

## RESEARCH ARTICLE

# Using streamflow observations to estimate the impact of hydrological regimes and anthropogenic water use on European stream macroinvertebrate occurrences

Sami Domisch<sup>1,2,3</sup>  | Felix T. Portmann<sup>1,2,4</sup> | Mathias Kuemmerlen<sup>1,2,5</sup> | Robert B. O'Hara<sup>1,6</sup>  | Richard K. Johnson<sup>7</sup> | John Davy-Bowker<sup>8,9</sup> | Torleif Bækken<sup>10</sup> | Carmen Zamora-Muñoz<sup>11</sup> | Marta Sáinz-Bariáin<sup>11</sup> | Núria Bonada<sup>12</sup> | Peter Haase<sup>1,2,13</sup> | Petra Döll<sup>1,4</sup>  | Sonja C. Jähnig<sup>1,2,3</sup>

<sup>1</sup>Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

<sup>2</sup>Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystraße 12, D-63571 Gelnhausen, Germany

<sup>3</sup>Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department of Ecosystem Research, Müggelseedamm 301, D-12587 Berlin, Germany

<sup>4</sup>Institute of Physical Geography, Goethe University Frankfurt, Altenhöferallee 1, D-60438 Frankfurt am Main, Germany

<sup>5</sup>Department Systems Analysis, Integrated Assessment and Modelling, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, 8600 Dübendorf, Switzerland

<sup>6</sup>Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway

<sup>7</sup>Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, P.O. Box 7050, SE-750 07 Uppsala, Sweden

<sup>8</sup>The Freshwater Biological Association, River Laboratory, East Stoke, Wareham, Dorset BH20 6BB, UK

<sup>9</sup>Natural History Museum, Cromwell Road, London, SW7 5BD, UK

<sup>10</sup>Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, NO-0349 OSLO, Norway

<sup>11</sup>Department of Zoology, Faculty of Science, University of Granada, C/ Severo Ochoa s/n, 18171 Granada, Spain

<sup>12</sup>Grup de Recerca Freshwater Ecology and Management (FEM), Departament d'Ecologia, Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Diagonal 643, 08028 Barcelona, Catalonia/Spain

<sup>13</sup>University of Duisburg-Essen, Department of River and Floodplain Ecology, Faculty of Biology, Universitätsstr, 5, 45141 Essen, Germany

## Correspondence

Sami Domisch, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department of Ecosystem Research, Müggelseedamm 301, D-12587 Berlin, Germany.  
Email: domisch@igb-berlin.de

## Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: DO 1880/1-1; LOEWE initiative; German Federal Ministry of Education and Research, Grant/Award Number: 01LN1320A; Spanish Ministry of Education and Science, Grant/Award Number: HID98-0323-C05 and REN2001-3438-C07; Spanish Ministry of Environment, Grant/Award Number: 039/2007

## Abstract

Understanding the drivers of stream macroinvertebrate distribution patterns—the most diverse animal group in freshwater ecosystems—is a major goal in freshwater biogeography. Climate and topography have been shown to explain species' distributions at continental scales, but the contribution of natural and anthropogenically altered streamflow is often omitted in large-scale analyses due to the lack of appropriate data. We test how macroinvertebrate occurrences can be linked to streamflow observations and evaluate the relative importance of streamflow regimes and water use for macroinvertebrate occurrences from 19 orders across Europe. We first paired species sampling locations with hydrological gauging stations considering 5 combinations of the geographic distance and difference in flow accumulation (upstream contributing area). We then used Generalized Linear Models to assess the influence of the streamflow regime, simulated water use, and climate and topography on the occurrence of macroinvertebrates. The pairing method that assigned species records to the closest gauging station in terms of both distance and flow accumulation performed best. Most of the species studied occurred preferentially in river habitats with low mean annual streamflow and streamflow variability, high winter streamflow, and low levels of water withdrawals for irrigation or manufacturing. We conclude that flow accumulation is a useful proxy to evaluate the proximity of species records to gauging stations, omitting species records that do not belong to a given stream reach. The strong contribution of streamflow and water use indicators on macroinvertebrate occurrences underline their importance for yielding robust occurrence estimates.

**Abbreviations:** avg, average; Q, river discharge; WW, water withdrawal.

## KEYWORDS

collinearity, environmental flow, flow accumulation, gauging station, hydrology, river, species distributions, stream

## 1 | INTRODUCTION

Freshwater ecosystems are unique biodiversity hotspots, harbouring 9.5% of all described animal species and covering less than 1% of Earth's surface (Balian, Lévêque, Segers, & Martens, 2008). These species-rich ecosystems are heavily impacted by multiple stressors, such as climate change, pollution, fragmentation, and invasive species (Strayer & Dudgeon, 2010). This underlines the need for a better understanding of freshwater species' biogeography and the environmental factors that shape species' distributions.

Discharge can be seen as a dominant variable in riverine ecosystems, affecting a multitude of subsequent abiotic and biotic parameters (Power, Sun, Parker, Dietrich, & Wootton, 1995). Here, the flow regime, that is, the temporal sequence of river discharge, plays a major role in shaping stream and river habitats (Dewson, James, & Death, 2007; Poff, 1997). A variety of hydrological metrics can be used to quantify different dimensions of the flow regime, such as the frequency, timing, duration, and rate of change of flows. In addition, indicators of flow magnitude can be considered, such as mean annual discharge and seasonal/monthly flows, and statistical low and high flows (McGarvey, 2014; Monk, Wood, Hannah, & Wilson, 2008; Olden & Poff, 2003). Riverine flow-ecology relationships seek to reveal links between the flow regime and the functional traits of instream biota (Tonkin, Jähnig, & Haase, 2014) and have mainly focused on fish as model organisms (but see Monk et al., 2008, Stewart-Koster, Olden, & Gido, 2014, Booker, Snelder, Greenwood, & Crow, 2015, Leigh & Datry, 2016 for macroinvertebrates). Hydrologic regimes and their alterations have been recognized as a major influence in shaping freshwater fish occurrence patterns at the basin scale (Oberdorff, Guégan, & Hugueny, 1995; Xenopoulos & Lodge, 2006), and large-scale discharge patterns have been shown to impact fish species richness and composition (McGarvey, 2014; Mims & Olden, 2013). Several factors have been identified as contributing to fish distribution patterns, such as short-term flow variation and the scale at which the patterns are analysed (Jackson, Peres-Neto, & Olden, 2001; Olden & Poff, 2003; Stewart-Koster et al., 2014).

Though fish are ecologically and economically important, about 60% of animal species in freshwater ecosystems are benthic stream macroinvertebrates (Balian et al., 2008). Understanding the distribution patterns of these highly diverse organisms is an important goal in freshwater biogeography and conservation (Dijkstra, Monaghan, & Pauls, 2014). Macroinvertebrates play a crucial role in freshwater ecosystems regarding food webs and species community structure and are sensitive to environmental influences. Moreover, they are used in the bioassessment of streams and rivers (Bonada, Prat, Resh, & Statzner, 2006; Johnson, Wiederholm, & Rosenberg, 1993), making them particularly useful for understanding the impact of environmental and hydrological factors. However, to fully understand broad patterns of species occurrences, analyses over a "large

geographical range" but at a "fine spatial grain," are needed. Local species assemblages can be shaped through a composite of basin, catchment, and site level habitat characteristics, including site-specific hydrological conditions (Poff, 1997, Thorp, 2014, Booker et al., 2015, Domisch, Jähnig, Simaika, Kuemmerlen, Stoll, 2015). It is unclear how important fine-scale hydrological regimes are when assessing general and large-extent patterns for a variety of stream macroinvertebrate distributions. Furthermore, it is unclear which streamflow and water use metrics are potentially most important across a continental scale.

