Improvement of Global Change Projections for Riverine Benthic Macroinvertebrates

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Katherine Sarah Irving

2019

This thesis work was conducted during the period 1st November 2015 – 07th August 2019, under the supervision of **Dr. Sonja C. Jähnig** (Leibniz Institute of Freshwater Ecology and Inland Fisheries), **Prof. Dr. Klement Tockner** (Freie Universität Berlin and Leibniz Institute of Freshwater Ecology and Inland Fisheries) and **Dr. Mathias Kuemmerlen** (Swiss Federal Institute of Aquatic Science and Technology and Senkenberg Research Institute and Natural History Museum) at Freie Universität Berlin and Leibniz Institute of Freshwater Ecology and Inland Fisheries Berlin.

1st Reviewer: **Dr. Sonja C. Jähnig**

2nd Reviewer: **Prof. Dr. Klement Tockner**

Date of defense: __22nd November 2019__________

I. Table of Contents

II. Summary

Species' distribution models (SDMs) are predictive models that are increasingly applied to river ecosystems as a complement to large scale observational analyses. Current ecological theory on river discharge acknowledges that hydrological flow regime is one of the most important drivers of lotic systems, influencing the abundance and distribution of river biota. However, few studies on stream SDMs incorporate specific data describing flow regime, with most studies only including data describing climate or river related surrogates. These hydrological variables have a significant impact however, they only partially represent the critical aspects of flow regime. This limitation is partially due to available hydrological data, which are often limited in their spatio-temporal extent and resolution for use in SDMs. Another major challenge in SDM studies is the selection of relevant environmental predictors, particularly when modeling large communities. Often, variable choice is made for an entire community and not for specific species, resulting in inappropriate predictors for at least some species, and affecting model performance and predicted distributions.

This thesis is method based, and my main goal was to advance the predictive ability of SDMs for riverine benthic macroinvertebrates by integrating hydrological predictors that describe flow regime. The thesis is divided into three parts: First, I developed a high resolution spatio-temporal dataset of streamflow, and a set of hydrological metrics for the German stream network. Second, I proposed a variable selection method to select the optimal environmental variables for use in SDMs, and I investigated the impact of predictor set choice in SDMs. Third, I disentangled the role of hydrology in SDMs by investigating the influence of climate and hydrology related datasets.

Using empirical streamflow data from gauging stations across Germany and modeled seasonal accumulated precipitation, I applied a weighted linear regression model to predict a continuous daily time series of streamflow (m^3 s^{-1}) spanning 64 years (1950-2013). The daily streamflow data were subsequently applied as input to successfully calculate 53 Indicators of Hydrologic Alteration (IHA), which describe the magnitude, frequency, duration, timing and change rate of high, low and average streamflow conditions. I performed temporal and spatial validations on the streamflow data, through which I confirmed that the predicted flow data are adequate for use in predictive ecological models. Both the IHA metrics and the streamflow datasets are available open access for use in predictive models.

By applying the IHA metrics, together with data describing climate, land-use, and topography in Boosted Regression Trees (BRTs), I created two predictor sets 1) a species-specific predictor set for each individual species and 2) a uniform predictor set for the community as a whole. Through this procedure, I highlighted a useful and effective method to impartially select highly relevant environmental variables. To investigate the impact of each predictor set on predictive ability, I applied SDMs on a community of macroinvertebrates. The SDMs rendered 10 species where the models increased in accuracy (Mean TSS = 0.59 ± 0.03) and 10 species where the models decreased in accuracy (Mean TSS =0.49 ± 0.04) with the species-specific predictor set. The 20 species, showed distinct differences in terms of their ecological traits, known occurrences, and preferred environmental conditions.

To investigate the separate influence of climate and hydrology, I calibrated SDMs on a community of macroinvertebrates with three datasets describing either 1) climate only (bioclimate), 2) hydrology only (hydrology) and 3) information on both climate and hydrology (hydroclimate), in four model configurations. SDMs applied with bioclimate and hydrology, performed significantly better overall (Mean TSS = 0.68 ± 0.02), exhibited the lowest unexplained variance (0.29), and predicted significantly larger range sizes (Mean no. of presences; 3482.6 ± 129.1). I found bioclimate to be the most important individual factor for species' distributions in terms of both variable importance and proportional explained variance. Despite the importance of bioclimate, hydrology contributed to a higher proportion of explained variance, unrivalled by other SDM configurations. The larger predicted range sizes may be due to the better description of the river discharge regime provided by the hydrological variables.

Through this thesis, I have created and integrated hydrological variables in SDMs, as well as developed and validated effective methods to improve prediction performance of riverine species' distribution to advance freshwater SDM research. The introduced methods can be applied in different geographical regions as well as under alternative time periods and spatial scales. Due to the implications associated with altered model accuracy and predicted range size, applying SDMs with hydrological variables has the potential to aid river management decisions and conservation efforts.

III. Zusammenfassung

Artverbreitungsmodelle (eng.: *species distribution models*; SDMs) werden zunehmend für Flussökosysteme angewandt um groß-skalige Analysen zu ergänzen. In der aktuellen ökologischen Theorie wird das Abflussverhalten als einer der wesentlichen Einflussfaktoren für das Vorkommen und die Verbreitung von Flusslebewesen beschrieben. Es gibt jedoch nur wenige Studien zur Modellierung der Verbreitung von Fließgewässerarten, die Daten berücksichtigen, die das Abflussverhalten detailliert beschreiben. Anstelle dessen, werden häufig Klimadaten, oder aber indirekte Indikatoren genutzt. Derartige indirekte hydrologische Indikatoren haben zwar einen großen Einfluss auf die Verbreitung von Fließgewässerarten, dennoch können sie die wesentlichen Faktoren des Abflussverhaltens nur teilweise abbilden. Dieses Vorgehen ist teilweise auf die Verfügbarkeit von geeigneten hydrologischen Daten für SDMs zurückzuführen, da diese meist in ihrer räumlichen und zeitlichen Ausdehnung und Auflösung limitiert sind. Eine weitere Herausforderung in der Anwendung von SDMs ist die Auswahl relevanter Umwelt-Prädiktoren bei der Modellierung großer Artgemeinschaften, da diese Entscheidung zumeist für die gesamte Artgemeinschaft vorgenommen wird und entsprechend nicht artspezifisch ist. Dies führt dazu, dass die Prädiktoren für einige Arten ungeeignet sind, was wiederum die Modellgüte und die vorhergesagten Verbreitungsmuster beeinflusst.

Das Hauptziel der vorliegenden methodischen Arbeit ist es, die Vorhersagekapazitäten von SDMs für benthische Makroinvertebraten durch Einbindung von hydrologischen Prädiktoren, die das Abflussverhalten beschreiben, zu verbessern. Die Arbeit besteht aus drei Teilen. Im ersten Teil habe ich einen zeitlich und räumlich (1 km^2) hoch aufgelösten Datensatz, der den Abfluss und eine Reihe weiterer hydrologischer Einflussgrößen beinhaltet, für Deutschland entwickelt. Im zweiten Teil habe ich eine Methode zur Ermittlung der optimalen Prädiktoren für den Einsatz in SDMs entwickelt und den Effekt der Auswahl der Prädiktoren auf SDMs untersucht. Im dritten Teil geht es um die Rolle der Hydrologie in SDMs, die ich über den Einfluss von klimatischen und hydrologischen Datensätzen untersucht habe.

Auf der Grundlage von deutschlandweit gemessenen Abflussdaten und modellierten Niederschlagsdaten, habe ich mittels gewichteter linearer Regression deutschlandweite tägliche Abflussdaten (m^3 s^{-1}) für einen Zeitraum von 64 Jahren (1950

bis 2013) erstellt. Im Anschluss wurden diese täglichen Abflussdaten verwendet, um 53 Indikatoren der hydrologischen Veränderung (IHA) zu berechnen, die die Stärke, Frequenz, Dauer, und Größe der Veränderung von Hoch- Niedrig-und Mittelwasser Ereignissen beschreiben. Die Abflussdaten wurden zeitlich und räumlich validiert, wodurch ich erfolgreich zeigen konnte, dass die modellierten IHA für SDMs genutzt werden können. Sowohl die IHA, als auch die modellierten Abflussdaten sind öffentlich verfügbar und können so für SDMs genutzt werden.

Unter Anwendung der modellierten IHA sowie der Klima-, Landnutzungs-, und topografischen Prädiktoren wurden zwei separate Sets an Prädiktoren mit Hilfe von Boosted Regression Trees (BRTs) erstellt. Ein Set war dabei artspezifisch (für jede der Arten individuell), das zweite Set war ein uniformes Set (für alle Arten gleich). Mit diesem Ansatz konnte ich die Anwendbarkeit und Effektivität der Methode aufzeigen, die eine Auswahl von Prädiktoren für individuelle Arten ermöglicht. Um den Effekt der unterschiedlichen Sets an Prädiktoren auf die Vorhersagekapazität zu untersuchen wurden diese auf eine Makroinvertebratengemeinschaft angewendet. Die individuellen Sets an Prädiktoren resultierten in einer deutlichen Verbesserung der Modellgüte für 10 der modellierten Arten (Mean TSS = 0.59 ± 0.03). Für 10 weitere Arten wurde allerdings eine deutliche Verschlechterung der Modellgüte aufgezeigt (Mean TSS =0.49 ± 0.04). Diese 20 Arten weisen sehr deutliche Unterschiede in Bezug auf ihre Traits, Vorkommenspunkte und die bevorzugten Habitateigenschaften auf.

Um die Einzeleffekte von Klima und Hydrologie auf SDMs und deren Vorhersagen abzuschätzen, habe ich für eine Makroinvertebratengemeinschaft drei verschiedene Sets von Prediktoren 1.) nur Klima, 2.) nur Hydrologie und 3.) eine Kombination aus Klima und Hydrologie in vier verschiedenen Konfigurationen untersucht. SDMs die mit einer Kombination aus nur klimatischen und nur hydrologischen Prediktoren kalibriert wurden, wiesen eine signifikant bessere Modellgüte (Mittlerer TSS = 0.68 ± 0.02) auf, hatten die kleinste unerklärte Varianz (0.29) und haben signifikant größere Verbreitungsgebiete für die einzelnen Arten vorhergesagt (Mittlere Anzahl der vorhergesagten Vorkommenspunkte 3482.6 ± 129.1). Sowohl, hinsichtlich der relativen Bedeutung der Prädiktoren, als auch in Bezug auf die erklärte Varianz in den Modellen, haben sich reine Klimaprädiktoren als wichtigste Einflussgrößen für die Modellierung der Verbreitungsgebiete der Makroinvertebraten herausgestellt. Neben der großen Bedeutung von Klimaprädiktoren zeigte sich, dass hydrologische Prädiktoren im Allgemeinen einen höheren Anteil zur erklärten Varianz beigetragen haben. Die größeren vorhergesagten Verbreitungsgebiete für die Arten basierend auf den ausschließlich hydrologischen Prädiktoren, deuten auf eine bessere Beschreibung des Abflussverhaltens durch die Prädiktoren hin.

In dieser Arbeit habe ich hydrologische Variablen für SDMs erstellt und implementiert, effektive Methoden zur Verbesserung der Vorhersagen der Verbreitung von Fließgewässerarten entwickelt und validiert und somit die Forschung im Bereich der SDMs vorangebracht. Die entwickelten Methoden können sowohl in unterschiedlichen geographischen Regionen als auch für verschiedene Zeitschnitte und räumliche Skalen angewendet werden. Durch die Verbesserung der Modellgenauigkeit und der vorhergesagten Verbreitung kann die Anwendung von SDMs somit dazu beitragen Managemententscheidungen und Naturschutzbestrebungen zu unterstützen.

IV. Thesis outline

This thesis is composed of three manuscripts (Chapters 2 to 4) that are either published, accepted or ready to be submitted to peer-reviewed journals. Chapter 1 is structured as a data descriptor, with an introduction, methods, technical validation, and usage notes. Chapters 3 & 4 are research articles, and each have an introduction, methods, results and discussion. The general introduction (Chapter 1) provides the general context of the thesis and the results of the thesis are discussed in the general discussion (Chapter 5). The layout of the three manuscripts was modified and figures, tables and paragraphs are numbered consistently through the thesis. The references of all chapters are situated at the end of each section. The research aims of Chapters 2, 3 and 4 are described in Paragraph 1.5.

Chapter 1:

General introduction

Chapter 2:

Irving K, Kuemmerlen M, Kiesel J, Kakouei K, Domisch S, Jähnig SC. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model [Data Descriptor]. *Scientific Data.* 5:180224. doi: 10.1038/sdata.2018.224

Author contributions: K.I designed the study, coded the model, computed the data, analyzed the results and wrote the manuscript. M.K, S.J, S.D & J.K co-designed the study. S.D, K.K & M.K helped draft the code. All authors advised on methodology, discussed the results and commented on the manuscript.

Chapter 3:

Irving K, Jähnig SC. Kuemmerlen M, 2019. Identifying and applying an optimum set of environmental variables in species distribution models. *Inland Waters.* Accepted. doi; 10.1080/20442041.2019.1653111

Author contributions: K.I designed the study, coded the models, analyzed the results and wrote the manuscript. M.K, S.J, co-designed the study, advised on methodology, discussed the results and commented on the manuscript.

Chapter 4:

Irving K, Jähnig SC. Kuemmerlen M, 2019. Disentangling the influence of climatic and hydrological predictor variables on benthic macroinvertebrate distribution. To be submitted.

Author contributions: K.I designed the study, coded the models, analyzed the results and wrote the manuscript. M.K, S.J, co-designed the study, advised on methodology, discussed the results and commented on the manuscript.

Chapter 5

General discussion

V. List of Figures

VI. List of Tables

1 Chapter 1: General Introduction

1.1 Background

Species' distribution models (SDMs) are increasingly applied to river ecosystems as a complement to large scale observational analyses and are used to aid conservation efforts. To establish adequate conversation solutions, it is important to include a full representation of the stream ecosystem in SDMs. Stream networks involve complex relationships with interacting factors such as land use, topography and geology (Allan 2004). Therefore, by including various aspects of the ecosystem, a more extensive impression of a species' environmental niche is depicted, resulting in higher predictive ability. Current ecological theory on river discharge acknowledges that hydrological flow regime is one of the most important drivers of lotic systems, influencing the abundance and distribution of river biota (Resh et al. 1988, Poff 1997, Bunn & Arthington, 2002). However, few studies on freshwater SDMs include specific data describing flow regime, which could be due, in part, to 1) insufficient availability of data describing the stream environment and 2) the complex interactions between the numerous driving factors of river ecosystems.

The implementation of SDMs has traditionally relied on surrogate variables, e.g. air temperature, as actual measurements of environmental conditions. Variables describing direct environmental conditions are rare for continuous networks, even at coarse spatial scales. Data at fine temporal resolutions is even less common. Therefore, a certain acceptance for surrogate data has been established, with many studies using alternative information such as precipitation and river basin characteristics as proxies for hydrology (Maloney et al. 2013, Zeng et al. 2015, Domisch et al. 2019). The consequence of these approximations is that SDM predictions have remained suboptimal. Novel efforts are targeting this issue, aiming at developing environmental datasets which are, at least partially, more realistic. Although, datasets describing the hydrological flow regime at large scales are still rather scarce.

Even with appropriate datasets, applying the most relevant assortment of variables in SDMs, is a considerable challenge. It is best practice to include the most optimal predictors, based on "true cause-effect relationships" (Araújo et al. 2019) between the species and its associated environmental conditions. Determining such variables, requires substantial knowledge regarding the described species, which may not be available, or feasible to collect. To aid optimal variable selection, numerous procedures are outlined in existing SDM literature; however, currently there is no general agreement on the most suitable process (Petitpierre et al. 2017). As a consequence, traditionally SDMs are applied with an identical set of environmental predictors for the entire community, which may be of minor relevance for a portion of species, leading to, further, suboptimal predictions. Based on this viewpoint, there is an urgent need to bridge some gaps in freshwater SDM research. This thesis aims to develop methodological approaches to integrate hydrological variables that describe flow regime in order to improve SDM performance and provide reliable distribution predictions by filling the following research gaps:

1. The limited availability of high-resolution data describing flow regime.
2. Applying the optimal environmental predictors to estimate species' distribution in stream networks.
3. The integration and role of variables describing flow regime in SDMs.

1.2 Species Distribution Models

Species' distribution models (SDMs) are predictive models that statistically relate the known occurrence of a species with its associated environmental conditions. Predictions on the distribution of the species are then made in geographical space under, for example, current or future time periods or in alternative geographical regions. A common application, for example, is predicting species' distribution under future climate change (e.g. Araujo et al. 2005, Elith and Leathwick 2009, Zimmermann et al. 2009, Balint et al. 2011, Domisch et al. 2013, Markovic et al. 2014). SDMs are increasingly used to complement large scale distribution analyses, where it is costly and time consuming to analyze extensive study areas.

In terrestrial research, SDMs have been a key part of ecological research for several decades, commonly applied to terrestrial plants and animals (Austin et al. 1990, Pearson and Dawson 2003, Araujo et al. 2005, Elith and Leathwick 2009). They are fast becoming a standard method to assist management decisions by, for example, identifying conservation areas, application as conservation planning tools as well as managing the conservation for species of special interest (Araújo et al. 2011, Guisan et al. 2013, Eaton et al. 2018). SDMs also play a large role in predicting potential range shifts as well as colonization of problematic invasive species (Jiménez-Valverde et al. 2011, Muha et al. 2017). Very recently these models been used in paleontological studies (Eduardo et al. 2018) and as a tool for insights into climate change related to the dinosaur extinction (Chiarenza et al. 2019).

It has only been relatively recently however, that SDMs have been applied in the aquatic realm, with the first applications appearing 15-20 years ago (Elith and Leathwick 2009). In recent years the use of SDMs in freshwater systems has significantly increased, for example; European distribution of macroinvertebrates under future climate change (Domisch et al. 2011), land use change intensification on macroinvertebrates (Kuemmerlen et al. 2015) and using eDNA to predict the distribution of freshwater invasive species (Muha et al. 2017).

SDMs are widely used as they have several major advantages. First, they are not demanding in terms of biological data. The basic input requirements are, georeferenced occurrences and absences for a species in binary format (i.e. 1=presence, 0= absence). In practice, true absence data (a location where the species is certain not to occur) is difficult to obtain, but a generalized alternative is the application of pseudo absences. Therefore, biological data for SDM studies can be sourced from databases such as museum archives (Elith et al. 2006) and biodiversity databases such as the Global Biodiversity Information Facility (GBIF.org 2019).

The models are also applicable over large scales, from catchments (e.g. Kuemmerlen et al. 2012), regional & continental (Araujo et al. 2005, Araujo and Luoto 2007, Domisch et al. 2013) and global (Ihlow et al. 2012) study areas. Predicting distributions over such scales is especially possible due to the increasing availability of open access global or large-scale environmental datasets e.g. WorldClim (climate, Hijmans et al. 2005, Lehner et al. 2008), EarthEnv (streams and topography, Domisch et al. 2015, Amatulli et al. 2018) and Corine (land use, land.copernicus.eu). Together with these datasets, it is also possible to predict species distribution through future climate change scenarios and under historical conditions.

SDMs are also known as Environmental Niche Models (Harrison 1997, Peterson et al. 1999) and Bioclimatic Envelope Models (Araújo and Peterson 2012) although, the correct use of the terminology is inconsistent (Peterson and Soberón 2012). For the benefit of this thesis, it is important to conceptually define the purpose of an SDM. The fundamental niche estimated by SDMs is defined through the abiotic factors, such as climate and hydrology that determine appropriate habitat (Soberón and Peterson 2005). However, biotic factors further retract the fundamental niche, termed the realized niche, caused by the species' inability to reach areas due to biogeographical barriers (Pearson and Dawson 2003, Guisan and Thuiller 2005), e.g. water barriers for land-based animals, as well as biological limits, e.g. dispersal ability, competition and predator/prey relationships. Therefore, when only abiotic factors are considered in the SDM, the fundamental niche is predicted (Austin et al. 1990, Guisan and Zimmermann 2000, Pearson and Dawson 2003). This thesis deals with the role of hydrological regimes in relation to other abiotic drivers in species' distribution therefore; the target output is the fundamental niche of a species.

1.3 River ecosystems and the hydrological regime

Precipitation is the first and foremost driver of streamflow, but flow run off patterns are guided through complex interactions between topography (e.g. hillslope), climate (e.g. temperature), geology (e.g. porous rock) and land use (e.g. forests) into the stream (Poff 1992, Lake 2000, Bunn and Arthington 2002, Xenopoulos et al. 2005). Hydrological flow regimes are a driving factor of lotic habitats (Bunn & Arthington, 2002), which have a major influence on species assemblage (Resh et al. 1988, Poff et al. 1997). Global change is predicted to significantly alter seasonal flow regimes (Döll and Zhang 2010) and increase the frequency and severity of extreme hydrological events such as floods and droughts (IPCC 2007, 2014). This shift is thought to be one of the most serious threats to the sustainability of rivers and their biodiversity (Dudgeon et al. 2006, Heino et al. 2009). Consequently, research exploring the relationship between streamflow and ecology has become increasingly prevalent over the last two decades (Tonkin et al. 2014).The hydrology of stream ecosystems can also be altered by agricultural land-use change, which impacts the stream channel as well as riparian habitats (Allan 2004, Schmalz et al. 2015). Due to their sensitivity to both temperature change (Heino et al., 2009) and altered hydrological regimes (Xenopoulos et al. 2005), species assemblages are predicted to shift (Sala et al. 2000, Dudgeon et al. 2006) in, for example, altitude (Domisch et al. 2013) and latitudinal range (Hickling et al. 2005, Heino et al. 2009). With this perspective, including hydrological variables in SDMs it is of utmost importance to make reliable estimates on riverine species' distributions.

1.3.1 Riverine benthic macroinvertebrates

Riverine benthic macroinvertebrates are an extremely diverse group encompassing 60% of all fauna in freshwater ecosystems (Balian et al. 2008). They inhabit the substratum of the river bed and play a key role in many ecological processes such as, nutrient recycling, food web dynamics and decomposition (Palmer 1997). Macroinvertebrates are mostly sedentary, so they represent site-specific ecological conditions (Metcalfe 1989). Due to their varied responses to anthropogenic disturbances, they are useful indicators of the general degradation of aquatic ecosystems (Metcalfe 1989, Covich et al. 1999). Consequently, they are some of the most frequently used bioindicators in freshwater biological monitoring (Li et al. 2010, Hussain and Pandit 2012) and are a regulated component of the European Water Framework Directive (WFD). Despite their use as bioindicators globally, SDMs have only been recently applied on macroinvertebrates (Domisch et al. 2011) most likely in accordance with the relatively recent application of SDMs in river systems.

1.4 Research gaps

In the following, three important gaps in freshwater SDMs are described. These gaps make up the bulk of this thesis and resemble the structure outlined in paragraph **1.5** "Thesis aims and structure".

1.4.1 Data availability

The limited application of hydrological variables in SDMs is the lack of suitable data describing flow regime. One way to include flow regime information into SDMs, is the implementation of the "Indicators of Hydrologic Alteration" (IHA, Richter et al. 1996, Olden and Poff 2003), which describe the magnitude, frequency, duration, timing and change rate of high, low and average streamflow conditions. These metrics are used frequently in, for example, studies investigating flow-ecology relationships (e.g. Poff et al. 2010, Peres and Cancelliere 2016, Kakouei et al. 2018) as they can provide vital information describing flow regime including the impact of anthropogenic disturbance. Consequently, they are suitable for use in SDMs as they can also be applied under future scenarios relating climate-induced hydrological changes to species' distribution. The tools required to calculate IHA are freely available (www.github.com/USGS-R/EflowStats, Henriksen et al. 2006, Archfield et al. 2014), however, the procedure requires streamflow data that is high in temporal resolution (daily, m^3s^{-1}), continuous (i.e. gapless in time and space) and regionally representative (large-scale). Available data are commonly limited to localized geographical points, i.e. gauging station sites, making it difficult to analyze a large-scale stream network. An effective way to obtain streamflow data is through the application of complex hydrological models, e.g. SWAT (Arnold et al. 1998) that can produce highly precise estimates. However, these models usually require a large amount of input data, which can be challenging to simulate over the large scale necessary for comprehensive predictions.

1.4.2 Variable selection approach

One of the major challenges of SDM studies in every realm is the selection of relevant environmental predictors applied in the model (Guisan and Zimmermann 2000, Araújo and Guisan 2006). The "Gold standard" for selection of variables according to Araújo et al. (2019) is "to use proximal variables exclusively, for which the effect on a species' distribution is well evidenced, so that a model builds on true cause-and-effect relationships". This standard corresponds to fundamental niche theory, which states that individual species differ in their environmental preferences, crucial for their reproduction and survival (Hutchinson 1957). To apply the optimal environmental variables, expert knowledge of the ecological preferences of the species is essential. It has been shown that models perform significantly better with a high level of expert knowledge (Reside et al. 2019). However, when predicting species' distribution of a large riverine macroinvertebrate community (e.g. 200+ species), gathering expert knowledge typically involves seminar meetings and extensive literature reviews even for a relatively small number of species. It is therefore not always feasible to conduct such schemes for large communities, over large scales, as it can be time consuming and financially impractical. As yet, there is no general consensus on the most appropriate variable selection process, although examples include; 1) applying variables known to have an influence, or are the most commonly applied on the study species, requiring it to be well-studied with an established variable choice (e.g. plants, Austin and Van Niel 2011), 2) applying all variables but ignoring multi-correlation, potentially inducing uncertainty (Braunisch et al. 2013), 3) choosing a predefined number of the best variables from several categories, which may result in omitting other important variables (e.g. Kuemmerlen et al. 2015) and 4) via statistical analysis, e.g., PCA (e.g. McKenna and Johnson 2011, Markovic et al. 2012, De Marco and Nóbrega 2018) or BRTs (e.g. Record et al. 2013) on a preselected set of variables. Consequently, it is common practice that the same set of environmental predictor variables is uniformly applied to every species in an entire community to predict their distribution. This common application could lead to less robust predictions, as the optimum set of predictors for the community may not be the optimum set for every species within that community, which could impact model performance.

1.4.3 Role of hydrology

Most SDM studies in river systems commonly focus on climatic conditions (e.g. Domisch et al. 2011, Markovic et al. 2014, Ruiz-Navarro et al. 2016, Kärcher et al. 2019, Rodríguez-Merino et al. 2019) by interpreting precipitation variables as a hydrological influence (e.g. Domisch et al. 2019), or use river basin characteristics (Maloney et al. 2013, Zeng et al. 2015). These hydrological variables have an important influence on species distribution; however, they only partially represent the flow regime that influences species distribution and abundance. Some recent attempts have been made to include, at least, some aspects of flow regime e.g. high flow days (Kuemmerlen et al. 2015), which was shown to be of high relevance to macroinvertebrate distribution.

Recently, data are becoming increasingly available that describe specific aspects of streams e.g. stream specific climate and land use (Domisch et al. 2015), hydrology (Barbarossa et al. 2018, Irving et al. 2018 developed as part of this thesis), river classification and characteristics (Hydrosheds, Lehner et al. 2008) as well as dams and reservoirs (globaldamwatch.org, Lehner et al. 2011). By describing a broader representation of the stream ecosystem these datasets are vital steps in improving predictive modeling approaches for river ecosystems leading to more robust conclusions. In addition, some stream-specific data combine aspects of both climate and hydrology i.e. earthenv.org (Domisch et al. 2015) by incorporating the flow accumulation, known to be highly correlated with stream-flow (Kuemmerlen et al. 2014), into climate data. These data include the information from the upper sub-catchment, which is an important aspect to consider when assessing the distribution of river biota (Vinson and Hawkins 1998, Malmqvist and Rundle 2002, Kuemmerlen et al. 2014). However, it is difficult to disentangle the separate influence of either climate or hydrology on river species distribution, leading to difficulties in interpretation.

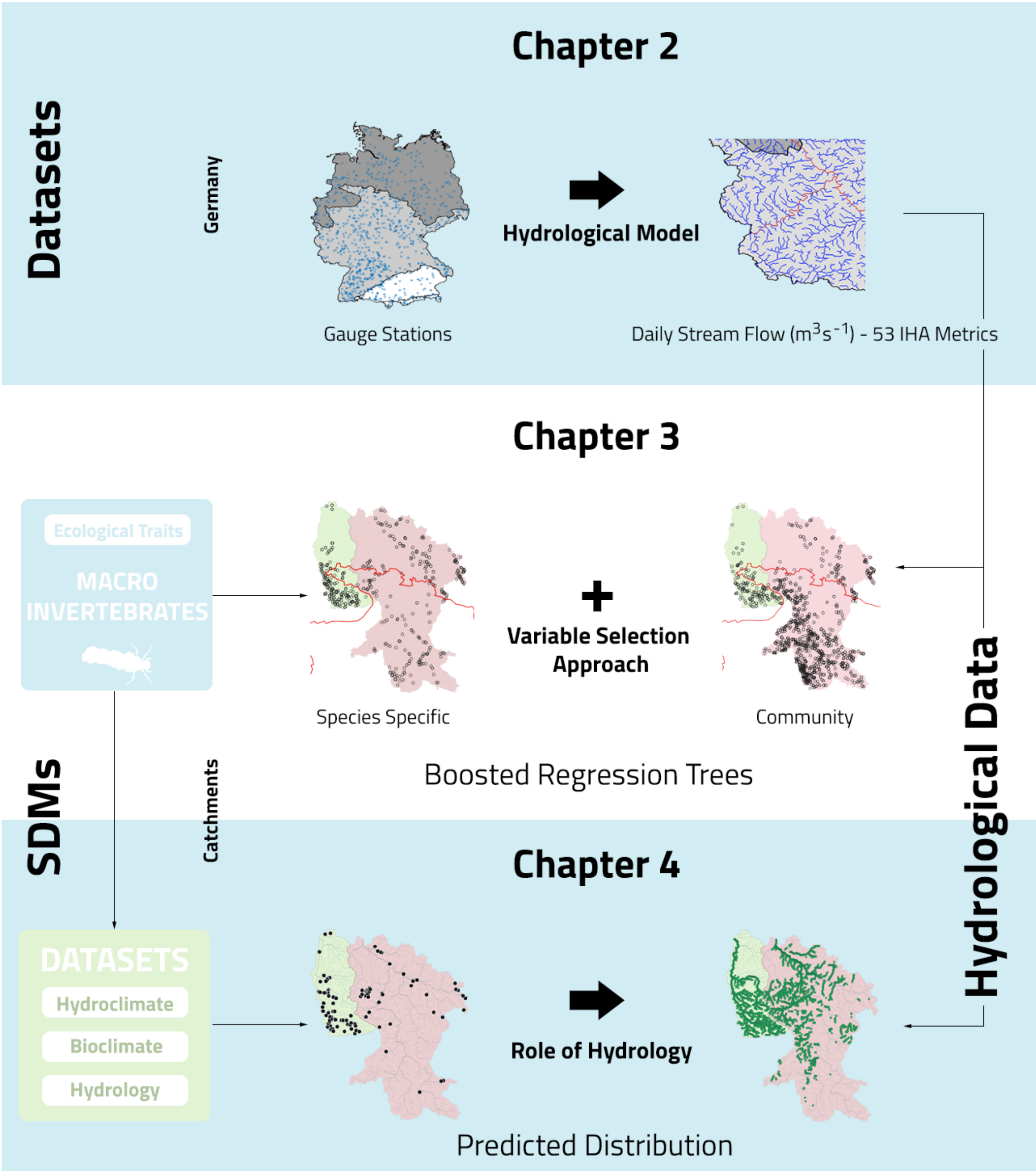


Figure 1.1: Conceptual overview of thesis structure.

1.5 Thesis aims & structure

SDMs are powerful tools for investigating global-change effects on river ecosystems however, the methodological concept of applying SDMs to predicted distributions of benthic macroinvertebrates is limited in providing a full representation of the stream ecosystem. My research is method based, and the primary focus is to develop the predictive ability of SDMs for riverine benthic macroinvertebrates, with an emphasis on integrating hydrological predictors that describe flow regime. The thesis is divided into three main components intended to fill the research gaps outlined above (Figure 1.1):

Chapter 2. **A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model**

Available hydrological data are often limited in their spatio-temporal extent and resolution for use in ecological applications such as predictive modeling. To overcome this limitation, I developed a simple hydrological model, to apply on a 1 km^2 gridded stream network of Germany to obtain a daily streamflow (m^3 s^{-1}) time series spanning a 64 years period (1950-2013). Accordingly, a high resolution spatio-temporal dataset of streamflow and a set of 53 hydrological metrics at 1 km^2 grid size are provided. The dataset is validated both spatially; in 70:30 ratio split, and temporally using Kling Gupta Efficiency measure. I also tested the metrics on 32 macroinvertebrate species using Generalized Linear Models (GLMs). I intended to keep the method simple and by exploiting globally available data, I aimed to ensure the model could be directly applied to alternate geographical regions or time periods. The datasets and data descriptor were published open access through the Nature Research Journal *Scientific Data*.

Chapter 3. **Identifying and applying an optimum set of environmental variables in species distribution models**

The common procedure of applying a uniform set of environmental predictors to an entire community may impact SDM performance. In **Chapter 3** I investigate whether applying a specific set of environmental predictors (hydrology, climate, land use and topography) to individual species within a community, as opposed to the community as a whole, will optimize macroinvertebrate distribution predictions through SDMs. In addition, I propose a variable selection process using Boosted Regression Trees (BRTs). I apply the method on 67 macroinvertebrate species within two large catchments in Germany. The catchments, the Ems and the Weser, are separated by a lowland/mountainous eco-regional divide. I identified the species that increased in model performance with a species-specific set of predictors, and the species that decreased in performance. To compare the differences between the increasing and decreasing species, I collated information on species' ecological traits and preferences and related them to the environmental conditions at the species known occurrences. I expected that model accuracy will increase for a subset of species within the community when a species-specific predictor set is applied, and that the ecological traits and spatial patterns of these species will differ to those of the community.

Chapter 4. **Disentangling the influence of climatic and hydrological predictor variables on benthic macroinvertebrate distributions**

With current available data, it is not always possible to disentangle the influence of climate and flow regime on benthic macroinvertebrate distribution. In **Chapter 4**, I compared the influence of three openly available datasets that can be applied in predictive modeling to describe: 1) climate only, 2) hydrology only and 3) information describing both climate and hydrology embedded within the data (hydroclimate). I applied the three datasets through SDMs on a community of 92 macroinvertebrate species in four model configurations representing different dataset combinations. I investigated the differences in each model configuration in terms of; 1) model performance, 2) prediction range size. I evaluated the influence of each predictor set on species' distribution by 1) investigating proportional variance explained by each dataset, 2) variable importance from SDM ensemble of each predictor set. I hypothesized that SDMs that include the hydrological dataset developed in **Chapter 2** will improve SDM performance. Further, I analyzed how the choice of predictor datasets influences the predicted distributions.

Chapter 5. **General Discussion**

I summarize the key findings of the three research studies outlined above and discuss the methodological limitations. I also discuss the potential applications of my findings in SDM research and future opportunities in freshwater SDMs. Finally, I make recommendations based on my research to further advance methods in SDMs.

1.6 References (Chapter 1)

Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution, and Systematics **35**:257-284.

Amatulli, G., S. Domisch, M.-N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, and W. Jetz. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific Data **5**:180040.

Araújo, M. B., D. Alagador, M. Cabeza, D. Nogués-Bravo, and W. Thuiller. 2011. Climate change threatens European conservation areas. Ecology Letters **14**:484-492.

Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, L. Maiorano, B. Naimi, R. B. O'Hara, N. E. Zimmermann, and C. Rahbek. 2019. Standards for distribution models in biodiversity assessments. Science Advances **5**:eaat4858.

Araújo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography **33**:1677-1688.

Araujo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography **16**:743-753.

Araújo, M. B., and A. T. Peterson. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology **93**:1527-1539.

Araujo, M. B., R. J. Whittaker, R. J. Ladle, and M. Erhard. 2005. Reducing uncertainty in projections of extinction risk from climate change. Global Ecology and Biogeography **14**:529-538.

Archfield, S. A., J. G. Kennen, D. M. Carlisle, and D. M. Wolock. 2014. An objective and parsimonious approach for classifying natural flow regimes at a continental scale. River Research and Applications **30**:1166-1183.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment - Part 1: Model development. Journal of the American Water Resources Association **34**:73-89.

Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. Ecological Monographs **60**:161-177.

Austin, M. P., and K. P. Van Niel. 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography **38**:1-8.

