 723–737. <https://doi.org/10.1002/rra.1389>
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modelling and assessment part I: Model development. *Journal of the American Water Resources Association*, 34(1), 73–89. <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>
- Beven, K. (2007). Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process. *Hydrology and Earth System Sciences*, 11, 460–467. <https://doi.org/10.5194/hess-11-460-2007>
- BGR (1995–2014). *Bodenübersichtskarte 1:200,000, Verbreitung der Bodengesellschaften*. Hannover: Bundesanstalt für Geowissenschaften und Rohstoffe.
- Ceola, S., Hödl, I., Adlboller, M., Singer, G., Bertuzzo, E., Mari, L., ... Rinaldo, A. (2013). Hydrologic variability affects invertebrate grazing on phototrophic biofilms in stream microcosms. *PLoS ONE*, 8(4), 1–11. <https://doi.org/10.1371/journal.pone.0060629>
- Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., ... Zeng, X. (2015). Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, 51(8), 5929–5956. <https://doi.org/10.1002/2015WR017096>
- Domisch, S., Portmann, F. T., Kuemmerlen, M., O'Hara, R. B., Johnson, R. K., Davy-Bowkery, J., ... Jähnig, S. C. (2017). Using streamflow observations to estimate the impact of hydrological regimes and anthropogenic water use on European stream macroinvertebrate occurrences. *Ecohydrology*, 10(8), 1–11. <https://doi.org/10.1002/eco.1895>
- DWD. 2016. Climate data from the German Weather Service (Deutscher Wetterdienst).
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11), 1–13. <https://doi.org/10.1029/2010WR010174>
- Fohrer, N., Schmalz, B. (2012). Das UNESCO Ökohydrologie-Referenzprojekt Kielstau-Einzugsgebiet–Nachhaltiges Wasserressourcenmanagement und Ausbildung im ländlichen Raum. *Hydrologie und Wasserbewirtschaftung* 56(4), 160–168. [https://doi.org/10.5675/HyWa\\_2012.4\\_1](https://doi.org/10.5675/HyWa_2012.4_1)
- Fohrer, N., Schmalz, B., Tavares, F., & Golon, J. (2007). Modelling the landscape water balance of mesoscale lowland catchments considering agricultural drainage systems. *Hydrologie und Wasserbewirtschaftung*, 51(4), 164–169.
- GeoBasis-DE/BKG. (2013). Atkis Basis-DLM vector data, version 6.0, 3m accuracy, Bundesamt für Kartografie und Geodesie.
- Google Earth. (2016). V.7.1.5.1557, DigitalGlobe 2016.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., & Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, 21, 5663–5679. <https://doi.org/10.5194/hess-21-5663-2017>
- Guse, B., Pfannerstill, M., Kiesel, J., Strauch, M., Volk, M., & Fohrer, N. (2019). Analysing spatio-temporal process and parameter dynamics in models to characterise contrasting catchments. *Journal of Hydrology*, 570, 863–874. <https://doi.org/10.1016/j.jhydrol.2018.12.050>
- Guse, B., Reusser, D. E., & Fohrer, N. (2014). How to improve the representation of hydrological processes in SWAT for a lowland catchment—Temporal analysis of parameter sensitivity and model performance. *Hydrological Processes*, 28, 2651–2670. <https://doi.org/10.1002/hyp.9777>
- Haase, P., Frenzel, M., Klotz, S., Musche, M., & Stoll, S. (2016). The long-term ecological research (LTER) network: Relevance, current status, future perspective and examples from marine, freshwater and terrestrial long-term observation. *Ecological Indicators*, 65, 1–3. <https://doi.org/10.1016/j.ecolind.2016.01.040>
- Haase, P., Pauls, S., Sundermann, A., & Zenker, A. (2004). Testing different sorting techniques in macroinvertebrate samples from running waters. *Limnologica*, 34(4), 366–378. [https://doi.org/10.1016/S0075-9511\(04\)80006-9](https://doi.org/10.1016/S0075-9511(04)80006-9)
- Hering, D., Borja, A., Carvalho, L., & Feld, C. K. (2013). Assessment and recovery of European water bodies: Key messages from the WISER project. *Hydrobiologia*, 704(1), 1–9. <https://doi.org/10.1007/s10750-012-1438-9>
- Hessian Administration for Soil Management and Geoinformation. (2011). DEM 25 m grid size. Hessische Verwaltung für Bodenmanagement und Geoinformation.