The relevance of particular predictors to species occurrences is usually assessed by regressing species presence-absence or presence-only data against environmental data at sites where the species were sampled (i.e., species distribution models [SDMs], Elith & Leathwick, 2009). Several issues need to be considered when carrying out these analyses: First, it is important to consider selecting predictors that are ecologically relevant for describing the occurrence of species, where direct measures should be preferred over proxies (Vaughan & Ormerod, 2003). Second, species occurrence data have to match spatially and temporally with the environmental data (DeWeber & Wagner, 2014; Monk et al., 2012). Hence, when using flow data, the appropriate gauging stations need to be carefully paired to each species occurrence due to the species data source heterogeneity (survey data vs. public biodiversity databases), as well as spatial and temporal discrepancies between flow and macroinvertebrate sampling (Leigh & Datry, 2016; Monk et al., 2012). Third, collinearity among predictors complicates assessment of relative importance and can lead to spurious conclusions. Dormann et al. (2013) have suggested that collinearity of  $|r| > 0.7$  warrants consideration of removing redundant predictors (Olden & Poff, 2003). Fourth, over large spatial scales (such as continental), species sampling effort may be very heterogeneous in terms of sampling effort and spatial coverage. Biased species data in geographic space has the high potential (but not necessarily) to also be biased in environmental space and could therefore lead to biased species-environment response curves as well. It may be therefore necessary to adjust for such a geographic bias in the distribution modelling (Araújo & Guisan, 2006; Wisz et al., 2008). In summary, carefully accounting for these issues provides an estimate of the most important contributing predictors that shape the distributions of the species.

In this study, we evaluate the relative importance of streamflow regimes and water use in relation to the occurrence of macroinvertebrates from 19 taxonomic orders across Europe. For streamflow regimes, we use gauging stations, whereas for water use, we consider simulated water use based on statistical data. The latter serves mainly as a proxy for water quality alterations (for which no suitable pan-European data set is currently available). As locations of streamflow gauging stations seldom coincide with macroinvertebrate sampling locations, we tested various pairing methods to link macroinvertebrate occurrences to streamflow observations. We then used a Generalized

Linear Modelling (GLM) approach to relate macroinvertebrate occurrences to a number of environmental predictors.

In the absence of wide-ranging instream environmental data, we acknowledge that a variety of methods have been used to estimate discharge at ungauged locations, for example, through scaling observed flow at gauged streams to ungauged streams based on the ratio of the drainage areas (McMahon, Fenton, Stewardson, Costelloe, & Finlayson, 2002), hydrological regionalization (Patil & Stieglitz, 2011, and reviewed in Olden, Kennard, & Pusey, 2012, Patil & Stieglitz, 2012), or by means of correlative analyses of daily discharge data (Yuan, 2013). Our primary aim, however, is to test a species sampling location-gauging station matching strategy that could be also used in less studied regions without high-resolution environmental instream data calibrating and validating the modelled (discharge) data. Specifically, we aimed to answer the following questions, with a focus on the methodology:

1. What is the optimal method to spatially pair macroinvertebrate sampling locations with streamflow gauging stations?
2. Which predictors of hydrological regime and water use contribute most to macroinvertebrate occurrences in Europe?
3. What is the effect of accounting for a geographic sampling bias on the predictor contributions?

## 2 | MATERIAL AND METHODS

### 2.1 | Species data

We used the [freshwaterecology.info](http://freshwaterecology.info) database (Schmidt-Kloiber & Hering, 2015) as a baseline for the taxonomic nomenclature of stream macroinvertebrates, and synonyms were unified to avoid redundancy. We then used this baseline to scan for macroinvertebrate presence records (geographic coordinates with their sampling date) in several databases across Europe (Table 1) to retrieve records for those species.

We removed duplicate records, and only those sampled after 1950 were considered; this matches the time period covered by the hydrological and climatic predictors. This procedure yielded a total of >3 million occurrence records, collected at 191,752 unique locations across Europe (Figure S1).

### 2.2 | Observed streamflow from gauging stations

Monthly discharge data from gauging stations since 1950 was provided by the Global Runoff Data Centre (<http://grdc.bafg.de>). If necessary, we adjusted locations to match a gridded flow accumulation based on a 15 arc-sec (0.00417 degree) digital elevation model (Lehner, Verdin, & Jarvis, 2008), resulting in 1,514 unique stations (see Supporting Information for a detailed description). Flow accumulation refers to the total number of grid cells that comprise the upstream contributing stream network relative to any specific point in the network. For each station, we applied two criteria: (a) Species sampling took place within a radius of 10 km, and (b) at least 10 years of monthly streamflow data was available. Accordingly, 757 stations across Europe were used in our analyses with 32,942 species sampling locations and 169,631 unique species records.

We calculated all hydrological predictors from monthly flow time series, for only complete calendar years of gauging records. Long-term averages were calculated from 10 to 30 years of observations and, where possible, for the time period 1961–1990 (the reference period of the WaterGAP model for the simulated water use data, see below). In the case of incomplete years, the next available complete year was merged to the given series, to allow for a minimum 10-year series (see Table 2 and Table S2).

### 2.3 | Simulated water use predictors

Water abstraction for domestic use, manufacturing, thermal power plant cooling, and irrigation was represented by simulated gridded water withdrawal data derived from the WaterGAP water use submodels (Water-Global Analysis and Prognosis; [www.uni-](http://www.uni-)

**TABLE 1** Data sources and the spatial coverage for retrieving species point records for the analyses

Source	Spatial coverage	Reference/download
Standardization of River Classifications (STAR)	Europe	(Furse et al., 2006)
River InVertebrate Prediction And Classification System (RIVPACS)	UK	(Wright, Sutcliffe, & Furse, 2000)
RElationships Between Ecological and Chemical stAtus in surface waters (REBECCA)	Norway	(Schartau, Moe, Sandin, McFarland, & Raddum, 2008)
Environmental database (MVM)	Sweden	(Fölster, Johnson, Futter, & Wilander, 2014)
GUADALMED project	Spain	(Bonada, Zamora-Muñoz, El Alami, Murría, & Prat, 2008; Bonada, Zamora-Muñoz, Rieradevall, & Prat, 2004)
Personal communication (C. Zamora-Muñoz, M. Sáinz-Bariáin, N. Bonada)	Mediterranean Basin	Unpublished
Environmental agencies: UBA, HLUG, LUBW	Germany	Unpublished
Global Biodiversity Information Facility (GBIF)	Europe	<a href="http://www.gbif.org">www.gbif.org</a>
Zoological-Botanical Database (ZOBODAT)	Europe	<a href="http://www.zobodat.at">www.zobodat.at</a>

Note. The GBIF and ZOBODAT databases served for range-filling procedure with the aim of obtaining additional records across the entire species range. Species records only derived from GBIF or ZOBODAT, but not from the other survey-based datasets, were not added to the analyses.