Balian, E. V., H. Segers, K. Martens, and C. Lévêque. 2008. An introduction to the Freshwater Animal Diversity Assessment (FADA) project. Pages 3-8 *in* E. V. Balian, C. Lévêque, H. Segers, and K. Martens, editors. Freshwater Animal Diversity Assessment. Springer Netherlands.

Balint, M., S. Domisch, C. H. M. Engelhardt, P. Haase, S. Lehrian, J. Sauer, K. Theissinger, S. U. Pauls, and C. Nowak. 2011. Cryptic biodiversity loss linked to global climate change. Nature Climate Change **1**:313-318.

Barbarossa, V., M. A. J. Huijbregts, A. H. W. Beusen, H. E. Beck, H. King, and A. M. Schipper. 2018. FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. Scientific Data **5**:180052.

Braunisch, V., J. Coppes, R. Arlettaz, R. Suchant, H. Schmid, and K. Bollmann. 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. Ecography **36**:971-983.

Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management **30**:492-507.

Chiarenza, A. A., P. D. Mannion, D. J. Lunt, A. Farnsworth, L. A. Jones, S.-J. Kelland, and P. A. Allison. 2019. Ecological niche modelling does not support climatically-driven dinosaur diversity decline before the Cretaceous/Paleogene mass extinction. Nature Communications **10**:1091.

Covich, A. P., M. A. Palmer, and T. A. Crowl. 1999. The role of benthic invertebrate species in freshwater ecosystems - Zoobenthic species influence energy flows and nutrient cycling. Bioscience **49**:119-127.

De Marco, P. J., and C. C. Nóbrega. 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. PLoS ONE **13**:e0202403.

Döll, P., and J. Zhang. 2010. Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations. Hydrology and Earth System Science. **14**:783-799.

Domisch, S., G. Amatulli, and W. Jetz. 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data **2**:150073.

Domisch, S., M. B. Araujo, N. Bonada, S. U. Pauls, S. C. Jähnig, and P. Haase. 2013. Modelling distribution in European stream macroinvertebrates under future climates. Global Change Biology **19**:752-762.

Domisch, S., M. Friedrichs, T. Hein, F. Borgwardt, A. Wetzig, S. C. Jähnig, and S. D. Langhans. 2019. Spatially explicit species distribution models: A missed opportunity in conservation planning? Diversity and Distributions **25**:758-769.

Domisch, S., S. C. Jähnig, and P. Haase. 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshwater Biology **56**:2009-2020.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z. I. Kawabata, D. J. Knowler, C. Leveque, R. J. Naiman, A. H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews **81**:163-182.

Eaton, S., C. Ellis, D. Genney, R. Thompson, R. Yahr, and D. T. Haydon. 2018. Adding small species to the big picture: Species distribution modelling in an age of landscape scale conservation. Biological Conservation **217**:251-258.

Eduardo, A. A., P. A. Martinez, S. F. Gouveia, F. d. S. Santos, W. S. d. Aragão, J. Morales-Barbero, L. Kerber, and A. Liparini. 2018. Extending the paleontology–biogeography reciprocity with SDMs: Exploring models and data in reducing fossil taxonomic uncertainty. PLoS ONE **13**:e0194725.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Pages 677-697. Annual Review of Ecology Evolution and Systematics.

GBIF.org. 2019. GBIF Home Page, Available from: https://www.gbif.org.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, T. G. Martin, J. R. Rhodes, R. Maggini, S. A. Setterfield, J. Elith, M. W. Schwartz, B. A. Wintle, O. Broennimann, M. Austin, S. Ferrier, M. R. Kearney, H. P. Possingham, and Y. M. Buckley. 2013. Predicting species distributions for conservation decisions. Ecology Letters **16**:1424-1435.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Harrison, S. 1997. How Natural Habitat Patchiness Affects the Distribution of Diversity in Californian Serpentine Chaparral. Ecology **78**:1898-1906.

Heino, J., R. Virkkala, and H. Toivonen. 2009. Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. Biological Reviews **84**:39-54.

Henriksen, J. A., J. Heasley, J. G. Kennen, and S. Nieswand. 2006. Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools).72.

Hickling, R., D. B. Roy, J. K. Hill, and C. D. Thomas. 2005. A northward shift of range margins in British Odonata. Global Change Biology **11**:502-506.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hussain, Q., A., and A. Pandit, K. 2012. Macroinvertebrates in streams: A review of some ecological factors. International Journal of Fisheries and Aquaculture. International Journal of Fisheries and Aquaculture **4**:114 - 123.

Hutchinson, G. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415–427.

Ihlow, F., J. Dambach, J. O. Engler, M. Flecks, T. Hartmann, S. Nekum, H. Rajaei, and D. Rödder. 2012. On the brink of extinction? How climate change may affect global chelonian species richness and distribution. Global Change Biology **18**:1520-1530.

IPCC. 2007. *Climate Change 2007: The Physical Science Basis.* WG1, Cambridge University Press.

IPCC. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.

Irving, K., M. Kuemmerlen, J. Kiesel, K. Kakouei, S. Domisch, and S. C. Jähnig. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model. Scientific Data **5**:180224.

Jiménez-Valverde, A., A. T. Peterson, J. Soberón, J. M. Overton, P. Aragón, and J. M. Lobo. 2011. Use of niche models in invasive species risk assessments. Biological Invasions **13**:2785-2797.

Kakouei, K., J. Kiesel, S. Domisch, K. S. Irving, S. C. Jähnig, and J. Kail. 2018. Projected effects of Climate-change-induced flow alterations on stream macroinvertebrate abundances. Ecology and Evolution **8**:3393-3409.

Kärcher, O., K. Frank, A. Walz, and D. Markovic. 2019. Scale effects on the performance of niche-based models of freshwater fish distributions. Ecological Modelling **405**:33-42.

Kuemmerlen, M., S. Domisch, B. Schmalz, Q. Cai, N. Fohrer, and S. C. Jähnig. 2012. Integrated modelling of aquatic ecosystems in China: Distribution of benthic macroinvertebrates at catchment scale. Hydrologie Und Wasserbewirtschaftung **56**:185-192.

Kuemmerlen, M., B. Schmalz, Q. Cai, P. Haase, N. Fohrer, and S. C. Jähnig. 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.

Kuemmerlen, M., B. Schmalz, B. Guse, Q. Cai, N. Fohrer, and S. C. Jähnig. 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. Ecological Modelling **277**:77-86.

Lake, P. S. 2000. Disturbance, patchiness, and diversity in streams. Journal of the North American Benthological Society **19**:573-592.

Lehner, B., C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J. C. Robertson, R. Rödel, N. Sindorf, and D. Wisser. 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment **9**:494-502.

Lehner, B., K. Verdin, and A. Jarvis. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Transactions American Geophysical Union **89**:93-94.

Li, L., B. Zheng, and L. Liu. 2010. Biomonitoring and Bioindicators Used for River Ecosystems: Definitions, Approaches and Trends. Procedia Environmental Sciences **2**:1510-1524.

Malmqvist, B., and S. Rundle. 2002. Threats to the running water ecosystems of the world. Environmental Conservation **29**:134-153.

Maloney, K. O., D. E. Weller, D. E. Michaelson, and P. J. Ciccotto. 2013. Species Distribution Models of Freshwater Stream Fishes in Maryland and Their Implications for Management. Environmental Modeling & Assessment **18**:1-12.

Markovic, D., S. Carrizo, J. Freyhof, N. Cid, S. Lengyel, M. Scholz, H. Kasperdius, and W. Darwall. 2014. Europe's freshwater biodiversity under climate change: distribution shifts and conservation needs. Diversity and Distributions **20**:1097-1107.

Markovic, D., J. Freyhof, and C. Wolter. 2012. Where Are All the Fish: Potential of Biogeographical Maps to Project Current and Future Distribution Patterns of Freshwater Species. PLoS ONE **7**:e40530.

McKenna, J., James E., and J. H. Johnson. 2011. Landscape Models of Brook Trout Abundance and Distribution in Lotic Habitat with Field Validation. North American Journal of Fisheries Management **31**:742-756.

Metcalfe, J. L. 1989. Biological water quality assessment of running waters based on macroinvertebrate communities: History and present status in Europe. Environmental Pollution **60**:101-139.

Muha, T. P., M. Rodríguez-Rey, M. Rolla, and E. Tricarico. 2017. Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. Frontiers in Ecology and Evolution **5**.

Olden, J. D., and N. L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications **19**:101-121.

Palmer, M. A. 1997. Biodiversity and Ecosystem Processes in Freshwater Sediments Author. Ambio **26**:571-577.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

Peres, D. J., and A. Cancelliere. 2016. Environmental Flow Assessment Based on Different Metrics of Hydrological Alteration. Water Resources Management **30**:5799-5817.

Peterson, A., and J. Soberón. 2012. Species Distribution Modeling and Ecological Niche Modeling?: Getting the Concepts Right. Natureza & Conservação **10**:1–6.

Peterson, A. T., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of Ecological Niches in Evolutionary Time. Science **285**:1265-1267.

Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A. 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. Global Ecology and Biogeography. 26(3):275-287.

Poff, N. 1992. Regional Hydrologic Response to Climate Change: An Ecological Perspective. Pages 88-115 *in* P. Firth and S. Fisher, editors. Global Climate Change and Freshwater Ecosystems. Springer New York.

Poff, N., J. Allan, M. Bain, J. Karr, K. Prestegaard, B. Richter, R. Sparks, and J. Stromberg. 1997. The natural flow regime. Bioscience **47**:769-784.

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

Record, S., N. D. Charney, R. M. Zakaria, and A. M. Ellison. 2013. Projecting global mangrove species and community distributions under climate change. Ecosphere **4**:23.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The Role of

Disturbance in Stream Ecology. Journal of the North American Benthological Society **7**:433-455.

Reside, A. E., K. Critchell, D. M. Crayn, M. Goosem, S. Goosem, C. J. Hoskin, T. Sydes, E. P. Vanderduys, and R. L. Pressey. 2019. Beyond the model: expert knowledge improves predictions of species' fates under climate change. Ecological Applications **29**:e01824.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun. 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. Conservation Biology **10**:1163-1174.

Rodríguez-Merino, A., R. Fernández-Zamudio, and P. García-Murillo. 2019. Identifying areas of aquatic plant richness in a Mediterranean hotspot to improve the conservation of freshwater ecosystems. Aquatic Conservation: Marine and Freshwater Ecosystems **29**:589-602.

Ruiz-Navarro, A., P. K. Gillingham, and J. R. Britton. 2016. Predicting shifts in the climate space of freshwater fishes in Great Britain due to climate change. Biological Conservation **203**:33-42.

Sala, O. E., F. S. Chapin, 3rd, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall. 2000. Global biodiversity scenarios for the year 2100. Science **287**:1770-1774.

Schmalz, B., M. Kuemmerlen, J. Kiesel, Q. Cai, S. C. Jähnig, and N. Fohrer. 2015. Impacts of land use changes on hydrological components and macroinvertebrate distributions in the Poyang lake area. Ecohydrology **8**:1119-1136.

Soberón, J. M., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics **2**:1-10.

Tonkin, J. D., S. C. Jähnig, and P. Haase. 2014. The Rise of Riverine Flow-ecology and Environmental Flow Research. Environmental Processes **1**:323-330.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: Variation at local, basin, and regional scales. Annual Review of Entomology **43**:271-293.

Xenopoulos, M. A., D. M. Lodge, J. Alcamo, M. Marker, K. Schulze, and D. P. Van Vuuren. 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. Global Change Biology **11**:1557-1564.

Zeng, Q., Y. Zhang, G. Sun, H. Duo, L. Wen, and G. Lei. 2015. Using Species Distribution Model to Estimate the Wintering Population Size of the Endangered Scaly-Sided Merganser in China. PLoS ONE **10**:e0117307.

Zimmermann, N. E., N. G. Yoccoz, T. C. Edwards, Jr., E. S. Meier, W. Thuiller, A. Guisan, D. R. Schmatz, and P. B. Pearman. 2009. Climatic extremes improve predictions of spatial patterns of tree species. Proceedings of the National Academy of Sciences of the United States of America **106**:19723-19728.

2 A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model

Katie Irving[1,2], Mathias Kuemmerlen[3], Jens Kiesel[1,4], Karan Kakouei[1,2], Sami Domisch[1], Sonja C. Jähnig[1]

[1]*Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 310, 12587 Berlin, Germany*

[2]*Department of Biology, Chemistry and Pharmacy, Freie University Berlin, Takustraße 3, 14195 Berlin, Germany*

[3]*Department Systems Analysis, Integrated Assessment and Modeling, Eawag, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland*

[4] *Christian-Albrechts-University Kiel, Institute for Natural Resource Conservation, Department of Hydrology and Water Resources Management, Germany*

Article published as:
Irving K, Kuemmerlen M, Kiesel J, Kakouei K, Domisch S, Jähnig SC. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model [Data Descriptor]. *Scientific Data.* 5:180224. doi: https://doi.org/10.1038/sdata.2018.224

Data published as:

Irving K, Kuemmerlen M, Kiesel J, Kakouei K, Domisch S, Jähnig SC. 2018. figshare https://doi.org/10.6084/m9.figshare.c.3906376 (2018).

2.1 Abstract

Hydrological variables are among the most influential when analyzing or modeling stream ecosystems. However, available hydrological data are often limited in their spatiotemporal scale and resolution for use in ecological applications such as predictive modeling of species distributions. To overcome this limitation, a regression model was applied to a 1 km gridded stream network of Germany to obtain estimated daily streamflow data ($m^3 s^{-1}$) spanning 64 years (1950-2013). The data are used as input to calculate hydrological indices characterizing streamflow regimes. Both temporal and spatial validations were performed. In addition, Generalized Linear Models (GLMs) using both the calculated and observed hydrological indices were compared, suggesting that the predicted flow data are adequate for use in predictive ecological models. Accordingly, we provide estimated streamflow as well as a set of 53 hydrological metrics at 1 km grid for the stream network of Germany. In addition, we provide an R script where the presented methodology is implemented, that uses globally available data and can be directly applied to any other geographical region.

2.2 Background & Summary

Natural flow regimes have a large influence in shaping biological communities (Bunn and Arthington 2002) and regulate numerous ecological processes in stream ecosystems (Poff and Ward 1989, Poff et al. 1997, Poff et al. 2010). Future flow regimes are predicted to alter significantly by, for example, an increase in the frequency and severity of floods and droughts (IPCC 2007, Döll and Zhang 2010). To understand the ecological consequences of these changes, it is important to study how the critical components of flow regimes affect stream ecosystems and to include this knowledge in the assessment of future global-change scenarios. However, information about flow regimes is often not available or sufficiently diverse for detailed modeling analyses such as Species' Distribution Models (SDMs; Jähnig et al. 2012, Domisch et al. 2015b), which are a common tool used in ecological analysis. SDMs relate known occurrences of species to their environmental conditions and predict species distributions in geographical space under current or future conditions.

One way to include flow regime information into SDMs is the implementation of the "Indicators of Hydrologic Alteration" (IHA) (Richter et al. 1996), which describe the magnitude, frequency, duration, timing and change rate of high, low and average streamflow conditions. The metrics can provide essential information on freshwater ecosystems in general and on the impact of human activities and may support river management and conservation. Additionally, they are well suited to be applied in predictive modeling (i.e. SDMs) and can be used under future hydrological scenarios to assess the effects of climate change on species distribution.

While the tools required to calculate IHA are freely available (i.e. www.github.com/USGS-R/EflowStats, Henriksen et al. 2006, Archfield et al. 2014), the streamflow data required for carrying out ecological predictions can be challenging to acquire, as it needs to be of high temporal resolution (daily, $m^3 s^{-1}$), continuous (i.e. gapless in time and space) and regionally representative. Such data are typically restricted to gauging stations i.e. only point localities being available for analysis, therefore it is not possible to analyze large sections of a stream network. An effective way to obtain spatially gapless data is through the application of hydrological models. However, predictive modeling applications such as SDMs in rivers often require environmental predictors at fine spatial resolutions (< 1 km^2), over a large spatial scale, in order to include the species' full range of occurrence for comprehensive predictions (Barbet-Massin et al. 2010). Depending on the complexity of hydrological model, such as SWAT (Arnold et al. 1998) & WaSIiM-ETH (Schulla and Jasper 2006), and the large amount of input data required, it can become tedious to simulate on these spatial scales. Given these limitations, in order to fill the much-needed data gap for ecological analyses, linear regression models are simple and fast methods that can be applied for the spatiotemporal extrapolation of streamflow (McIntyre et al. 2007, Seelbach et al. 2011). Following these considerations, the regression model developed here used two freely available data components: 1) observed gauging data from the Global Runoff Data Centre (GRDC, www.bafg.de/GRDC/), 2) downstream accumulated precipitation data along a river network at high resolution (Domisch et al. 2015a). The low data requirements of our model, together with the simple modeling approach, render it an inexpensive, easy-to-use tool which can be applied to any geographical scale, and/or time period.

We applied the model to a 1 km gridded stream network in Germany (n=85,363 1 km grid cells) to create a continuous daily time series of streamflow ($m^3 s^{-1}$) spanning 64 years (1950-2013). The estimated daily streamflow data were used as input to create a set of IHA metrics for the German stream network. From the 165 metrics that were tested, 53 were validated successfully: predominantly metrics describing mean values of streamflow, e.g. mean monthly flow.

To test the results for their usability in ecological applications, we predicted the occurrence of 34 benthic macroinvertebrate species with GLMs using the validated metrics with either observed or simulated values. Results from both yielded equally good predictions, showing that data predicted from this study is adequate for the purpose of ecological predictive modeling.

We provide both the simulated daily streamflow dataset and the 53 IHA metrics as downloadable files (Data Citation 1) which can be used "as is". In addition, we provide R scripts that allow users to apply the model to other geographical regions or calculate the hydrological metrics for different time periods.

2.3 Methods

The primary goal of this study was to create a much-needed dataset of hydrological variables for use in ecological predictive modeling. There is a high demand (Barbarossa et al. 2018) for large scale hydrological data at high spatiotemporal resolutions to be used as input in models such as SDMs. Such data should be either widely available or easy to reproduce. For this reason, we propose to estimate streamflow for entire stream networks using a linear regression model with only one predictor: the accumulative precipitation in the upper subcatchment. While there are several other relevant processes influencing discharge (e.g. infiltration, groundwater storage, evapotranspiration, etc.), adding further predictors significantly increases model complexity and computation time, particularly if models are of large scale and high spatiotemporal resolution. We are aware that applying a simple model over a large scale and fine resolution comes at the cost of lower prediction accuracy, as it does not fully consider important natural and anthropogenic influences (e.g. water abstraction and river management such as dams). Nevertheless, a model which can be readily implemented will help fill the demand for large scale, continuous, high spatiotemporal resolution data to be used in ecological predictive modeling, specifically SDMs.

Moreover, our proposed modeling framework should be easily applied globally at any scale, which is why open and near global data sources have been chosen. Here, observed daily streamflow is extrapolated from gauging stations to all grid cells on the

German stream network using weighted linear regression and subcatchment-accumulated precipitation as a predictor. The predicted daily streamflow data was used as input data to calculate the IHA metrics set out in Olden and Poff (2003) for every grid cell (see Figure 2.1).

In a last step, the level of accuracy of the model predictions for the purpose of ecological modeling were tested by implementing Generalized Linear Models (GLMs) and comparing species occurrence predictions of 34 stream macroinvertebrates at gauging stations using both observed and predicted streamflow.

2.3.1 Base layer & study area

The area of study is the stream network of Germany. An openly available, modeled 1 km gridded stream network of Germany was used as a base layer, taken from earthEnv.org/streams (Domisch et al. 2015a), originally derived from the modeled HydroSHEDS dataset (www.hydrosheds.org, 30-arc-second spatial grain, Lehner et al. 2008), which in turn is derived from SRTM (www.srtm.csi.cgiar.org, Jarvis et al. 2008) and available in GEOTiff raster file format (Figure 2.1).

2.3.2 Data collection: streamflow

Daily streamflow data (m^3 s^{-1}) were collected from 1,065 gauging stations in Germany from the GRDC and the Federal Environment Agencies of Germany (Table 2.1). While all 1,065 sites collectively covered 64 years, each individual site had to contain at least 10 years of continuous data between 1st January 1950 until 31st December 2013 to be considered in order to maximize input and to standardize the dataset. Due to possible simplifications in the 1 km stream network, the gauging station sites within a 3 km buffer were moved to the next grid cell of the stream network base layer in QGIS (www.qgis.org, QGIS Development Team 2016); any stations beyond the buffered network were excluded from the analysis.

2.3.3 Data collection: predictors

We wanted to use freely and globally available data as model predictors, as we aim to produce a framework that can be applied in other geographical regions. The predictors had to also match the high spatial resolution of 1 km grid cells to be able to be applied.

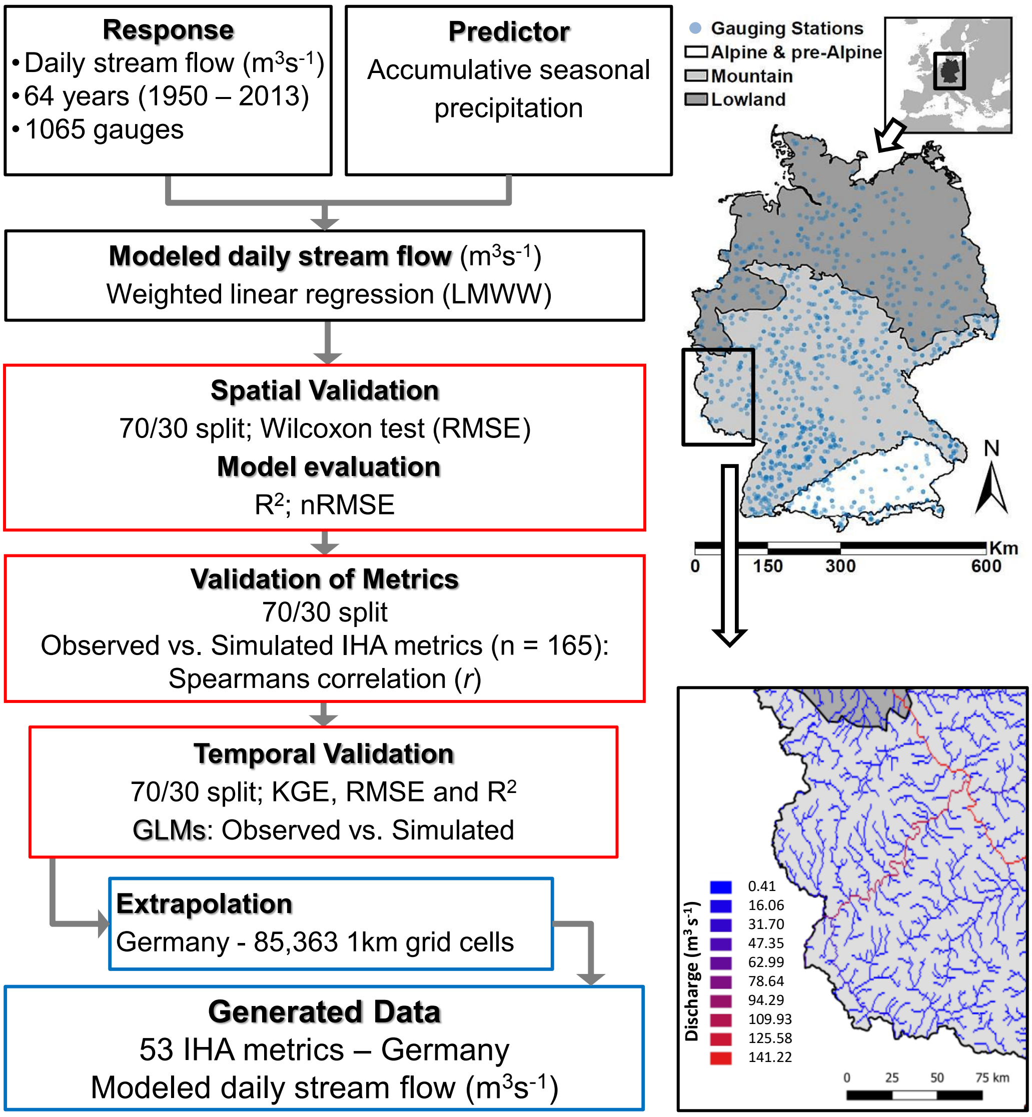


Figure 2.1: Workflow of the modeling and validation procedure.

Observed daily streamflow data was used as the response variable in a weighted linear regression (LMWW). Seasonal accumulative precipitation data was used as the predictor variable. The model was validated spatially using Wilcoxon tests on the RMSE. The IHA metrics were calculated from observed and simulated streamflow and validated through Spearman's correlation. The model was validated temporally with KGE, RMSE and R2, and GLMs were performed on 34 species and compared with IHA metrics calculated from observed and simulated flow data, respectively. The daily streamflow data were extrapolated to the entire stream network of Germany resulting in a time series covering 64 years. The validated IHA indices were then calculated for all grid cells on the stream network. Top right map is Germany's location in Europe, and underneath, the study area of Germany with distribution of gauging stations. Bottom right map shows the modeled daily streamflow ($m^3\ s^{-1}$) of 11th February 1950.

Subcatchment-accumulated precipitation (hereafter referred to as Precipitation) data were taken from www.earthenv.org/streams (downloaded 2016, Domisch et al. 2015a), which in turn was derived from WorldClim database (Hijmans et al. 2005). It consists of 12 average monthly downstream precipitation values (Jan-Dec) for each grid cell along the stream network, where each monthly value is a 50 year average (1950-2000) (Domisch et al. 2015a). The data in raster format represent the stream network of the study area outlined above (Hijmans et al. 2015). In this dataset, each grid cell of the stream network represents the added precipitation of all grid cells contained in the contributing area (i.e. upper subcatchment) of that specific cell. The accumulative nature of the data takes into account influencing upstream processes, important for stream systems (Allan 2004).

Table 2.1: Gauging stations. Min, max and median number of gauging stations, caused by temporal gauge data availability, used in models throughout the time series (n=23,376 models) applied within each region

Spatial region	Min	Max	Median
Alpine	6	112	97
Lowland	22	254	145
Mountain	43	584	418
Germany	71	944	685

The predictor variable precipitation was determined by calculating the mean monthly precipitation of four sets of three months to represent the annual seasons: Dec-Feb (Winter), Mar-May (Spring), Jun-Aug (Summer) and Sep-Nov (Fall). The streamflow data began on 1st Jan 1950 and ended on 31st Dec 2013. Therefore, the seasonal precipitation for winter of 1950 consisted only of Jan and Feb, with Dec 1949 being omitted. Each winter thereafter included all 3 months (Dec-Feb) and is referred to as winter of the year containing the information from January and February (i.e. Dec 1983 + Jan and Feb 1984 = winter 1984). Similarly, only data from December 2013 were available for winter 2014.

2.3.4 Preliminary analysis

The preliminary analysis tested model performance of various configurations of predictors and spatial scale, the purpose of which was to find the better performing model at the spatial scale and resolution we required. Therefore, we ran the analysis by applying three different predictors in the regression model: 1) flow accumulation, 2) monthly precipitation and 3) seasonal precipitation. The model trials were then applied on four different regional extents as to detect regional differences in model performance: 1) Germany in its entirety, 2) Germany sub divided into three regions i.e. Central Plains, Central Highlands and Alpine (sensu Illies 1967). This procedure resulted in 12 model configurations in total, which were compared in their capability of predicting daily streamflow.

Flow accumulation is the sum of contributing grid cells from the upper subcatchment that naturally flow into one grid cell and is known to be highly correlated with streamflow (Kuemmerlen et al. 2014, Kuemmerlen et al. 2015). Flow accumulation was taken from www.earthenv.org/streams, based on the HydroSHEDS digital elevation model (DEM, Lehner et al. 2008), which in turn is derived from (SRTM, Jarvis et al. 2008). Here, flow direction was used as a base layer for routing and delineating the upper stream network in grass GIS (see Domisch et al. 2015a for further details). It is important to note that due to the precipitation variables being accumulative, flow accumulation information is contained within the precipitation. This induces correlation between both and therefore the variables were tested in separate models.

Three performance metrics were used to compare model performance and validate the model: 1) Root mean square error (RMSE), 2) normalized root mean square error (nRMSE) and 3) coefficient of determination (R^2). The R^2 of each regression model was used for evaluation and to compare between model configurations (reported as mean R^2 +/- SE) and referred to as explained variance. Both RMSE and nRMSE are measurements of each model's predictive ability and were calculated from the observed and simulated flow values using the R package HydroGof (Zambrano-Bigiarini 2014).

It is important to note that the performance metrics are derived from the regression model applied on each individual daily time step and were used to compare between each model configuration. This is a comparison of model goodness of fit, which was used to assess the model spatially across ecoregions and using different predictors. To effectively visualize the comparison between model configurations/spatial regions with differing variations of streamflow and orders of

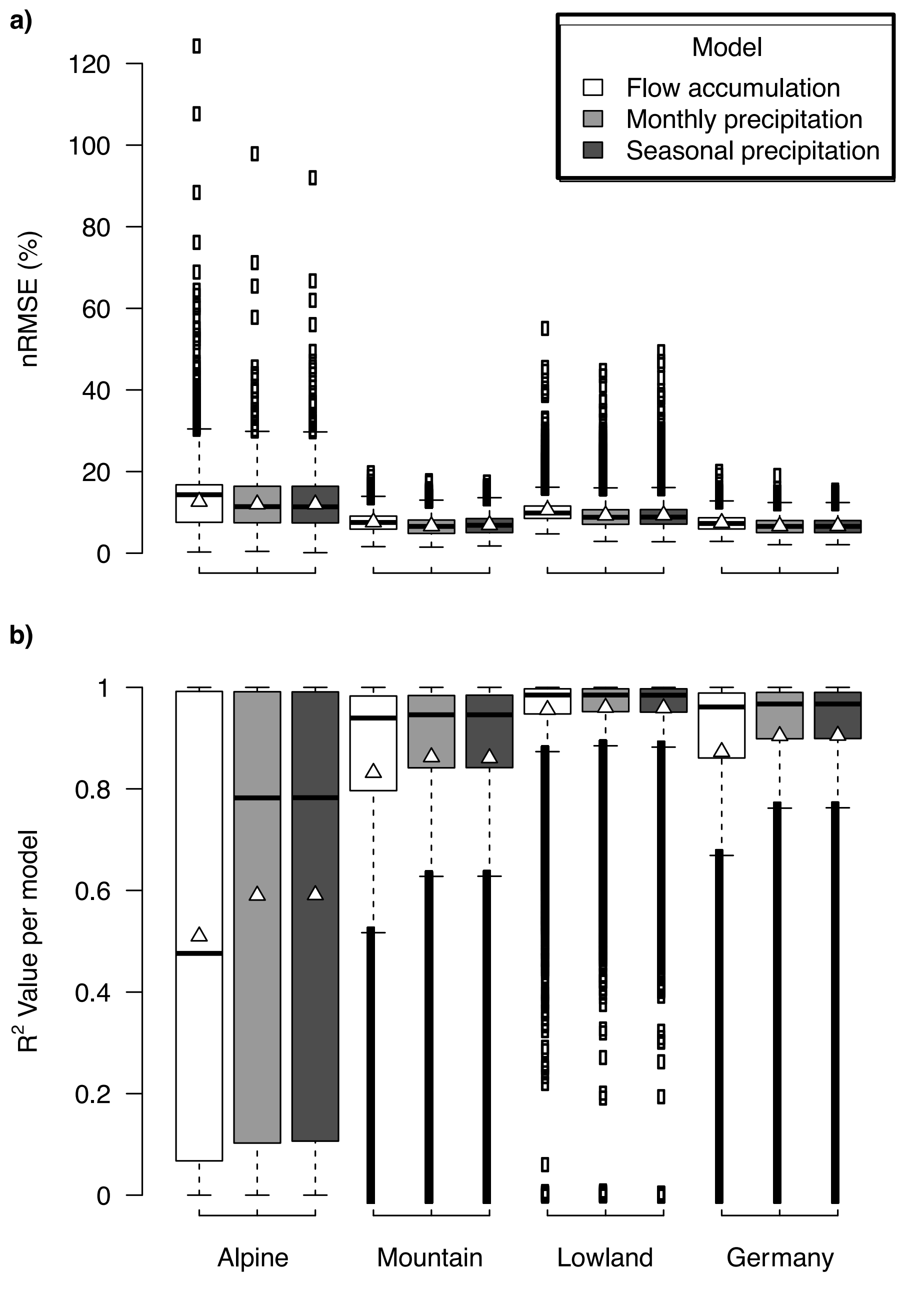


Figure 2.2: Comparison between each model configuration. Performance metric a) normalized rooted mean square error (nRMSE) and b) R squared statistic (R^2). Points represent each model (day) over the time series (n=23,376 models). Boxplots (bar = median, box = IQR, whiskers = 1.5 x IQR and outliers. Triangles = mean (Table 2).

magnitude, the RMSE was normalized (nRMSE). This was calculated manually by dividing the RMSE by the difference in the maximum and minimum observed streamflow values.

Wilcoxon tests were applied to compare the RMSE values of each model configuration. Any individual dataset was large (n = 23,376 days), so in order to avoid possible type 1 errors (false positives), the data were split into yearly subsets (n = 365 *64). Each set of 64 Wilcoxon tests are reported as a percentage (%) of tests that show a significant difference between the two predictions tested. A higher percentage of significance indicates a larger difference between configurations. For comparison, each configuration was also tested using the entire dataset (n=23,376 days), without sub-dividing into annual components.

Table 2.2. Overall statistics. RMSE, nRMSE & R^2 of each model configuration (mean ± SE, n=23,376)

	RMSE ($m^3 s^{-1}$)	nRMSE (%)	R^2
Flow Accumulation			
Alpine	86.35 ± 0.39	12.67 ± 0.04	0.51 ± 0.0027
Lowland	249.16 ± 1.0	10.58 ± 0.02	0.96 ± 0.0005
Mountain	153.45 ± 0.50	7.63 ± 0.02	0.83 ± 0.0016
Germany	173.69 ± 0.59	7.44 ± 0.01	0.87 ± 0.0014
Monthly Precipitation			
Alpine	81.83 ± 0.37	12.04 ± 0.03	0.59 ± 0.0027
Lowland	217.96 ± 0.91	9.31 ± 0.02	0.96 ± 0.0005
Mountain	129.79 ± 0.44	6.67 ± 0.02	0.86 ± 0.0014
Germany	154.62 ± 0.53	6.68 ± 0.01	0.90 ± 0.0011
Seasonal Precipitation			
Alpine	81.63 ± 0.37	12.02 ± 0.03	0.59 ± 0.0027
Lowland	218.48 ± 0.91	9.34 ± 0.02	0.96 ± 0.0005
Mountain	138.82 ± 0.47	6.95 ± 0.02	0.86 ± 0.0014
Germany	154.78 ± 0.53	6.68 ± 0.01	0.91 ± 0.0011

Precipitation performed better as a predictor than flow accumulation for both nRMSE (Figure 2.2a, Table 2.2) and R^2 (Figure 2.2b, Table 2.2) across all regions. There was no significant difference between monthly and seasonal precipitation predictors (Figure 2.2b, Wilcoxon Test: Table 2.3), with the only exception in the mountain region. Models applied with flow accumulation only performed significantly worse than any of the precipitation predictors (Wilcoxon Test: Table 2.3) and were not applied in further analyses. Models in the lowland region had the highest R^2 values (mean R^2= 0.96, Table 2.2, Figure 2.2a), and performed better in terms of explained variance for both monthly and seasonal precipitation predictors. However, for nRMSE, the models spanning the whole of Germany performed best (mean nRMSE = 6.68%, Table 2.2, Figure 2.2a) compared to the lowland and mountain regions for seasonal precipitation. Models applied in the alpine region displayed the highest variation in R^2 values (Table 2.2, Figure 2.2b). The models from all regions varied significantly from each other in terms of RMSE (Wilcoxon Test: Table 2.4, Figure 2.2 a&b), except for tests of difference that showed a lower percentage of significance (Wilcoxon test: 68.75%, $p < 0.05$) between mountain and Germany regions with seasonal precipitation. The models applied to Germany in its entirety performed well. The contribution of both mountain (area = 164,544 km^2 (46%)) and lowland region (area = 153,952 km^2 (43%)) dominates the landscape (89%) of the entire area of Germany (355,926 km^2).

Table 2.3. Wilcoxon test of difference of predicted discharge RMSE between predictors. Flow accumulation (FL), monthly precipitation (MP), seasonal precipitation (SP), and spatial region. Numbers indicate % of years (total=64 years, n=365 days) with significant difference (p<0.05), overall significance when using total data set (n=23,376 days) indicated by*.