- Kakouei, K., Kiesel, J., Domisch, S., Irving, K., Jähnig, S., & Kail, J. (2018). Projected effects of climate change-induced flow alterations on stream macroinvertebrate abundances. *Ecology and Evolution*, 2018(00), 1–17. <https://doi.org/10.1002/ece3.3907>
- Kakouei, K., Kiesel, J., Kail, J., Pusch, M., & Jähnig, S. C. (2017). Quantitative hydrologic preferences of benthic stream invertebrates in Germany. *Ecological Indicators*, 79, 163–172. <https://doi.org/10.1016/j.ecolind.2017.04.029>
- Kiesel, J., Guse, B., Pfannerstill, M., Kakouei, K., Jähnig, S. C., & Fohrer, N. (2017). Improving hydrological model optimization for riverine species. *Ecological Indicators*, 80, 376–385. <https://doi.org/10.1016/j.ecolind.2017.04.032>
- Kiesel, J., Schröder, M., Hering, D., Schmalz, B., Hörmann, G., Jähnig, S. C., & Fohrer, N. (2015). A new model linking macroinvertebrate assemblages to habitat composition in rivers: development, sensitivity and univariate application. *Fundamental and*

- Applied Limnology*, 186(1-2), 117–133. <https://doi.org/10.1127/fal/2015/0625>
- Kling, H., Fuchs, M., & Paulin, M. (2012). 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>
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23, 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- KTBL. 2009. Faustzahlen für die Landwirtschaft. 14. Auflage, Kuratorium für Technik und Bauwesen in der Landwirtschaft eV. 1–1180. ISBN 978-3-939371-91-5.
- Lorenz, A., Hering, D., Feld, C. K., & Rolaufts, P. (2004). A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. *Hydrobiologia*, 516, 107–127. <https://doi.org/10.1023/B:HYDR.0000025261.79761.b3>
- LVA (1992–2004). DEM 25 m grid size. Kiel: Landesvermessungsamt Schleswig-Holstein.
- Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M. P., & Uijlenhoet, R. (2016). Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-alpine basin. *Hydrology and Earth System Sciences*, 20, 2207–2226. <https://doi.org/10.5194/hess-20-2207-2016>
- Mondy, C. P., & Schuwirth, N. (2017). Integrating ecological theories and traits in process-based modeling of macroinvertebrate community dynamics in streams. *Ecological Applications*, 27(4), 1365–1377. <https://doi.org/10.1002/eap.1530>
- Moriasi, D. N., Arnold, J. G., VanLiew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Nash, J. E., & Sutcliffe, J. V. (1970). 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](https://doi.org/10.1016/0022-1694(70)90255-6)
- O’Keeffe, J., Piniewski, M., Szcześniak, M., Oglęcki, P., Parasiewicz, P., & Okruszko, T. (2018). Index based analysis of climate change impact on streamflow conditions important for Northern Pike, Chub and Atlantic salmon. *Fisheries Management and Ecology*, 2018, 474–485. <https://doi.org/10.1111/fme.12316>
- Olden, J. D., & Poff, N. L. (2003). 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>
- Pfannerstill, M., Bieger, K., Guse, B., Bosch, D., Fohrer, N., & Arnold, J. G. (2017). How to constrain multi-objective calibrations of the SWAT model using water balance components. *Journal of the American Water Resources Association*, 53(3), 532–546. <https://doi.org/10.1111/1752-1688.12524>
- Pfannerstill, M., Guse, B., & Fohrer, N. (2014a). A multi-storage groundwater concept for the SWAT model to emphasize nonlinear groundwater dynamics in lowland basins. *Hydrological Processes*, 28(22), 5599–5612. <https://doi.org/10.1002/hyp.10062>
- Pfannerstill, M., Guse, B., & Fohrer, N. (2014b). 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>
- Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegard, K. L., Richter, B. D., ... Stromberg, J. C. (1997). The natural flow regime. A paradigm for river conservation and restoration. *Bioscience*, 47(11), 769–784. <https://doi.org/10.2307/1313099>