**TABLE 2** The predictor codes, explanations, data units, temporal scale, spatial grain, and the data source used in the analyses

Predictor code	Predictor explanation	Data unit	Temporal scale	Spatial grain	Data source
Q mean	Avg. annual discharge	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
Q winter	Avg. discharge Dec + Jan + Feb/Q mean	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
Q spring	Avg. discharge Mar + Apr + May/Q mean	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
Q autumn	Avg. discharge Sep + Oct + Nov/Q mean	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
Q dynamics	Avg. monthly Q min–Q max/Q mean	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
Q 90events	Avg. number of events where monthly Q < = Q 90	Counts	10–30 yr	Point	GRDC
Q 90months	Avg. number of consecutive months per year where monthly Q < = Q 90	No. of months	10–30 yr avg.	Point	GRDC
Q 90std	Monthly low flow Q (Q 90/Q mean)	m <sup>3</sup> /s	10–30 yr avg.	Point	GRDC
WW domestic	Domestic water withdrawal	m <sup>3</sup> /yr	1961–1990	0.5°	WaterGAP
WW manufacturing	Manufacturing industry water withdrawal	m <sup>3</sup> /yr	1961–1990	0.5°	WaterGAP
WW electric	Water withdrawal for electricity generation in thermal power plants	m <sup>3</sup> /yr	1961–1990	0.5°	WaterGAP
WW irrigation	Irrigation water withdrawal	m <sup>3</sup> /yr	1961–1990	0.5°	WaterGAP
Temperature	Annual mean air temperature	°C*10	1950–2000	0.00833°	WorldClim
Slope	Terrain slope representative for stream slope	Degree of slope	–	0.00833°	Hydro1k

Note. Avg = average; Q = discharge; WW = water withdrawal. Data sources are Global Runoff Data Centre (GRDC, [www.bafg.de/GRDC](http://www.bafg.de/GRDC)), WaterGAP ([www.uni-frankfurt.de/45218063/WaterGAP](http://www.uni-frankfurt.de/45218063/WaterGAP)), WorldClim ([www.worldclim.org](http://www.worldclim.org)), and Hydro1k ([ita.cr.usgs.gov/HYDRO1K](http://ita.cr.usgs.gov/HYDRO1K)). Q90 indicates that during 90% of the months, the observed streamflow is higher than this value.

[frankfurt.de/45218063/WaterGAP](http://frankfurt.de/45218063/WaterGAP); Flörke et al., 2013; Müller Schmied et al., 2014). WaterGAP simulates the macro-scale behaviour of the terrestrial water cycle, including the anthropogenic impact. These data at 0.5° spatial grain were averaged to mean annual water withdrawals for the period 1961–1990 (Table 2) and extracted for each gauging station.

We note that return flows, resulting from water withdrawal, generally decrease natural streamflow that are in turn reflected by streamflow observation in a feedback mechanism (Döll, Fiedler, & Zhang, 2009). Thus, the observed hydrological regimes are already affected by water withdrawals. However, return flows, for example, in the form of untreated wastewater, may also cause an alteration of water quality and thus affect stream macroinvertebrate occurrences. For example, rural agriculture is globally a source of contaminants, for example, from dissolved fertilizers and nutrients, chemical run-off (e.g., pesticides), and livestock manure (Corcoran, 2010; FAO, 2012). About 80% to 95% of irrigation water is currently assumed to be return flow (Döll, Müller Schmied, Schuh, Portmann, & Eicker, 2014; Döll et al., 2012). Untreated wastewater from the domestic and manufacturing sectors can be contaminated by a variety of harmful dissolved or suspended matter (Flörke et al., 2013). Given the lack of consistency in water quality data at the European scale (Sato, Qadir, Yamamoto, Endo, & Zahoor, 2013), we found it useful to substitute water quality predictors by gridded estimates of water withdrawals (Flörke et al., 2013, Müller Schmied et al., 2014).

## 2.4 | Climatic and topographic predictors

Models were supplemented with gridded 30 arc-sec (0.00833 degree) climate and topography (a) to fill the possible gap in unexplained variance of hydrological predictors, and (b) to serve as a reference to judge the relative importance of the hydrological predictors. Mean annual air temperature and slope were extracted from the WorldClim and

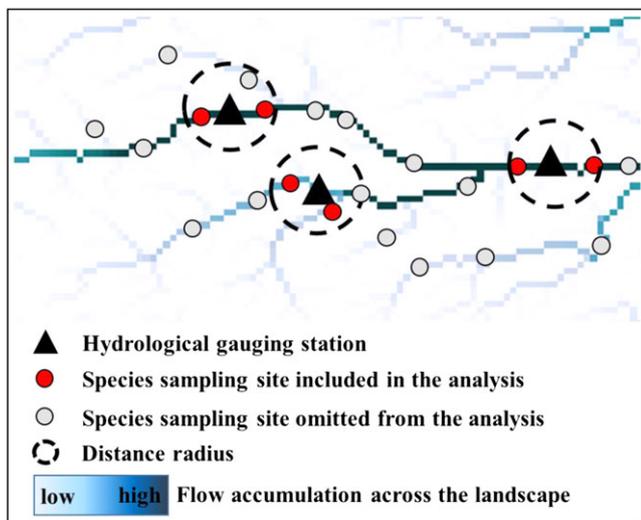
Hydro1K databases (USGS; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) for each macroinvertebrate sampling location (Table 2). Air temperature and slope were used as surrogates for water temperature (Caissie, 2006) and flow velocity, respectively.

## 2.5 | Pairing of sampling locations with streamflow gauging stations

We related the species sampling locations to the respective discharge gauging station using a combination of the Euclidian (straight line) distance and the difference in flow accumulation ( $\Delta$ F<sub>A</sub>) between the gauging station and the species sampling locations. Here,  $\Delta$ F<sub>A</sub> helped to identify cases where species and hydrological data were likely to be collected in river reaches of similar size. We tested five pairing methods (Figure 1) and only species occurrences found within these ranges were used (Tables S1 and S2). Species occurrences needed to be within a distance of (a) 10 km to the nearest gauging station with a maximum  $\Delta$ F<sub>A</sub> of  $\pm 15\%$ , (b) 10 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub>, (c) 3 km and  $\pm 15\%$   $\Delta$ F<sub>A</sub>, (d) 3 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub>, and (5) 3-km distance. These distance values are in line with studies analysing data using pairing distances from 200 m up to 25 km (Leigh, 2013; Leigh & Datry, 2016; Monk et al., 2012).

Essentially, the different pairing methods applied a filter on the macroinvertebrate sampling locations that were considered in the models (Figure 1, Table S2). For instance, the “10 km and  $\pm 15\%$   $\Delta$ F<sub>A</sub>” combination allows most species sampling locations in tributaries to be included in the analysis. In contrast, the most conservative method (3 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub>) omits multiple species sampling locations that are not located on the main stem.

In addition, we created two species subsets having at least 50 and 20 unique geographic occurrences, respectively (Stockwell & Peterson, 2002). The smaller set of 57 species (50 unique occurrences) was considered more robust and was therefore used for identifying the optimal



**FIGURE 1** Scheme of the pairing methods based on the distance (in km) and difference in flow accumulation ( $\Delta$ F<sub>A</sub>) between stream macroinvertebrate sampling locations and discharge gauging stations. We first checked if species sampling locations fall within a radius (3 or 10 km) of the gauging station, and then if the  $\Delta$ F<sub>A</sub> between the gauging station and the species sampling locations did not exceed 5% or 15%. In total, we used five combinations of distance and  $\Delta$ F<sub>A</sub>

pairing method, with 459 available gauging stations (Figure 2). For all further analyses, we relaxed the number of unique occurrences to 20 and aimed to use a larger set of 151 species with 487 available gauging stations (Figure 2, Tables S1 and S2).