Spatial Region	FL x MP	FL x SP	MP x SP
Lowland	96.9*	96.9*	0
Alpine	14.1*	15.6*	0
Mountain	100*	100*	50*
Germany	98.4*	100*	0

Table 2.4. Wilcoxon test of difference of predicted discharge RMSE between spatial regions. Alpine (AL), Mountain (MM), Lowland (LL), Germany (DE), and predictor. Numbers indicate % of years (total =64) with significant difference (p<0.05), overall significance when using total data set (n=23,376) indicated by*.

Predictor	AL x MM	AL x LL	AL x DE	MM x LL	MM x DE	LL x DE
Monthly precipitation	100*	100*	100*	98.4*	95.3*	95.3*
Seasonal precipitation	100*	100*	100*	96.9*	68.8*	95.3*

Therefore, the lower performing models applied in the alpine region (area=37,430 km^2 (11%)), do not seem to have an influencing factor on the overall performance of the models applied throughout Germany, presumably due to their low contribution to the average.

Accordingly, the seasonal precipitation model applied throughout Germany performed best and was therefore used to predict daily streamflow as input data for the IHA metrics.

Table 2.5. Spatial comparison (mean ± SE) of regression model. Training and testing data (70:30) and between model methods LMWW & lmRob; Wilcoxon test percentage of significance (%) of $p< 0.05$ concluding that the means are not significantly different (i.e. observed and predicted values are similar).

	RMSE ($m^3 s^{-1}$)	nRMSE (%)	R^2	% of significance
Training (70%)				Training vs. testing
LMWW	154.21± 0.55	6.89 ± 0.02	0.83± 0.0005	4.70%
lmRob	152.77± 0.54	6.82 ± 0.01	0.83± 0.0005	
Testing (30%)				LMWW vs. lmRob
LMWW	151.61 ± 0.58	7.35 ± 0.02	0.82± 0.0009	0%
lmRob	150.33± 0.58	7.29 ± 0.02	0.82± 0.0009	

2.3.5 Modeling method

The first step in the modeling process was to extract the daily streamflow data from the gauging sites on a day by day basis for the entire study area (i.e. first day 1st Jan 1950). All gauging sites with discharge data available for the same day (n=varies dependent on day and regional extents) were used as the response variable input for the model of that specific day. A linear model was performed to estimate the streamflow for that particular day only. This procedure was repeated 23,376 times from the first day (1st Jan 1950) until the last day (31st Dec 2013) to create the daily time series of 64 years. In other words, the discharge is predicted on a daily basis for all of the study area. Unlike common hydrological modeling approaches, discharge predictions here are performed on the spatial dimension (i.e. as raster layers). Discharge time series are later aggregated by stacking the daily spatial predictions (i.e. stacking gridded datasets). We acknowledge that, although daily streamflow is used as the response variable in the model, we use a coarser (seasonal/monthly) resolution to predict the high resolution (daily) time series. We understand that achieving high accuracy from this type of model input is challenging. However, these data are readily available for the spatial scale and resolution we input into the model and were tested to provide adequate precision for use in ecological predictive models.

Exploratory analysis of the data showed that the distribution of the empirical streamflow data was heavily tailed due to several outliers, which violated the assumption of normal distribution. Therefore, a robust linear model (lmRob), less sensitive to a non-normal distribution (Fox and Weisberg 2011), was applied as implemented in the Robustbase R package (Maechler et al. 2016). This method uses maximum likelihood estimation to apply a weight system reducing the impact of outliers, while still including the benefits of simple linear regression and has been effectively used previously (Venables, W.N, personal communication, 2017). However, the performance metrics to determine goodness of model fit (e.g. R^2) are easier and more intuitive to extract from simple linear models. Therefore, the weights were first calculated through the lmRob function and later introduced into a simple linear model to produce a linear model with weights (LMWW) (Ronchetti et al. 1997). The lmRob function calculates the weights using the method of MM-estimation, a development of Huber's M-estimation (Huber 1964), which returns highly robust and efficient estimators. Further descriptions are outlined in Yohai (1987). For validation purposes, the spatial predictions from the LMWW were compared, using Wilcoxon tests on the RMSE, to those of the lmRob and no difference between models was found (Table 2.5). The resulting weighted linear regression (LMWW) is described in equation 1:

$$Eq. 1) \qquad Q_s = \sum_{s=1}^{n} W_s \, (Q_s - a - \beta x_s)^2$$

where Q_s is the discharge, x is the predictor, W is the weight at the *sth* gauging site, α is the intercept and β is the slope of the model. The configuration of gauges with streamflow data varied daily, therefore the calculated weights also differed daily.

2.3.6 Application of the model

The model was used to extrapolate predicted streamflow values to each 1 km grid cell on the stream network raster of Germany (n=85,363). This was done for every day in the time series from 01/01/1950 to 31/12/2013 (n=23,376), creating a 64 year dataset of daily streamflow data, covering the entire stream network of Germany. Due to the nature of linear models, daily streamflow predictions included a number of negative discharge values, particularly in the headwater region of the stream network, where streamflow is lowest (for more details see "Limitations"). In the final dataset, any negative streamflow predictions were removed and replaced with the minimum predicted value across the entire time period for that grid cell. Next to the adjusted dataset, an additional dataset of the negative values is provided.

Table 2.6. Wilcoxon test of significance. Training and test datasets of modeling methods; lmRob – robust linear model, LMWW – linear model with weights. Years that showed a significant difference.

Training mean	Testing mean	W value	P value	Method	Year
178.85	167.89	72922.00	0.03	lmRob	1950
222.04	214.89	72618.00	0.05	LMWW	1952
173.13	164.56	73816.50	0.02	LMWW	2000
194.89	182.15	74903.50	0.00	LMWW	2013

2.4 Code availability

The R scripts are available online from www.github.com/ksirving/stream_flow. There are three scripts in total, used for different steps in the modeling procedure.

1) The weighted linear regression model script is used to predict streamflow in any geographical region, given the appropriate data is available. The user needs to provide a data frame with daily streamflow input data for that specific region (e.g. GRDC www.bafg.de/GRDC/) and precipitation data as a GeoTIFF file that can be downloaded from www.earthenv.org/streams (Domisch et al. 2015a). The output is a dataframe for each day of the simulation time containing streamflow values for all grid cells.
2) Format streamflow data calculated in script (1), including replacement of the negative values with the minimum value for that grid cell and structuring the data frame for input into IHA calculations and calculation of the IHA metrics.
3) Format and structure the provided streamflow NetCDF files for input into IHA calculations and calculation of the IHA metrics.

For script (2) & (3), the user needs to download the necessary IHA functions from www.github.com/USGS-R/EflowStats (2016) (Henriksen et al. 2006, Archfield et al. 2014) .

2.5 Data Records

The modeled streamflow ($m^3 s^{-1}$) dataset contains daily (n=23,376) data over 64 years (1950-2013) for every 1 km grid cell (n = 85,363) in the German stream network. The dataset is available as NetCDF files and is available for download (Data Citation 1). Each individual raster layer represents one day in the time series, which are available as annual raster stacks (n=64). The user can subset the time series to the required period and follow script (3) to structure and format the data for input into the IHA calculations. All negative values have been replaced. However, annual NetCDF files (n=64) containing all original values are provided as an additional dataset. Here, each individual raster layer (n=1,621) of the NetCDF file represents the day in the time series and contains only the negative values. The 53 IHA metrics that were validated successfully are also available for download (Data Citation 1) for the same stream network (Germany, grid cells n=85,363). The IHA metrics are available as GeoTIFF files, with each layer representing one metric. All NetCDF and GeoTiff files are in WGS84 coordinate system with an extent of 55°N to 47°S latitude and 5°E to 15°W longitude. All layers contain 914 rows and 1,100 columns. To reduce the file size, all values have been multiplied by 10,000 in order to achieve an integer format without precision loss. Potential users therefore need to convert data back to the original units by dividing each raster file by 10,000 (data type = Int4S, NoData value = -999).

2.6 Technical Validation

2.6.1 Spatial Validation

To validate the newly-developed dataset (Data Citation 1), we split the flow data of each individual model (n=23,376 days) spatially into 70% training and 30% testing data sets. Each model was then built using the training data and then assessed on how well it predicts the independent testing data.

Wilcoxon tests were then applied to test for any difference between the root mean square error (RMSE) values of the training and testing datasets. The RMSE is a measurement of the model's predictive ability. RMSE is an absolute measure of fit, calculated through the comparison of the observed and predicted streamflow values from corresponding sites (equation (2)):

$$Eq. 2) \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{obs,i} - y_{sim,i}\right)^2}{n}}$$

where n is the number of sites, $y_{obs,i}$ is the observed discharge value at the *ith* site and $y_{sim,i}$ is the simulated discharge at the *ith* site.

RMSE is reported in the same units as the response variable (i.e. m^3 s^{-1}) and is an important measurement of fit when the model is used for prediction. Zero indicates the best possible fit. The RMSE was calculated through the HydroGof package (Zambrano-Bigiarini 2014).

An indication of good model performance was considered to be the lack of significant differences between the RMSE values of the training and testing datasets. A method of subsetting the large dataset into yearly subsamples was applied (n=365 *64). Each set of 64 Wilcoxon tests are reported as percentage (%) of tests that show a significant difference.

According to the RMSE comparison between the training and testing datasets, the majority (95%, n=61, $p > 0.05$)) of tests showed no significant difference between datasets (Table 2.5, LMWW: % of significance). The three years that showed a significant difference are listed in Table 2.6.

2.6.2 Calculation and validation of hydrological metrics

There are 171 IHA metrics that describe the frequency, magnitude, duration and timing of streamflow events set out in Olden and Poff (2003) (full descriptions in Supplementary Table S2.1 in the supplementary material). Of these, 165 were chosen and a correlation analysis performed. The six that were omitted included drainage area variables, such as "mean annual runoff", an aspect which is beyond the scope of this study. The hydrological metrics were calculated using the functions available from the R package EflowStats (www.github.com/USGS-R/EflowStats) (Henriksen et al. 2006, Archfield et al. 2014). For validation, the observed data were randomly and spatially split into 70% training and 30% testing datasets. The model was built using the 70%

training data, with its subsequent predictions calculated for the 30% testing data, which were used as input for the metrics calculation of the simulated data. Any simulated values that were below zero were replaced with the lowest value throughout the time series for that grid cell.

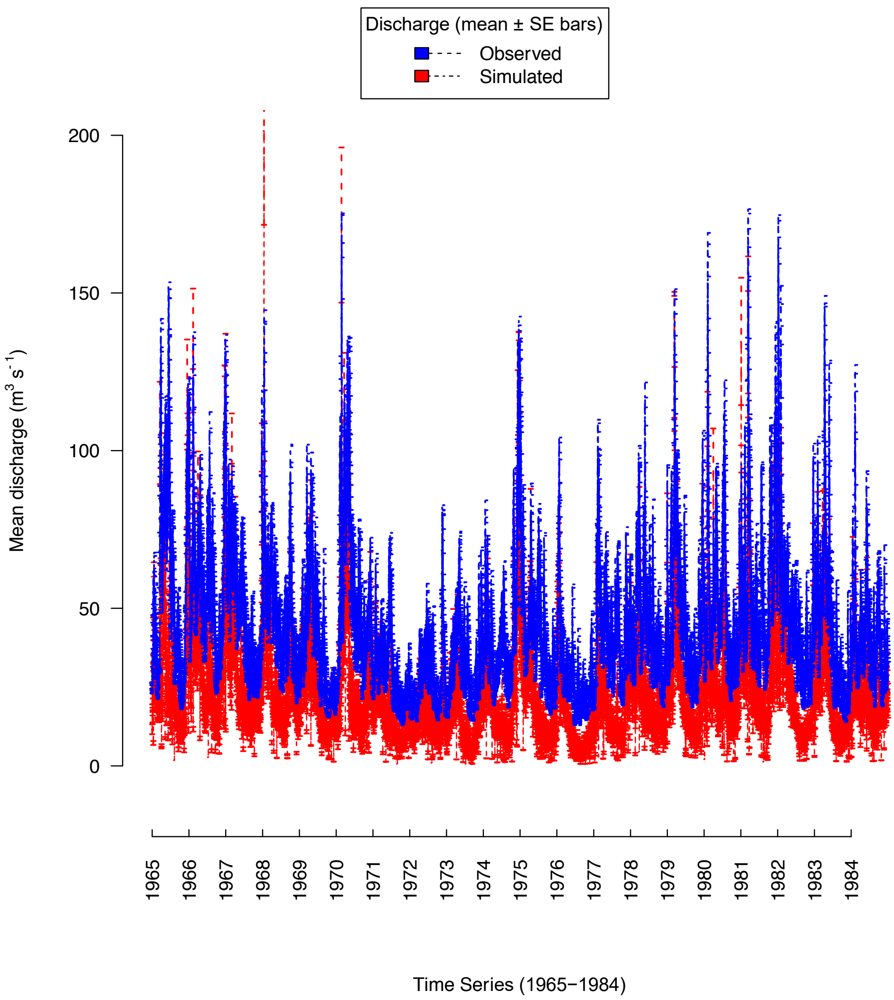


Figure 2.3: Time series of observed and simulated daily mean discharge. Taken over time period of 2 0 years (1965-1984), at gauge sites ± SE bars (n=518 gauges).

A continuous (gapless) daily time series was needed to calculate the IHA metrics. Due to the nature of the data splitting, the testing dataset was not continuous over the entire time series for every gauge. To test the most truthful values of observed streamflow, instead of interpolation, and to reduce computing time, this required subsetting the observed flow data into a smaller dataset including 518 gauging sites, across 20 years (1965-1984). This subset was then compared with metrics calculated from the above described simulated flow data for the same time period and grid cells matching the gauge sites in the observed dataset. Julian day and hydrological year information were added to each dataset, which were then arranged to match the format outlined through EflowStats. The 165 metrics calculated from the observed and simulated streamflow data were then compared using Spearman's Rank correlation coefficient (r). The specific metrics that were correlated (r=>0.5) were calculated for all grid cells in the stream network (n=85,363).

Spearman's correlations of the observed and simulated hydrological metrics for 518 sites are set out in the supplementary material (Supplementary Table S2.1). Of the 165 metrics, 53 were sufficiently (r >=0.5) positively correlated with their observed equivalent. Positive correlations were predominantly found for metrics describing mean values of streamflow, e.g. mean monthly flow. Figure 2.3 shows the mean streamflow for the same 518 sites and time period (1965-1984).

While the simulated streamflow was consistently lower than the observed one throughout the time series, there is a good match in the recurring temporal trends of the streamflow. It is this match in the long term trend which indicates that the streamflow predicted here can be used to derive hydrological metrics for large spatial scales, at high spatiotemporal resolutions.

2.6.3 Temporal validation

To validate the predictions temporally, the data were split as above and Kling-Gupta Efficiency (KGE, Kling et al. 2012), RMSE and the coefficient of determination (R^2) were calculated on the observed vs. simulated values in the testing dataset, across the entire time series of 64 years through the R package HydroGof (Zambrano-Bigiarini 2014).

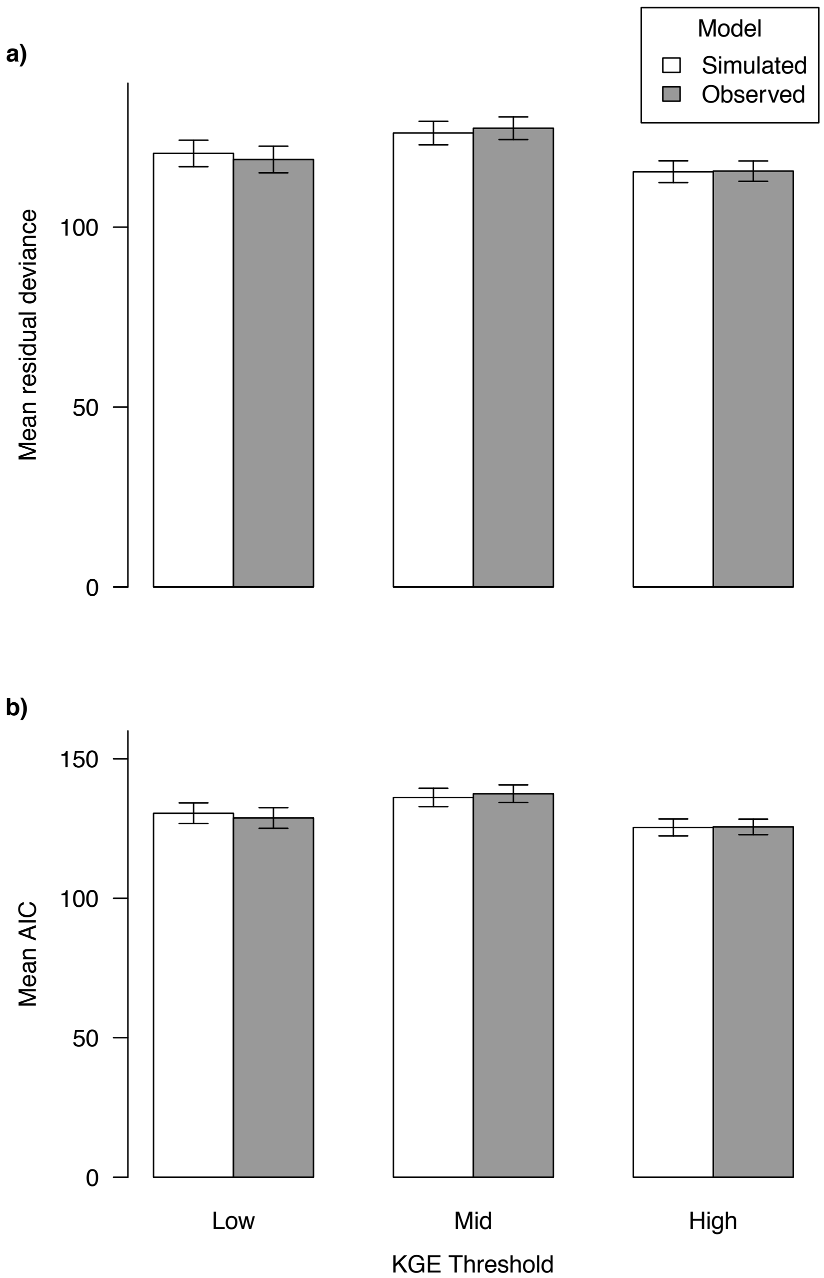
From a total of 1,014 sites, 69 had KGE over 0.6 (mean = -1.58, n=1,014) and 352 had R^2 over 0.5 (mean = 0.4, n=1,014). In addition, RMSE for the testing data shows a mean of 29.67 (m^3 s^{-1}) and the training data a mean of 29.67 (m^3 s^{-1}). Full values are shown in the supplementary material (Supplementary Table S2.2).

We are aware that the majority of these results do not meet the requirements for most hydrological applications (Moriasi et al. 2007). However, the objective of this study was to produce streamflow data to be used for ecological predictive models, therefore we performed GLM models on metrics calculated with both observed and simulated values.

For this purpose, benthic macro-invertebrate occurrence data for Germany were collected from federal state environment agencies (see supplementary material for full species list, Supplementary Table S2.3). Species presences were defined as those with at least one occurrence per site over the period 2005-2013 and occurring at more than 19 sampling sites across all Germany. The gauging sites were paired with species sampling sites in QGIS (QGIS Development Team 2016) within a buffer of 3 km. To ensure the sample sites were placed on the original stream, the original flow accumulation value had to be within 10% of the flow accumulation of the newly allocated grid cell (Domisch et al. 2017). A total of 327 sites remained that had associated observed and simulated discharge, as well as species presence data. A total of 34 species with over 20 presences were modeled. The gauging sites were split into 3 categories of ranging KGE values: low (KGE < 0, no. of sites= 108, no of species =20), mid (KGE > 0 < 0.4 no of sites= 116, no of species =32) and high (KGE > 0.4, no of sites= 103, no of species =27). Two sets of four IHA metrics (TA1, TA2, MH21, MH8, see Supplementary Table S2.1 for full descriptions (Colwell 1974, Hughes and James 1989, Olden and Poff 2003) were calculated with observed and simulated discharge, respectively.

For each KGE category we performed a GLM to predict species distribution for Germany. The comparison of the skill, which is defined as the residual deviance and Akaike Information Criterion (AIC), from each GLM is illustrated in Figure 2.4 with full results shown in the supplementary material (Supplementary Table S2.3), where lower values indicate a better fit.

The validation of the ecological models showed that the simulated streamflow yielded almost the same model accuracy as the ones calculated with observed streamflow. A Wilcoxon test was applied as a statistical test for difference on both the AIC and the residual deviance. Overall, there was no difference between models applied with observed or with simulated predictors (low: Wilcoxon test, $p = 0.69$; mid: Wilcoxon test, $p = 0.85$; high: Wilcoxon test, $p = 0.96$). The results of the GLM suggest that the data predicted from this study is adequate for the purpose of ecological predictive modeling at the scale of Germany, encompassing steep environmental and hydrological gradients that facilitate the model to discriminate between suitable and non-suitable hydrological conditions for the species considered here.



.

Figure 2.4: Comparison of observed and simulated GLMs across 3 KGE thresholds. Low (KGE < 0, no of sites= 108), mid (KGE>0<0.4 no of sites= 116) and high (KGE > 0.4, no of sites= 103). The columns represent a) mean deviance, b) mean AIC, on total n=34 species, sub-divided per KGE category; low (n=20), mid (n=32), high (n=27) with standard error bars.

2.7 Usage Notes

Obtaining streamflow data for ecological modeling over large scales with high spatial and temporal resolution is challenging in terms of data availability and computing effort. Our simple regression model was able to overcome these challenges while still rendering adequate predictions regarding the occurrence of benthic macro-invertebrate species predicted using SDMs. The set of 53 IHA metrics provided as downloadable GeoTIFF files (Data Citation 1) describe important aspects of the flow regime and has the potential to be applied in a number of freshwater investigations such as predictive modeling (e.g. SDMs) (Domisch et al. 2016) that are relevant for conservation and river management plans. The 64 year time series of simulated daily streamflow data are also provided as downloadable NetCDF files (Data Citation 1). Thus, the user can calculate IHA metrics for any other time period within the 64 years provided for any catchment or any river section of the 1 km German stream network.

Available data from GRDC and EarthEnv (Domisch et al. 2015a), together with the procedure provided through an R script, creates an accessible method for calculating both daily streamflow data and IHA metrics within any other geographical region where gauged streamflow data are available. This is especially helpful in areas with insufficient resources to implement complex models.

2.7.1 Limitations

We note that the daily streamflow values are estimated assuming conditions where most of the discharge is driven by precipitation. Hence, the user should be aware of the hydrological processes that drive the streams within the study region of interest such as the influence of groundwater (Guse et al. 2014) and soil infiltration processes. The user should also validate the results if applying the model in headwaters and during extreme events.

From the preliminary analysis we note that our model worked very well when applied in lowland regions; however, the models applied within the alpine region performed least well. Rivers in lowland regions are typically fed by ground water (Guse et al. 2014). Through soil infiltration and groundwater processes, streamflow has a slower response to precipitation. As we applied a seasonal resolution of precipitation, the delayed response time is incorporated within the model. The relatively low number of gauges used within the alpine model (Table 2.1) compared with the mountain and lowland regions, may partially explain its relatively low performance.

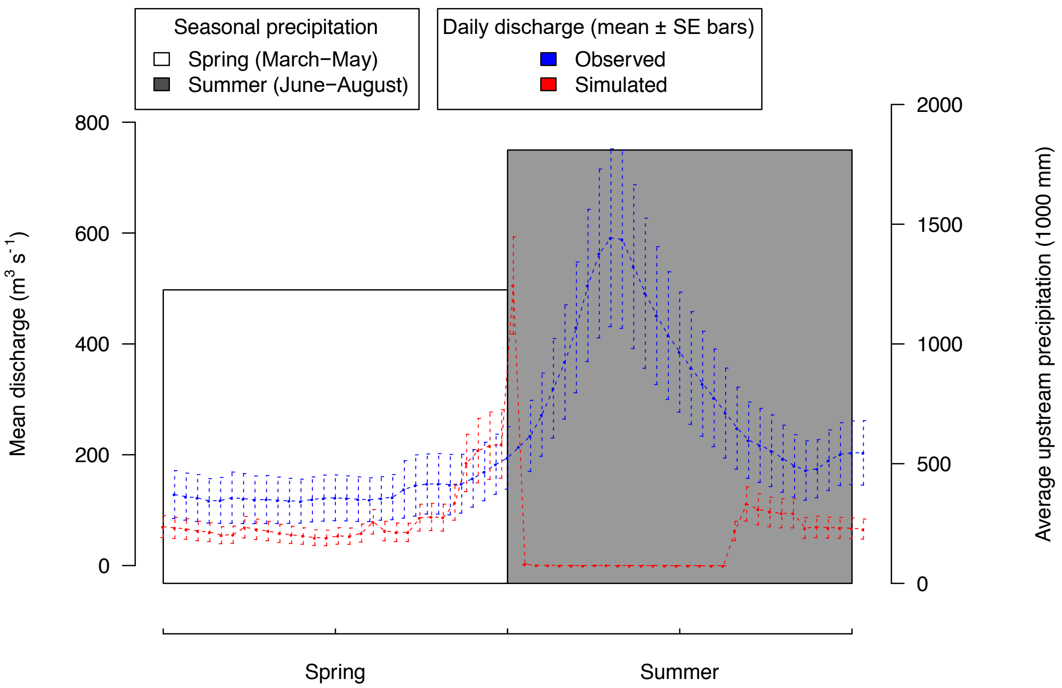


Figure 2.5: Time series of observed and simulated discharge (mean ± SE bars) over lowland ecoregion in May & June 2013. Columns; seasonal precipitation value for that period, points; daily mean ± SE bars observed and simulated discharge values.

Another explanation could be that the highly varied topography creates a complex landscape and hence complex precipitation-run off interactions that are highly impacted by daily events. Here, attributes such as altitude and steep mountain slopes are major factors in determining streamflow regimes by distributing rainfall to streams much quicker than in lowland areas. In addition, a prominent feature of alpine regions is the existence of glaciers and increased snow cover. These features largely control flow regimes through the periodical storage and melting of rainfall, which is released on various time scales from days to years (Jansson et al. 2003). This time lag, together with highly varied precipitation patterns (Isotta et al. 2014) and daily fluctuations in the melting of snow (Warscher et al. 2013), is not reflected in the seasonal precipitation of our model and is therefore not captured.

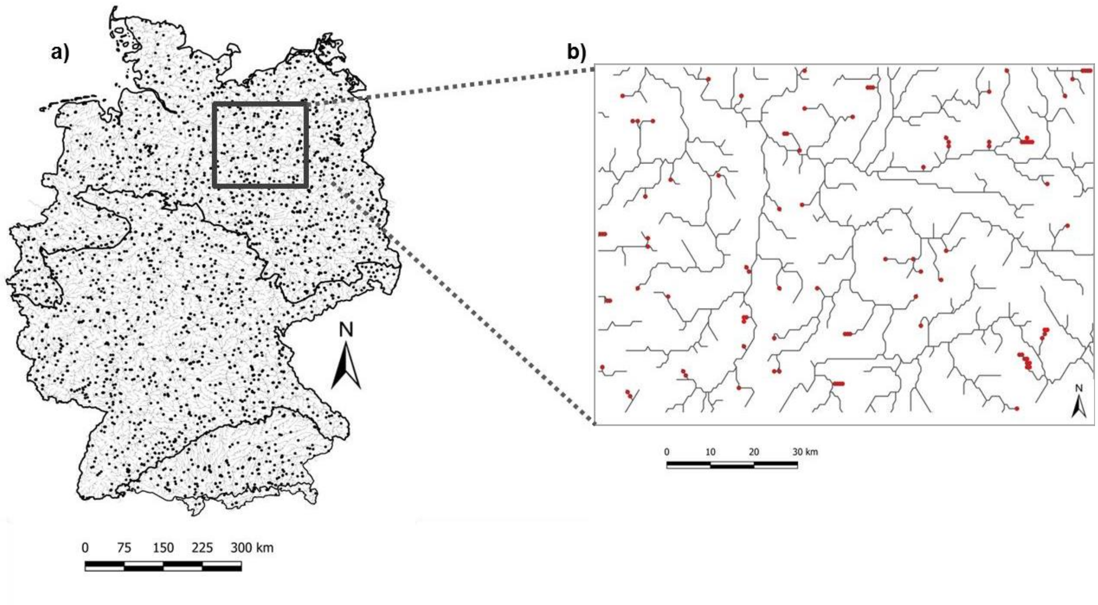


Figure 2.6: Distribution of predicted negative discharge values per grid cell. a) all grid cells that show > 5% negative values over the time series, b) subsetted grid cells > 5% negative values over the time series. Total grid cells = 85,363; grid cells > 5% negative values over days in time series (n=23,376) = 2,645 (red dots). Max % negatives = 6.93%.

The outliers that are apparent in Figure 2.2b (lowland) represent low performing models (mean R^2 = 0.003 +/- 0.0005, 1st-18th June 2013). This time period coincides with an extreme flooding event in June 2013 (The German Federal Institute of Hydrology (BfG) 2013) (Figure 2.5). There was a very apparent peak in mean observed discharge (n=77), which rose from 157.3 $m^3 s^{-1}$ on 27th May to over 588.5 $m^3 s^{-1}$ on 9th June (14 days). The outlier on May 31st represents data from fewer gauges (n=24) than available during the remaining days (n=77 for all days), possibly because several stations temporarily ceased operating during the flood. It is evident that extreme flooding events are difficult to capture and predict in the lowland region with any modeling approach (Figure 2.2). The flooding event in June 2013 was a result of heavy rainfall between 31st May and 18th June. This heavy rainfall was not reflected in the 50-year average precipitation value for that month in our model, which resulted in the model failing to converge, thus producing low performing predictions. In contrast, models applied in both mountain and Germany regions performed well for the same time period (mountain mean R^2 =0.99, Germany mean R^2 = 0.99).

A number of negative discharge values were predicted, which arise when the linear model produces a negative intercept, corresponding to the value of flow when the precipitation is equal to zero. Predicted negative discharge values are reported as a percentage of all sites over the entire time series. The total number of negative streamflow values was 0.99% of the entire extrapolated dataset (n = 1,995,445,488 grid cells over the time series), considered a negligible amount. Overall, only 3.09% (n=2,645) of all grid cells for the stream network of Germany (total n=85,363) yielded more than 5% negative values (maximum for one cell 6.93%) over the entire time series (n=23,376 days). The distribution of grid cells with negative flow values (Figure 2.6a) showed a strong pattern towards the headwaters of the stream network (Figure 2.6b).

Although obvious, it is important to note that negative flow values are impossible and are to be understood as regions where no discharge is predicted. However, the simplicity of the model together with basic hydrological theory can, to a great extent, explain this issue. On occasions, the model produces a strong slope for the linear model, which could be induced by the difficulty of capturing the time lag of precipitation into rivers with high volume flow. The strong slope results from the direct association of flow with precipitation, without considering other hydrological processes such as (ground)water storage, evaporation and evapotranspiration from soil (Beven 2004), interception (Brutsaert 1982) and surface depression storage (Kiesel et al. 2010). If these factors were to be included, the model would likely show a non-linear response, e.g. causing zero flow for regions below a certain flow accumulation threshold. This could potentially decrease the number of negative flow values. However, such relationships are only possible with significantly more complex models and beyond the goal of our approach to use a simple model.

2.8 Acknowledgements

This study was undertaken as part of the project “Global Change Effects in River Ecosystems” (GLANCE, no. 01LN1320A), funded by the German Federal Ministry of Education and Research (BMBF). We thank the German Federal State Environmental agencies for providing streamflow data and the German Working Group on Water Issues of the Federal States and the Federal Government (LAWA) for providing biological data. For useful advice on statistical analysis we thank B. O’Hara, G. Singer, and B. Venables. We thank T. Mehner and the participants of the workshop “Scientific

Writing" at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries for helpful discussions on an early stage of the manuscript. We also thank Tony Hodgson for proof reading the final manuscript and Judith Mahnkopf for help with data formatting. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

2.9 References (Chapter 2)

Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution, and Systematics **35**:257-284.

Archfield, S. A., J. G. Kennen, D. M. Carlisle, and D. M. Wolock. 2014. An objective and parsimonious approach for classifying natural flow regimes at a continental scale. River Research and Applications **30**:1166-1183.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment - Part 1: Model development. Journal of the American Water Resources Association **34**:73-89.

Barbarossa, V., M. A. J. Huijbregts, A. H. W. Beusen, H. E. Beck, H. King, and A. M. Schipper. 2018. FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. Scientific Data **5**:180052.

Barbet-Massin, M., W. Thuiller, and F. Jiguet. 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? Ecography **33**:878-886.

Beven, K. 2004. Robert E. Horton's perceptual model of infiltration processes. Hydrological Processes **18**:3447-3460.

Brutsaert, W. 1982. Evaporation into the atmosphere: Theory, History and Applications. Springer, Netherlands, Dordrecht.

Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management **30**:492-507.

Colwell, R. K. 1974. Predictability, Constancy, and Contingency of Periodic Phenomena. Ecology **55**:1148-1153.

Döll, P., and J. Zhang. 2010. Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations. Hydrology and Earth System Science. **14**:783-799.

Domisch, S., G. Amatulli, and W. Jetz. 2015a. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data **2**:150073.

Domisch, S., S. C. Jahnig, J. P. Simaika, M. Kuemmerlen, and S. Stoll. 2015b. 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.

Domisch, S., F. T. Portmann, M. Kuemmerlen, R. B. O'Hara, R. K. Johnson, J. Davy-Bowker, T. Baekken, C. Zamora-Munoz, M. Sainz-Bariain, N. Bonada, P. Haase, P. Doll, and S. C. Jahnig. 2017. Using streamflow observations to estimate the impact of hydrological regimes and anthropogenic water use on European stream macroinvertebrate occurrences. Ecohydrology **10**:11.

Domisch, S., A. M. Wilson, and W. Jetz. 2016. Model-based integration of observed and expert-based information for assessing the geographic and environmental distribution of freshwater species. Ecography **39**:1078-1088.

Fox, J., and S. Weisberg. 2011. Robust Regression in R.*in* K. V, editor. An Appendix to An R Companion to Applied Regression, Second Edition. Sage Publishing, Thousand Oaks, CA.

Guse, B., D. E. Reusser, and N. Fohrer. 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.

Henriksen, J. A., J. Heasley, J. G. Kennen, and S. Nieswand. 2006. Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools).72.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hijmans, R. J., J. v. Etten, J. Cheng, M. Mattiuzzi, M. Sumner, J. A. Greenberg, O. P. Lamigueiro, A. Bevan, E. B. Racine, A. Shortridge, and A. Ghosh. 2015. raster:

Geographic Data Analysis and Modeling, version 2.5-2. *R Package*. https://CRAN.R-project.org/package=raster.

Huber, P. J. 1964. Robust Estimation of a Location Parameter. Annals of Mathematical Statistics. **35**:73-101.

Hughes, J., and B. James. 1989. A hydrological regionalization of streams in Victoria, Australia, with implications for stream Ecology. Marine and Freshwater Research **40**:303-326.

Illies, J. 1967. Limnofauna Europaea: Stuttgart 1967 ; Prospekt. Gustav Fisher Verlag, Stuttgart.

IPCC. 2007. Climate Change 2007: The Physical Science Basis. WG1, Cambridge University Press.

Isotta, F. A., C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, and G. Vertačnik. 2014. The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. International Journal of Climatology **34**:1657-1675.

Jähnig, S. C., M. Kuemmerlen, J. Kiesel, S. Domisch, Q. Cai, B. Schmalz, and N. Fohrer. 2012. Modelling of riverine ecosystems by integrating models: conceptual approach, a case study and research agenda. Journal of Biogeography **39**:2253-2263.

Jansson, P., R. Hock, and T. Schneider. 2003. The concept of glacier storage: a review. Journal of Hydrology **282**:116-129.

Jarvis, A., H.I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Kiesel, J., N. Fohrer, B. Schmalz, and M. J. White. 2010. Incorporating landscape depressions and tile drainages of a northern German lowland catchment into a semi-distributed model. Hydrological Processes **24**:1472-1486.

Kling, H., M. Fuchs, and M. Paulin. 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology **424–425**:264-277.

Kuemmerlen, M., B. Schmalz, Q. Cai, P. Haase, N. Fohrer, and S. C. Jähnig. 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.