- Poff, N. L., & Zimmerman, J. K. H. (2010). 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>
- Pool, S., Vis, M. J. P., Knight, R. R., & Seibert, J. (2017). 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>
- Pyne, M. I., & Poff, N. L. (2017). Vulnerability of stream community composition and function to projected thermal warming and hydrologic change across ecoregions in the western United States. *Global Change Biology*, 23, 77–93. <https://doi.org/10.1111/gcb.13437>
- Qu, Y., Wu, N., Guse, B., & Fohrer, N. (2018). Riverine phytoplankton shifting along a lentic–lotic continuum under hydrological, physiochemical conditions and species dispersal. *Science of the Total Environment*, 619–620, 1628–1636. <https://doi.org/10.1016/j.scitotenv.2017.10.139>
- Reusser, D. E., Blume, T., Schaeffli, B., & Zehe, E. (2009). Analysing the temporal dynamics of model performance for hydrological models. *Hydrology and Earth System Sciences*, 13, 999–1018. <https://doi.org/10.5194/hess-13-999-2009>
- Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, 10(4), 1163–1174. <https://doi.org/10.2307/2387152>
- Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33–45. <https://doi.org/10.1016/j.jhydrol.2012.12.004>
- Rolaufts, P., Hering, D., Sommerhäuser, M., Rödiger, S., & Jähnig, S. (2003). Entwicklung eines leitbildorientierten Saprobienindex für die biologische Fließgewässerbewertung, Forschungsbericht 200 24 227. Umweltbundesamt, 1–140.
- Schröder, M., Kiesel, J., Schattmann, A., Jähnig, S. C., Lorenz, A. W., Kramm, S., ... Hering, D. (2013). Substratum associations of benthic invertebrates in lowland and mountain streams. *Ecological Indicators*, 30, 178–189. <https://doi.org/10.1016/j.ecolind.2013.02.012>
- Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multi-criteria model calibration. *Water Resources Research*, 38(11), 1241–23–14. <https://doi.org/10.1029/2001WR000978>
- Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. J. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, 32, 1120–1125. <https://doi.org/10.1002/hyp.11746>
- Shafii, M., & Tolson, B. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, 51, 3796–3814. <https://doi.org/10.1002/2014WR016520>
- Soetaert, K., & Petzoldt, T. (2010). Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *Journal of Statistical Software*, 33(3), 1–28.
- Stubbington, R., Greenwood, A. M., Wood, P. J., Armitage, P. D., Gunn, J., & Robertson, A. L. (2009). The response of perennial and temporary headwater stream invertebrate communities to hydrological extremes. *Hydrobiologia*, 630, 299–312. <https://doi.org/10.1007/s10750-009-9823-8>
- Vigiak, O., Lutz, S., Mentzafou, A., Chiogna, G., Tuo, Y., Majoneg, B., ... Pistocchi, A. (2018). Uncertainty of modelled flow regime for flow-ecological assessment in Southern Europe. *Science of the Total Environment*, 615, 1028–1047. <https://doi.org/10.1016/j.scitotenv.2017.09.295>
- Wagner, P. D., Hörmann, G., Schmalz, B., & Fohrer, N. (2018). Characterisation of the water and nutrient balance in the rural lowland catchment of the Kielstau. *Hydrologie und Wasserbewirtschaftung*, 62(H3), 145–158. [https://doi.org/10.5675/HyWa\\_2018,3\\_2](https://doi.org/10.5675/HyWa_2018,3_2)
- Wu, N., Qu, Y., Guse, B., Makarevičiūtė, K., To, S., Riis, T., & Fohrer, N. (2018). Hydrological and environmental variables outperform spatial

factors in structuring species, trait composition, and beta diversity of pelagic algae. *Ecology and Evolution*, 8(5), 2947–2961. <https://doi.org/10.1002/ece3.3903>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kiesel J, Kakouei K, Guse B, Fohrer N, Jähnig SC. When is a hydrological model sufficiently calibrated to depict flow preferences of riverine species? *Ecohydrology*. 2020;13:e2193. <https://doi.org/10.1002/eco.2193>