## 2.6 | Preselection of predictors

From an initial pool of 39 predictors (Table S3), we selected 14, avoiding high collinearity among predictors and selecting those we

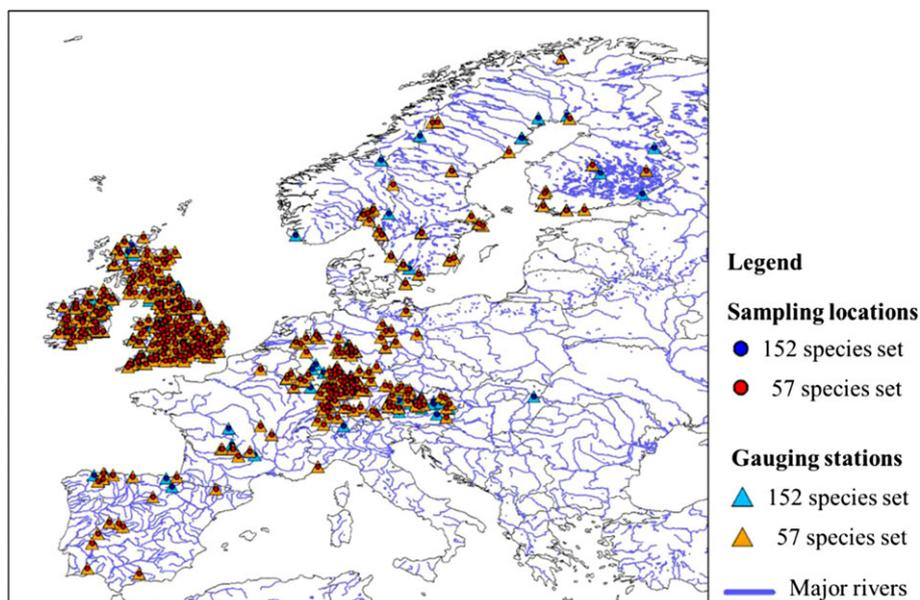
deemed relevant for our study (Table 1;  $r < |0.7|$ , Dormann et al., 2013). The only exception was the relative magnitude of low flows ( $Q_{90std}$ ), which was the only predictor representing low flow conditions, and which was correlated with seasonal dynamics ( $Q_{dynamics}$ ,  $r = 0.76$ ). As the species encompass a wide taxonomic diversity, we kept a heterogeneous set of predictors to enable building models with only one significant predictor. In other words, we calibrated a model for each species using all 14 predictors and then identified the most important predictor for each. We scaled and standardized all predictors (centred to zero mean and unit variance) to make the regression coefficients and thus the predictor contribution comparable between models.

## 2.7 | Accounting for sampling bias

The spatial density of macroinvertebrate sampling locations differed among countries, with a high density in the British Isles (Figure 1). We tested for bias-derived effects using the sampling locations “as is,” and weighting each location to decrease the bias (weighted method sensu Araújo & Guisan, 2006). Weighting factors were calculated for each pairing method as the inverse square root of the spatial density of sampling locations per country (or country group), ranging between 1.0 for high data density countries (UK) and 40.2 for countries with sparse data (e.g., Eastern Europe, Table S4).

## 2.8 | Statistical modelling

Species occurrence (i.e., presence) was treated as a Bernoulli variable, where absences were locations where other but the target species were recorded. As most species records originated from survey data with explicit species-specific absence records, this method was considered



**FIGURE 2** Spatial distribution of macroinvertebrate sampling locations and gauging stations across Europe under the most conservative pairing method of 3 km and a difference in  $\pm 5\%$  flow accumulation ( $\Delta$ F<sub>A</sub>) that was used for model comparison. Flow accumulation refers to the upstream contributing area of a stream reach. Blue points and triangles mark the data with 151 species (min. 20 occurrences), whereas red points and orange triangles represent the data with 57 species (min. 50 occurrences). Note that behind every red point/orange triangle is a blue one

more appropriate than randomly drawing pseudoabsences. For each species, data were split into 70% calibration and 30% validation subsets.

We fit a GLM for each species separately, assuming a binomial error distribution, with an elastic net regularization (Zou & Hastie, 2005), using the “cv.glmnet” function within the “glmnet” package (ver. 1.9–5) in R (ver. 3.0.2) (Friedman, Hastie, & Tibshirani, 2010; R Development Core Team, 2017). The elastic net regularization penalizes the coefficients, so that the coefficients for predictors that have little or no effect get shrunk to zero. This is useful when many predictors are thought to have a negligible effect, and also when the number of predictors is higher than the number of species presence points (Friedman et al., 2010). We set the tuning parameter lambda to automatic, ran a 10-fold internal cross-validation (which randomly selects parts of the dataset for its analysis and selects the best model), and let the GLM run 10 repetitions for each species (see the example R-code in the Supporting Information). The glmnet function screens internally at every lambda step and checks after convergence if any violations occurred. To evaluate the models, we extracted the regression coefficients and the deviance for each repetition. The deviance is minus twice the log-likelihood of the model fitted to the data, so a smaller deviance indicated a better-fitting model. Across the 10 repetitions for each species, we calculated coefficient means, standard deviation, minimum and maximum, the final number of non-zero predictors, and also mean and standard deviation of the deviance. Accounting for the sampling bias was done using the “weights” option in the “cv.glmnet” function. Overall, the described modelling approach yielded faster converging results and a more conservative selection of significant predictors than preliminary tests with a longer list of predictors and with a GLM fitted by maximum likelihood.

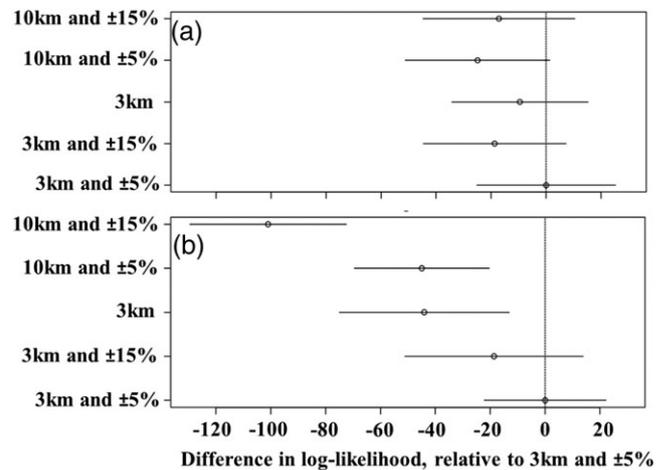
We summarized the mean, maximum, and minimum predictor coefficients across all species to assess the relative importance of each predictor. To distinguish positive and negative predictor effects across species, we summed the mean positive or negative coefficients and standardized them (i.e., dividing by the number of species). We also counted the number of positive and negative predictor counts with non-zero coefficients. Finally, we ranked the predictors according to the sum of absolute (i.e., the positive and negative) mean coefficients. In summary, we built 57 models to determine the optimal pairing method and a further 151 models to examine the key coefficients.

### 3 | RESULTS

#### 3.1 | Optimal pairing method based on model evaluation

Among the five pairing methods, and whether ignoring or accounting for a sampling bias, the most conservative pairing method with 3 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub> (Figure 1c) performed best in terms of model evaluation (i.e., deviance) based on held out validation data (Figure 3). The subsequent results are therefore based on this pairing method.