Kuemmerlen, M., B. Schmalz, B. Guse, Q. Cai, N. Fohrer, and S. C. Jähnig. 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. Ecological Modelling **277**:77-86.

Lehner, B., K. Verdin, and A. Jarvis. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Transactions American Geophysical Union **89**:93-94.

Maechler, M., P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, and M. A. d. Palma. 2016. robustbase: Basic Robust Statistics, version 0.92-7. R package. http://CRAN.R-project.org/package=robustbase.

McIntyre, N., A. Al-Qurashi, and H. Wheater. 2007. Regression analysis of rainfall-runoff data from an arid catchment in Oman. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques **52**:1103-1118.

Michigan Department of Natural Resources. 2011. Fisheries Research Report. 2095, Lansing.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the Asabe **50**:885-900.

Olden, J. D., and N. L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications **19**:101-121.

Poff, N., J. Allan, M. Bain, J. Karr, K. Prestegaard, B. Richter, R. Sparks, and J. Stromberg. 1997. The natural flow regime. Bioscience **47**:769-784.

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

Poff, N. L., and J. V. Ward. 1989. Implications of streamflow variability and predictability for lotic community structure - a regional-analysis of streamflow patterns. Canadian Journal of Fisheries and Aquatic Sciences **46**:1805-1818.

QGIS Development Team. 2016. QGIS Geographic Information System, version 2.18.10. *Open Source Geospatial Foundation Project*. https://qgis.org/en/site/.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun. 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. Conservation Biology **10**:1163-1174.

Ronchetti, E., C. Field, and W. Blanchard. 1997. Robust linear model selection by cross-validation. Journal of the American Statistical Association **92**:1017-1023.

Seelbach PW, Hinz LC, Wiley MJ, Cooper. AR. 2011. Michigan Department of Natural Resources,. Lansing. 2095.

Schulla, J., and K. Jasper. 2006. Model description WaSiM-ETH. – Internal report. Institute for Atmospheric and Climate Science, ETH Zürich

The German Federal Institute of Hydrology (BfG). 2013. The June 2013 floodings in Germany – an overview. Bericht BfG-1797, Federal Institute of Hydrology, Koblenz.

Warscher, M., U. Strasser, G. Kraller, T. Marke, H. Franz, and H. Kunstmann. 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. Water Resources Research **49**:2619-2637.

Yohai, V. J. 1987. High Breakdown-Point and High Efficiency Robust Estimates for Regression. Annals of Statistics. **15**:642-656.

Zambrano-Bigiarini, M. 2014. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, version 0.3-8. *R Package*. https://CRAN.R-project.org/package=hydroGOF.

3 Identifying and applying an optimum set of environmental variables in species distribution models

Katie Irving[1,2], Sonja C. Jähnig[1] & Mathias Kuemmerlen[3*]

[1]*Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 310, 12587 Berlin, Germany.*

[2]*Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Takustraße 3, 14195 Berlin, Germany.*

[3]*Department Systems Analysis, Integrated Assessment and Modeling, Eawag, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.*

This is a modified version of an Accepted Manuscript of an article published by Taylor & Francis in Inland Waters on 3rd December 2019, available online: https://doi.org/10.1080/20442041.2019.1653111

Irving K, Kuemmerlen M, Jähnig SC. 2019.Identifying and applying an optimum set of environmental variables in species distribution models, Inland Waters. doi; 10.1080/20442041.2019.1653111

This chapter is missing in the online version due to copyright reasons.

4 Disentangling the influence of climatic and hydrological predictor variables on benthic macroinvertebrate distributions

Katie Irving[1,2], Sonja C. Jähnig[1] & Mathias Kuemmerlen[3*]

[1]*Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 310, 12587 Berlin, Germany.*

[2]*Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Takustraße 3, 14195 Berlin, Germany.*

[3]*Department Systems Analysis, Integrated Assessment and Modeling, Eawag, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.*

The following version has been modified from the ready to submit version

Irving K, Kuemmerlen M, Jähnig SC. 2019.Disentangling the influence of climatic and hydrological predictor variables on benthic macroinvertebrate distributions. (to be submitted)

4.1 Abstract

For lotic freshwater Species' Distribution Models (SDMs), applying environmental variables that describe flow regime is a priority, to be consistent with flow-ecology theory. However, most studies to date only include data describing climate or stream related surrogates. We calibrated four SDMs using different combinations of three predictor sets on 92 macroinvertebrate species. We compared model performance (TSS) of all combinations of predictor sets, i.e. four model configurations. The relative influence of each predictor set on the spatial distribution of the community was obtained from both the influence of individual predictors (relative importance) from SDMs and a variance partitioning analysis. SDMs with bioclimate and hydrology configurations performed significantly better overall (Mean TSS = 0.68 ± 0.02), demonstrating the lowest unexplained variance (0.29) and predicted significantly larger range sizes (Mean no. of presences; 3482.6 ± 129.1) with a range overlap of 47.6 ± 1.6 to 59.1 ± 2.1%. In terms of both variable importance and proportional variance, bioclimate was found here to be the most important factor for species' distributions. Despite the importance of bioclimate, hydrology contributed to a higher proportion of explained variance, unrivalled by other SDM configurations Individually, however, the hydrology data implemented here had the lowest influence on species' distribution most likely due to scale-dependency. Hydrology describes the discharge regime, which highly influences macroinvertebrate distribution and may have resulted in larger predicted range sizes. The impact of including hydrology in SDMs, on predicted range size, has important implications for river management decisions.

4.2 Introduction

Species distribution models (SDMs) are ecological predictive models that are increasingly used to inform and complement large scale distribution analyses to aid conservation efforts (Araújo et al. 2011, Guisan et al. 2013, Eaton et al. 2018). In river ecosystems, however, SDMs have only relatively recently been applied due, in part, to 1) complex interactions between the numerous driving factors of river systems and 2) insufficient data describing the stream environment.

Hydrological flow regime is said to be the "master variable" (Power et al. 1995) of lotic habitats and critical to the ecological stability of river ecosystems (Poff et al. 1997). It is highly variable, both spatially and temporally, making it a core driver of the physical structure of river habitats and a regulator of species distribution and abundance (Resh et al. 1988, Poff et al. 1997). In most regions on earth, precipitation is the first and foremost driver of streamflow, but flow run off patterns are shaped by complex interactions between topography (e.g. hillslope), climate (e.g. temperature), geology (e.g. porous rock) and land use (e.g. forests) throughout the stream network.

Many river species are dependent on undisturbed flow regimes for either all, or a very important part of their life history (Lytle and Poff 2004). Numerous macroinvertebrate species depend on flow related cues that directly or indirectly initiate, for example, breeding period (Hancock and Bunn 1997), development (Gray 1981) and emergence & metamorphosis (Peckarsky et al. 2000, Lytle 2002). Therefore, species have evolved to the habitat heterogeneity in rivers caused by the variability in flow regime e.g. free flowing water, pools & riffles, low/high flows, intermittent and ephemeral flows. With the increasing changes in hydrological regime due to climate change, e.g. severity and frequency of floods and droughts, it is essential to understand the influence streamflow has on the distribution of species, so that suitable recommendations can be made to restore or conserve river systems successfully.

Recently, data describing stream ecosystems are becoming available e.g. stream specific climate and land use (Domisch et al. 2015a), hydrology (Barbarossa et al. 2018, Irving et al. 2018), river classification and characteristics (Hydrosheds, Ouellet Dallaire et al. 2019) as well as dams and reservoirs (globaldamwatch.org, Lehner et al. 2011). Despite this increasing availability, there are still a limited number of studies that directly investigate the influence of flow regime on riverine species distribution. Of these, most focus on climate (e.g. Domisch et al. 2011, Ihlow et al. 2012, Markovic et al. 2014, Ruiz-Navarro et al. 2016, Kärcher et al. 2019, Rodríguez-Merino et al. 2019) and implement precipitation variables as surrogates of hydrological variables (e.g. Domisch et al. 2019). While climate is certainly a dominant factor in driving species distribution and abundance, its sole use may be misrepresenting the effect on species' distributions due to correlating factors, such as topography, which may impact predictions, leading to ambiguous conclusions (Real et al. 2013).

In addition to climate, some studies have used watershed characteristics such as river density to model the distribution of water birds (Zeng et al 2015), and river drainage area for fish distribution (Maloney et al 2013). These examples show that river related topography has an influence on species' distribution; however, they only marginally describe the flow regime that influences species' distribution and abundance.

Some recent attempts have been made to include, at least, some aspects of flow regime e.g. high flow days (Kuemmerlen et al. 2015a) as well as aggregated flow statistics , e.g., mean annual flow (Kuemmerlen et al. 2015b, Pyne and Poff 2017), which were shown to be the of high relevance to macroinvertebrate distribution.

There are 19 bioclimatic variables openly available from worldclim.org (Hijmans et al. 2005, Lehner et al. 2008, Fick and Hijmans 2017) which are applied frequently in SDMs and other predictive modeling studies. These data are local grid-cell based, and include variables describing temperature and precipitation. Although these data are useful in predictive modeling, they do not describe the stream ecosystem due to the lack of a catchment perspective that provides information on upstream zones. The environmental conditions of the upstream environment have important consequences for the species living downstream (Allan 2004) therefore, it is important to include such aspects when predicting species' distribution. The dataset from www.earthenv.org/streams (Domisch et al. 2015a) is based on, among others, the 19 bioclimatic variables from WorldClim. However, it differs in that the bioclimatic information is accumulated down the stream network, accumulating information from the upper subcatchment in every point along the stream network. This dataset, therefore, includes information specific to the stream environment, an aspect which has proven to be highly relevant to distribution predictions of freshwater biodiversity (Vinson and Hawkins 1998, Malmqvist and Rundle 2002, Kuemmerlen et al. 2014). However, because flow accumulation is used as the mechanism to relate environmental variables with the stream network (see Domisch et al. 2015a for details), the accumulative feature causes high correlated among many of the variables and with streamflow (Kuemmerlen et al. 2014, Kuemmerlen et al. 2015a). Because such data include aspects of both climate and hydrology, disentangling the relative influence of either factor on species' distribution becomes problematic.

The dataset from Irving et al. (2018), includes 53 of the Indicators of Hydrological Alteration (IHA) Metrics that describe the magnitude, frequency, timing, duration and rate of change of flow events (Olden and Poff 2003). IHA metrics are commonly used in flow-ecology assessments (e.g. Kakouei et al. 2018) and environmental flow research (Poff et al. 2010, Peres and Cancelliere 2016) as they comprehensively describe hydrological flow regime. These metrics, however, are rarely included in river SDMs (but see Irving et al, 2019). As the information contained in the

IHA metrics is directly related to the hydrological regime of rivers, it is logical to suggest that this data could positively impact predicted species' distributions.

The three datasets described above are important to the distribution of river biota; however, the separate influence of each driver, remains unknown. By disentangling the influence these factors, we can build on existing knowledge regarding the abiotic drivers of species distribution with a view to inform management decisions in formulating wise conservation or restoration strategies.

4.2.1 Aims and objectives

To investigate the separate influence of climate and hydrology, we applied SDMs on a community of benthic macroinvertebrates with three datasets describing either 1) climate only, 2) hydrology only and 3) information describing both climate and hydrology (hydroclimate). We evaluate the influence of each predictor set on species' distribution by 1) investigating individual variance explained as well as shared variance explained by each dataset, 2) assessing the variable importance from the SDM of each predictor set on the community. In addition, we compare how well each model configuration performs by evaluating the differences in 1) model performance, 2) predicted range size as well as range overlap of the predictions.

By comparing the differences in model performance, variable importance and explained variance, we expect to determine the individual influence of both climate and hydrology, and to what extent these datasets influence predicted species' distributions. We hypothesize that SDMs including the recently developed hydrological dataset will improve SDM performance. Further, we analyze how the choice of predictor datasets influences the predicted distributions.

4.3 Method

4.3.1 Study area

The study area was the Ems (17,934 km^2) and Weser (46,306 km^2) catchments located in Germany (Figure 4.1). Due to restricted access to biological data outside Germany, these catchments were chosen as they fully lie within Germany. The two catchments are adjacent to each other and cover two ecoregions: Central Plains (lowland) and Central Highlands (mountain, sensu Illies 1967).

The stream network for the study area is based on a layer in GEOTiff raster format of a modeled 1 km^2 gridded stream network with a total of 13,749 cells. The network was downloaded from earthenv.org/streams, which was derived from Hydrosheds (www.hydrosheds.org, 30-arc-second spatial grain Lehner et al. 2008), which in turn is based on the SRTM dataset (www.srtm.csi.cgiar.org, Jarvis et al. 2008).

4.3.2 Biological data

Macroinvertebrate species were sourced from German federal state agencies and collected following the sampling protocol outlined in Haase et al. (2004). To be included in the study, each species had to be identified to species level and have at least 20 occurrences within the study area. A total of 92 species occurrences at 1258 sites were available through this process, in a presence only format and covering the period between 2005 and 2013.

4.3.3 Model set-up

Three predictor sets containing information that describe 1) climate (bioclimate, bC), 2) flow regime (hydrology, H), 3) climate and hydrology combined (hydroclimate, hC) were applied in this study. Hence, we set up four model configurations to compare all eventualities: 1) hydroclimate & hydrology (hC-H), 2) bioclimate & hydrology (bC-H), 3) hydroclimate & bioclimate (hC-bC), 4) hydroclimate, bioclimate & hydrology (hC-bC-H). By comparing each predictor data set in the relative influence of their individual predictors within SDMs, (relative importance) and the explained variance of each predictor set from the variance partitioning analysis, we can identify the influence of each climate or hydrology predictor set on the spatial distribution of the community. Here, the full model (hC-bC-H) represents the full coverage of environmental predictors used in our study; therefore, it is intended for comparative purposes only.

4.3.4 Environmental Predictors

The predictor variables for hydrology, hydroclimate and bioclimate used in this study are all openly sourced data and freely available to the user. All datasets are available in raster GEOTiff format.

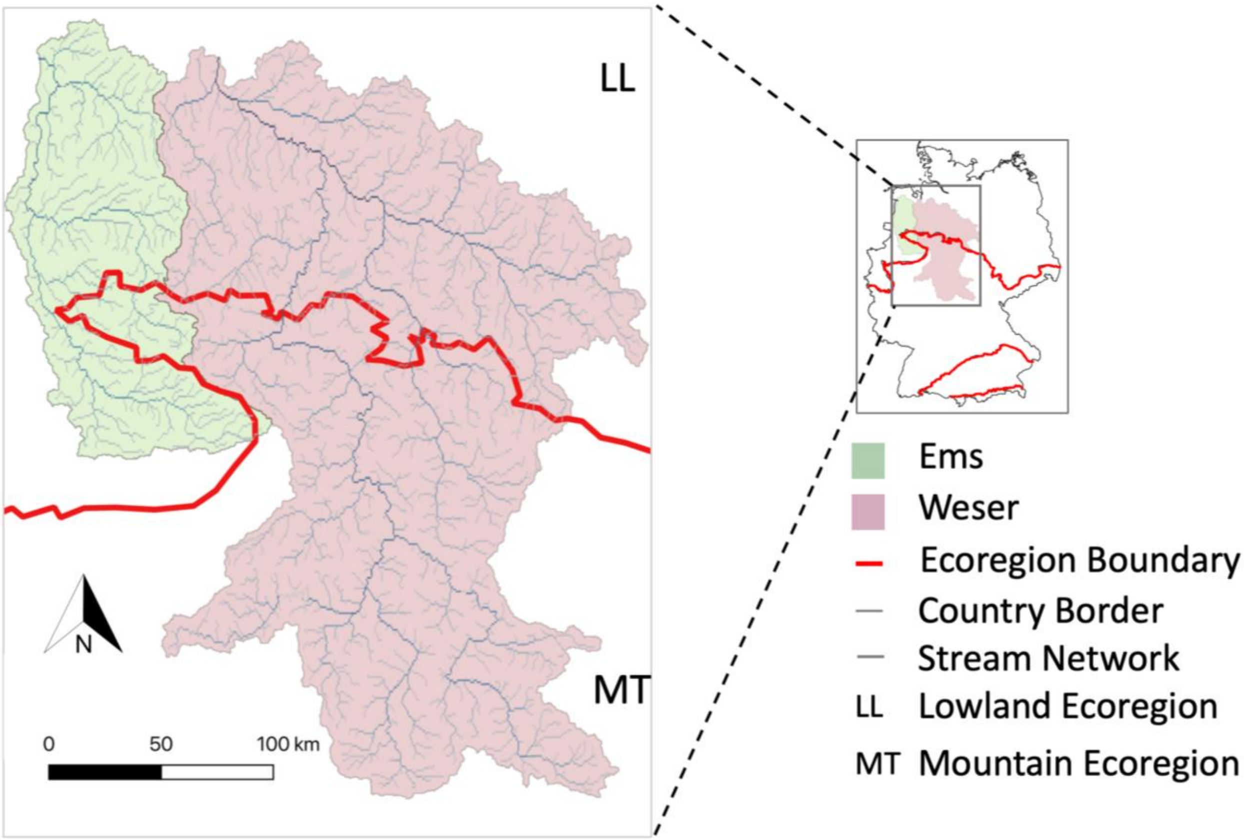


Figure 4.1: Study area of Ems (green) and Weser (pink) catchments. Location within Germany, along with ecoregion boundaries (red lines).

4.3.4.1 Hydrology

The 53 hydrological metrics derived from Irving et al. (2018) were applied in this study. These variables are based on the Indicators of Hydrological Alteration (IHA metrics) outlined in Olden and Poff (2003), which describe the various aspects of flow regime: duration, timing, magnitude, frequency and rate of change of flow events.

In brief, streamflow was extrapolated for the German stream network through a weighted linear regression using accumulated seasonal precipitation from earthenv.org/streams (Domisch et al. 2015a). The daily streamflow (m^3 s^{-1}) was then applied as input to calculate the IHA metrics via Eflowstats (www.github.com/USGS-R/EflowStats, Henriksen et al. 2006, Archfield et al. 2014). The data have been found to be affective for use in predictive modeling (see Irving et al. 2018 for further details). All 53 IHA metrics were available for the time period 1985-2013 for the Ems and Weser catchments.

Table 4.1: Predictors used in each model configuration as chosen by BRT variable selection. hC = Hydroclimate, bC = Bioclimate, H = Hydrology

hC-H	bC-H	hC-bC	hC-bC-H
Bio 08 (hC)	Bio 02 (bC)	Bio 02 (bC)	Bio 02 (bC)
Bio 09 (hC)	Bio 04 (bC)	Bio 04 (bC)	Bio 04 (bC)
Bio 12 (hC)	Bio 08 (bC)	Bio 08 (bC)	Bio 08 (bC)
MH21	Bio 09 (bC)	Bio 09 (bC)	Bio 09 (bC)
DH1	Bio 15 (bC)	Bio 15 (bC)	Bio 15 (bC)
	MH21	Bio 08 (hC)	Bio 08 (hC)
	DH1	Bio 09 (hC)	Bio 09 (hC)
		Bio 12 (hC)	Bio 12 (hC)
			MH21
			DH1

These metrics were originally derived from both gauging station daily streamflow data and seasonal precipitation also sourced from earthenv.org/streams (Domisch et al. 2015a). It is therefore important to note that some degree of correlation is to be expected with the hydroclimate variables used in this study (Bio12-19). Nonetheless, the additional high-resolution streamflow data adequately adds various aspects of flow regime at a high temporal resolution and therefore better represents hydrology.

4.3.4.2 Hydroclimate

Hydroclimate variables were downloaded from earthenv.org/streams (Domisch et al. 2015a), which are based on the 19 bioclimatic variables available from Worldclim (Hijmans et al. 2005, Lehner et al. 2008) and contain variables describing temperature and precipitation. The bioclimatic variables accumulate down the stream network. Therefore, each individual grid cell includes information from all upstream cells. The accumulative nature of the data is developed through flow accumulation, which correlates with streamflow, and hence includes an aspect of hydrological information. This dataset therefore contains information describing both climate and hydrology, embedded within the values.

4.3.4.3 Bioclimate

The 19 bioclimatic variables available through Worldclim (Hijmans et al. 2005, Lehner et al. 2008) were downloaded in 1 km (30 arc secs) resolution. These variables are commonly used in predictive modeling applications. All 19 variables were masked to the base layer 1 km^2 stream network described above. The bioclimate dataset differs from the hydroclimate dataset in that it does not include the accumulative aspect originating from flow accumulation. Therefore the bioclimate data is measured at local grid cell scale and represents our measurement of climate.

4.3.5 Variable selection process

The variables were selected following the procedure from (Irving et al. 2019) using Boosted Regression Trees (BRTs). In brief, BRTs were applied in a two-step process. First, each predictor set were applied in BRTs separately (hydroclimate n=19, bioclimate n=19, hydrology n=53) for every species within the community (n=92). BRTs calculate the variable importance of each predictor from the number of times each variable was chosen by the algorithm (Elith et al. 2008). The variable importance was averaged (mean) across all species to find the variable importance for the community. The average variable importance was used to determine the most important (30%) individual variables from each predictor set (hydroclimate n=6, bioclimate n=6, hydrology n=19). The remaining variables from all predictor sets (n=31) were then applied collectively into the 2nd run of BRTs with the same criteria as above. The purpose of the 2nd BRT run was to impartially select the most relevant predictors for this community and spatial scale, without forcing a specific number of variables from each category. This forcing results in limited consideration of all the most relevant variables from each predictor category, potentially resulting in a bias outcome.

A pair-wise Pearson's correlation analysis was undertaken for each model configuration with the threshold 0.7 (Dormann et al. 2013). The variable importance from the 2nd run of BRTs was used to determine which of the correlated variables were chosen to remain in analysis. Variables chosen for each model are outlined in Table 4.1. As the variables are related, i.e. hydroclimate is derived from the bioclimate dataset, and hydrology was derived, in part, from the hydroclimate dataset, it was likely that a high level of correlation would be observed between datasets. Therefore, the correlation analysis was undertaken with great caution. (see Table S4.1 for correlation matrix).

The final predictor set (n=10) was applied as the full 3-set model configuration hC-bC-H (hydroclimate, bioclimate and hydrology). The variables contained in the full configuration were then distributed according to the remaining 2-set model configurations: hC-H (hydroclimate and hydrology), bC-H (bioclimate and hydrology), hC-bC (hydroclimate and bioclimate). This method of predictor distribution was applied to maintain consistency throughout model configurations. Here, we can intuitively evaluate the relative influence of each predictor set, even though model configurations have differing numbers of predictors.

4.3.6 Species distribution models

All SDM analysis was undertaken in R sdm package (Naimi and Araujo 2016). We applied the four model configurations outlined above to each species within the community in separate SDMs. Each SDM was applied with an ensemble of 5 algorithms: Artificial Neural Network (ANN), Generalized Linear Model (GLM), Flexible Discriminant Analysis (FDA), Boosted Regression Tree (BRT), and Classification Tree Analysis (TREE). As the species data are presence only, we applied 2000 randomly placed pseudo absences in geographical space as back ground absences. Each model was repeated ten times by bootstrapping. This resulted in 50 models per species, per configuration. For validation, the data were randomly split into training and testing datasets in 70:30 ratios. The True Skills Statistic (TSS) and the Sensitivity values were derived from the validation as a measurement of model performance. Sensitivity is a measure of true positives, i.e. where the model correctly places a presence, and the TSS is derived from both the sensitivity and the specificity (the number of true negatives). The TSS values are reported as weighted mean ± standard error (mean ± se).

The ensemble predictions for each species were calculated though the weighted TSS. As output, each model produced a probability map of occurrence for the entire study area. To convert the probability map to binary presence/absence predictions (1,0), we applied a threshold determined from maximizing sensitivity and specificity (Lui et al 2005, 2013).

4.3.7 Statistical Analysis

4.3.7.1 Model performance and variable importance

All analysis was undertaken in R version 3.5.2 (R Core Team 2018). Model performance was analyzed using the TSS values calculated through the training and testing validation datasets. Pairwise Wilcoxon tests were applied to the 50 TSS values of each species to test for differences between model configurations. Any values $p < 0.05$ were considered significantly different. Each configuration modelled 92 species resulting in a total of 276 Wilcox tests. We summarize the outcome as a percentage of significance for each model configuration i.e. % S = (number of significant models/276)*100.

The variable importance from each SDM was extracted for every species per configuration to assess how important each predictor category was to the community. The variable importance is a measurement of correlation (Naimi and Araujo 2016), which is not directly comparable between configurations. Therefore, to allow for comparison across the different model configurations, a relevance metric (following Irving et al. 2019) was applied. Here, each variable was assigned a value 1-10 depending on its ordinal position, i.e. the variable with the highest importance was assigned 10 points, the 2nd assigned 9 points and so on. We used values between 1 and 10 owing to 10 representing the maximum number of variables applied in the model configurations (see Table 4.1, hC-bC-H). This variable importance ranking system results in comparability across model configurations.

4.3.7.2 Predicted distribution

The predicted distributions were compared according to range size and percentage overlap. Range size was defined as the number of presences within the study area predicted by the model, after converting the predicted probabilities to binary presence/absence predictions (1,0). Range size was determined by counting the number of species presences predicted by each model configuration. To test for differences in range size between model configurations, pairwise paired Wilcoxon tests were applied. To compare the predicted distribution of the community, and how they were similar or different in geographical space, pairwise range overlap values were calculated by counting the number of grid cells that contained a same species' presence predicted by each respective model configuration. The proportion of shared grid cells in relation to the predicted range of the pairwise model configurations was calculated as percentage overlap (mean ± se) for the community.

4.3.7.3 Variance partitioning

Variance partitioning analysis was applied on all model configurations for each species. First, Generalized Linear Models (GLMs) were performed on all binary predictions (i.e. presence/absence) to determine the proportional variance for each predictor set separately, then collectively according to the model configuration to ascertain the shared variance. Due to the nature of logistic regression, i.e. GLM, the standard coefficient of determination (R^2) cannot be derived from the model. Therefore, from each GLM a pseudo R^2 value was calculated through the nagelkerke function in the rcompanion package (Mangiafico 2019) using the McFadden method (de Araujo et al. 2014). It is important to note that pseudo R^2 values cannot be interpreted in same manner as other regression techniques, i.e. ordinary least squares (OLS); the amount of variance in the response variable explained by the predictor variable. The pseudo R^2 value in GLM context is a relative measure between models of the same type describing how well the model explains the data (http://rcompanion.org/handbook/G_10.html, online book, accessed 13/06/2019). The pseudo R^2 values were used here to determine the proportional contribution of variance as explained by the model, the total variance explained by the model being 1. This proportion was calculated following de Araujo et al. (2014), where the proportion of variance explained by the first predictor set can be described as total variance explained minus the proportion of variance explained by the second predictor set in the configuration.

The shared variance of both/all predictor sets can be described as the total variance explained minus the sum of both/all predictor sets in the configuration. For example, the proportional variance of the hydroclimate data in the hydroclimate and hydrology configuration would be; $R^2_{\text{hydroclimate}} = 1 - R^2_{\text{hydrology}}$, the proportional variance of hydrology; $R^2_{\text{hydrology}} = 1 - R^2_{\text{hydroclimate}}$, and the amount of shared variance: $R^2_{\text{shared}} = 1 - (R^2_{\text{hydrology}} + R^2_{\text{hydroclimate}})$. These proportional variance values were then used as input into the varPart function in modEva package (Barbosa et al. 2016) to calculate the proportional variance partition. This procedure was applied to every species and for each model configuration then the proportional variance was averaged across all species and reported as community proportional variance.

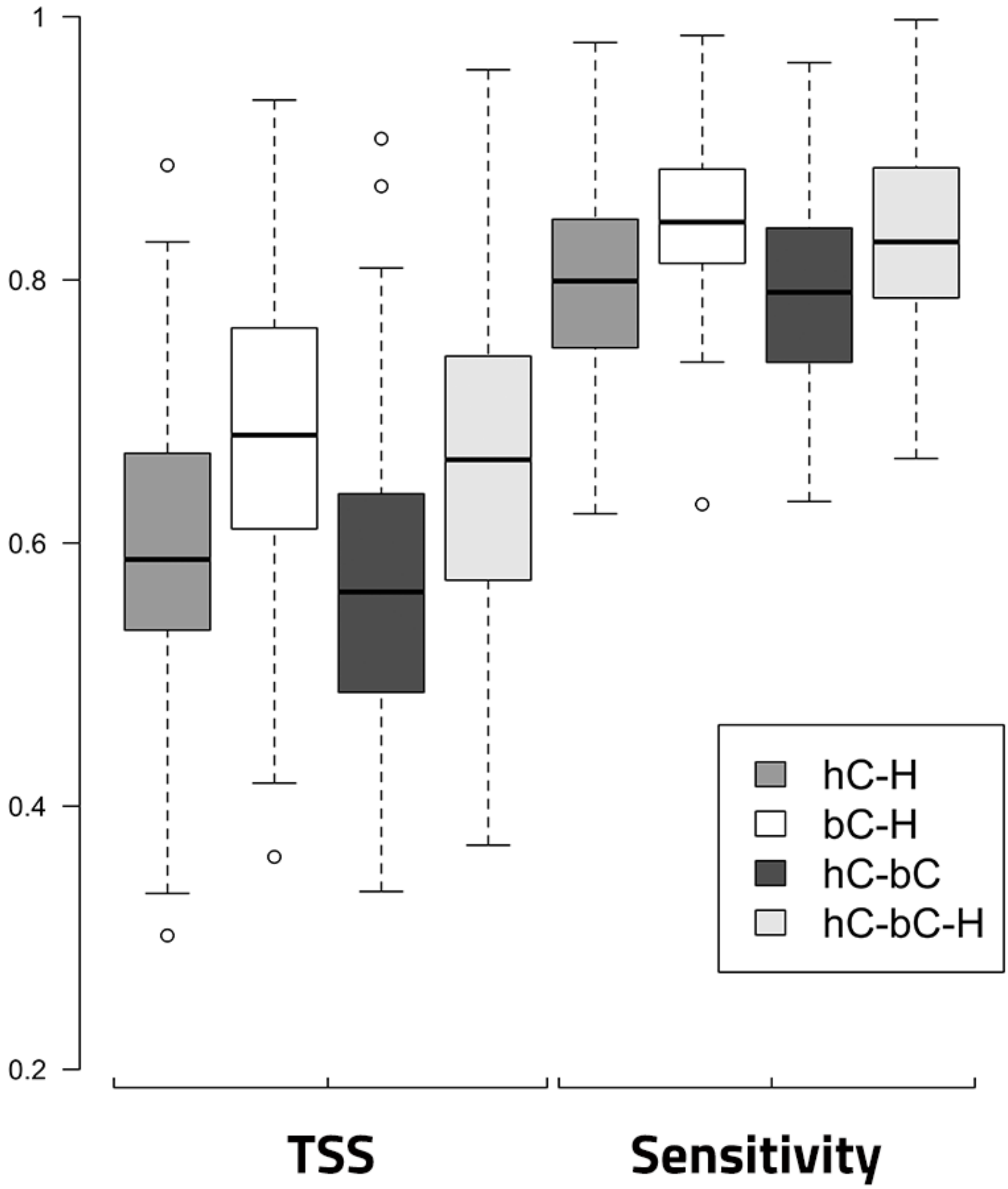


Figure 4.2: Comparison of TSS and Sensitivity values across 4 models (variable combinations): 1) hC-H; hydroclimate and hydrology, 2) bC-H; bioclimate & hydrology, 3) hC-bC; hydroclimate & bioclimate, 4) hC-bC-H; hydroclimate, bioclimate and hydrology. Boxplots (bar = median, box = IQR, whiskers = 1.5×IQR and outliers).

4.4 Results

4.4.1 Variable selection process

The variable selection process resulted in 10 predictors (Table S4.2 for BRT coefficients) applied in the full model configuration and distributed to each 2-set model (Table 4.1). Interestingly, variables describing mean temperature of both wettest quarter (Bio 08) and of driest quarter (Bio 09) from both the hydroclimate and the bioclimate predictor sets were included in the model (see Table 4.1). It could be expected that the same variable from both climate-related predictor sets would be highly correlated, however the correlation was negligible: Bio 08 from bC and hC, corr = 0.37, Bio 09 from bC and hC, corr =0.3 (Table S4.1). This furthers the notion that even though the datasets are derived from the same source, they contain different information describing distinctive aspects of the environment and therefore have a different influence on species' distribution.

The BRT selection process chose five bioclimate, three hydroclimate and two hydrology variables. This disproportionate number of variables from each predictor set will have an influence on variable importance ranking and variance partitioning.

4.4.2 Model performance

The SDMs performed well over all (mean ± se: hC-H; TSS = 0.59 ± 0.02, S = 0.80 ± 0.02, bC-H; TSS = 0.68 ± 0.02, S = 0.85 ± 0.02, hC-bC; TSS = 0.57 ± 0.01, S = 0.79 ± 0.02, hC-bC-H; TSS = 0.66 ± 0.01, S = 0.83 ± 0.02, Table S4.3, Figure 4.2). The Wilcoxon tests of difference showed that model configuration bC-H performed better overall (Figure 4.2). The percentage of significance for model configuration bC-H (66%, n=181) is greater than all remaining model configurations: hC-H (18.5%, n=51), hC-bC (12.7%, n=35) and hC-bC-H (49.6%, n=137). See Table 4.2 for pairwise totals of significantly better models/species and % of significance. Bioclimate & hydrology performs moderately better than the full model (hC-bC-H), i.e. significantly better for 38 species (41%). While this model produced the highest TSS values of all model configurations (Figure 4.2, max TSS = 0.96) and performed significantly better in 49.6% of pair wise comparisons, it performed better than bioclimate and hydrology (bC-H) for 25 species only (27%).

Table 4.2: Comparison of model configurations. Total number of species that performed significantly better ($p<0.05$) in pairwise Wilcox tests. Model bC-H outperforms the others. MC = model configuration (hC-H; hydroclimate and hydrology, bC-H; bioclimate & hydrology, hC-bC; hydroclimate & bioclimate, hC-bC-H; hydroclimate, bioclimate and hydrology). S = number of significantly better performing species (n=92). Overall % determined by number of significantly better performing species from total number of pairwise comparisons (n=276).

Pairwise Wilcox test MC1 vs. MC2		S	% significance
hC-H	bC-H	7	7.6
	hC-bC	31	33.7
	hC-bC-H	13	14.1
Total (n=276)		51	**Overall %** = 18.5
bC-H	hC-H	63	68.5
	hC-bC	80	87.0
	hC-bC-H	38	41.3
Total (n=276)		181	**Overall %** = 66.0
hC-bC	hC-H	26	27.2
	bC-H	3	3.3
	hC-bC-H	6	6.5
Total (n=276)		35	**Overall %** = 12.7
hC-bC-H	hC-H	57	62.0
	bC-H	25	27.2
	hC-bC	55	59.8
Total (n=276)		137	**Overall %** = 49.6

4.4.3 Variable importance ranking and variance partitioning

Bioclimate variables showed the highest variable importance across all models (Figure 4.3), considerably more than hydrology and hydroclimate in every model configuration: in model configuration bC-H, bioclimate variables showed relative importance of 73.6% compared with hydrology 26.4% and in hC-bC-H, bioclimate had relative importance of 51.4%, compared with hydrology 18.0% and hydroclimate 30.5%. For hC-bC, bioclimate had importance of 62% compared with hydroclimate 38%. Hydroclimate represented 61.2% of the variable importance of hC-H, compared with hydrology 38.8%.

We applied variance partitioning to the predicted distributions from the SDMs (presence/absence). Overall, bioclimate shows the highest explained variance over the whole community in the 3-set model configuration hC-bC-H (0.698 Figure 4.4d). Hydrology and hydroclimate are comparable to each other but show a lesser amount of explained variance, 0.435 and 0.467, respectively (Figure 4.4d), than bioclimate. The variance partitioning of the 2-set model configurations shows that bioclimate also had the highest explained variance compared with hydroclimate (0.409, Figure 4.4c) and hydrology (0.608, Figure 4.4b).