The model deviances, described by the spread of log-likelihood values around the mean, show how the different models within the nonweighted model type generally perform better (Figure 3a, spread



**FIGURE 3** Raw log-likelihoods derived from the set of 57 species (used to test the goodness of fit of the different pairing method schemes) and the different pairing methods, with the range from minimum to maximum centred around the mean relative to pairing method 3 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub>; (a) derived from the data “as is,” and (b) after accounting for a sampling bias (weighted)

closer to the best-performing pairing method of 3 km and  $\pm 5\%$   $\Delta$ F<sub>A</sub>) than those in the weighted model type (Figure 3b).

#### 3.2 | Predictor ranking (larger species set of 151 species)

The following predictors ranked highest with regard to counts and absolute coefficient sums, meaning that these were selected most often as important predictors: domestic ( $WW_{\text{domestic}}$ ), manufacturing ( $WW_{\text{manufacturing}}$ ) and irrigation ( $WW_{\text{irrigation}}$ ) water withdrawals, seasonal dynamics ( $Q_{\text{dynamics}}$ ), and discharge of the autumn months ( $Q_{\text{autumn}}$ , Table 3, Figure 4).

Among the hydrological indicators,  $Q_{\text{autumn}}$  and  $Q_{\text{dynamics}}$  showed the strongest, mostly negative model coefficients, whereas mean annual streamflow ( $Q_{\text{mean}}$ ) was a significant predictor for many species but contributed less than  $Q_{\text{dynamics}}$ . Most macroinvertebrate species occurred preferentially in habitats with a low  $Q_{\text{mean}}$ , a low  $Q_{\text{dynamics}}$ , or a high winter streamflow ( $Q_{\text{winter}}$ , Figure 4). The strongest model coefficients were found for  $WW_{\text{irrigation}}$ ,  $WW_{\text{domestic}}$ , and  $WW_{\text{manufacturing}}$ , with most macroinvertebrate species occurring preferentially at locations with small  $WW_{\text{irrigation}}$  or  $WW_{\text{manufacturing}}$ , and some species reacting positively and others negatively to  $WW_{\text{domestic}}$  (Table 3, Figure 4).

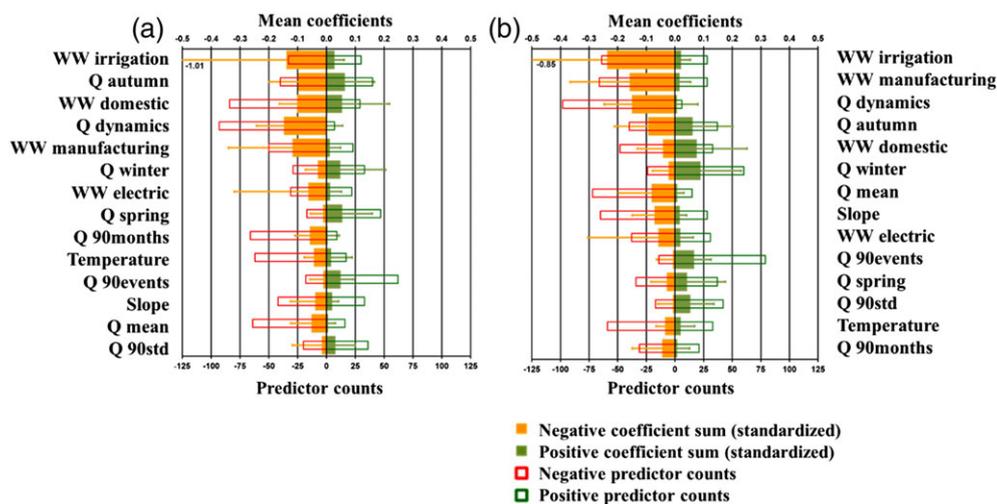
#### 3.3 | Effect of bias reduction on predictor ranking

The nonweighted analysis produced a different predictor ranking, with the highest ranking  $WW_{\text{domestic}}$  appearing in the nonweighted models more often, than the highest ranking  $Q_{\text{dynamics}}$  in the weighted models (Table 3). The largest difference regarding the mean coefficient was observed for  $WW_{\text{irrigation}}$  (nonweighted models: 0.167; weighted models: 0.260). Essentially, applying weights in the model yielded a slightly different set of predictors that were selected most often.

**TABLE 3** Predictor count and the sum of absolute model coefficients for the optimal pairing method 3 km and  $\pm 5\%$   $\Delta$ Fa for the nonweighted and weighted analyses for the larger set of all 151 species

Predictor	Nonweighted		Weighted	
	Predictor count	Sum abs. coefficients	Predictor count	Sum abs. coefficients
Q mean	<b>80</b>	0.058	87	0.089
Q winter	61	0.079	83	0.112
Q spring	63	0.067	70	0.069
Q autumn	<b>79</b>	<b>0.163</b>	77	<b>0.156</b>
Q dynamics	<b>100</b>	<b>0.152</b>	<b>104</b>	<b>0.159</b>
Q 90events	79	0.061	92	0.075
Q 90months	75	0.063	52	0.053
Q 90std	56	0.048	59	0.062
WW domestic	<b>113</b>	<b>0.156</b>	81	<b>0.118</b>
WW manufacturing	72	<b>0.130</b>	<b>94</b>	<b>0.176</b>
WW electric	53	0.076	69	0.078
WW irrigation	63	<b>0.167</b>	<b>91</b>	<b>0.260</b>
Temperature	79	0.061	92	0.055
Slope	74	0.059	92	0.088

Note. The sum of absolute model coefficients is standardized by the total number of species. The five highest counts and coefficient sums are in bold. The order of the predictors follows Table 2. Q = discharge; WW = water withdrawal.

**FIGURE 4** Ranking of the 14 predictors for the larger species set (151) under the most conservative pairing method of 3 km and  $\pm 5\%$   $\Delta$ Fa. (a) Using the data “as is” (nonweighted), and (b) after accounting for a sampling bias (weighted). Red and green bars indicate negative and positive coefficients, respectively. Solid bars indicate the standardized sum of positive (green) and negative (orange) coefficients (top x-axis), where error bars represent the standard deviation across all species. Hollow bars indicate the positive and negative predictor counts in the models (bottom x-axis). Predictors are sorted by the absolute magnitude of the standardized coefficients. Q = discharge; WW = water withdrawal

## 4 | DISCUSSION

Our study quantifies the relative influence of observed decadal hydrology, simulated water use, and climate and topography on stream macroinvertebrates distributions at the European scale. The most important hydrological predictors underline the contribution of seasonal effects of stream flow, and simulated domestic, manufacturing, and irrigation water use on macroinvertebrate occurrences. From a methodological point of view, species records can be related to the appropriate gauging stations in a post-sampling process using flow accumulation (i.e., size of the upstream contributing area) as a simple proxy for the proximity between the

locations. Accounting for a geographic sampling bias did not change the main outcome.

### 4.1 | Optimal pairing method of sampling and gauging locations

When species records are sampled independently from the hydrological data, the flow accumulation serves as a method for assigning species records to the gauging station. Here, the Euclidian distance is important; however, better performance in terms of model deviance was achieved by allowing the species sampling locations to be only “a flow accumulation of  $\pm 5\%$  away” in addition to the distance. This

largely eliminates sampling locations that do not belong to the same river reach as the gauging station. This is not surprising, as the environmental conditions closest to the stations should best reflect those at species' sampling locations. However, even the pairing method with the most restrictive distance of 3 km alone was not able to yield the same robustness of the results as when flow accumulation was used as an additional criterion. This has important implications for freshwater biogeographic studies and SDMs where usually a "snapping tolerance" of 1 to 3 km is used to assign and move species records to the stream network (Benstead & Leigh, 2012).