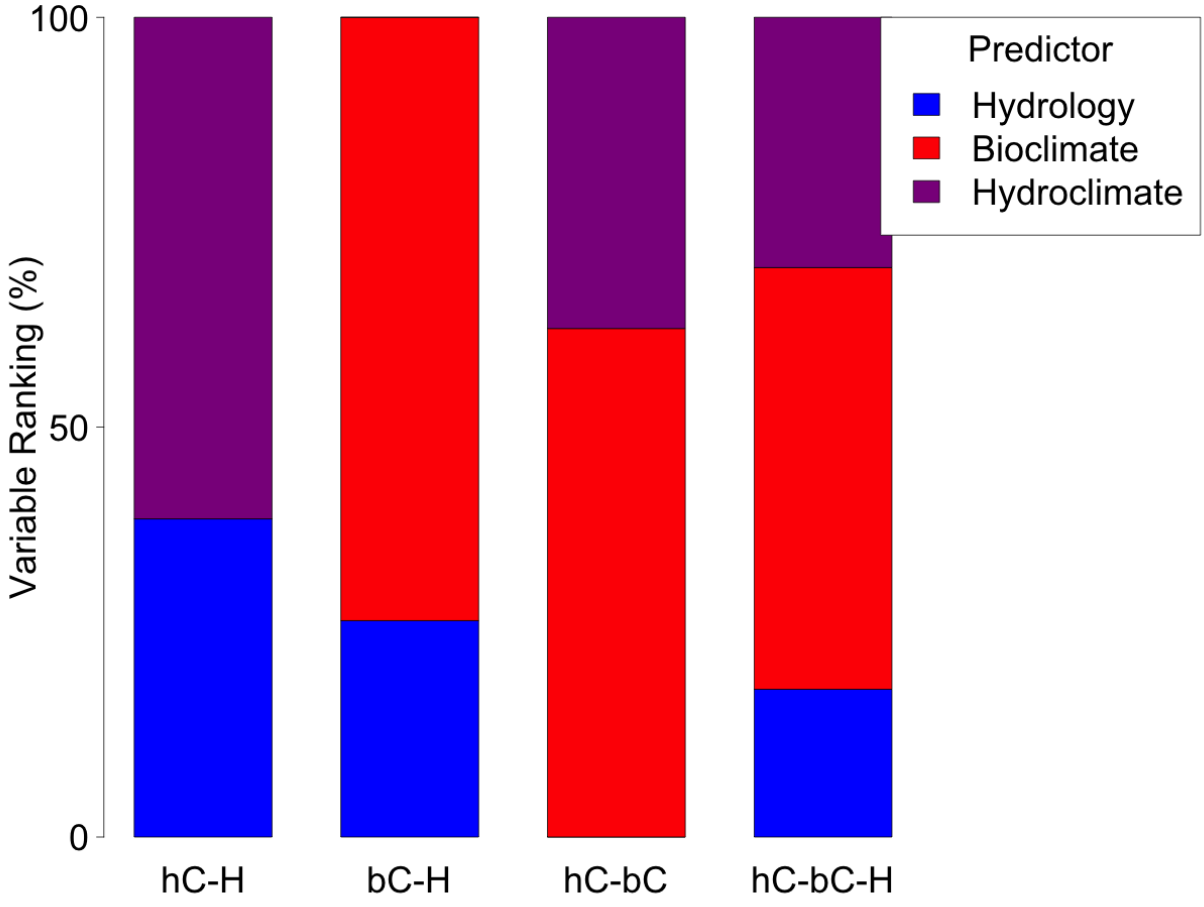


Figure 4.3: Comparison of variable importance across 4 models: 1) hC-H; hydroclimate and hydrology, 2) bC-H; bioclimate & hydrology, 3) hC-bC; hydroclimate & bioclimate, 4) hC-bC-H; hydroclimate, bioclimate and hydrology.

Hydrology shows the lowest amount of explained variance in all 2-set models, compared with hydroclimate (0.229, Figure 4.4a) and bioclimate (0.166, Figure 4.4b). Hydroclimate shows lower explained variance than bioclimate (0.142, Figure 4.4c) but higher explained variance than hydrology (0.457, Figure 4.4a).

The shared variance in all 2-set models is relatively low (Figure 4.4a-c). This demonstrates that the predictor sets have a separate influence on species' distribution. Model configurations hC-H and bC-H show a negative shared variance meaning that the two predictor sets explain the variance in different directions i.e. both a positive and negative relationship. The model configuration hC-bC-H shows a shared variance of 0.32 between all three predictor sets (Figure 4.4d). It is not possible to interpret the exact meaning of the shared variance from this analysis. It is potentially an interaction between datasets, yet without further analysis we cannot determine whether it is, for example, an amplified, additive, or an interacting effect at all.

Unexplained variance is lowest in bC-H model configuration (0.29, Figure 4.4b), compared with hC-H (0.38, Figure 4.4a) and hC-bC (0.36, Figure 4.4c). It is important to note that due to the explained variance here being proportional, the variance of the full model equals 1, and therefore, the unexplained variance is 0.

4.4.4 Predicted distributions

4.4.4.1 Range size

Range size is defined as the area of occurrence predicted by the model once the predicted probabilities had been converted to binary. Model configuration bC-H predicted on average, larger range sizes (Mean no. of presences; 3482.6 ± 129.1, Figure 4.5 & 4.6) than all other models (Mean no. of presences; hC-H; 2914.9 ± 156.1; hC-bC; 1916.4± 149.6, hC-bC-H; 2716.2 ± 168.9, Figure 4.5 & 4.6). Model configuration hC-bC predicted on average, smaller range sizes than all other configurations. All pairwise Wilcoxon tests showed a significant difference ($p< 0.0002$), with the exception of hC-H vs. hC-bC-H ($p=0.23$).

The smallest percentage range overlap was between model configurations hC-H and hC-bC (44.4 ± 1.8%, n = 1187, Table 4.3). The largest range overlap was between model configurations hC-bC and hC-bC-H (60.4 ± 2.3%, n=1501, Table 4.3). The differences in predicted range are clear when mapping individual species' distribution (species richness in Figure 4.5, single species example in Figure 4.6) in geographical space.

Table 4.3: Range overlap from predicted range size of model configurations. Lower left table is mean percentage overlap in geographical space across all species (n=92) ± standard error. Upper right table is absolute averaged number of overlapping predicted presences.

	HC-H	**BC-H**	**HC-BC**	**HC-BC-H**
HC-H		1638	1187	1532
BC-H	47.6 ± 1.6		1520	1973
HC-BC	44.4 ± 1.8	49.4 ± 2.4		1501
HC-BC-H	51.1 ± 1.8	59.1 ± 2.1	60.4 ± 2.3	

4.5 Discussion

We compared three datasets, combined in four dataset configurations, to evaluate their influence on macroinvertebrate distribution using SDMs. We found that the predictor set describing climate has the most influence on the model in terms of variable importance and proportional variance. The configuration combining bioclimate and hydrology (bC-H) performed best, while the hydroclimate and bioclimate (hC-bC) configuration performing least well. The model configurations that include the data describing hydrology only (bC-H, hC-H & bC-hC-H), predicted significantly larger range sizes. Additionally, the different model configurations did not always predict geographically analogous species' distributions.

4.5.1 Variable selection (BRTs)

The predictors mean temperature of wettest month (Bio08) and mean temperature of driest month (Bio09), were selected from both the hydroclimate and bioclimate datasets. Because these predictors were not highly correlated, their contribution in terms of environmental information can be assumed to be dissimilar and relevant for the community distribution. In addition to those predictors describing temperature, two precipitation variables are included in each model configuration i.e. annual precipitation (Bio12) is included from the hydroclimate dataset and precipitation seasonality (Bio15) from the bioclimate dataset. The inclusion of these variables indicates that precipitation is an important aspect to include in SDMs as a separate entity to hydrology, suggesting that local precipitation should not be used as a substitute for hydrology related variables (e.g. Domisch et al. 2019). The precipitation and hydrology data are also not correlated, despite being derived, in part, from one another, suggesting that the data contain, to a great extent, separate and equally justified information. This is supported by the low amount of shared variance demonstrated in all model configurations through variance partitioning.

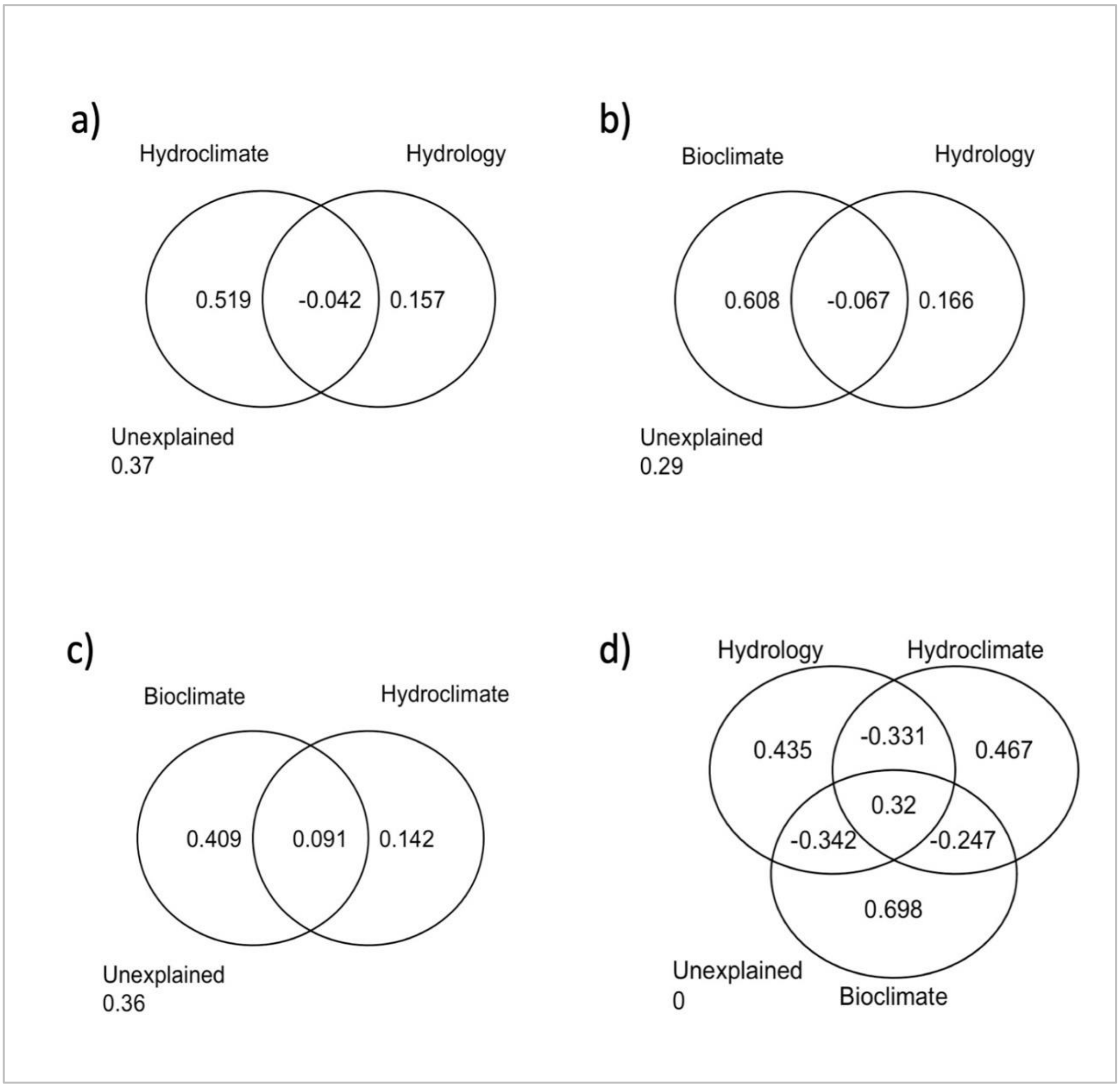


Figure 4.4: Proportional variance partitioning of all 4 models; a) hC-H; hydroclimate and hydrology, b) bC-H; bioclimate & hydrology, c) hC-bC; hydroclimate & bioclimate, d) hC-bC-H; hydroclimate, bioclimate and hydrology.

4.5.2 Model performance

The bioclimate and hydrology (bC-H) model configuration performed best overall. To a lesser extent, the full model (hC-bC-H) also performed well. It was expected that the full model would perform better overall as it contained the most information, i.e. 10 variables. Accordingly, we propose that SDMs should be applied on this community using bioclimate and hydrology (bC-H) at this spatial scale, because it outperformed others when taking into account the entire community. In addition, the increased performance was attained using 7 variables only, insuring model simplicity. We therefore argue that these two predictor sets complement each other well and that adding hydrology into the model configuration can improve model performance significantly.

The only model configuration that did not contain hydrology, i.e. hydroclimate and bioclimate (hC-bC), performed least well overall. The lower performance of this model configuration may be explained, in part, by the specifics of the hydroclimate data. Because hydroclimate is derived from modelled climate data (i.e. bioclimate) and flow accumulation, some important factors that control river hydrology are not incorporated (e.g. percolation or evapotranspiration). Hydroclimate summarizes rainfall information from the upper grid cells, therefore including a strong spatial relationship in the data. In contrast, the hydrology dataset is partially derived from discharge gauging stations, which measure real-time streamflow ($m^3 s^{-1}$), and therefore account, to some extent, for those important factors that control river hydrology, which are not covered by the hydroclimate data. Furthermore, the fact that bioclimate and hydrology data are far less related to each other than to hydroclimate, may explain their complementing nature and superior combined performance.

4.5.3 Variable importance & variance partitioning

Variable importance ranking and variance partitioning of the individual predictor sets both suggest that bioclimate has the most influence on the studied macroinvertebrate community, at this scale. While climate is an important factor, freshwater ecological literature indicates that hydrology can be expected to have a large influence on species' distribution. Our finding is not consistent with current flow-ecology theory that suggests hydrology to be as important as climate in determining species distribution (Pyne and Poff 2017). The scale at which the predictors are applied, may partially explain our results (Pearson and Dawson 2003, Randin et al. 2009, Lenoir et al. 2013, Domisch et al. 2015b, Record et al. 2018). Climate and geology are the drivers of broad aspects of hydrological regime, i.e. floods and droughts, at large scales (i.e. catchment scale) (Poff 1997). Whereas at small scales, i.e. reach scale, hydrological regimes are influenced by, e.g., hydraulics, riverbed substrates and stream channel morphology, to mention only a few (Allen and Vaughn 2010, Soranno et al. 2010). Nonetheless, large-scale variables describing hydrological regime, such as flood and droughts, are able to induce changes in river biota communities by influencing local-scale habitat. However, local-scale hydraulics, e.g. pool/riffles, and their resultant impact on physical habitat, e.g. creation of refugia, influence the distribution of biota, reducing the effect of large-scale drivers: a theory known as the "Landscape Filters Hypothesis" (Poff 1997). This scale-dependency is a recognized challenge in SDM research (Domisch et al. 2015b).

Despite this challenge, incorporating hydrology at this scale does not hinder predictive ability (Araújo et al. 2019) and resulted in an improvement in SDM performance as well as a significant impact on predicted range size. We do, however, suggest that applying the same predictors at a smaller spatial (e.g. < 2500km^2) and finer resolution (e.g. <100m^2) (Kuemmerlen et al. 2014) could result in hydrology demonstrating a higher variable importance, and hence influence on macroinvertebrate distribution. A smaller scale and finer resolution was not possible in this study due to the requirement of spatially analogous biological response data and environmental variables in SDMs, i.e., hydrology, bioclimate and hydroclimate at 1 km (Araújo et al. 2019).

4.5.4 Predicted distributions

The bioclimate and hydrology (bC-H) model configuration predicted the largest range size overall. Conversely, the hydroclimate and bioclimate (hC-bC) model predicted the smallest range size. Interestingly, linking range size to TSS values suggests that model performance has a positive relationship with range size. As far as we are aware, there is no established link connecting TSS values with predicted range size. In addition to different range size predictions, the model configurations are regularly predicting species' distributions in different geographical locations as suggested by the range overlap between model configurations.

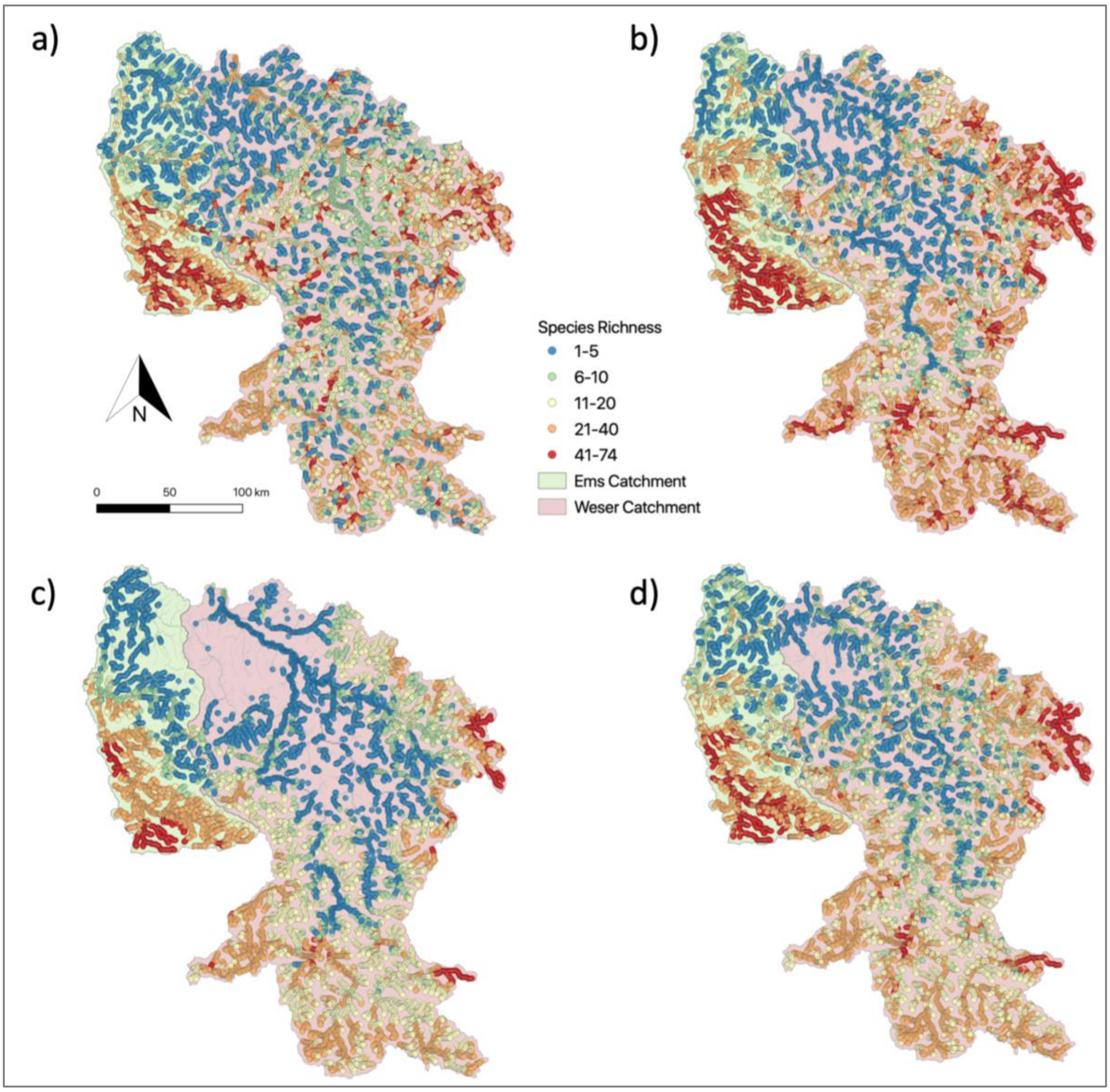


Figure 4.5: Distribution of all species predicted by each model: a) hC-H; hydroclimate and hydrology, b) bC-H; bioclimate & hydrology, c) hC-bC; hydroclimate & bioclimate, d) hC-bC-H; hydroclimate, bioclimate and hydrology. Points represent locations, colors represent number of species predicted presence at point locations.

As we discovered through the correlation analysis and variance partitioning, the datasets contain, to some extent, distinct environmental information, thus the differences in predicted range size and location are not surprising. Here, the SDMs are relating different environmental conditions to the species' known occurrences and predicting suitable habitat accordingly. The range sizes predicted by model configurations containing hydrology are significantly larger than the configuration that does not include hydrology, i.e. hydroclimate and bioclimate (hC-bC). We can only hypothesize on the reasons for this; however, it could be related to the specific information regarding flow regime contained in the hydrology dataset, delivered by physical hydrologic gauging stations used in its calculation.

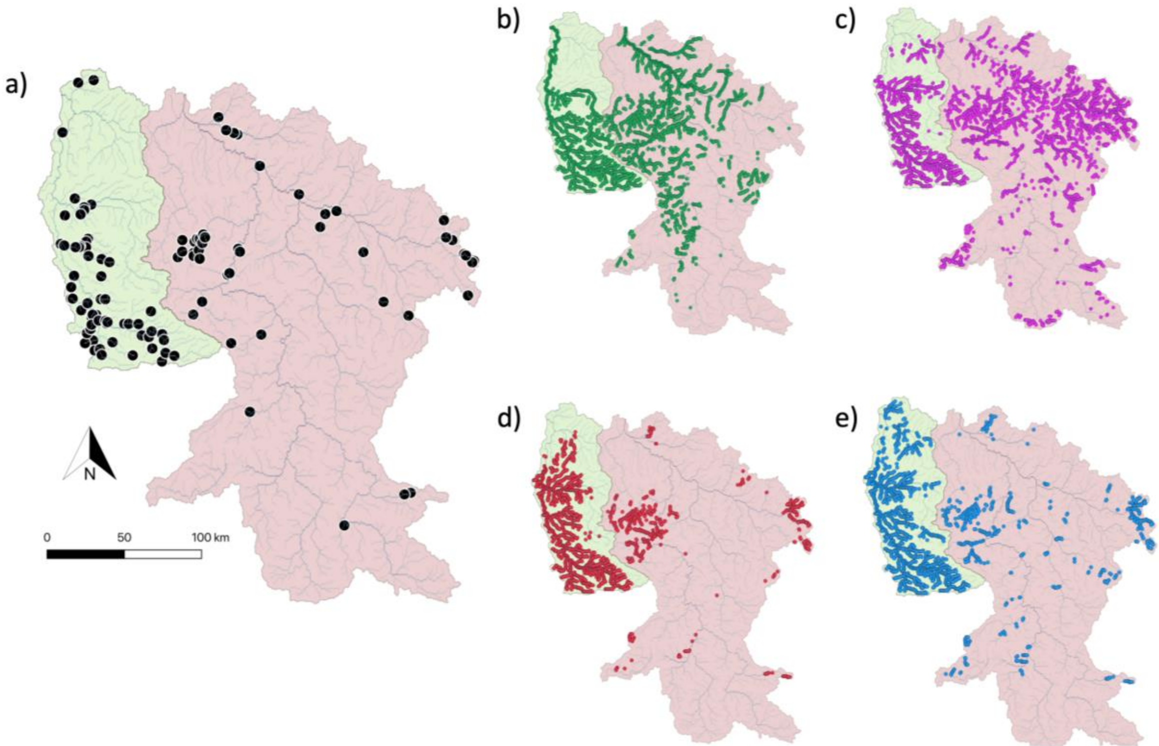


Figure 4.6: Example of geographical locations in study area of a) known occurrence (n=143) and predicted distribution of *Caenis horaria* by four model configurations; b) hC-H; hydroclimate and hydrology, c) bC-H; bioclimate & hydrology, d) hC-bC; hydroclimate & bioclimate, e) hC-bC-H; hydroclimate, bioclimate and hydrology.

The hydrologic variables (and the anthropogenic disturbance to hydrological regimes) are likely to better depict the varied influences on the physical habitat of the river biota (Resh et al. 1988, Poff et al. 1997), which shapes the structure and function of river ecosystems. In contrast, the climate datasets (bioclimate and hydroclimate) are applied as indirect surrogates for water temperature (Moore et al. 2005, Caissie 2006) and hydrology (e.g. Domisch et al. 2019). Hence, the models estimate expanded suitable habitat when applied with hydrological variables (Figure 4.5 & 4.6).

4.5.5 The role of hydrology in SDMs

Our findings have important implications when applying such models to inform conservation efforts: omitting flow regime variables in SDMs may lead to an underrepresentation of macroinvertebrate distributional range. For example, SDMs applied under current, and future, climate conditions have predicted range shifts of e.g. macroinvertebrate distribution to higher altitudes (Domisch et al 2011), several riverine taxonomic groups with regards to protected areas (Markovic et al 2014), freshwater fish (Ruiz-Navarro et al 2016) and chelonian species (Ihlow et al 2011), as well as to identify priority conservation areas with high aquatic plant diversity (Rodriguez-Merino et al 2019). These studies refer to changes in species' distributional range under changes in climate, not implemented predictors describing hydrology. Nonetheless, adding complementary factors describing flow regime may result in deviating species' predicted range, both for current and future potential distributions, subsequently adjusting mitigation strategies for conservation efforts.

4.6 Conclusion

Bioclimate was found to be the most important factor influencing macroinvertebrate distributions in the Ems and Weser basins, at this scale. The bioclimate and hydrology datasets appear to complement each other well compared to other configurations, as they show the highest model performance, largest range sizes, and the lowest unexplained variance. The model configuration including the both climate-related predictor sets, performed least well, and predicted smaller range sizes.

The main findings of our study suggest that by including environmental predictors describing flow regime, SDMs applied on macroinvertebrates can potentially; 1) increase model performance, and 2) expand predicted distribution, despite a low contribution of hydrology to explained variance and variable importance. The IHA metrics applied in this study are partially derived from real-time streamflow data from gauging stations, which incorporate an element of the principal factors, such as evapotranspiration and percolation that control river hydrology. In addition, the metrics describe direct influencing factors of river habitat, including some aspects of anthropogenic disturbance. These characteristics are not described by either of the climate-related predictors included in our study, therefore it is beneficial to include IHA metrics in SDMs on river species.

We are aware that scale-dependency is a limiting factor in our study. The scale (13,749 km^2) of our study area and its resolution (1 km grid cells), may be too coarse to fully capture the influence of all the dimensions of the hydrological regime on macroinvertebrate distribution. The IHA metrics chosen by the variable selection process, i.e. high flow volume (MH21) and the annual maxima of 1 day means of daily discharge (DH1) represent the broader aspects of hydrological regime, i.e. flooding, driven by climate and geology. These particular metrics may not comprehensively depict changes in hydrological variability that are important for macroinvertebrate communities on smaller scales, therefore, some of the flow regime metrics may not be relative to coarser measurements of bioclimate at the scale of our study area. Accordingly, it may be advisable to apply aggregated hydrologic variables, such as mean annual flow, for larger scale studies (Pyne and Poff 2017), however these types of variables may not describe sufficient information for studies directly investigating flow regime on species' distribution.

We conclude that that freshwater SDMs can profit from the inclusion of hydrological variables. Our study highlights improvements to predictive ability as well as how hydrological variables can influence the physical differences in predicted species' distribution. We recommend that given its fundamental importance, variables describing flow regime must be considered in SDM studies applied on river biota.

4.7 Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the project "Global Change Effects in River Ecosystems" (GLANCE, no. 01LN1320A) project. We thank the German Working Group on Water Issues of the Federal States and the Federal Government (LAWA) for providing biological data. For useful advice on the SDM package we would like to thank Babak Naimi.

4.8 References (Chapter 4)

Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution, and Systematics **35**:257-284.

Allen, D. C., and C. C. Vaughn. 2010. Complex hydraulic and substrate variables limit freshwater mussel species richness and abundance. Journal of the North American Benthological Society **29**:383-394, 312.

Araújo, M. B., D. Alagador, M. Cabeza, D. Nogués-Bravo, and W. Thuiller. 2011. Climate change threatens European conservation areas. Ecology Letters **14**:484-492.

Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, L. Maiorano, B. Naimi, R. B. O'Hara, N. E. Zimmermann, and C. Rahbek. 2019. Standards for distribution models in biodiversity assessments. Science Advances **5**:eaat4858.

Archfield, S. A., J. G. Kennen, D. M. Carlisle, and D. M. Wolock. 2014. An objective and parsimonious approach for classifying natural flow regimes at a continental scale. River Research and Applications **30**:1166-1183.

Barbarossa, V., M. A. J. Huijbregts, A. H. W. Beusen, H. E. Beck, H. King, and A. M. Schipper. 2018. FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. Scientific Data **5**:180052.

Barbosa, A. M., J. A. Brown, A. Jimenez-Valverde, and R. Real. 2016. modEvA: Model Evaluation and Analysis. R package version 1.3.2. https://CRAN.R-project.org/package=modEvA.

Caissie, D. 2006. The thermal regime of rivers: a review. Freshwater Biology **51**:1389-1406.

de Araujo, C. B., L. O. Marcondes-Machado, and G. C. Costa. 2014. The importance of biotic interactions in species distribution models: a test of the Eltonian noise hypothesis using parrots. Journal of Biogeography **41**:513-523.

Domisch, S., G. Amatulli, and W. Jetz. 2015a. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data **2**:150073.

Domisch, S., M. Friedrichs, T. Hein, F. Borgwardt, A. Wetzig, S. C. Jähnig, and S. D. Langhans. 2019. Spatially explicit species distribution models: A missed opportunity in conservation planning? Diversity and Distributions **25**:758-769.

Domisch, S., S. C. Jähnig, and P. Haase. 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshwater Biology **56**:2009-2020.

Domisch, S., S. C. Jahnig, J. P. Simaika, M. Kuemmerlen, and S. Stoll. 2015b. 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.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. G. Marquez, B. Gruber, B. Lafourcade, P. J. Leitao, T. Munkemuller, C. McClean, P. E. Osborne, B. Reineking, B. Schroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**:27-46.

Eaton, S., C. Ellis, D. Genney, R. Thompson, R. Yahr, and D. T. Haydon. 2018. Adding small species to the big picture: Species distribution modelling in an age of landscape scale conservation. Biological Conservation **217**:251-258.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802-813.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Gray, L. J. 1981. Species Composition and Life Histories of Aquatic Insects in a Lowland Sonoran Desert Stream. The American Midland Naturalist **106**:229-242.

Guisan, A., R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, T. G. Martin, J. R. Rhodes, R. Maggini, S. A. Setterfield, J. Elith, M. W. Schwartz, B. A. Wintle, O. Broennimann, M. Austin, S. Ferrier, M. R. Kearney, H. P. Possingham, and Y. M. Buckley. 2013. Predicting species distributions for conservation decisions. Ecology Letters **16**:1424-1435.

Haase, P., S. Lohse, S. Pauls, K. Schindehütte, A. Sundermann, P. Rolauffs, and D. Hering. 2004. Assessing streams in Germany with benthic invertebrates: development of a practical standardized protocol for macroinvertebrate sampling and sorting. Limnologica **34**:349-365.

Hancock, M. A., and S. E. Bunn. 1997. Population dynamics and life history of Paratya australiensis Kemp, 1917 (Decapoda: Atyidae) in upland rainforest streams, south-eastern Queensland, Australia. Marine and Freshwater Research **48**:361-369.

Henriksen, J. A., J. Heasley, J. G. Kennen, and S. Nieswand. 2006. Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools).72.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Ihlow, F., J. Dambach, J. O. Engler, M. Flecks, T. Hartmann, S. Nekum, H. Rajaei, and D. Rödder. 2012. On the brink of extinction? How climate change may affect global chelonian species richness and distribution. Global Change Biology **18**:1520-1530.

Illies, J. 1967. Limnofauna Europaea: Stuttgart 1967 ; Prospekt. Gustav Fisher Verlag, Stuttgart.

Irving, K., M. Kuemmerlen, J. Kiesel, K. Kakouei, S. Domisch, and S. C. Jähnig. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model. Scientific Data **5**:180224.

Irving, K., S. C, Jähnig & M. Kuemmerlen. 2019. Identifying and applying an optimum set of environmental variables in species distribution models, Inland Waters, DOI: 10.1080/20442041.2019.1653111

Jarvis, A., H.I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Kakouei, K., J. Kiesel, S. Domisch, K. S. Irving, S. C. Jähnig, and J. Kail. 2018. Projected effects of Climate-change-induced flow alterations on stream macroinvertebrate abundances. Ecology and Evolution **8**:3393-3409.

Kärcher, O., K. Frank, A. Walz, and D. Markovic. 2019. Scale effects on the performance of niche-based models of freshwater fish distributions. Ecological Modelling **405**:33-42.

Kuemmerlen, M., B. Schmalz, Q. Cai, P. Haase, N. Fohrer, and S. C. Jähnig. 2015a. An attack on two fronts: predicting how changes in land use and climate affect the distribution of stream macroinvertebrates. Freshwater Biology **60**:1443-1458.

Kuemmerlen, M., B. Schmalz, B. Guse, Q. Cai, N. Fohrer, and S. C. Jähnig. 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. Ecological Modelling **277**:77-86.

Kuemmerlen, M., S. Stoll, A. Sundermann, and P. Haase. 2015b. Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. Ecological Indicators. 65:122-132

Lehner, B., C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J. C. Robertson, R. Rödel, N. Sindorf, and D. Wisser. 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment **9**:494-502.

Lehner, B., K. Verdin, and A. Jarvis. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Transactions American Geophysical Union **89**:93-94.

Lenoir, J., B. J. Graae, P. A. Aarrestad, I. G. Alsos, W. S. Armbruster, G. Austrheim, C. Bergendorff, H. J. B. Birks, K. A. Bråthen, J. Brunet, H. H. Bruun, C. J. Dahlberg, G. Decocq, M. Diekmann, M. Dynesius, R. Ejrnæs, J.-A. Grytnes, K. Hylander, K. Klanderud, M. Luoto, A. Milbau, M. Moora, B. Nygaard, A. Odland, V. T. Ravolainen, S. Reinhardt, S. M. Sandvik, F. H. Schei, J. D. M. Speed, L. U. Tveraabak, V. Vandvik, L. G. Velle, R. Virtanen, M. Zobel, and J.-C. Svenning. 2013. Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across Northern Europe. Global Change Biology **19**:1470-1481.

Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28**:385-393.

Liu, C., M. White, and G. Newell. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography **40**:778-789.

Lytle, D. A. 2002. Flash floods and aquatic insect life-history evolution: evaluation of multiple models. Ecology **83**:370-385.

Lytle, D. A., and N. L. Poff. 2004. Adaptation to natural flow regimes. Trends in Ecology & Evolution **19**:94-100.

Malmqvist, B., and S. Rundle. 2002. Threats to the running water ecosystems of the world. Environmental Conservation **29**:134-153.

Mangiafico, S. 2019. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.1.1. https://CRAN.R-project.org/package=rcompanion.

Markovic, D., S. Carrizo, J. Freyhof, N. Cid, S. Lengyel, M. Scholz, H. Kasperdius, and W. Darwall. 2014. Europe's freshwater biodiversity under climate change: distribution shifts and conservation needs. Diversity and Distributions **20**:1097-1107.

Moore, R. D., P. Sutherland, T. Gomi, and A. Dhakal. 2005. Thermal regime of a headwater stream within a clear-cut, coastal British Columbia, Canada. Hydrological Processes **19**:2591-2608.

Naimi, B., and M. Araujo. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography **39**:368-375.

Olden, J. D., and N. L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications **19**:101-121.

Ouellet Dallaire, C., B. Lehner, R. Sayre, and M. Thieme. 2019. A multidisciplinary framework to derive global river reach classifications at high spatial resolution. Environmental Research Letters **14**:024003.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

Peckarsky, B. L., B. W. Taylor, and C. C. Caudill. 2000. Hydrologic and behavioral constraints on oviposition of stream insects: implications for adult dispersal. Oecologia **125**:186-200.

Peres, D. J., and A. Cancelliere. 2016. Environmental Flow Assessment Based on Different Metrics of Hydrological Alteration. Water Resources Management **30**:5799-5817.

Poff, N., J. Allan, M. Bain, J. Karr, K. Prestegaard, B. Richter, R. Sparks, and J. Stromberg. 1997. The natural flow regime. Bioscience **47**:769-784.

Poff, N. L. 1997. Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. Journal of the North American Benthological Society **16**:391-409.

Poff, N. L., B. D. Richter, A. H. Arthington, S. E. Bunn, R. J. Naiman, E. Kendy, M. Acreman, C. Apse, B. P. Bledsoe, M. C. Freeman, J. Henriksen, R. B. Jacobson, J. G. Kennen, D. M. Merritt, J. H. O'Keeffe, J. D. Olden, K. Rogers, R. E. Tharme, and A. Warner. 2010. The ecological limits of hydrologic alteration

(ELOHA): a new framework for developing regional environmental flow standards. Freshwater Biology **55**:147-170.

Power, M. E., A. Sun, G. Parker, W. E. Dietrich, and J. T. Wootton. 1995. Hydraulic Food-Chain Models. Bioscience **45**:159-167.

Pyne, M. I., and N. L. Poff. 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.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, https://www.R-project.org/.