We note that for a successful model comparison, the same set of data for each pairing method was required. Following Liebig's Law of the minimum, the 57 species selected had a minimum of 50 unique occurrences under the most conservative pairing method. While plenty of species records were available across Europe (see Figure S1), the species set was further narrowed by the availability of representative streamflow data and hence the gauging stations (Figure 2).

An alternative procedure could have consisted of using only those species records where corresponding rivers could be correctly identified in the first place (i.e., both species and hydrological data were sampled simultaneously). This procedure was, however, considered to be subject to at least three limitations. First, most species occurrences would have been omitted from the analyses due to the smaller data set, yielding not only a smaller set of species but spatially and environmentally nonrepresentative species distributions. Second, the ability to match the names of sampling sites and gauging stations would have been limited, due to naming ambiguity (incorrect or multiple spellings of river names, multiple rivers having the same name or changing names after confluences). Last, for taxa, which are strong flight dispersers, the occurrence in a nearby tributary water body may be still associated with flow characteristics of the main gauged river. We are aware that for Europe, river names and topological Pfafstetter codes exist, starting at the river mouth, whereas Strahler codes start at headwater river reaches (De Jager & Vogt, 2010). As we wanted to design a matching strategy applicable also for other continents than Europe, this source would have been only of limited use to pair sampling locations and gauging stations.

## 4.2 | Predictor ranking

Predictor rankings indicate the importance of seasonal streamflow variability, autumn discharge, and water withdrawals for domestic, manufacturing, and irrigation use. These results highlight the impact of both natural and anthropogenically altered streamflow. The latter may be seen as a proxy for different types and degrees of water pollutants that are introduced into rivers by the return flows of these sectoral water uses. Thus, many streams and rivers are anthropogenically impacted, which determines the present-day distribution of macroinvertebrates across Europe (Döll et al., 2009). The effect of water withdrawals for cooling of thermal power plants ( $WW_{\text{electric}}$ ) was small compared to the magnitude of the other water use indicators. This may be due to the more local and patchy distribution of such thermal pollution sources, thus influencing macroinvertebrate occurrences less than chemical pollution.

Macroecological predictors in freshwater studies, represented here by air temperature and slope, had an intermediate influence in shaping the distribution of macroinvertebrate species. While such predictors are valuable in assessing the "big picture" on coarse spatial grains (Domisch, Jähnig, et al., 2015), characterization of in-stream habitat can be further improved by adding hydrological regime and anthropogenic water usage (Figure 4). Thus, our study shows that accounting for the composite of scale-dependent characteristics yields valuable information regarding freshwater species occurrence patterns, as done here with climate (basin), hydrology and water use (catchment), and local topography (site) (Poff, 1997, Thorp, 2014, Domisch, Jähnig, et al., 2015).

We note that using modelled in-stream conditions provides an alternative to account for the small-scale variation in abiotic conditions and has been demonstrated in previous studies (Chinnayakanahalli, Hawkins, Tarboton, & Hill, 2011; Hill & Hawkins, 2014; Jähnig et al., 2012; McMahon et al., 2002). Such modelled predictors may provide well-performing models; however as a trade-off, modelled predictors may also introduce additional uncertainties especially when applied to a multitude of predictors (similar to downscaling coarser grain data over considerable distances to create range-wide input data).

The importance of seasonal flow dynamics ( $Q_{\text{dynamics}}$ ) highlights the analogy to the so-called bioclimatic predictors (Busby, 1991) that are frequently used in terrestrial biogeographic studies and are based on SDMs. Monthly hydrological metrics can therefore be useful in freshwater predictive modelling studies, once aggregated to, for example, quarterly "hydroclim" predictors (Kuemmerlen et al., 2012, Domisch, Amatulli, Jetz, 2015).

Several studies found the short-term streamflow of equally major relevance. For instance, Wagner, Marxsen, Zwick, and Cox (2011) reported that monthly discharge patterns influence the occurrence of macroinvertebrate species in a small catchment of central Germany, and Stewart-Koster et al. (2014) showed that the occurrence of fish in 32 North American streams depended on the discharge the year before the sampling event. Kuemmerlen, Schmalz, et al. (2015) found that among the best predictors for macroinvertebrate in SDMs was the number of days in the year with flow exceeding mean annual discharge by 150% and by 200% in the within the Chinese Changjiang catchment and in the German (Hesse) Kinzig catchment, respectively (Kuemmerlen, Stoll, Sundermann, Haase, 2015). These studies were performed at smaller spatial scales or single rivers, and such short-term variables (accounting for the discharge in previous year) were not available for the majority of the sampling locations at the European scale and were therefore omitted from the analyses prior to modelling.

It is important to bear in mind that a large portion of the 151 species are likely to be generalist species, and our results could emphasize those hydrologic predictors that influence the distribution of these wide spread species. In addition to a minimum of 20 unique occurrences to yield robust model estimates, the macroinvertebrate sampling locations needed to be in the vicinity of the gauging stations (which are mostly located along larger rivers, and scarce in Mediterranean intermittent rivers), we narrowed the ecological preferences of the species to only those that were considered that can thrive at these hydrological conditions in the first place (i.e., where the species occur).

### 4.3 | Effect of bias reduction on predictor ranking

In contrast to our expectations, accounting for sampling bias only marginally changed the predictor ranking. On the one hand, the model deviance of the different pairing methods using the “raw” data was scattered less around the mean deviance derived from the optimal pairing method than those using weighted data (and when all other factors were kept constant, Figure 3a,b). We interpret this as an effect of the geographic bias of the data, where the models are largely trained on the data from the British Isles, and thus spatial autocorrelation is more likely to lead to overfitting (e.g., Wenger & Olden, 2012). On the other hand, the larger ranges of deviance for the weighted models potentially better represent the more variable conditions of the predictors across Europe, as the weights facilitate the discrimination of the data. While any bias correction needs to be assessed with care, in our case, both options yielded almost identical results in terms of predictor rankings.

### 4.4 | Conclusions

At large spatial scales, the “best case scenario” of simultaneously sampled biological and hydrological data is seldom available. Linking the two independently collected data types together in a post-sampling step increases the sample size and may add to the understanding of freshwater species distribution patterns and provide a new dimension for freshwater biogeographic questions. The hydrological regime and simulated water use ranked higher than temperature and slope, potentially influencing the species occurrences most in our study. This emphasizes the importance of the hydrological metrics in large-scale freshwater biogeographic studies that are usually neglected. While this study is focused on the methodology, our findings are valuable for potential follow-up studies, such as in-depth analyses regarding the causal relationship between species and hydrological data, and regarding species group- and trait-specific preferences (via species occurrences) related to environmental conditions at large spatial scales.

### ACKNOWLEDGEMENTS

We thank the Global Runoff Data Centre for generously providing river discharge data, D. M. Hannah for his consent to use EWA data within EURO-FRIEND, and Bernhard Lehner for generating provisional discharge estimates and flow accumulation for HydroSHEDS. We thank all contributors of the freshwater ecology.info database, and all species collectors of the databases we used in this study. We thank Marie-Pierre Gosselin for valuable comments on an earlier version of the manuscript. S. D. acknowledges funding by the German Research Foundation DFG (grant DO 1880/1-1), C. Z. M. funding by the Autonomous Organism of National Parks (OAPN) of the Spanish Ministry of Environment (ref. 039/2007), N. B. funding the GUADALMED project (HID98-0323-C05 and REN2001-3438-C07) funded by the Spanish Ministry of Education and Science, and S. C. J. funding by the German Federal Ministry of Education and Research (BMBF) for funding “GLANCE” (Global change effects in river ecosystems; 01LN1320A). The Hessian state supported the work within the LOEWE initiative funding of the Biodiversity and Climate

Research Centre. We wish to thank the reviewers for their constructive comments on an earlier version of the manuscript.

### ORCID

Sami Domisch  <http://orcid.org/0000-0002-8127-9335>

Robert B. O'Hara  <http://orcid.org/0000-0001-9737-3724>

Petra Döll  <http://orcid.org/0000-0003-2238-4546>

### REFERENCES

- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Balian, E. V., Lévêque, C., Segers, H., & Martens, K. (2008). Freshwater animal diversity assessment. 198. DOI: ISBN: 1402082592
- Benstead, J. P., & Leigh, D. S. (2012). An expanded role for river networks. *Nature Geoscience*, 5, 678–679. <https://doi.org/10.1038/ngeo1593>
- Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. *Annual Review of Entomology*, 51, 495–523. <https://doi.org/10.1146/annurev.ento.51.110104.151124>
- Bonada, N., Zamora-Muñoz, C., El Alami, M., Múrria, C., & Prat, N. (2008). New records of trichoptera in reference mediterranean-climate rivers of the Iberian Peninsula and north of Africa: Taxonomical, faunistic and ecological aspects. *Graellsia*, 64, 189–208. ISSN: 0367-5041
- Bonada, N., Zamora-Muñoz, C., Rieradevall, M., & Prat, N. (2004). Trichoptera (Insecta) collected in Mediterranean river basins of the Iberian Peninsula: Taxonomic remarks and notes on ecology. *Graellsia*, 41–69. ISSN: 0367-5041
- Booker, D., Snelder, T., Greenwood, M., & Crow, S. (2015). Relationships between invertebrate communities and both hydrological regime and other environmental factors across New Zealand's rivers. *Ecohydrology*, 8, 13–32. <https://doi.org/10.1002/eco.1481>
- Busby, J. R. (1991). BIOCLIM—a bioclimate analysis and prediction system. *Plant Protection Quarterly (Australia)*. ISSN0815-2195
- Caissie, D. (2006). The thermal regime of rivers: A review. *Freshw Biol*, 51, 1389–1406. <https://doi.org/10.1111/j.1365-2427.2006.01597.x>
- Chinnayakanahalli, K. J., Hawkins, C. P., Tarboton, D. G., & Hill, R. A. (2011). Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. *Freshwater Biology*, 56, 1248–1265. <https://doi.org/10.1111/j.1365-2427.2010.02560.x>
- Corcoran E 2010. Sick water? The central role of wastewater management in sustainable development: A rapid response assessment. United Nations Environment Programme, UN-HABITAT, GRID-Arendal. www.grida.no. ISBN: 978-82-7701-075-5
- De Jager, A., & Vogt, J. (2010). Development and demonstration of a structured hydrological feature coding system for Europe. *Hydrological Sciences Journal*, 55, 661–675. <https://doi.org/10.1080/02626667.2010.490786>
- DeWeber, J. T., & Wagner, T. (2014). A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology*, 517, 187–200. <https://doi.org/10.1016/j.jhydrol.2014.05.006>
- Dewson, Z. S., James, A. B., & Death, R. G. (2007). A review of the consequences of decreased flow for instream habitat and macroinvertebrates. *Journal of the North American Benthological Society*, 26, 401–415. <https://doi.org/10.1899/06-110.1>
- Dijkstra, K.-D. B., Monaghan, M. T., & Pauls, S. U. (2014). Freshwater biodiversity and aquatic insect diversification. *Annual Review of Entomology*, 59, 143–163. <https://doi.org/10.1146/annurev-ento-011613-161958>
- Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, 13, 2413–2432. <https://doi.org/10.5194/hess-13-2413-2009>

- Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., ... Scanlon, B. R. (2012). Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, 59–60, 143–156. <https://doi.org/10.1016/j.jog.2011.05.001>
- Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, 50, 5698–5720. <https://doi.org/10.1002/2014WR015595>
- Domisch, S., Amatulli, G., & Jetz, W. (2015). Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scientific Data*, 2, 150073. <https://doi.org/10.1038/sdata.2015.73>
- Domisch, S., Jähnig, S. C., Simaika, J. P., Kuemmerlen, M., & Stoll, S. (2015). Application of species distribution models in stream ecosystems: The challenges of spatial and temporal scale, environmental predictors and species occurrence data. *Fundamental and Applied Limnology*, 186, 45–61. <https://doi.org/10.1127/fal/2015/0627>
- Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 027–046. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- FAO 2012. Water pollution from agriculture: A review. Online at [http://www.fao.org/fsnforum/cfs-hlpe/sites/cfs-hlpe/files/resources/Water%20Quality%20and%20Agriculture\\_%20Draft%2003%20final\\_0.pdf](http://www.fao.org/fsnforum/cfs-hlpe/sites/cfs-hlpe/files/resources/Water%20Quality%20and%20Agriculture_%20Draft%2003%20final_0.pdf)
- Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., & Alcamo, J. (2013). Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change*, 23, 144–156. <https://doi.org/10.1016/j.gloenvcha.2012.10.018>
- Fölster, J., Johnson, R. K., Futter, M. N., & Wilander, A. (2014). The Swedish monitoring of surface waters: 50 years of adaptive monitoring. *Ambio*, 43, 3–18. <https://doi.org/10.1007/s13280-014-0558-z>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33, 1. <http://www.jstatsoft.org/v33/i01/>
- Furse, M., Hering, D., Moog, O., Verdonschot, P., Johnson, R. K., Brabec, K., ... Krno, I. (2006). The STAR project: Context, objectives and approaches. *Hydrobiologia*, 566, 3–29. <https://doi.org/10.1007/s10750-006-0067-6>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. <https://doi.org/10.1002/Joc.1276>
- Hill, R. A., & Hawkins, C. P. (2014). Using modelled stream temperatures to predict macro-spatial patterns of stream invertebrate biodiversity. *Freshwater Biology*, 59, 2632–2644. <https://doi.org/10.1111/fwb.12459>
- Jackson, D. A., Peres-Neto, P. R., & Olden, J. D. (2001). What controls who is where in freshwater fish communities—The roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 58, 157–170. <https://doi.org/10.1139/cjfas-58-1-157>
- Jähnig, S., Kuemmerlen, M., Kiesel, J., Domisch, S., Cai, Q., Schmalz, B., & Fohrer, N. (2012). Modelling of riverine ecosystems by integrating models: Conceptual approach, a case study and research agenda. *Journal of Biogeography*, 39, 2253–2263. <https://doi.org/10.1111/jbi.12009>
- Johnson, R. K., Wiederholm, T., & Rosenberg, D. M. (1993, 1993). Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macroinvertebrates. In D. M. Rosenberg, & V. H. Resh (Eds.), *Freshwater biomonitoring and benthic macroinvertebrates* (pp. 40, 488–158) Chapman & Hall.
- Kuemmerlen, M., Domisch, S., Schmalz, B., Cai, Q., Fohrer, N., & Jähnig, S. C. (2012). Integrierte Modellierung von aquatischen Ökosystemen in China: Arealbestimmung von Makrozoobenthos auf Einzugsgebiete-sebene. *Hydrologie und Wasserbewirtschaftung*, 56, 185–192. [https://doi.org/10.5675/HyWa\\_2012.4\\_3](https://doi.org/10.5675/HyWa_2012.4_3)
- Kuemmerlen, M., Schmalz, B., Cai, Q., Haase, P., Fohrer, N., & Jähnig, S. C. (2015). An attack on two fronts: Predicting how changes in land use and climate affect the distribution of stream macroinvertebrates. *Freshwater Biology*, 60, 1443–1458. <https://doi.org/10.1111/fwb.12580>
- Kuemmerlen, M., Stoll, S., Sundermann, A., & Haase, P. (2015). Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. *Ecological Indicators* <https://doi.org/10.1016/j.ecolind.2015.08.008>
- Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89, 93–94. <https://doi.org/10.1029/2008EO100001>
- Leigh, C. (2013). Dry-season changes in macroinvertebrate assemblages of highly seasonal rivers: Responses to low flow, no flow and antecedent hydrology. *Hydrobiologia*, 703, 95–112. <https://doi.org/10.1007/s10750-012-1347-y>
- Leigh, C., & Datry, T. (2016). Drying as a primary hydrological determinant of biodiversity in river systems: A broad-scale analysis. *Ecography*. <https://doi.org/10.1111/ecog.02230>
- McGarvey, D. J. (2014). Moving beyond species-discharge relationships to a flow-mediated, macroecological theory of fish species richness. *Freshwater Science*, 33, 18–31. <https://doi.org/10.1086/674967>
- McMahon, T. A., Fenton, J. D., Stewardson, M. J., Costelloe, J., & Finlayson, B. L. (2002). Estimating discharge at an ungauged site. *Australasian Journal of Water Resources*, 5, 113–117. <https://doi.org/10.1080/13241583.2002.11465197>
- Mims, M. C., & Olden, J. D. (2013). Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies. *Freshwater Biology*, 58, 50–62. <https://doi.org/10.1111/fwb.12037>
- Monk, W. A., Wood, P. J., Hannah, D. M., Extence, C. A., Chadd, R. P., & Dunbar, M. J. (2012). How does macroinvertebrate taxonomic resolution influence ecohydrological relationships in riverine ecosystems. *Ecohydrology*, 5, 36–45. <https://doi.org/10.1002/eco.192>
- Monk, W. A., Wood, P. J., Hannah, D. M., & Wilson, D. A. (2008). Macroinvertebrate community response to inter-annual and regional river flow regime dynamics. *River Research and Applications*, 24, 988–1001. <https://doi.org/10.1002/rra.1120>
- Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, 18, 3511–3538. <https://doi.org/10.5194/hess-18-3511-2014>
- Oberdorff, T., Guégan, J. F., & Hugueny, B. (1995). Global scale patterns of fish species richness in rivers. *Ecography*, 18, 345–352. <http://www.jstor.org/stable/3682818>
- Olden, J. D., Kennard, M. J., & Pusey, B. J. (2012). A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology*, 5, 503–518. <https://doi.org/10.1002/eco.251>
- 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>
- Patil, S., & Stieglitz, M. (2011). Hydrologic similarity among catchments under variable flow conditions. *Hydrology and Earth System Sciences*, 15, 989–997. <https://doi.org/10.5194/hess-15-989-2011>
- Patil, S., & Stieglitz, M. (2012). Controls on hydrologic similarity: Role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences*, 16, 551–562. <https://doi.org/10.5194/hess-16-551-2012>