Randin, C. F., R. Engler, S. Normand, M. Zappa, N. Zimmermann, P. B. Pearman, P. Vittoz, W. Thuiller, and A. Guisan. 2009. Climate change and plant distribution: local models predict high-elevation persistence. Global Change Biology**15**:1557-1569.

Real, R., D. Romero, J. Olivero, A. Estrada, and A. L. Márquez. 2013. Estimating How Inflated or Obscured Effects of Climate Affect Forecasted Species Distribution. PLoS ONE **8**:e53646.

Record, S., A. Strecker, M.-N. Tuanmu, L. Beaudrot, P. Zarnetske, J. Belmaker, and B. Gerstner. 2018. Does scale matter? A systematic review of incorporating biological realism when predicting changes in species distributions. PLoS ONE **13**:e0194650.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The Role of Disturbance in Stream Ecology. Journal of the North American Benthological Society **7**:433-455.

Rodríguez-Merino, A., R. Fernández-Zamudio, and P. García-Murillo. 2019. Identifying areas of aquatic plant richness in a Mediterranean hotspot to improve the conservation of freshwater ecosystems. Aquatic Conservation: Marine and Freshwater Ecosystems **29**:589-602.

Ruiz-Navarro, A., P. K. Gillingham, and J. R. Britton. 2016. Predicting shifts in the climate space of freshwater fishes in Great Britain due to climate change. Biological Conservation **203**:33-42.

Soranno, P. A., K. S. Cheruvelil, K. E. Webster, M. T. Bremigan, T. Wagner, and C. A. Stow. 2010. Using Landscape Limnology to Classify Freshwater Ecosystems for Multi-ecosystem Management and Conservation. Bioscience **60**:440-454.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: Variation at local, basin, and regional scales. Annual Review of Entomology **43**:271-293.

5 General discussion

Species distribution models (SDMs) are employed to analyze large scale changes in species distribution to, for example, aid conservation decision making. In order to find the most effective conservation solutions, understanding the distribution of species must coincide with determining their most influential environmental factors. It is widely accepted that the flow regime is one of the most important factors in determining species distribution and abundance in rivers (Resh et al. 1988, Poff et al. 1997, Bunn and Arthington 2002). Despite this, very few studies assess the influence of these variables on riverine species distributions because SDMs are limited in including such descriptive variables. This thesis deals with the role of hydrology in SDM predictions in general: developing a hydrologic dataset, optimizing variable selection and testing existing hydrological datasets. I provide some necessary steps to include these variables in order to improve prediction performance of riverine species' distribution.

5.1 Key research findings

I demonstrate in **Chapter 2**, that by applying a simple hydrological model it is possible to predict streamflow over a large spatial extent at a fine spatial resolution (1 km^2). The 53 Indicators of Hydrologic Alteration (IHA, Richter et al. 1996, Olden & Poff. 2003) metrics created from this procedure describe the various aspects of flow regime that are widely established in flow–ecology assessments, i.e. the duration, magnitude, timing, frequency, and rate of change of flow events. IHA metrics are well-established metrics for environmental flow assessments to aid management decisions (e.g. Peres and Cancelliere 2016) and are also applied in predictive modeling (e.g. Kakouei et al. 2018). The metrics were calculated from 64 years of modeled daily streamflow data for the entire stream network of Germany at 1 km spatial resolution. Both the IHA metrics and the streamflow datasets are available open access for use in predictive models. The modeling method was intentionally simple and exploits openly sourced accessible data thus, the model can be applied in alternative geographical regions or time periods (Irving et al. 2018). These datasets have bridged a recognized data gap in available high-resolution hydrological data and have been successfully applied in predictive modeling of river ecosystems.

Applying the data created in **Chapter 2**, together with data describing climate, land-use and topography in Boosted Regression Trees (BRTs), **Chapter 3** highlights a useful and effective method (Figure 5.1) to impartially select highly relevant environmental variables as an alternative to time consuming extensive literature reviews and collating expert knowledge of large communities. Applying a species-specific custom-made set of predictors is viable for uncommon, specialist or invasive species, whereas the common uniform predictor approach is appropriate for widespread, generalist species. The variables chosen by the variable selection process were validated through investigating species traits from online databases and cross referenced with the environmental conditions at the known occurrence sites. Variables describing bioclimate were, overall, the most important for this community. Nonetheless, variables describing hydrology were deemed important by the variable selection process, and differences in species groups, with regards to hydrological preferences, were validated through their known occurrence sites and ecological traits (Irving et al. 2019).

Chapter 4 investigated the influence of the IHA metrics created in **Chapter 2**, against datasets commonly used as surrogates to hydrology in predictive modeling. From the three datasets i.e. 1) bioclimate, 2) hydrology and 3) hydroclimate, bioclimate exhibited the most influence in terms of variable importance and variance partitioning. Hydrology exhibited the lowest variable importance ranking, against both hydroclimate and bioclimate. The best performing model configuration was the bioclimate and hydrology model, which also produced, on average, significantly larger range sizes than all model other configurations. An important finding was that even though the hydrology data explained a relatively low proportion of variance, the data influenced the predicted distribution and model performance significantly.

5.1.1 High resolution hydrological datasets

Using empirical streamflow data from gauging stations across Germany and modeled seasonal accumulated precipitation, I applied a weighted linear regression model to predict a continuous daily time series of streamflow in $m^3 s^{-1}$ spanning 64 years (1950-2013). This simple modeling approach requires only two components, which are openly available: 1) observed gauging data are available worldwide, at least in part, from the Global Runoff Data Centre (GRDC, 2016), 2) accumulative precipitation available at high resolution on a near global scale (Domisch et al. 2015a).

The daily streamflow data were subsequently applied as input to successfully calculate 53 IHA metrics (Olden and Poff 2003). Of these metrics, the indices that describe average flow conditions e.g. mean daily flows, were produced to a high level of accuracy. However, indices that rely on variability in flow, e.g. variability in daily flows, were not described sufficiently well by the model. Nonetheless, the successfully validated variables provide an adequate description of flow regime, e.g., predictability of flow, high flow volume, rise and fall rate. The IHA metrics were validated successfully through Generalized Linear Models (GLMs), a commonly used algorithm in ensemble SDMs (Thuiller et al. 2014, Naimi and Araujo 2016), on 32 macroinvertebrate species. Further, the metrics were applied as input into the SDMs of **Chapters 3 and 4.**

The weighted linear regression models did not perform as well in alpine regions as those applied in lowland regions. The difference in performance of the models can be explained, in part, to the hydrological processes of these regions. Lowland rivers are typically fed by ground water (Guse et al. 2014). Through soil infiltration and groundwater processes, streamflow has a slower response to precipitation. As I applied precipitation on seasonal resolution, the delayed response time is captured within the model. In contrast, alpine rivers distribute precipitation into streams much faster than in lowland regions through several influencing factors; complex topography, high altitude and steep hillslope, the periodic storage and melting of snow and glaciers, as well as highly varied precipitation patterns (Jansson et al. 2003, Warscher et al. 2013, Isotta et al. 2014). Accordingly, the model is limited for use in these complex environments. Streamflow data in alpine regions could benefit from applying higher temporal resolution input data, i.e. daily precipitation (e.g. DWD, www.esrl.noaa.gov), to capture the complexity and variability of the regions' intricate precipitation-run off patterns. Additionally, the higher variability in the streamflow data produced by higher resolution input data could result in an increased number of IHA metrics that describe flow variability being validated successfully, which could not be validated in **Chapter 2**.

The models applied in lowland regions produced low performing models on a small number of occasions. These low performing models coincide with an extreme flooding event in June 2013 (The German Federal Institute of Hydrology (BfG) 2013) as a result of heavy rainfall. Here, a number of gauges ceased to operate, reducing the amount of streamflow data available for the model. Additionally, the heavy rainfall of 2013 was not captured in the 50-year (1950-2000) average of the precipitation data. These restrictions in the data resulted in the model failing to converge, thus producing low performing models for that time period.

A number of negative flow values (0.99% of the total dataset produced, considered a negligible amount) were predicted on occasions when the linear model produced a negative intercept, which corresponds to the value of flow when the precipitation is equal to zero, i.e., a flow value of zero. These values are a result of the direct association of flow from gauging stations with modeled precipitation, without considering other hydrological processes such as (ground) water storage, evaporation and evapotranspiration from soil, and interception (Brutsaert 1982, Beven 2004, Kiesel et al. 2010), which reduces the amount of precipitation that physically converts to streamflow. If these factors were to be included (e.g. potential evapotranspiration at 1 km^2, cgiarcsi.org) the model may predict zero flows more accurately, reducing the number of negative flow values. However, by including the necessary information to describe such relationships, the model would significantly increase in complexity, thus would contradict my goal of applying a simple modeling approach.

5.1.2 Variable selection method and species-specific predictor sets

SDMs predicting large communities tend to use the same set of predictors for the entire community (Markovic et al. 2012, Kuemmerlen et al. 2015, Domisch et al. 2019). Since individual species have adapted and evolved to respond differently to environmental conditions (Cox and Rutherford 2000, Lytle and Poff 2004, Fenoglio et al. 2007, Kroll et al. 2017), this uniform approach may affect model performance, as the variables chosen may not be relevant to a subset of species within the community. In **Chapter 3,** I proposed a variable selection process using BRTs to select the optimum set of predictors from four predictor categories (climate, hydrology, land use and topography, Figure 5.1). I tested the common uniform approach against a species-specific custom approach in SDMs and demonstrated how variable selection approach can impact different guilds of macroinvertebrates species.

The SDMs applied on 10 species increased in accuracy (Mean TSS = 0.59 ± 0.03) and the models applied on 10 species decreased in accuracy (Mean TSS =0.49 ± 0.04) with the species-specific predictor set. The 20 species, separated into an increased group and decreased group, showed distinct differences in terms of their ecological traits, known occurrences and preferred environmental conditions.

The increased species group were typically lowland species. Their associated environmental conditions include a higher coverage of agriculture, urban and barren land, higher stream-related variables; high flow volume and mean monthly flow, while also exhibiting a lower coverage of forested area than the decreased group sites. These conditions correspond to the species' downstream zonation preferences and high representation of functional group "gatherers", an indication of downstream zone preference (River Continuum Concept, Vannote et al. 1980). The increased group also included two invasive species *Gammarus tigrinus* (Sexton, 1939) and *Potamothrix moldaviensis* (Vejdovský & Mrazek, 1902). As some invasive macroinvertebrates, including *G.tigrinus*, have been documented to have a more specialized ecological niche than their native counterparts (Herkül et al. 2016), I suggest that these species show more specialist preferences. The decreased species traits also coincide with their environmental conditions and known occurrence. For example, the species prefer faster flowing water (rheophilic), which corresponds to the steeper hill slope, and hence high flow velocity (Austin 2007, Domisch et al. 2011) of their occurrence sites.

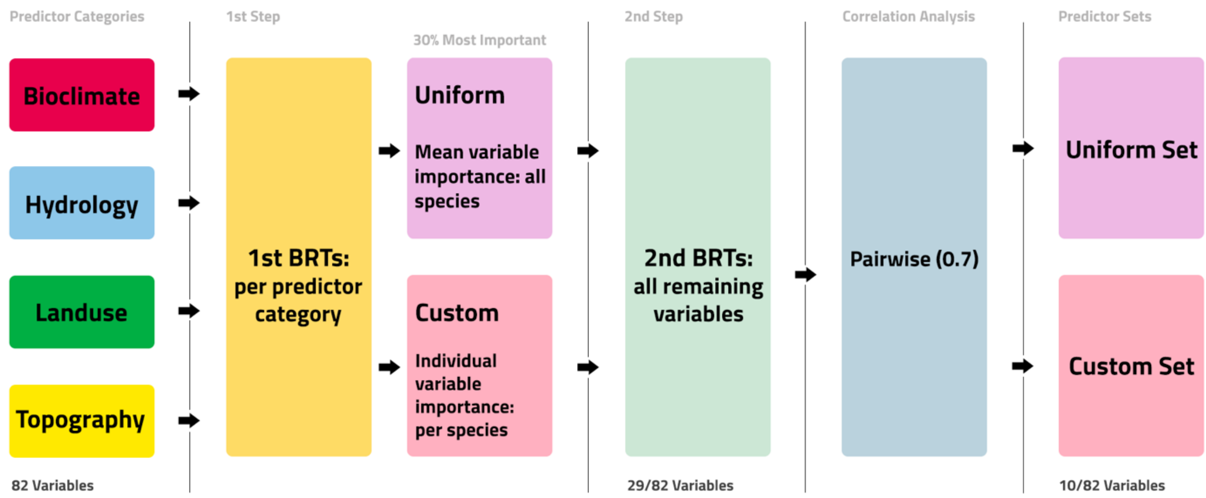


Figure 5.1: Schematic overview of the 2-step variable selection process using Boosted Regression Trees (BRTs) outlined in **Chapter 3.**

Chapter 3 produced 20 significant results from a total of 67 species. The remaining 47 species showed no change when applied with a custom or uniform set of predictors, even though all custom predictor sets varied, at least slightly, from the uniform set. Thus, these 47 species may be successfully modeled with any of the predictors in the study and show no difference in model performance. However, if variables describing, for example biotic information/limits, were included, it is possible that these species may show improvements in model performance. It would be interesting to investigate the ecological traits of the 47 species, however the focus of **Chapter 3** was to compare the significant differences between the custom and the uniform predictor sets.

In terms of variable importance, bioclimate was overall the most important predictor category for the entire community and for each model type (uniform and custom). Some of the variables that were overlooked by the uniform predictor set, resulted in a redistribution of variable importance in the increased group, which appeared to be the most important variables for some species in the custom model i.e. max temperature of warmest month & predictability of flow.

The hydrology predictor category showed a relatively low variable importance ranking in the study. This low ranking is possibly due to accumulative nature of the bioclimate data, which uses flow accumulation to route the information through the stream network, incorporating all information from the upper sub catchment. Flow accumulation is highly correlated with streamflow (Kuemmerlen et al. 2014), therefore the bioclimate data contains information describing aspects of both climate and hydrology. It may be possible that bioclimate is dominating the influence of hydrology and/or correlating highly, resulting in hydrology variables being removed during the correlation analysis. The disentangling of the bioclimate data to determine the separate influence of the variables, climate and hydrology, was investigated in **Chapter 4**.

5.1.3 The role of hydrology in SDMs

Although climate is a dominant factor of riverine species' distribution, it is important to include variables describing flow regime in SDMs (Collins and McIntyre 2015). However, many river SDM studies primarily focus on climate (e.g. Markovic et al. 2012, Kärcher et al. 2019) and interpret precipitation as hydrological influence (e.g. Domisch et al. 2019)..

Implementing the IHA metrics created in **Chapter 2** and applying the variable selection method outlined in **Chapter 3** (Figure 5.1), I assessed three hydrology or hydrology-related datasets in SDMs applied in four model configurations; 1) hydroclimate & hydrology, 2) bioclimate & hydrology, 3) hydroclimate & bioclimate, 4) hydroclimate, bioclimate & hydrology. The hydroclimate variables referred to in **Chapter 4**, are the same accumulated climate data applied in **Chapter 3** as 'bioclimate'. However, it was necessary to distinguish between the two climate related datasets applied in **Chapter 4**, i.e., bioclimate and hydroclimate. **Chapter 4** reflects that climate is the principal predictor set in terms of variable importance and proportional variance; however, the relatively low proportion of variance explained by hydrology can significantly alter species' predicted distributions

The best performing model configuration was bioclimate and hydrology, which also predicted the largest distributional range size. Model configuration bioclimate and hydroclimate performed least well and predicted the smallest distributional range size. These differences could be explained by the information contained within each dataset. Hydroclimate describes a spatial aspect of river discharge, i.e. accumulating rainfall, however it is originally derived from the modeled bioclimate data and flow accumulation, which do not incorporate specific aspects of hydrology, e.g. percolation and evapotranspiration. In contrast, the hydrology dataset is partially derived from physical discharge gauging stations, which measure real-time streamflow in $m^3\ s^{-1}$ (**Chapter 2**, Irving et al. 2018) and may account for some additional stream-related elements (not contained in the hydroclimate data), resulting in higher model performance. Furthermore, the bioclimate and hydrology are far less related to one another, which may explain their complementing nature and high performance. The empirical streamflow data contribute directly-related aspects of hydrological regimes, i.e. "proximal variables" (Araújo et al. 2019) to the hydrological variables, which are likely to better describe the varied influences on the physical habitat of the river biota (Resh et al. 1988, Poff et al. 1997), potentially resulting in larger predicted range sizes. On the contrary, both hydroclimate and bioclimate describe indirect factors influencing stream ecosystems, in the form of surrogate information. Specifically, air temperature is applied as a proxy for water temperature (Moore et al. 2005, Caissie 2006) and precipitation a proxy for hydrology (e.g. Domisch et al. 2019).

The main goal of **Chapter 4** was to include flow regime variables into SDMs and compare with hydrology-related datasets that are commonly applied as descriptors of the stream ecosystem. For this reason, additional variables known to be important to the stream environment, such as topography of the landscape and catchment land use (**Chapter 3**, Moore et al. 1991, Lake 2000, Schmalz et al. 2015, Amatulli et al. 2018) were omitted from the study. The results of **Chapter 4** demonstrate that including hydrology impacts predicted distribution; however, for a fuller representation of the stream ecosystem addition stream-related factors must be included.

5.2 Potential applications

The hydrological datasets created in **Chapter 2** can be used effectively in SDMs and other predictive modeling studies. The data are readily available for Germany (https://doi.org/10.6084/m9.figshare.c.3906376, Irving et al. 2018) at 1 km^2 spatial resolution covering 64 years (1950-2013). In addition, R scripts are available to; first, create daily streamflow from downloadable openly sourced data, in alternative geographic regions or time periods, which can be used as input to, second, create all or a subset of IHA metrics (https://github.com/ksirving/stream_flow). The daily flow data could also be aggregated to create descriptive flow metrics such as mean annual flow, or minimum/maximum monthly flow values.

The IHA metrics are based on Hydrosheds 1 km gridded stream network, so can be used simultaneously and without scaling, together with climate data e.g. WorldClim (Hijmans et al. 2005) and earthenv.org/streams (Jarvis et al. 2008, Lehner et al. 2008, Domisch et al. 2015a), land use and topography (Domisch et al. 2015a, Amatulli et al. 2018) as well as Hydrosheds products (e.g. river classification, Ouellet Dallaire et al. 2019). All of which are commonly used in SDMs. In addition to predictive modeling, depending on the applicability of reference sites, the metrics can be used in environmental flows assessment such as the ELOHA framework (Poff et al. 2010), in analysis of flow alteration. Stream flow is calculated from historical gauging data; however, it could also be used in conjunction with future projections of stream flow from e.g. SWAT models (Arnold et al. 1998), as a method of extrapolation under future scenarios. The method could be applied instead of the common method of pairing sampling sites (e.g. Domisch et al. 2017, Kakouei et al. 2018) with gauging sites.

The BRT variable selection process outlined in **Chapter 3** has been used successfully in other of predictive modeling studies (Record et al. 2013, Tonkin et al. 2015) and can be applied in any ecosystem. The BRT method is a highly beneficial method to employ in situations where expert knowledge is limited for a large community and can impartially select a predictor set, without bias, by, e.g., forcing a specific number of variables from each category (Kuemmerlen et al. 2015).

The custom-made predictor set can be applied to evaluate species which are less common in the study area, specialist or invasive species (**Chapter 3**). The ecological niche of invasive species could be well known in their native range however, their non-native range is relatively unfamiliar. Therefore, it can be difficult to fully identify the extent of which invasive species adapt to their new environment, and hence, the environmental conditions that are most influential. The custom-made predictor approach could be an informative way to identify key problematic areas where invasions may take place and provide ideas on how to mitigate the invasion and conserve the native species.

My findings from **Chapter 4** potentially have important implications when applying SDMs to inform river conservation efforts, since omitting flow regime variables may lead to an underrepresentation of macroinvertebrate distributional range. For example, SDMs applied under current, and future, climate conditions have investigated; macroinvertebrate (Domisch et al. 2013), freshwater fish (Markovic et al. 2012) and chelonian (Ihlow et al. 2012) range shifts, as well as the assessment of protected and conservation areas (Ruiz-Navarro et al. 2016, Rodríguez-Merino et al. 2019). While these studies refer to changes in species' distributional range under changes in climate, adding complementary factors describing flow regime may result in deviations to predicted range, subsequently adjusting mitigation strategies for conservation efforts.

The IHA metric high flow volume (**Chapter 2**) was deemed important to the macroinvertebrate community in **Chapters 3 & 4.** This metric is a measure (days) describing high magnitude flow events (Olden and Poff 2003). High flow volume could therefore be a useful predictor to include when investigating extreme aspects of flow regime on macroinvertebrate communities at catchment scale.

5.3 Limitations of the SDM methodology

The main goal of this thesis was to improve macroinvertebrate predicted distributions by including variables describing flow regime into SDMs. The IHA metrics (**Chapter 2**) showed a relatively low influence in terms of variable importance (**Chapters 3 & 4**) and explained variance (**Chapter 4**) on macroinvertebrate distribution. This could be explained by the scale at which the predictors were applied (Pearson and Dawson 2003, Randin et al. 2009, Lenoir et al. 2013, Domisch et al. 2015b, Record et al. 2018). On large scales, such as the catchment scale of these studies, climate and geology are the main drivers of the broad aspects of hydrology, such as floods and droughts (Poff 1997). These variables are able to depict changes in river biota communities by influencing local-scale habitat; nevertheless, their effect is reduced by local-scale characteristics, e.g. hydraulics (pool/riffles), and their resultant impact on physical habitat, e.g. creation of refugia. Therefore, on small scales, i.e. reach scale; hydrological regimes are influenced by, e.g., hydraulics, substrate and lateral characteristics such as stream width (Allen and Vaughn 2010, Soranno et al. 2010). This effect is pertinent to the "Landscape Filters Hypothesis" (Poff 1997) and is a recognized challenge in SDM research (Domisch et al. 2015b). Therefore, the scale of my study area (13,749 km^2) and its resolution (1 km^2 grid cells) may be too coarse to fully capture the influence of all the dimensions of the hydrological regime on macroinvertebrate distribution. Applying IHA metrics on a smaller scale (e.g. < 2500km^2) and finer resolution (e.g. <100m^2; Kuemmerlen et al. 2014), may result in hydrology demonstrating higher variable importance and explained variance, hence impacting predicted distribution. SDMs require spatially analogous environmental predictor and biological response variables (Araújo et al. 2019) therefore, it was not possible to apply the hydrology variables on a finer resolution due to the 1 km grid cell resolution of bioclimate, hydroclimate, and species' occurrence data assigned onto the base layer stream network.

Despite the challenge of scale dependency in **Chapters 3 & 4,** I demonstrate in **Chapter 4** that applying hydrology at this scale does not reduce predictive ability (Araújo et al. 2019) and in fact improves SDM performance as well as induces a significant impact on predicted range size. Therefore, given their fundamental importance, the IHA metrics should be considered in river SDMs.

BRTs were applied as the variable selection process in **Chapter 3 & 4** because they have several advantages in their application which include; robustness to co-linearity of variables, their ability to handle outliers as well as different units of measurements without having to standardize data (Friedman 2001, Elith et al. 2008). However, there are some limitations. For example, due to the stochastic nature of the algorithm it is possible that the BRTs would produce different results with each model run (Elith et al. 2008). The cross-validation process applied through R package 'dismo' is a method of mitigating this issue so the outcome may only differ slightly (Elith et al. 2008, Elith and Leathwick 2017, Hijmans et al. 2017). In addition, the predictors chosen by the variable selection process, may be different than those shown to be important by the SDM analysis. This is due to the differing ways in which, 1) the individual algorithms are processed, e.g. the Generalized Linear Models (GLMs) used in the ensemble SDMs do not include the element of stochasticity in BRTs and may not be as efficient in handling outliers (Elith et al. 2008), 2) the methods differ in their calculation of variable importance from the BRTs in the standalone algorithm (number of times the variable is chosen, Elith et al. 2008), and the BRTs as part of the ensemble SDM (through correlation, Naimi and Araujo 2016).

The biological data for **Chapters 2, 3 & 4** were collated from sixteen separate German federal state agencies. The sampling method was standardized following the procedure outlined in (Haase et al. 2004). However, the quality and quantity of biological data is a perpetual concern in any predictive modeling study. Potential challenges include; human error in the identification of species (Haase et al. 2010) as well as sampling bias in geographical space. These challenges may introduce some uncertainty within the model (Miller et al. 2012, Araújo et al. 2019). My research is methodological based; thus, I did not formulate conclusions to, e.g., identify specific areas of conservation, therefore, uncertainty regarding biological data concerns were not considered. Nonetheless, this uncertainty would have to be dealt with appropriately (Chapman 2005) before SDMs are applied in empirical investigations.

5.4 Further Opportunities in river SDM research

Despite the improvements made through this thesis, SDM procedures for riverine macroinvertebrates are still limited. Including variables that describe biotic conditions would retract the predicted fundamental niche into the more ecologically conclusive 'realized' niche (Austin et al. 1990, Guisan and Zimmermann 2000, Pearson and Dawson 2003). In **Chapter 3,** I was able to validate the environmental conditions of species known occurrence sites with their ecological traits. Ecological traits of macroinvertebrates have been used numerous times in river ecology studies (Poff and Zimmerman 2010 and references therein). Fundamentally, it is the traits of a species that determines its distribution, due to their adaptations in terms of, for example; phenology, life history, and growth, in response to changes in environmental conditions (Arthur et al. 1982, Poff and Zimmerman 2010, Alba-Tercedor et al. 2017). Consequently, a further opportunity to improve SDMs would be to include ecological traits when building the model to aid predictive accuracy.

Species' ecological traits could be applied in several forms. Phenology, how a life cycle of a species is associated with seasonal changes in environmental conditions, is known to be an important factor controlling species abundance and distribution (Verberk et al. 2008, Porst et al. 2012). Because macroinvertebrates depend on flow regime for some, or all of their life (Lytle and Poff 2004), phenological traits such as timing of reproduction periods and embryonic development or emergence (Gray 1981, Arthur et al. 1982, Hancock and Bunn 1997, Peckarsky et al. 2000, Lytle 2002, Haidekker and Hering 2008) should be taken into consideration when predicting distribution of species under climate change, and climate-induced changes in flow regime. Recent attempts to include phenology in terrestrial SDMs have been made (Smeraldo et al. 2018).

Nonetheless, because SDMs are a static snapshot in time, incorporating phenology is no easy task, although recent work has been done to attempt to include temporal variability (e.g. Devisser et al. 2010). Additionally, the IHA metrics created in **Chapter 2** describe the timing, frequency, duration and rate of change of flow events (Olden and Poff 2003), relate, to an extent, to seasonal flow changes, therefore could be utilized in SDMs developed to incorporate phenological traits.

Biotic interactions have been successfully introduced into terrestrial SDMs (Guisan and Thuiller 2005, Araujo and Luoto 2007, de Araujo et al. 2014), however it is only recently that attempts have been made in the aquatic realm. For example, the predicted occurrence, or species richness, of prey species can be applied as a predictor variable (Pletterbauer et al. 2016, Gherghel et al. 2018). In addition, Joint SDMs (JSDMs) can incorporate the co-occurrence of multiple species (D'Amen et al. 2018, Zurell et al. 2018). In addition to negative interactions, exploring positive biotic interactions such as the mutualistic relationship of *Potamothrix moldaviensis* and *Tubifex tubifex* that feed on fecal bacteria of one another (Milbrink 1993) would be an interesting avenue of research. The occurrence of one mutual species could be included as a predictor variable in an SDM, or both species' co-occurrence applied in a JSDM.

Species' dispersal also has a pronounced influence on species distribution and is one of the most important factors known to retract the fundamental niche into a realized niche (Austin et al. 1990, Guisan and Zimmermann 2000, Pearson and Dawson 2003). Methods to include dispersal are becoming established in the terrestrial realm (Bateman et al. 2013). In terms of freshwaters or rivers, quantitative dispersal abilities for fish have been documented (Radinger et al. 2014, Radinger and Wolter 2014, Comte and Olden 2018) and could be included either as a predictor variable or during post processing to retract a species' range. Dispersal ability for some macroinvertebrate groups have been proposed using functional traits and life cycle information (Usseglio-Polatera et al. 2000, Poff et al. 2006). Macroinvertebrates exhibit varied dispersal pathways, e.g. passive or active, aquatic or aerial (David T. Bilton et al. 2001, Graf et al. 2008, Graf et al. 2009, Li et al. 2018). Metrics based on these pathways (e.g. Li et al. 2016) would be highly beneficial to incorporate into SDMs.

5.5 Conclusion & outlook

Through this thesis, I have created, tested and integrated hydrological variables in SDMs, as well as developed and validated effective methods to improve prediction performance of riverine species' distribution to advance freshwater SDM research. Based on my results, I fully advocate the integration of hydrological variables into freshwater SDMs and believe this area of research should be further pursued.

The method applied in **Chapter 2** has bridged a recognized gap concerning the availability of high resolution, large-scale hydrological data and are well suited for use in predictive ecological modeling (**Chapter 3 & 4**). The 53 validated IHA metrics provide an adequate description of flow regime, e.g., predictability of flow, high flow volume, rise and fall rate. As the model was simple, it has some limitations, i.e. capturing the variability of flow and limited application in complex regions. Nonetheless, the straight-forward modeling procedure and low data requirements render it an uncomplicated method to apply in alternative geographical regions and time periods.

My application of BRTs (**Chapter 3**) outlines a beneficial and impartial variable selection method that can be applied in any system. It was my intent, to provide a relatively straight-forward procedure that could be implemented by ecology practitioners. It is most useful for communities with a large number of individual species, where it would be tedious and time consuming to collect knowledge of their individual ecological tendencies. The species-specific variable selection approach (**Chapter 3**) could help gain insights into potential species' invasions as well as the ecological requirements of specialist species of interest, where species' habitat preferences may be poorly understood. Furthermore, I have confirmed the traditional "uniform" predictor approach for a widespread species, for which I have shown that the community-wide predictor choice (selected by BRTs) corresponds to the ecological preferences of these species.

The IHA metrics created and applied in this thesis, should be included in SDMs for riverine species, as they include information on the hydrological regime directly related to stream habitat, which has a substantial impact on predicted distributions (**Chapter 4**). As an example, high flow volume captures an aspect of extreme flow events, which is shown to be particularly important for riverine benthic macroinvertebrates at catchment scale

(**Chapters 3 & 4**) as it represents the broader aspects of hydrological regime, i.e. flooding, driven by climate and geology. These flow-related disturbances change the physical habitat of the stream bed which impacts benthic macroinvertebrate assemblage (Statzner et al. 1988, Lake 2000), and hence, distribution. SDMs applied with this metric could help gain insights into the relationship between species distribution and flow disturbances.

Based on my results, my recommendations are twofold. First, my developed methods (**Chapter 2** & **3**) can be used as-is by water managers & practitioners, as well as students and academics interested in species' distribution. The hydrological data created in **Chapter 2**, is readily available for Germany, and is supplied with R scripts (github.com/ksirving/stream_flow) for straight forward application. Additionally, R scripts are also provided for the BRT variable selection procedure (**Chapter 2**, github.com/ksirving/variable_selection). I encourage water managers to consider investigating species traits in both current and future distribution analyses to ensure the predicted habitat suitability coincides with species ecological preferences. I also encourage the development and expansion of online trait databases. Species trait data are limited for some taxon and geographic areas (Naeem and Bunker 2009), however, overall, these databases provide vital information to incorporate into species' distribution analysis.

Second, to advance this research from an academic perspective, the hydrological data created in **Chapter 2** can be improved to incorporate more complex regions such as alpine. In addition, the data could be modeled on various spatial resolutions (e.g. $90m^2$) to depict the multi-scale influence of various environmental drivers of river systems (Domisch et al. 2015, Kärcher et al. 2019) and, potentially, capture the variability in flow that could not be validated in **Chapter 2.** In addition, this data could also be modeled on a global scale: with recent technological advances and increased accessible computing power, this objective is eminently achievable. In terms of SDM methodology, I encourage the further development of SDM methods to include biotic factors, e.g. interactions, dispersal and ecological traits with an applied-science perspective for straight-forward implementation by water managers and practitioners. The application of such advances would indeed be complex; however, by providing user-friendly tools together with practical guidelines, implementation outside academia would be feasible.

The recent Living Planet Report identified an 83% decline in global freshwater biodiversity (WWF 2018). This current status, added to the growing threat of anthropogenic global change, such as future climate change, dam expansion and land use change (Meybeck 2003, Vorosmarty et al. 2010), determines freshwater ecosystems as some of the most vulnerable ecosystems on earth (Dudgeon et al. 2006). To reliably estimate future instabilities, it is of utmost importance to further develop modeling methodology and exploit the predictive power of prognostic analyses such as SDMs. By building on existing methods, we can further increase our theoretical knowledge and understanding of the ecological drivers of species' distribution. Such advancements have the power to advise water managers in formulating wise mitigation and conservation decisions.

5.6 References (Chapter 5)

Alba-Tercedor, J., M. Sáinz-Bariáin, J. M. Poquet, and R. Rodríguez-López. 2017. Predicting River Macroinvertebrate Communities Distributional Shifts under Future Global Change Scenarios in the Spanish Mediterranean Area. PLoS ONE **12**:e0167904.

Allen, D. C., and C. C. Vaughn. 2010. Complex hydraulic and substrate variables limit freshwater mussel species richness and abundance. Journal of the North American Benthological Society **29**:383-394, 312.

Amatulli, G., S. Domisch, M.-N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, and W. Jetz. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific Data **5**:180040.

Araújo, M. B., R. P. Anderson, A. Márcia Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, L. Maiorano, B. Naimi, R. B. O'Hara, N. E. Zimmermann, and C. Rahbek. 2019. Standards for distribution models in biodiversity assessments. Science Advances **5**:eaat4858.

Araujo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography **16**:743-753.

Araújo, M. B., and A. T. Peterson. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology **93**:1527-1539.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large area hydrologic modeling and assessment - Part 1: Model development. Journal of the American Water Resources Association **34**:73-89.

Arthur, J. W., J. A. Zischke, and G. L. Ericksen. 1982. Effect of elevated water temperature on macroinvertebrate communities in outdoor experimental channels. Water Research **16**:1465-1477.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling **200**:1-19.

Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species. Ecological Monographs **60**:161-177.

Bateman, B. L., H. T. Murphy, A. E. Reside, K. Mokany, and J. VanDerWal. 2013. Appropriateness of full-, partial- and no-dispersal scenarios in climate change impact modelling. Diversity and Distributions **19**:1224-1234.

Beven, K. 2004. Robert E. Horton's perceptual model of infiltration processes. Hydrological Processes **18**:3447-3460.

Brutsaert, W. 1982. Evaporation into the atmosphere: Theory, History and Applications. Springer, Netherlands, Dordrecht.

Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management **30**:492-507.

Caissie, D. 2006. The thermal regime of rivers: a review. Freshwater Biology **51**:1389-1406.

Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data. Copenhagen.

Collins, S. D., and N. E. McIntyre. 2015. Modeling the distribution of odonates: a review. Freshwater Science **34**:1144-1158.

Comte, L., and J. D. Olden. 2018. Evidence for dispersal syndromes in freshwater fishes. Proceedings of the Royal Society B: Biological Sciences **285**:20172214.

Cox, T. J., and J. C. Rutherford. 2000. Thermal tolerances of two stream invertebrates exposed to diumally varying temperature. New Zealand Journal of Marine and Freshwater Research **34**:203-208.

D'Amen, M., H. K. Mod, N. J. Gotelli, and A. Guisan. 2018. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. Ecography **41**:1233-1244.

David T. Bilton, Joanna R. Freeland, and B. Okamura. 2001. Dispersal in Freshwater Invertebrates. Annual Review of Ecology and Systematics **32**:159-181.

de Araujo, C. B., L. O. Marcondes-Machado, and G. C. Costa. 2014. The importance of biotic interactions in species distribution models: a test of the Eltonian noise hypothesis using parrots. Journal of Biogeography **41**:513-523.

Devisser, M. H., J. P. Messina, N. J. Moore, D. P. Lusch, and J. Maitima. 2010. A dynamic species distribution model of Glossina subgenus Morsitans: The identification of tsetse reservoirs and refugia. Ecosphere **1**:art6.

Domisch, S., G. Amatulli, and W. Jetz. 2015a. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data **2**:150073.

Domisch, S., M. B. Araujo, N. Bonada, S. U. Pauls, S. C. Jähnig, and P. Haase. 2013. Modelling distribution in European stream macroinvertebrates under future climates. Global Change Biology **19**:752-762.

Domisch, S., M. Friedrichs, T. Hein, F. Borgwardt, A. Wetzig, S. C. Jähnig, and S. D. Langhans. 2019. Spatially explicit species distribution models: A missed opportunity in conservation planning? Diversity and Distributions **25**:758-769.

Domisch, S., S. C. Jähnig, and P. Haase. 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshwater Biology **56**:2009-2020.

Domisch, S., F. T. Portmann, M. Kuemmerlen, R. B. O'Hara, R. K. Johnson, J. Davy-Bowker, T. Baekken, C. Zamora-Munoz, M. Sainz-Bariain, N. Bonada, P. Haase, P. Doll, and S. C. Jahnig. 2017. Using streamflow observations to estimate the impact of hydrological regimes and anthropogenic water use on European stream macroinvertebrate occurrences. Ecohydrology **10**:11.

Domisch, S., S. C. Jahnig, J. P. Simaika, M. Kuemmerlen, and S. Stoll. 2015b. 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.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z. I. Kawabata, D. J. Knowler, C. Leveque, R. J. Naiman, A. H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews **81**:163-182.

Elith, J., and J. R. Leathwick. 2017. Boosted Regression Trees for Ecological Modelling. http://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802-813.

Fenoglio, S., T. Bo, M. Cucco, and G. Malacarne. 2007. Response of benthic invertebrate assemblages to varying drought conditions in the Po river (NW Italy). Italian Journal of Zoology **74**:191-201.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics. **29**:1189-1232.

Gherghel, I., F. Brischoux, and M. Papeş. 2018. Using biotic interactions in broad-scale estimates of species' distributions. Journal of Biogeography **45**:2216-2225.

Graf, W., A. Lorenz, J. Tierno de Figueroa, S. Lücke, M. López-Rodríguez, and C. Davies. 2009. Distribution and Ecological Preferences of European Freshwater Organisms. Volume 2 - Plecoptera. Pensoft Publishers, Sofia-Moscow.

Graf, W., J. Murphy, J. Dahl, C. Zamora-Muñoz, and M. López-Rodríguez. 2008. Distribution and Ecological Preferences of European Freshwater Organisms. Volume 1 - Trichoptera. Pensoft Publishers, Sofia-Moscow.

Gray, L. J. 1981. Species Composition and Life Histories of Aquatic Insects in a Lowland Sonoran Desert Stream. The American Midland Naturalist **106**:229-242.

GRDC. 2016. The Global Runoff Data Centre. 56068 Koblenz, Germany.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling **135**:147-186.

Guse, B., D. E. Reusser, and N. Fohrer. 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.

Haase, P., S. Lohse, S. Pauls, K. Schindehütte, A. Sundermann, P. Rolauffs, and D. Hering. 2004. Assessing streams in Germany with benthic invertebrates: development of a practical standardized protocol for macroinvertebrate sampling and sorting. Limnologica **34**:349-365.

Haase, P., S. U. Pauls, K. Schindehütte, and A. Sundermann. 2010. First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. BIOONE.

Haidekker, A., and D. Hering. 2008. Relationship between benthic insects (Ephemeroptera, Plecoptera, Coleoptera, Trichoptera) and temperature in small and medium-sized streams in Germany: A multivariate study. Aquatic Ecology **42**:463-481.

Hancock, M. A., and S. E. Bunn. 1997. Population dynamics and life history of Paratya australiensis. Kemp, 1917 (Decapoda: Atyidae) in upland rainforest streams, south-eastern Queensland, Australia. Marine and Freshwater Research **48**:361-369.

Herkül, K., V. Lauringson, and J. Kotta. 2016. Specialization among amphipods: the invasive Gammarus tigrinus has narrower niche space compared to native gammarids. Ecosphere **7**:e01306.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Ihlow, F., J. Dambach, J. O. Engler, M. Flecks, T. Hartmann, S. Nekum, H. Rajaei, and D. Rödder. 2012. On the brink of extinction? How climate change may affect global chelonian species richness and distribution. Global Change Biology **18**:1520-1530.

Irving, K., M. Kuemmerlen, J. Kiesel, K. Kakouei, S. Domisch, and S. C. Jähnig. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model. Scientific Data **5**:180224.

Irving, K., S. C, Jähnig & M. Kuemmerlen. 2019. Identifying and applying an optimum set of environmental variables in species distribution models, Inland Waters, DOI: 10.1080/20442041.2019.1653111

Isotta, F. A., C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassègues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, M. Munari, S. Micheletti, V. Bonati, C. Lussana, C. Ronchi, E. Panettieri, G. Marigo, and G. Vertačnik. 2014. The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. International Journal of Climatology **34**:1657-1675.

Jansson, P., R. Hock, and T. Schneider. 2003. The concept of glacier storage: a review. Journal of Hydrology **282**:116-129.

Jarvis, A., H.I. Reuter, A. Nelson, and E. Guevara. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Kakouei, K., J. Kiesel, S. Domisch, K. S. Irving, S. C. Jähnig, and J. Kail. 2018. Projected effects of Climate-change-induced flow alterations on stream macroinvertebrate abundances. Ecology and Evolution **8**:3393-3409.

Kärcher, O., K. Frank, A. Walz, and D. Markovic. 2019. Scale effects on the performance of niche-based models of freshwater fish distributions. Ecological Modelling **405**:33-42.

Kiesel, J., N. Fohrer, B. Schmalz, and M. J. White. 2010. Incorporating landscape depressions and tile drainages of a northern German lowland catchment into a semi-distributed model. Hydrological Processes **24**:1472-1486.

Kroll, S. A., N. H. Ringler, M. d. l. C. Cano Costa, and J. De las Heras Ibañez. 2017. Macroinvertebrates on the front lines: projected community response to temperature and precipitation changes in Mediterranean streams. Journal of Freshwater Ecology **32**:513-528.

Kuemmerlen, M., B. Schmalz, Q. Cai, P. Haase, N. Fohrer, and S. C. Jähnig. 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.

Kuemmerlen, M., B. Schmalz, B. Guse, Q. Cai, N. Fohrer, and S. C. Jähnig. 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. Ecological Modelling **277**:77-86.

Lake, P. S. 2000. Disturbance, patchiness, and diversity in streams. Journal of the North American Benthological Society **19**:573-592.

Lehner, B., K. Verdin, and A. Jarvis. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Transactions American Geophysical Union **89**:93-94.

Lenoir, J., B. J. Graae, P. A. Aarrestad, I. G. Alsos, W. S. Armbruster, G. Austrheim, C. Bergendorff, H. J. B. Birks, K. A. Bråthen, J. Brunet, H. H. Bruun, C. J. Dahlberg, G. Decocq, M. Diekmann, M. Dynesius, R. Ejrnæs, J.-A. Grytnes, K. Hylander, K. Klanderud, M. Luoto, A. Milbau, M. Moora, B. Nygaard, A. Odland, V. T. Ravolainen, S. Reinhardt, S. M. Sandvik, F. H. Schei, J. D. M. Speed, L. U.

Tveraabak, V. Vandvik, L. G. Velle, R. Virtanen, M. Zobel, and J.-C. Svenning. 2013. Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across Northern Europe. Global Change Biology **19**:1470-1481.

Li, F., A. Sundermann, S. Stoll, and P. Haase. 2016. A newly developed dispersal metric indicates the succession of benthic invertebrates in restored rivers. Science of the Total Environment **569-570**:1570-1578.

Li, F., J. D. Tonkin, and P. Haase. 2018. Dispersal capacity and broad-scale landscape structure shape benthic invertebrate communities along stream networks. Limnologica **71**:68-74.

Lytle, D. A. 2002. Flash floods and aquatic insect life-history evolution: evaluation of multiple models. Ecology **83**:370-385.

Lytle, D. A., and N. L. Poff. 2004. Adaptation to natural flow regimes. Trends in Ecology & Evolution **19**:94-100.

Markovic, D., J. Freyhof, and C. Wolter. 2012. Where Are All the Fish: Potential of Biogeographical Maps to Project Current and Future Distribution Patterns of Freshwater Species. PLoS ONE **7**:e40530.

Meybeck, M. 2003. Global analysis of river systems: from Earth system controls to Anthropocene syndromes. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **358**:1935-1955.

Milbrink, G. 1993. Evidence for Mutualistic Interactions in Freshwater Oligochaete Communities. Oikos **68**:317-322.

Miller, D. A. W., L. A. Weir, B. T. McClintock, E. H. C. Grant, L. L. Bailey, and T. R. Simons. 2012. Experimental investigation of false positive errors in auditory species occurrence surveys. Ecological Applications **22**:1665-1674.

Moore, I. D., R. B. Grayson, and A. R. Ladson. 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes **5**:3-30.

Moore, R. D., P. Sutherland, T. Gomi, and A. Dhakal. 2005. Thermal regime of a headwater stream within a clear-cut, coastal British Columbia, Canada. Hydrological Processes **19**:2591-2608.

Naimi, B., and M. Araujo. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography **39**:368-375.

Naeem, S., and D. E. Bunker. 2009. TraitNet: furthering biodiversity research through the curation, discovery, and sharing of species trait data. Biodiversity, Ecosystem Functioning, and Human Wellbeing **1**:281-290.

Olden, J. D., and N. L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications **19**:101-121.

Ouellet Dallaire, C., B. Lehner, R. Sayre, and M. Thieme. 2019. A multidisciplinary framework to derive global river reach classifications at high spatial resolution. Environmental Research Letters **14**:024003.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

Peckarsky, B. L., B. W. Taylor, and C. C. Caudill. 2000. Hydrologic and behavioral constraints on oviposition of stream insects: implications for adult dispersal. Oecologia **125**:186-200.

Peres, D. J., and A. Cancelliere. 2016. Environmental Flow Assessment Based on Different Metrics of Hydrological Alteration. Water Resources Management **30**:5799-5817.

Olden, J. D., and N. L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications **19**:101-121.

Pletterbauer, F., W. Graf, and S. Schmutz. 2016. Effect of biotic dependencies in species distribution models: The future distribution of Thymallus thymallus under consideration of Allogamus auricollis. Ecological Modelling **327**:95-104.

Poff, N., J. Allan, M. Bain, J. Karr, K. Prestegaard, B. Richter, R. Sparks, and J. Stromberg. 1997. The natural flow regime. Bioscience **47**:769-784.

Poff, N. L. 1997. Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. Journal of the North American Benthological Society **16**:391-409.

Poff, N. L., J. D. Olden, N. K. M. Vieira, D. S. Finn, M. P. Simmons, and B. C. Kondratieff. 2006. Functional trait niches of North American lotic insects: traits-based ecological applications in light of phylogenetic relationships. BIOONE.

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

Poff, N. L., and J. V. Ward. 1989. Implications of streamflow variability and predictability for lotic community structure - a regional-analysis of streamflow patterns. Canadian Journal of Fisheries and Aquatic Sciences **46**:1805-1818.

Poff, N. L., and J. K. H. Zimmerman. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology **55**:194-205.

Porst, G., O. Naughton, L. Gill, P. Johnston, and K. Irvine. 2012. Adaptation, phenology and disturbance of macroinvertebrates in temporary water bodies. Hydrobiologia **696**:47-62.

Pyne, M. I., and N. L. Poff. 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.

Radinger, J., J. Kail, and C. Wolter. 2014. FIDIMO — A free and open source GIS based dispersal model for riverine fish. Ecological Informatics **24**:238-247.

Radinger, J., and C. Wolter. 2014. Patterns and predictors of fish dispersal in rivers. Fish and Fisheries **15**:456-473.

Randin, C. F., R. Engler, S. Normand, M. Zappa, N. Zimmermann, P. B. Pearman, P. Vittoz, W. Thuiller, and A. Guisan. 2009. Climate change and plant distribution: local models predict high-elevation persistence. Global Change Biology **15**:1557-1569.

Record, S., N. D. Charney, R. M. Zakaria, and A. M. Ellison. 2013. Projecting global mangrove species and community distributions under climate change. Ecosphere **4**:23.

Record, S., A. Strecker, M.-N. Tuanmu, L. Beaudrot, P. Zarnetske, J. Belmaker, and B. Gerstner. 2018. Does scale matter? A systematic review of incorporating biological realism when predicting changes in species distributions. PLoS ONE **13**:e0194650.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar. 1988. The Role of

Disturbance in Stream Ecology. Journal of the North American Benthological Society **7**:433-455.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun. 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. Conservation Biology **10**:1163-1174.

Rodríguez-Merino, A., R. Fernández-Zamudio, and P. García-Murillo. 2019. Identifying areas of aquatic plant richness in a Mediterranean hotspot to improve the conservation of freshwater ecosystems. Aquatic Conservation: Marine and Freshwater Ecosystems **29**:589-602.

Ruiz-Navarro, A., P. K. Gillingham, and J. R. Britton. 2016. Predicting shifts in the climate space of freshwater fishes in Great Britain due to climate change. Biological Conservation **203**:33-42.

Schmalz, B., M. Kuemmerlen, J. Kiesel, Q. Cai, S. C. Jähnig, and N. Fohrer. 2015. Impacts of land use changes on hydrological components and macroinvertebrate distributions in the Poyang lake area. Ecohydrology **8**:1119-1136.

Smeraldo, S., M. Di Febbraro, L. Bosso, C. Flaquer, D. Guixé, F. Lisón, A. Meschede, J. Juste, J. Prüger, X. Puig-Montserrat, and D. Russo. 2018. Ignoring seasonal changes in the ecological niche of non-migratory species may lead to biases in potential distribution models: lessons from bats. Biodiversity and Conservation **27**:2425-2441.

Soranno, P. A., K. S. Cheruvelil, K. E. Webster, M. T. Bremigan, T. Wagner, and C. A. Stow. 2010. Using Landscape Limnology to Classify Freshwater Ecosystems for Multi-ecosystem Management and Conservation. Bioscience **60**:440-454.

Statzner, B., J. A. Gore, and V. H. Resh. 1988. Hydraulic stream ecology - observed patterns and potential applications. Journal of the North American Benthological Society **7**:307-360.

The German Federal Institute of Hydrology (BfG). 2013. The June 2013 floodings in Germany – an overview. Bericht BfG-1797, Federal Institute of Hydrology, Koblenz.

Thuiller, W., D. Georges, and R. Engler. 2014. biomod2. Ensemble platform for species distribution modeling. R package version 3.1-64.

Tonkin, J. D., D. N. Shah, M. Kuemmerlen, F. Li, Q. Cai, P. Haase, and S. C. Jahnig. 2015. Climatic and Catchment-Scale Predictors of Chinese Stream Insect Richness Differ between Taxonomic Groups. PLoS ONE **10**:e0123250.

Usseglio-Polatera, P., M. Bournaud, P. Richoux, and H. Tachet. 2000. Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits. Freshwater Biology **43**:175-205.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences **37**:130-137.

Verberk, W. c. E. P., H. Siepel, and H. Esselink. 2008. Life-history strategies in freshwater macroinvertebrates. Freshwater Biology **53**:1722-1738.

Vorosmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global threats to human water security and river biodiversity Nature **467**:555-561.

Warscher, M., U. Strasser, G. Kraller, T. Marke, H. Franz, and H. Kunstmann. 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. Water Resources Research **49**:2619-2637.

WWF. 2018. Living Planet Report 2018: Aiming higher. Gland, Switzerland.

Zurell, D., L. J. Pollock, and W. Thuiller. 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? Ecography **41**:1812-1819.

Appendix A: Supporting information for Chapter 2

Table S2.1: Spearman's correlation (*r*) observed vs. simulated IHA metrics (n=165), names and descriptions as per Olden & Poff (2003). Light grey =$r > 0.5$ (positive correlation), no color = $r <= 0.50$ (no correlation).

IHA	r	Metric	Description
DH1	0.7	Annual maxima of 1 day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH2	0.71	Annual maxima of 3 day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH3	0.71	Annual maxima of 7 day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH4	0.72	Annual maxima of 30 day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH5	0.72	Annual maxima of 90 day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH6	0.16	Variability in annual maxima of 1 day means of daily discharge	Coefficient of variation in DH1–5
DH7	0.12	Variability in annual maxima of 3 day means of daily discharge	Coefficient of variation in DH1–5
DH8	0.07	Variability in annual maxima of 7 day means of daily discharge	Coefficient of variation in DH1–5
DH9	-0.07	Variability in annual maxima of 30 day means of daily discharge	Coefficient of variation in DH1–5
DH10	0.02	Variability in annual maxima of 90 day means of daily discharge	Coefficient of variation in DH1–5
DH11	0.27	Means of 1 day maxima of daily discharge	Mean annual 1-day maximum, respectively, divided by median flow

DH12	0.31	Means of 7 day maxima of daily discharge	Mean annual 7-day maximum, respectively, divided by median flow
DH13	0.33	Means of 30 day maxima of daily discharge	Mean annual 30-day maximum, respectively, divided by median flow
DH14	0.23	Flood duration 1	Monthly flow equaled or exceeded 95% of the time divided by mean monthly flow
DH15	-0.14	High flow pulse duration	Mean duration of FH1
DH16	0.05	Variability in high flow pulse duration	Coefficient of variation in DH15
DH17	0.07	High flow duration 1	See DH15, where the upper threshold is defined as 1 times median flows, and the value is represented as an average instead of a tabulated count
DH18	0.03	High flow duration 1	See DH15, where the upper threshold is defined as 3 times median flows, and the value is represented as an average instead of a tabulated count
DH19	0.12	High flow duration 1	See DH15, where the upper threshold is defined as 7 times median flows, and the value is represented as an average instead of tabulated count
DH20	-0.03	High flow duration 2	See DH17–19, where the upper threshold is defined as the 25th percentile of median flows
DH21	0.06	High flow duration 2	See DH17–19, where the upper threshold is defined as the 75th percentile of median flows
DH22	-0.07	Flood interval	Mean annual median interval in days between floods over all years
DH23	-0.24	Flood duration 2	Mean annual number of days that flows remain above the flood threshold averaged across all year
DH24	0.09	Flood free days	Mean annual maximum number of 365 days over all water years during which no floods occurred over all years

DL1	0.72	Annual minima of 1 day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL2	0.72	Annual minima of 3 day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL3	0.72	Annual minima of 7 day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL4	0.72	Annual minima of 30 day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL5	0.73	Annual minima of 90 day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL6	-0.14	Variability in annual minima of 1 day means of daily discharge	Coefficient of variation in DL1–5
DL7	-0.2	Variability in annual minima of 3 day means of daily discharge	Coefficient of variation in DL1–5
DL8	-0.23	Variability in annual minima of 7 day means of daily discharge	Coefficient of variation in DL1–5
DL9	-0.22	Variability in annual minima of 30 day means of daily discharge	Coefficient of variation in DL1–5
DL10	0.1	Variability in annual minima of 90 day means of daily discharge	Coefficient of variation in DL1–5
DL11	-0.07	Means of 1 day minima of daily discharge	Mean annual 1-day minimum, divided by median flow
DL12	-0.03	Means of 7 day minima of daily discharge	Mean annual 7-day minimum, respectively, divided by median flow
DL13	-0.09	Means of 30 day minima of daily discharge	Mean annual 30-day minimum, respectively, divided by median flow
DL14	0	Low exceedance flows	Mean magnitude of flows exceeded 75% of the time (calculated from the flow duration curve) divided by median daily flow, respectively, over all years

DL15	-0.07	Low exceedance flows	Mean magnitude of flows exceeded 90% of the time (calculated from the flow duration curve) divided by median daily flow, respectively, over all years
DL16	-0.11	Low flow pulse duration	Mean duration of FL1
DL17	0.09	Variability in low flow pulse duration	Coefficient of variation in DL16
FH1	-0.05	High flood pulse count 1	See FL1, where the high pulse is defined as the 75th percentile
FH2	-0.01	Variability in high flood pulse count 1	Coefficient of variation in FH1
FH3	0.27	High flood pulse count 2	See FH1, where the upper threshold is defined as 3 times median daily flow, and the value is represented as an average instead of a tabulated count
FH4	0.24	High flood pulse count 2	See FH1, where the upper threshold is defined as 7 times median daily flow, and the value is represented as an average instead of a tabulated count
FH5	-0.01	Flood frequency 1	Mean number of high flow events per year using an upper threshold of 1 times median flow over all years
FH6	-0.02	Flood frequency 1	Mean number of high flow events per year using an upper threshold of 1 times median flow over all years
FH7	0.24	Flood frequency 1	Mean number of high flow events per year using an upper threshold of 7 times median flow over all years
FH8	-0.05	Flood frequency 2	See FH5–7, where the 25th percentile are used as the upper threshold
FH9	-0.09	Flood frequency 2	See FH5–7, where the 75th percentile are used as the upper threshold
FH10	0.01	Flood frequency 3	See FH5–7, where the median of the annual minima is used as the upper threshold
FH11	-0.24	Flood frequency 4	Mean number of discrete flood events per year
FL1	-0.09	Low flood pulse count	Number of annual occurrences during which the magnitude of flow remains below a lower threshold. Hydrologic pulses are defined as those periods within a year in which the flow drops below the 25th percentile (low pulse) of all daily values for the time period
FL2	0.05	Variability in low flood pulse count	Coefficient of variation in FL1

FL3	0.03	Frequency of low flow spells	Total number of low flow spells (threshold equal to 5% of mean daily flow) divided by the record length in years
MA1	0.73	Mean daily flows	Mean daily flows
MA2	0.73	Median daily flows	Median daily flows
MA3	0.25	Variability in daily flows 1	Coefficient of variation in daily flows
MA4	0.28	Variability in daily flows 2	Coefficient of variation of the logs in daily flows corresponding to the (5th, 10th, 15th, . . . , 85th, 90th 95th) percentiles
MA5	0.29	Skewness in daily flows	Mean daily flows divided by median daily flows
MA6	0.2	Ranges in daily flows	Ratio of 10th/90th percentiles in daily flows over all years
MA7	0.23	Ranges in daily flows	Ratio of 20th/80th percentiles in daily flows over all years
MA8	0.23	Ranges in daily flows	Ratio of 25th/75th percentiles in daily flows over all years
MA9	0.28	Spreads in daily flows	Ranges in daily flows (MA6–8) divided by median daily flows
MA10	0.27	Spreads in daily flows	Ranges in daily flows (MA6–8) divided by median daily flows
MA11	0.27	Spreads in daily flows	Ranges in daily flows (MA6–8) divided by median daily flows
MA12	0.74	Mean monthly flows	Mean monthly flow for all months
MA13	0.74	Mean monthly flows	Mean monthly flow for all months
MA14	0.73	Mean monthly flows	Mean monthly flow for all months
MA15	0.71	Mean monthly flows	Mean monthly flow for all months
MA16	0.71	Mean monthly flows	Mean monthly flow for all months
MA17	0.72	Mean monthly flows	Mean monthly flow for all months
MA18	0.71	Mean monthly flows	Mean monthly flow for all months
MA19	0.72	Mean monthly flows	Mean monthly flow for all months
MA20	0.71	Mean monthly flows	Mean monthly flow for all months
MA21	0.72	Mean monthly flows	Mean monthly flow for all months

MA22	0.73	Mean monthly flows	Mean monthly flow for all months
MA23	0.73	Mean monthly flows	Mean monthly flow for all months
MA24	0.17	Mean monthly flows	Mean monthly flow for all months
MA25	0.16	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA26	0.1	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA27	-0.02	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA28	0.03	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA29	-0.02	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA30	-0.06	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA31	-0.02	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA32	-0.09	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA33	0.01	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA34	0.16	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA35	0.16	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA36	0.25	Variability across monthly flows 1	Variability in monthly flows divided by median monthly flows, where variability is calculated as range
MA37	0.34	Variability across monthly flows 1	Variability in monthly flows divided by median monthly flows, where variability is calculated as interquartile
MA38	0.32	Variability across monthly flows 1	Variability in monthly flows divided by median monthly flows, where variability is calculated as 90th–10th percentile.
MA39	0.29	Variability across monthly flows 2	Coefficient of variation in mean monthly flows
MA40	0.31	Variability across monthly flows 2	Coefficient of variation in mean monthly flows 4 MA40 6 M Skewness in monthly flows (Mean monthly flow—median monthly flow)/median monthly flow

MH1	0.71	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH2	0.72	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH3	0.71	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH4	0.7	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH5	0.7	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH6	0.7	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH7	0.69	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH8	0.69	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH9	0.68	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH10	0.7	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH11	0.71	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH12	0.7	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH13	0.27	Variability across maximum monthly flows	Coefficient of variation in mean maximum monthly flows
MH14	0.2	Median of annual maximum flows	Median of the highest annual daily flow divided by the median annual daily flow averaged across all years
MH15	0.27	High flow discharge	Mean of the 1st, 10th and 25th percentile from the flow duration curve divided by median daily flow across all years
MH18	0.28	Variability across annual maximum flows	Coefficient of variation of logarithmic annual maximum flows
MH19	0.28	Skewness in annual maximum flows	See Hughes and James (1989)
MH21	0.64	High flow volume	Mean of the high flow volume (calculated as the area between the hydrograph and the upper threshold during the high flow event) divided by median annual daily flow across all years. The upper threshold is defined as 1 times median annual flow

MH22	0.06	High flow volume	Mean of the high flow volume (calculated as the area between the hydrograph and the upper threshold during the high flow event) divided by median annual daily flow across all years. The upper threshold is defined as 3 times median annual flow
MH23	0.15	High flow volume	Mean of the high flow volume (calculated as the area between the hydrograph and the upper threshold during the high flow event) divided by median annual daily flow across all years. The upper threshold is defined as 7 times median annual flow
MH24	0.25	High peak flow 1	Mean of the high peak flow during the high flow event (defined by the upper threshold) divided by median annual daily flow. The upper threshold is defined as 1 times median annual flow
MH25	0.24	High peak flow 1	Mean of the high peak flow during the high flow event (defined by the upper threshold) divided by median annual daily flow. The upper threshold is defined as 3 times median annual flow
MH26	0.27	High peak flow 1	Mean of the high peak flow during the high flow event (defined by the upper threshold) divided by median annual daily flow. The upper threshold is defined as 7 times median annual flow
MH27	0.31	High peak flow 2	See MH24–26, where the upper threshold is defined as the 25th percentile from the flow duration curve
ML1	0.74	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML2	0.74	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML3	0.74	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML4	0.72	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML5	0.71	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML6	0.71	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML7	0.72	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML8	0.72	Mean minimum monthly flows	Mean minimum monthly flow for all months

ML9	0.71	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML10	0.72	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML11	0.73	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML12	0.74	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML13	-0.06	Variability across minimum monthly	Coefficient of variation in minimum monthly flows
ML14	-0.07	Mean of annual minimum flows	Mean of the lowest annual daily flow divided by median
ML15	-0.07	Low flow index	Mean of the lowest annual daily flow divided by mean annual
ML16	-0.09	Median of annual minimum flows 2	Median of the lowest annual daily flows divided by median
ML17	0.05	Baseflow index 1	Seven-day minimum flow divided by mean annual daily flows averaged across all years
ML18	-0.29	Variability in Baseflow Index 1	Coefficient of variation in ML17
ML19	-0.07	Baseflow index 2	Mean of the ratio of the lowest annual daily flow to the mean annual daily flow times 100 averaged across all years
ML20	-0.02	Baseflow index 3	Ratio of baseflow volume to total flow volume
ML21	-0.13	Variability across annual minimum	Coefficient of variation in annual minimum flows averaged across all years
RA1	0.64	Rise rate	Mean rate of positive changes in flow from one day to the next
RA2	0.33	Variability in rise rate	Coefficient of variation in RA1
RA3	0.61	Fall rate	Mean rate of negative changes in flow from one day to the next
RA4	0.26	Variability in fall rate	Coefficient of variation in RA3
RA5	0.24	No day rises	Ratio of days where the flow is higher than the previous day
RA6	-0.06	Change of flow	Median of difference between natural logarithm of flows between two consecutive days with increasing/decreasing flow
RA7	-0.03	Change of flow	Median of difference between natural logarithm of flows between two consecutive days with increasing/decreasing flow

RA8	-0.04	Reversals	Number of negative and positive changes in water conditions from one day to the next
RA9	0.02	Variability in reversals	Coefficient of variation in RA8
TA1	0.58	Constancy	See Colwell (1974)
TA2	0.58	Predictability of flow	Composed of two independent, additive components: constancy (a measure of temporal invariance) and contingency (a measure of periodicity)
TA3	0.32	Seasonal predictability of flooding	Maximum proportion of all floods over the period of record that fall in any one of six 60-day ‘seasonal’ windows
TH1	0.16	Julian date of annual maximum	The mean Julian date of the 1-day annual maximum flow over all years
TH2	0.08	Variability in Julian date of annual maximum	Coefficient of variation in TH1
TH3	0.18	Seasonal predictability of non-flooding	Maximum proportion of the year (number of days/365) during which no floods have ever occurred over the period of record
TL1	0.06	Julian date of annual minimum	The mean Julian date of the 1-day annual minimum flow over all years
TL2	0.03	Variability in Julian date of annual minimum	Coefficient of variation in Tl1
TL3	0.19	Seasonal predictability of low flow	Proportion of low-flow events 5-year magnitude falling in a 60-day ‘seasonal’ window
TL4	-0.01	Seasonal predictability of non-low flow	Maximum proportion of the year (number of days/365) during which no 5-year + low flows have ever occurred over the entire period of record

Table S2.2: Temporal validation full list of KGE, R2 and RMSE results and coordinates of sites

Site	lon	lat	r2	KGE	RMSE_test	RMSE_train
1	9.704167	50.47083	0.69	0.78	0.56	0.63
2	8.5375	51.54583333	0.66	0.77	0.93	0.95
3	11.7	50.06	0.59	0.76	0.71	0.69
4	8.7375	51.74583333	0.72	0.76	0.86	0.89
5	9.8125	50.44583	0.76	0.76	0.54	0.55
6	10.60416667	52.57083333	0.62	0.75	1.16	1.24
7	10.158	50.631	0.65	0.74	0.66	0.65
8	11.02	53.32	0.58	0.74	2.2	2.27
9	8.31	53	0.62	0.74	6.25	6.39
10	9.8125	50.6375	0.64	0.74	0.56	0.61
11	11.0625	50.30416667	0.6	0.73	0.7	0.66
12	11.684707	51.226506	0.54	0.73	15.74	15.54
13	10.2375	52.62917	0.61	0.72	1.21	1.13
14	8.4125	52.65416667	0.64	0.72	3.48	3.44
15	6.470833333	51.8375	0.66	0.71	1.33	1.32
16	10.13	51.94	0.62	0.7	0.86	0.85
17	6.970833333	51.22083333	0.52	0.7	0.57	0.55
18	8.479166667	52.82916667	0.61	0.7	5.65	5.7
19	8.98	49.61	0.6	0.7	0.62	0.65
20	9.704166667	52.6875	0.52	0.7	23.42	23.25
21	12.17083	53.80417	0.51	0.69	1.74	1.79
22	9.5625	51.1125	0.65	0.69	0.72	0.71
23	9.929166667	51.40416667	0.65	0.69	1.45	1.69
24	12.05417	51.30417	0.61	0.68	34.21	38.11
25	8.48	52.83	0.58	0.68	5.95	5.78
26	8.670833	51.17083	0.62	0.68	0.55	0.56
27	8.529167	51.5375	0.63	0.67	0.99	0.91
28	12.01	53.15	0.59	0.66	1.88	1.89
29	6.370833	50.90417	0.46	0.66	7.51	7.51
30	8.045833333	49.2125	0.51	0.66	0.94	0.96
31	8.104166667	47.6625	0.63	0.66	1.45	1.57
32	8.520833	48.77917	0.49	0.66	0.62	0.62
33	9.970833	51.12083	0.45	0.66	1.1	0.85
34	10.04	52.74	0.55	0.65	2.56	2.64
35	10.29583333	53.5375	0.56	0.65	1.14	1.1
36	7.929167	51.62917	0.68	0.65	0.89	0.92
37	8.38	51.16	0.48	0.65	0.99	0.95

38	9.37	48.51	0.58	0.65	0.83	0.83
39	9.7375	50.62917	0.68	0.65	0.91	1
40	10.05416667	52.62083333	0.54	0.64	18.36	18.43
41	10.7125	51.10416667	0.49	0.64	2.58	2.79
42	9.612062	51.944106	0.51	0.64	0.77	0.81
43	9.6125	49.05416667	0.54	0.64	0.71	0.7
44	11.1875	51.8875	0.57	0.63	4.6	4.57
45	11.68	50.04	0.46	0.63	0.88	0.87
46	7.254166667	51.7875	0.48	0.63	1.67	1.71
47	8.39	48.47	0.44	0.63	0.67	0.67
48	8.571186	49.668736	0.56	0.63	1.78	1.73
49	8.7875	50.27916667	0.57	0.63	4.16	4.24
50	9.275496	48.083736	0.43	0.63	2.23	2.32
51	10.79583333	50.52083333	0.4	0.62	0.91	0.86
52	10.9625	51.35416667	0.61	0.62	2.03	1.93
53	11.8375	53.7375	0.47	0.62	1.37	1.4
54	13.395112	51.518745	0.39	0.62	9.97	9.26
55	9.695833333	52.10416667	0.58	0.62	1.04	1.1
56	12.32083333	50.0125	0.69	0.61	1.17	1.18
57	6.62	50.24	0.56	0.61	0.87	0.87
58	8.06	50.24	0.65	0.61	1.01	0.99
59	8.19	47.93	0.52	0.61	1.19	1.11
60	8.979166667	49.97083333	0.62	0.61	2.03	1.93
61	9.283338	50.853971	0.63	0.61	1.23	1.11
62	11.478205	49.028364	0.61	0.6	9.17	9.19
63	11.69	49.97	0.45	0.6	0.99	3.41
64	7.38	50.37	0.58	0.6	1.55	1.57
65	7.5125	49.9875	0.51	0.6	0.8	0.75
66	8.57	49.67	0.51	0.6	1.92	1.68
67	8.7375	50.17917	0.57	0.6	6.67	6.79
68	9.454166667	48.57083333	0.53	0.6	0.84	0.85
69	10.32083333	51.5625	0.46	0.59	0.73	0.71
70	11.061	50.548	0.46	0.59	0.88	0.84
71	11.12083333	53.85416667	0.44	0.59	2.28	2.28
72	11.8375	51.1625	0.58	0.59	39.78	37.26
73	12.2625	50.32083333	0.46	0.59	1.21	1.28
74	13.07917	51.77083	0.41	0.59	12.79	12.82
75	13.36	48.83	0.51	0.59	2.36	2.54
76	7.345833333	52.6875	0.58	0.59	15.74	15.8
77	7.43	50.42	0.54	0.59	1.76	1.76
78	8.000165	52.308177	0.46	0.59	2.52	2.3

79	8.145833	50.3625	0.54	0.59	1.33	1.43
80	9.2875	50.85416667	0.51	0.59	1.41	1.5
81	9.64	49.04	0.58	0.59	0.89	0.92
82	9.75	52.89	0.5	0.59	1.66	1.65
83	9.9375	51.1375	0.64	0.59	1.08	0.82
84	10.8875	53.2375	0.36	0.58	400.7	397.83
85	11.19583333	49.8125	0.65	0.58	0.96	0.97
86	11.5125	50.35416667	0.47	0.58	1.07	1
87	11.52083333	50.17916667	0.55	0.58	0.9	0.91
88	7.020833	52.22083	0.61	0.58	1.48	1.62
89	8.304166667	51.8625	0.49	0.58	2.54	2.43
90	9.104167	50.2875	0.62	0.58	0.84	0.88
91	9.2125	49.82916667	0.51	0.58	0.95	0.93
92	9.3875	52.7875	0.56	0.58	62.35	63.11
93	10.6875	51.74583	0.35	0.57	0.8	0.83
94	11.83	49.63	0.45	0.57	0.82	0.74
95	6.545833333	50.55416667	0.56	0.57	1.34	1.37
96	7.42	52.43	0.46	0.57	4.31	4.18
97	8.004166667	47.62916667	0.42	0.57	0.82	0.81
98	8.2125	50.75417	0.72	0.57	0.9	0.95
99	8.720833	51.6625	0.67	0.57	1.55	1.55
100	8.795833	50.47917	0.63	0.57	0.83	0.82
101	9.0875	52.2875	0.5	0.57	0.89	0.85
102	9.154166667	48.09583333	0.41	0.57	1.28	1.2
103	9.77	51.81	0.7	0.57	1.39	1.32
104	9.845833	52.97917	0.48	0.57	0.62	0.68
105	9.879166667	53.72083333	0.33	0.57	0.77	0.75
106	10.32	52.57	0.57	0.56	0.91	0.89
107	10.6375	50.4875	0.64	0.56	2.33	2.19
108	10.79583	51.69583	0.5	0.56	0.88	0.85
109	11.02083333	49.60416667	0.59	0.56	1.29	1.22
110	13.3125	53.7875	0.37	0.56	4.23	4.15
111	14.19	50.93	0.34	0.56	1.04	1.11
112	6.345833	50.92917	0.54	0.56	9.31	9.4
113	9.995833	51.17083	0.62	0.56	2.2	1.83
114	10.2375	53.50416667	0.57	0.55	1.58	1.62
115	10.3625	52.0125	0.45	0.55	1.99	1.92
116	11.64583333	52.12916667	0.39	0.55	326.72	330.46
117	11.9875	52.54583333	0.39	0.55	340.94	337.28
118	7.120833333	51.35416667	0.34	0.55	1.09	1.22
119	7.6375	52.72917	0.33	0.55	1	0.92