- Poff, N. L. (1997). Landscape filters and species traits: Towards mechanistic understanding and prediction in stream ecology. *J N Am Benthol Soc*, 16, 391. <https://doi.org/10.2307/1468026>
- Power, M. E., Sun, A., Parker, G., Dietrich, W. E., & Wootton, J. T. (1995). Hydraulic food-chain models. *Bioscience*, 45, 159–167.
- R Development Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN: 3-900051-07-0. <http://www.R-project.org>
- Sato, T., Qadir, M., Yamamoto, S., Endo, T., & Zahoor, A. (2013). Global, regional, and country level need for data on wastewater generation, treatment, and use. *Agricultural Water Management*, 130, 1–13. <https://doi.org/10.1016/j.agwat.2013.08.007>
- Schartau, A. K., Moe, S. J., Sandin, L., McFarland, B., & Raddum, G. G. (2008). Macroinvertebrate indicators of lake acidification: Analysis of monitoring data from UK, Norway and Sweden. *Aquatic Ecology*, 42, 293–305. <https://doi.org/10.1007/s10452-008-9190-y>
- Schmidt-Kloiber, A., & Hering, D. (2015). [www.Freshwaterecology.info](http://www.Freshwaterecology.info)—An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecological Indicators*, 53, 271–282. <https://doi.org/10.1016/j.ecolind.2015.02.007>
- Stewart-Koster, B., Olden, J. D., & Gido, K. B. (2014). Quantifying flow-ecology relationships with functional linear models. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 59, 629–644. <https://doi.org/10.1080/02626667.2013.860231>
- Stockwell, D. R. B., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1–13. [https://doi.org/10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X)
- Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: Recent progress and future challenges. *Journal of the North American Benthological Society*, 29, 344–358. <https://doi.org/10.1899/08-171.1>
- Thorp, J. (2014). Metamorphosis in river ecology: From reaches to macrosystems. *Freshwater Biology*, 59, 200–210. <https://doi.org/10.1111/fwb.12237>
- Tonkin, J. D., Jähnig, S. C., & Haase, P. (2014). The rise of riverine flow-ecology and environmental flow research. *Environmental Processes*, 1, 323–330. <https://doi.org/10.1007/s40710-014-0024-8>
- USGS U.S. Geological Survey Center for Earth Resources Observation and Science (EROS). HYDRO1k Elevation Derivative Database. LP DAAC, Sioux Falls, South Dakota. Available at: <http://eros.usgs.gov/>. USGS – U.S. Geological Survey Center
- Vaughan, I., & Ormerod, S. (2003). Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology*, 17, 1601–1611.
- Wagner, R., Marxsen, J., Zwick, P., & Cox, E. J. (2011). *Central European stream ecosystems: The long term study of the Breitenbach* John Wiley & Sons.
- Wenger, S., & Olden, J. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3, 260–267. <https://doi.org/10.1111/j.2041-210X.2011.00170.x>
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14, 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>
- Wright JF, Sutcliffe DW, Furse MT 2000. Assessing the biological quality of freshwaters. *RIVPACS and other techniques*. *Freshwater Biological Association, Ambleside, England*. Freshwater Biological Association, Ambleside, Cumbria, UK. ISBN: 0-900386-62-2
- Xenopoulos, M., & Lodge, D. (2006). Going with the flow: Using species–discharge relationships to forecast losses in fish biodiversity. *Ecology*, 87, 1907–1914. [https://doi.org/10.1890/0012-9658\(2006\)87\[1907:GWTFUS\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[1907:GWTFUS]2.0.CO;2)
- Yuan, L. L. (2013). Using correlation of daily flows to identify index gauges for ungauged streams. *Water Resources Research*, 49, 604–613. <https://doi.org/10.1002/wrcr.20070>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320. <http://www.jstor.org/stable/3647580>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Domisch S, Portmann FT, Kuemmerlen M, et al. Using streamflow observations to estimate the impact of hydrological regimes and anthropogenic water use on European stream macroinvertebrate occurrences. *Ecohydrology*. 2017:e1895. <https://doi.org/10.1002/eco.1895>