120	9.020833	50.4375	0.7	0.55	1.03	1.11
121	9.479167	50.32083	0.67	0.55	1.26	1.28
122	9.79	50.31	0.69	0.55	1.16	1.12
123	10.47	53.15	0.44	0.54	4.41	4.38
124	10.95416667	50.8375	0.61	0.54	1.48	1.46
125	10.957	50.921	0.6	0.54	1.99	2.08
126	10.99	49.44	0.42	0.54	7.96	7.19
127	11.16	50.432	0.46	0.54	0.96	1.07
128	11.2125	51.82917	0.46	0.54	3.82	3.81
129	11.8875	51.9875	0.39	0.54	336.82	340.17
130	13.37	48.63	0.47	0.54	1.58	1.56
131	6.94	49.83	0.6	0.54	1.49	1.56
132	7.495833	52.4375	0.61	0.54	2.47	2.35
133	7.9625	52.5625	0.59	0.54	5.94	5.84
134	8.004166667	50.9625	0.56	0.54	1.33	1.33
135	8.345833	50.69583	0.72	0.54	1.18	1.29
136	9.020833333	54.70416667	0.41	0.54	2.93	2.98
137	9.731472	50.805051	0.59	0.54	2.5	2.65
138	9.761058	50.022368	0.54	0.54	1.95	1.97
139	10.10416667	51.6875	0.53	0.53	6.7	6.88
140	6.1875	50.72917	0.52	0.53	0.75	0.8
141	6.970833333	52.50416667	0.5	0.53	6.96	6.97
142	9.125	52.964	0.58	0.53	182.19	178.36
143	9.345833333	51.07916667	0.5	0.53	5.07	5
144	10.746	50.509	0.61	0.52	1.71	1.63
145	10.8375	50.00416667	0.54	0.52	1.96	1.89
146	12.4125	48.9125	0.31	0.52	1.84	1.97
147	6.4375	50.22083	0.59	0.52	0.96	1.01
148	8.47	48.55	0.41	0.52	0.85	0.85
149	8.59	48.77	0.49	0.52	1.23	0.95
150	8.83	49.81	0.68	0.52	1.31	1.3
151	8.88	50.23	0.63	0.52	2.27	2.24
152	9.2875	50.75417	0.58	0.52	1.04	1.23
153	11.441377	49.106945	0.49	0.51	1.09	1.08
154	12.32083333	52.60416667	0.41	0.51	45.77	47.16
155	13.62083	52.29583	0.32	0.51	1.35	1.36
156	7.898373	51.972621	0.52	0.51	10.87	11.15
157	7.954166667	47.9375	0.45	0.51	1.15	1.12
158	8.3375	50.45417	0.7	0.51	1.72	1.78
159	8.515945	49.008091	0.33	0.51	1.72	1.59
160	8.895833333	52.60416667	0.56	0.51	5.58	5.94

161	8.897686	48.724098	0.27	0.51	1.49	1.48
162	8.902825	52.601321	0.55	0.51	5.79	5.44
163	8.954167	52.02083	0.69	0.51	0.71	0.74
164	9.379166667	54.62083333	0.41	0.51	1.61	1.57
165	10.02916667	48.89583333	0.68	0.5	1.21	1.22
166	10.64583333	51.45416667	0.45	0.5	2.12	1.78
167	11.4375	49.1125	0.5	0.5	0.69	0.62
168	12	50.13	0.78	0.5	1.08	1.19
169	7.6625	49.52917	0.31	0.5	1.45	1.39
170	7.84	52.72	0.58	0.5	10.5	10.74
171	8.7625	49.6375	0.57	0.5	0.63	0.62
172	9.211	52.852	0.61	0.5	116.29	117.51
173	11.39583333	50.9375	0.57	0.49	3.24	3.02
174	11.579321	51.071839	0.54	0.49	3.9	4.12
175	12.1125	50.02083333	0.69	0.49	1.24	1.26
176	12.2625	49.05416667	0.48	0.49	0.9	0.82
177	12.32083333	52.6125	0.38	0.49	47.78	47.27
178	12.40416667	51.20416667	0.32	0.49	4.74	4.65
179	12.40416667	51.2125	0.31	0.49	4.5	4.85
180	6.85115	51.673858	0.55	0.49	29.45	28.89
181	7.1875	51.7375	0.59	0.49	26.28	26.04
182	7.804166667	50.27916667	0.59	0.49	0.88	0.93
183	8.01	52.71	0.58	0.49	3.28	3.46
184	8.36	48.61	0.42	0.49	2	2.37
185	8.5125	49.00416667	0.33	0.49	1.69	1.58
186	8.62	48.14	0.48	0.49	2.43	2.97
187	9.24493	47.729498	0.32	0.49	2.12	2.03
188	9.731446	48.266271	0.41	0.49	24.96	25.49
189	10.47917	52.95417	0.53	0.48	1.05	1.03
190	10.79	51.6	0.36	0.48	1.17	1.09
191	10.89583333	49.79583333	0.58	0.48	1.32	1.51
192	11.206	50.405	0.48	0.48	1.1	1.03
193	11.67	50.87	0.43	0.48	2.07	1.82
194	12.65416667	51.8625	0.31	0.48	235.03	240.27
195	6.445833	50.37083	0.6	0.48	1.22	1.2
196	7.670833333	52.70416667	0.59	0.48	11.46	11.31
197	7.779167	51.90417	0.61	0.48	1.62	1.64
198	7.8375	49.77916667	0.43	0.48	1.89	1.73
199	8.54	48.72	0.61	0.48	1.39	1.36
200	8.630535	51.748833	0.58	0.48	7.32	7.34
201	8.6375	49.15416667	0.44	0.48	1.67	1.62

202	9.204167	50.7625	0.53	0.48	0.69	0.71
203	9.279166667	50.20416667	0.62	0.48	0.69	0.71
204	9.729166667	49.05416667	0.4	0.48	1.14	0.99
205	10.35	50.98	0.36	0.47	2.13	1.96
206	11.38	49.366979	0.53	0.47	1.66	1.72
207	11.5	48.73	0.32	0.47	4.81	4.73
208	12.845305	52.478085	0.34	0.47	41.08	39.87
209	13.010575	51.558968	0.33	0.47	218.42	224.27
210	13.44583333	48.72916667	0.49	0.47	3.29	2.95
211	13.7375	51.4625	0.32	0.47	4.51	4.15
212	6.095833	50.87083	0.32	0.47	1.3	1.31
213	7.254166667	49.67083333	0.55	0.47	1.21	1.16
214	7.279166667	51.0125	0.54	0.47	0.94	0.98
215	7.604166667	52.6875	0.53	0.47	14.02	13.75
216	8.37	48.31	0.69	0.47	1.63	1.71
217	9.77	48.68	0.51	0.47	1.2	1.1
218	10.087194	53.655835	0.5	0.46	2.31	2.29
219	10.8454	53.418912	0.58	0.46	2.09	2.03
220	10.90416667	50.42083333	0.51	0.46	1.06	0.99
221	10.989	50.926	0.53	0.46	4.28	4.39
222	13.57	48.82	0.23	0.46	1.29	1.22
223	13.77296	51.198328	0.36	0.46	1.83	1.98
224	8.070833333	47.82083333	0.35	0.46	1.29	1.27
225	8.862	52.176	0.61	0.46	100.48	103.53
226	9.09	52.15	0.45	0.46	1.37	1.39
227	9.120833333	52.59583333	0.62	0.46	113.45	113.42
228	9.248	47.7296	0.34	0.46	2.43	2.4
229	9.92	51.55	0.45	0.46	4.03	3.64
230	10.41	48.84	0.55	0.45	1.62	1.63
231	10.913473	48.953204	0.48	0.45	5.66	5.43
232	13.7375	51.05416667	0.36	0.45	218.65	220.11
233	14.1625	52.70416667	0.39	0.45	1.72	1.72
234	6.1875	51.02083333	0.25	0.45	2.14	2.14
235	6.804166667	49.94583333	0.62	0.45	1.67	1.92
236	7.229166667	52.7375	0.53	0.45	54.76	53.46
237	7.36	52.11	0.52	0.45	1.45	1.3
238	8.2	48.28	0.43	0.45	4.31	4.53
239	8.2875	50.7625	0.73	0.45	1.37	1.35
240	8.84	47.74	0.34	0.45	1.9	2.04
241	8.922	52.249	0.61	0.45	112.14	112.58
242	9.13	52.96	0.61	0.45	320.94	327.24

243	9.396086	48.072465	0.41	0.45	20.17	20.4
244	9.665742	49.625874	0.43	0.45	8.56	7.3
245	9.83	48.24	0.36	0.45	4.24	5.28
246	9.831213	48.238604	0.22	0.45	2.72	2.64
247	10.315221	50.980701	0.57	0.44	2.69	2.84
248	10.94583333	49.8375	0.48	0.44	33.76	32.69
249	12.059195	51.86316	0.32	0.44	285.36	291.6
250	7.1875	51.7375	0.54	0.44	29.82	29.87
251	7.2375	52.72916667	0.6	0.44	54.05	51.51
252	7.868864	48.340419	0.35	0.44	1.29	1.39
253	8.399592	50.136749	0.56	0.44	0.86	0.85
254	9.18	49.16	0.54	0.44	1.63	1.75
255	9.603865	50.009995	0.6	0.44	92.59	97.35
256	9.904166667	48.2625	0.28	0.44	2.08	2.09
257	10.400272	50.718201	0.65	0.43	1.56	1.77
258	13.61	50.8	0.21	0.43	1.11	1.16
259	6.454167	50.69583	0.44	0.43	7.59	7.47
260	7.24	52.73	0.51	0.43	54.75	53.62
261	8.029166667	50.89583333	0.56	0.43	2.34	2.56
262	8.720833333	52.5625	0.54	0.43	3.86	3.7
263	8.73	48.54	0.52	0.43	1.38	1.39
264	9.229166667	49.7125	0.62	0.43	106.07	112.35
265	9.318533	48.695677	0.24	0.43	1.37	1.44
266	9.438	51.648	0.6	0.43	87.9	87.4
267	9.469	51.227	0.51	0.43	39.95	40.74
268	9.516	51.973	0.57	0.43	96.49	91.36
269	10.57916667	51.45416667	0.45	0.42	0.93	0.96
270	13.07916667	49.17916667	0.24	0.42	1.12	1
271	13.993832	51.022092	0.31	0.42	1.77	1.81
272	14.32083333	51.30416667	0.29	0.42	1.1	1.09
273	7.24852	52.595595	0.58	0.42	34.25	34.06
274	7.7875	47.85416667	0.41	0.42	1.45	1.52
275	8.7125	50.1125	0.6	0.42	127.91	125.25
276	8.870833333	48.25416667	0.43	0.42	1.08	1.02
277	8.92	51.5	0.64	0.42	3.68	3.76
278	9.46	49.24	0.41	0.42	1.76	1.58
279	9.679167	47.90417	0.28	0.42	2.08	2.12
280	10.01	51.87	0.47	0.41	0.88	0.87
281	10.17	48.92	0.53	0.41	0.87	0.89
282	10.22	51.57	0.54	0.41	1.09	1.18
283	10.7125	48.60416667	0.2	0.41	2.86	2.68

284	11.196753	48.888543	0.49	0.41	7.9	7.38
285	11.70417	51.1125	0.5	0.41	26.64	26.16
286	6.9625	52.55416667	0.42	0.41	7.97	7.58
287	7.429167	49.80417	0.58	0.41	2.2	1.94
288	8.43	47.72	0.44	0.41	5.57	5.32
289	8.804166667	51.97916667	0.47	0.41	1.89	2.02
290	9.154166667	48.0875	0.52	0.41	0.95	0.9
291	9.22	49.69	0.62	0.41	3.15	3.2
292	9.2625	48.57083333	0.5	0.41	1.89	1.91
293	9.404167	52.72917	0.34	0.41	0.82	0.82
294	13.35416667	48.8375	0.35	0.4	1.59	1.62
295	7.920833333	51.75416667	0.44	0.4	0.78	0.82
296	8.145833333	48.69583333	0.46	0.4	0.84	0.86
297	8.4625	48.0625	0.53	0.4	1.97	1.79
298	9.1375	51.70417	0.61	0.4	1.31	1.37
299	9.456837	49.249141	0.39	0.4	1.79	1.69
300	9.595833333	52.67916667	0.54	0.4	41.6	40.19
301	9.641	51.426	0.59	0.4	75.28	77.39
302	9.676	52.388	0.56	0.4	34.21	32.63
303	10.72	48.61	0.32	0.39	3.63	3.73
304	11.75416667	50.57083333	0.37	0.39	1.57	1.7
305	7.08	50.07	0.52	0.39	2.25	2.23
306	7.854167	49.8875	0.53	0.39	1.14	1.08
307	8.454167	50.8875	0.62	0.39	1.69	1.29
308	8.8875	48.3875	0.31	0.39	1.56	1.6
309	9.520833333	51.62916667	0.59	0.39	81.69	78.05
310	9.6875	51.25417	0.56	0.39	0.63	0.61
311	9.895833333	48.27083333	0.17	0.39	1.96	1.88
312	10.19583333	53.3375	0.44	0.38	2.29	2.27
313	11.12083	51.45417	0.49	0.38	5.72	5.46
314	11.14	53.84	0.53	0.38	1.44	1.48
315	11.38	49.37	0.56	0.38	1.84	1.48
316	6.46575	49.888832	0.52	0.38	2.8	2.83
317	7.434	52.288	0.54	0.38	29.49	30.98
318	8.720833	50.65417	0.56	0.38	0.98	1.13
319	9.170121	53.339366	0.52	0.38	4.57	4.72
320	9.379167	53.29583	0.54	0.38	2.54	2.56
321	9.6	51.63	0.61	0.38	2.44	2.4
322	10.16	48.63	0.32	0.37	2.63	2.72
323	10.36	47.63	0.16	0.37	1.21	1.19
324	11.92083333	50.32083333	0.62	0.37	4.14	4.34

325	12.432924	50.889017	0.28	0.37	1.92	2.1
326	6.46	49.88	0.38	0.37	3.13	2.82
327	7.82	49.61	0.44	0.37	1.26	1.08
328	8.39	51.04	0.58	0.37	2.28	2.24
329	9.054167	49.8375	0.68	0.37	2.57	2.43
330	9.205753	53.075773	0.54	0.37	6.54	6.78
331	10.220815	50.031196	0.56	0.36	74.36	78.06
332	10.23	48.58	0.25	0.36	2.02	2.32
333	10.27916667	52.4625	0.21	0.36	0.87	0.83
334	10.483	50.917	0.56	0.36	0.67	0.64
335	11.02083	51.7375	0.43	0.36	3.79	3.83
336	11.58	47.74	0.14	0.36	1.13	1.14
337	11.75	50.57	0.36	0.36	1.92	1.7
338	14.37	51.577	0.25	0.36	10.06	10.02
339	14.574671	51.405506	0.2	0.36	3.78	3.48
340	6.779166667	52.07916667	0.53	0.36	3.25	3.28
341	7.495833	52.6375	0.6	0.36	0.73	0.76
342	8.36	48.32	0.59	0.36	1.46	1.29
343	9.78	49.5	0.52	0.36	7.74	9.35
344	9.904166667	49.05416667	0.53	0.36	2.15	2.34
345	10.2125	52.32916667	0.49	0.35	1.73	1.73
346	10.42083333	47.60416667	0.16	0.35	1.21	1.17
347	11.02083333	50.70416667	0.57	0.35	2.26	2.32
348	11.2625	51.8375	0.55	0.35	1.82	1.72
349	11.37	50.3	0.53	0.35	1.91	1.81
350	12.32083	53.6875	0.37	0.35	0.81	0.82
351	14.30417	51.20417	0.16	0.35	0.91	0.8
352	8.089404	51.662965	0.56	0.35	16.93	16.49
353	8.770833333	48.40416667	0.33	0.35	3.18	3.38
354	9.21	53.08	0.53	0.35	6.96	6.57
355	9.709	51.408	0.61	0.35	34.39	34.82
356	10.65667	48.28903	0.13	0.34	1.46	1.39
357	10.80417	51.7375	0.34	0.34	2.46	2.45
358	10.89	47.73	0.14	0.34	1.02	1.03
359	6.48	50.58	0.5	0.34	4.94	4.95
360	7.603	52.094	0.48	0.34	25.24	24.57
361	8.79	51.1	0.63	0.34	2.84	2.93
362	9.151828	48.51144	0.36	0.34	1.77	1.85
363	9.7375	47.64583333	0.29	0.34	5.1	5.09
364	9.773611	49.4925	0.45	0.34	7.12	6.94
365	9.86	48.81	0.46	0.34	1.55	1.73

366	9.974833	51.270925	0.61	0.34	32.75	31.88
367	10.52083333	47.55416667	0.14	0.33	1.02	1.06
368	11.52083333	48.42916667	0.19	0.33	2.97	2.77
369	11.578935	49.947823	0.61	0.33	3.23	3.04
370	12.0875	53.59583	0.31	0.33	0.84	0.83
371	12.432	50.892	0.23	0.33	2.15	2.03
372	12.66502	54.11993	0.36	0.33	1.89	1.99
373	13.412422	48.938176	0.18	0.33	0.84	0.85
374	6.2875	50.8125	0.49	0.33	2.8	2.86
375	7.720833333	51.8875	0.46	0.33	3.87	3.56
376	8.325545	47.623306	0.41	0.33	8.41	7.64
377	8.603951	48.586485	0.59	0.33	2.17	2.12
378	8.89	48.39	0.39	0.33	1.65	1.72
379	9.2625	48.5625	0.44	0.33	2.03	2.04
380	9.740556	48.985379	0.5	0.33	2.03	2.15
381	10.254582	51.651	0.48	0.32	2.66	2.93
382	14.12	51.81	0.19	0.32	1.52	1.46
383	14.944133	51.064069	0.24	0.32	1.6	1.65
384	7.85	48.34	0.27	0.32	1.8	1.4
385	8.7375	51.3875	0.84	0.32	3.12	2.76
386	10.78383	51.506747	0.55	0.31	3.94	3.56
387	11.936343	49.118264	0.55	0.31	34.53	36.85
388	11.94	49.12	0.52	0.31	40.2	39.42
389	12.665	54.1199	0.29	0.31	1.83	1.74
390	8.2875	51.05416667	0.56	0.31	2.96	2.73
391	8.3875	51.02917	0.54	0.31	2.35	2.37
392	8.43	50.5	0.58	0.31	0.84	0.89
393	9.54	50.03	0.67	0.31	2.43	2.46
394	10.1375	48.9375	0.49	0.3	2.66	2.68
395	10.7625	49.92916667	0.57	0.3	82.83	80.12
396	12.3625	48.45416667	0.21	0.3	2.76	3.46
397	12.747222	48.879753	0.38	0.3	293.84	296.74
398	8.795833	51.15417	0.67	0.3	3.82	3.69
399	9.128615	49.345874	0.64	0.3	2.19	2.47
400	9.317255	54.514342	0.4	0.3	4.64	4.65
401	9.445833333	48.70416667	0.4	0.3	3.1	2.98
402	9.495833333	51.1875	0.52	0.3	22.72	21.39
403	9.945833	48.0875	0.16	0.3	0.81	0.82
404	10.0125	50.65416667	0.66	0.29	2.66	2.49
405	10.197	51.125	0.58	0.29	30.54	29.26
406	11.33351	50.716568	0.45	0.29	20.22	20.8

407	6.395395	51.756918	0.45	0.29	1535.2	1513.05
408	8.629166667	48.17083333	0.39	0.29	5.98	6.62
409	8.764489	50.798716	0.44	0.29	16.57	15.96
410	8.84	49.31	0.5	0.29	2.73	2.56
411	9.07	48.49	0.36	0.29	1.78	1.54
412	9.56	49.29	0.44	0.29	1.26	1.2
413	9.72	51.004	0.51	0.29	19.44	20.2
414	9.945833333	51.8625	0.49	0.29	24.95	24.33
415	10.0875	51.70416667	0.55	0.28	11.25	10.91
416	12.04	48.25	0.14	0.28	1.21	1.29
417	12.138699	49.023576	0.38	0.28	288.91	292.58
418	6.7625	51.22916667	0.45	0.28	1436.46	1465.7
419	7.910784	49.911575	0.37	0.28	34.68	31.83
420	8.355248	50.545162	0.45	0.28	35.56	34.77
421	8.728344	51.38311	0.15	0.28	1.85	1.73
422	9.4625	48.9875	0.48	0.28	2.71	3.08
423	9.595833	47.77917	0.29	0.28	5.94	6.15
424	10.92083	51.7375	0.29	0.27	3.67	3.69
425	12.18	47.75	0.07	0.27	1.23	1.18
426	13.21	50.68	0.3	0.27	1.96	2.08
427	13.24	49.01	0.45	0.27	2.16	2.14
428	14.406338	51.164241	0.28	0.27	2.53	2.57
429	6.970833333	50.9375	0.45	0.27	1436.65	1429.94
430	7.88	51.84	0.51	0.27	0.98	0.97
431	8.386279	51.031445	0.51	0.27	2.5	2.43
432	8.47	50.92	0.6	0.27	2.07	1.95
433	8.56	52.36	0.08	0.27	1.35	1.29
434	9.86	49.13	0.44	0.27	1.86	1.88
435	10.46	50.72	0.52	0.26	1.6	1.55
436	11.87083	52.07917	0.43	0.26	1.06	1.03
437	13.07852	50.631093	0.33	0.26	2.66	2.76
438	6.151	51.684	0.62	0.26	4.93	4.69
439	7.581097	51.435987	0.49	0.26	25.8	25.86
440	8.138863	49.350089	0.3	0.26	1.58	1.58
441	8.68	49.29	0.28	0.26	0.84	0.81
442	8.729503	48.897062	0.49	0.26	14.36	14.87
443	9.62	50.2	0.72	0.26	5.04	4.95
444	9.851501	49.140837	0.49	0.26	3.18	3.31
445	10.0375	47.87083333	0.38	0.25	3.93	3.95
446	10.692384	48.783235	0.4	0.25	13.74	12.98
447	11.53	48.43	0.21	0.25	4.15	3.95

448	12.693803	48.606248	0.19	0.25	5.73	5.73
449	13.115161	48.676623	0.33	0.25	422.81	428.69
450	13.36914	48.62954	0.58	0.25	3.21	3.53
451	14.42916667	51.07916667	0.27	0.25	1.94	1.9
452	7.39205	50.443386	0.42	0.25	1386.13	1391.32
453	7.654167	49.6875	0.4	0.25	10.73	11.09
454	8.82	49.31	0.57	0.25	2.36	2.5
455	9.08	48.96	0.23	0.25	1.83	1.8
456	9.720833333	53.3875	0.36	0.25	0.98	0.98
457	9.9875	50.82083333	0.59	0.25	4.62	4.3
458	10.04	47.87	0.38	0.24	4.06	4.04
459	10.47083333	50.54583333	0.66	0.24	3.72	3.57
460	10.65	48.29	0.16	0.24	2.16	2.46
461	14.42083333	51.17916667	0.22	0.24	2.74	2.17
462	7.579167	51.4375	0.51	0.24	27.09	26.77
463	8.1	51.11	0.6	0.24	4.23	4.36
464	8.709661	48.823801	0.53	0.24	7.93	7.28
465	8.945812	50.132082	0.63	0.24	8.86	8.65
466	9.458716	48.984121	0.44	0.24	2.8	2.46
467	9.5625	49.2875	0.46	0.24	1.11	1.12
468	9.74	48.99	0.57	0.24	2.76	2.94
469	12.43	48.92	0.44	0.23	6.02	3.89
470	7.5625	50.6375	0.7	0.23	2.62	2.84
471	8.3375	47.62916667	0.44	0.23	8.89	8.23
472	8.73	48.9	0.44	0.23	15.16	15.32
473	9.39	48.67	0.26	0.23	1.91	1.76
474	9.77	49.5	0.34	0.23	8.24	8.69
475	12.014935	48.946957	0.27	0.22	240.98	240.36
476	12.37916667	48.47083333	0.24	0.22	4.22	4.79
477	12.79583333	47.70416667	0.11	0.22	1.58	1.6
478	14.89583333	51.27083333	0.26	0.22	1.39	1.32
479	7.229167	52.29583	0.43	0.22	1.32	1.38
480	7.764967	50.085444	0.34	0.22	1149.93	1160.54
481	8.279166667	47.69583333	0.53	0.22	2.09	2.01
482	9.129167	51.9125	0.52	0.22	4.48	4.29
483	10.83	47.65	0.06	0.21	1.2	1.11
484	10.94583333	52.45416667	0.5	0.21	2.39	2.33
485	12.53	48.28	0.22	0.21	4.97	5.77
486	13.09	48.8	0.58	0.21	0.88	0.86
487	13.3875	50.8875	0.37	0.21	3.22	3.32
488	13.93	51.26	0.26	0.21	1.05	1.08

489	6.857059	52.602223	0.51	0.21	17.02	17.24
490	8.295833333	48.00416667	0.47	0.21	0.77	0.78
491	8.4625	50.5875	0.6	0.21	10.37	10.74
492	9.0875	51.1625	0.26	0.21	18.59	18.93
493	11.05416667	50.55416667	0.49	0.2	2.8	2.59
494	11.21	50.23	0.64	0.2	3.76	3.72
495	6.604166667	50.12916667	0.51	0.2	7.1	7.59
496	6.89	50.14	0.49	0.2	0.74	0.73
497	7.1875	49.1375	0.44	0.2	16.77	16.2
498	8.620833	52.07083	0.54	0.2	2.82	3.06
499	9.5348	47.6732	0.29	0.2	9.16	8.29
500	9.970453	51.378703	0.48	0.2	3.59	3.67
501	10.50053	48.568397	0.27	0.19	114.23	111.88
502	13.48	51.56	0.46	0.19	2.46	2.42
503	7.404167	50.20417	0.43	0.19	0.96	0.89
504	7.445735	50.50212	0.6	0.19	8.24	9.06
505	8.22	48.77	0.08	0.19	1.81	1.87
506	8.275313	50.003988	0.3	0.19	1135.31	1135.75
507	8.29	47.88	0.48	0.19	0.75	0.81
508	8.670932	52.133419	0.46	0.19	10.5	10.25
509	9.005032	49.438249	0.41	0.19	118.93	123.13
510	9.219941	49.268549	0.49	0.19	18.22	18.22
511	9.6375	50.59583333	0.62	0.19	6.65	6.4
512	9.6625	54.80416667	0.13	0.19	0.67	0.67
513	9.720833	50.8625	0.46	0.19	22.38	19.49
514	12.02	50.25	0.56	0.18	0.78	0.75
515	12.15395	49.535609	0.55	0.18	18.19	17.41
516	13.22	52.13	0.36	0.18	1.17	1.13
517	6.620833333	49.72916667	0.48	0.18	274.86	285.82
518	7.73	52.77	0.26	0.18	1.01	1.01
519	8.57	48.14	0.63	0.18	3.72	3.85
520	8.679166667	49.55416667	0.54	0.18	2.98	3
521	8.78	50.33	0.35	0.18	1.46	1.28
522	8.8625	47.7875	0.41	0.18	1.9	1.85
523	8.945833333	51.7625	0.19	0.18	0.88	0.82
524	9.3875	53.5375	0.39	0.18	0.71	0.61
525	10.0625	47.7625	0.38	0.17	2.05	2.1
526	10.96	50.26	0.66	0.17	4.92	4.4
527	11.865229	48.460684	0.16	0.17	30.45	30.25
528	7.7125	49.77916667	0.39	0.17	32.73	33.34
529	8.36	48.51	0.57	0.17	2.58	2.58

530	9.129167	50.2375	0.51	0.17	0.66	0.61
531	9.545833333	47.67916667	0.29	0.17	9.35	9.29
532	9.6625	47.92916667	0.11	0.17	2.24	2.27
533	9.6875	48.2125	0.18	0.17	0.88	0.85
534	12.15416667	49.5375	0.5	0.16	21.81	20.76
535	6.592735	49.342302	0.37	0.16	17.65	16.14
536	6.91	49.41	0.52	0.16	3.73	3.89
537	7.168276	50.143351	0.48	0.16	321.38	309.35
538	8.229166667	52.0375	0.46	0.16	0.73	0.67
539	9.534877	47.672577	0.25	0.16	9.37	9.04
540	9.83928	48.622759	0.56	0.16	2.55	2.47
541	10.69583333	48.7875	0.31	0.15	19.1	17.83
542	13.27	51.06	0.39	0.15	6.32	6.46
543	7.2375	52.0625	0.27	0.15	0.63	0.67
544	9.829166667	48.62916667	0.44	0.15	2.84	2.89
545	10.72	50.5	0.62	0.14	4.53	4.61
546	10.87083333	49.94583333	0.54	0.14	44.45	44.01
547	12.27916667	49.72083333	0.25	0.14	2.04	2.13
548	6.6	49.35	0.37	0.14	17.45	17.04
549	6.648422	49.408927	0.37	0.14	80.35	78.08
550	7.3875	52.4125	0.48	0.14	2.82	2.82
551	8.22	47.86	0.52	0.14	2.74	2.5
552	8.280575	51.347759	0.57	0.14	8.56	8.41
553	8.72	48.56	0.47	0.14	2.07	2.07
554	9.3125	48.85416667	0.41	0.14	7.47	9.66
555	9.382104	48.675446	0.31	0.14	37.27	35.46
556	10.58	47.75	0.26	0.13	2.7	2.75
557	10.81	51.48	0.34	0.13	1.71	1.82
558	7.5625	51.22083	0.54	0.13	2.85	2.9
559	8.129166667	49.35416667	0.27	0.13	1.68	1.65
560	8.627985	49.813865	0.45	0.13	0.81	0.83
561	8.7125	52.19583	0.48	0.13	17.47	16.42
562	9.1875	50.19583	0.61	0.13	8.55	8.41
563	9.315816	48.84736	0.42	0.13	7.18	7.52
564	9.419	48.707	0.35	0.13	45.5	45.46
565	10.41	50.787	0.42	0.12	0.7	0.71
566	10.4125	50.57083333	0.61	0.12	14	13.57
567	10.74	53.39	0.55	0.12	1.18	1.11
568	7.1625	51.40416667	0.49	0.12	72.8	73.41
569	8.53	50.91	0.57	0.12	7.18	7.08
570	9.154166667	48.0125	0.34	0.12	1.85	1.85

571	9.5625	53.27917	0.44	0.12	0.72	0.72
572	9.75	48.69	0.61	0.12	5.67	5.53
573	10.722448	50.496464	0.55	0.11	4.7	4.49
574	11.78	47.69	0.06	0.11	1.73	1.7
575	11.87916667	48.4625	0.11	0.11	34.22	37.3
576	6.7625	51.7375	0.42	0.11	0.84	0.81
577	8.1375	51.39583333	0.53	0.11	15.83	15.66
578	8.154166667	47.90416667	0.48	0.11	1.05	1.05
579	8.212933	48.24	0.62	0.11	4.26	4.27
580	8.579166667	48.37083333	0.56	0.11	5.4	4.93
581	8.679166667	47.92083333	0.44	0.11	13.47	14.05
582	8.898182	51.155679	0.48	0.11	24.31	23.49
583	9.71	49.53	0.49	0.11	0.87	0.82
584	11.422	48.754	0.21	0.1	230.32	226.79
585	12.129441	49.127955	0.36	0.1	32.27	32.87
586	7.445833333	50.50416667	0.49	0.1	11.82	11.2
587	8.61	49.12	0.12	0.1	1.39	1.39
588	9.8625	48.25416667	0.22	0.1	2.82	2.71
589	14.159127	52.693351	0.37	0.09	2.32	2.23
590	8.125898	48.584384	0.41	0.09	2.24	1.85
591	9.261206	48.959297	0.34	0.09	7.08	7.4
592	9.3875	47.94583333	0.34	0.09	0.98	0.98
593	9.82	48.63	0.39	0.09	3.6	3.28
594	6.369035	49.472575	0.39	0.08	173.23	163.84
595	9.079166667	54.72083333	0.39	0.08	1.1	1.09
596	9.120833	50.4625	0.58	0.08	0.57	0.58
597	12.581	51.591	0.32	0.07	65.64	62.84
598	8.376019	49.64112	0.2	0.07	1052.81	1045.01
599	9.9875	53.3125	0.46	0.07	0.84	0.82
600	12.13	49.13	0.28	0.06	34.23	35.01
601	13.25416667	51.32083333	0.27	0.06	1.29	1.29
602	14.79583333	50.89583333	0.42	0.06	4.46	5.16
603	7.1625	51.25416667	0.46	0.06	8.03	7.9
604	7.970833333	50.84583333	0.49	0.06	10.26	11.74
605	8.354166667	47.9875	0.53	0.06	5.65	5.56
606	8.623725	51.035299	0.5	0.06	13.16	15.34
607	8.7	48.88	0.45	0.06	17.88	18.26
608	14.802623	50.891976	0.36	0.05	4.79	4.53
609	6.7875	49.37083333	0.49	0.05	12.42	12.71
610	8.58	48.81	0.16	0.05	5.12	5.01
611	9.1125	49.0625	0.22	0.05	2.85	3.26

612	10.41	47.59	0.13	0.04	1.96	1.9
613	11.642973	50.870692	0.15	0.04	1.53	1.48
614	13.853698	50.953832	0.19	0.04	4.2	3.08
615	14.1375	50.94583333	0.33	0.04	3.08	2.98
616	7.52	51.33	0.57	0.04	6.05	6.04
617	8.245833333	47.72916667	0.39	0.04	3.14	3.23
618	9.52	48.81	0.23	0.04	6.63	7.12
619	10.0148	48.9211	0.41	0.03	1.03	0.94
620	11.226675	50.181522	0.53	0.03	13.08	13.6
621	13.16	50.78	0.3	0.03	9.47	9.64
622	13.237261	50.703663	0.31	0.03	6.55	6.69
623	13.23	49.03	0.26	0.02	4.68	4.62
624	13.2375	50.70416667	0.3	0.02	6.61	6.54
625	14.13386	50.941862	0.31	0.02	3.1	3.2
626	6.78	49.37	0.45	0.02	13.87	13.6
627	7.329166667	50.7875	0.6	0.02	5.55	5.2
628	8.145405	48.490087	0.6	0.02	3.86	4.84
629	9.768198	49.001552	0.53	0.02	12.41	12.11
630	7.270833333	52.2875	0.52	0.01	2.38	2.3
631	8.6125	48.5875	0.28	0.01	4.9	4.47
632	9.1625	48.00416667	0.38	0.01	1.84	1.88
633	9.23	49.28	0.39	0.01	27.03	27.89
634	9.287113	49.258277	0.42	0.01	29.91	26.91
635	9.6	47.82	0.29	0.01	0.98	0.88
636	11.32083	52.0125	0.24	0	12.31	12.09
637	12.493608	50.740902	0.27	0	14.99	16.13
638	7.4375	50.77916667	0.48	0	34.89	35.73
639	7.442713	50.776982	0.47	0	34.72	37.61
640	7.91	47.61	0.31	0	4.51	4.66
641	8.1	47.88	0.21	0	0.95	0.93
642	8.745833	51.72917	0.25	0	3.66	3.66
643	8.94	50.8	0.56	0	8.62	8.49
644	10.3	47.39	0	-0.01	1.91	1.9
645	11.3625	52.85417	0.38	-0.01	0.63	0.66
646	12.9125	49.0625	0.23	-0.01	2.57	2.75
647	8.217126	48.317024	0.56	-0.01	5.31	5.06
648	8.9625	54.77083333	0.26	-0.01	1.96	1.91
649	11.12083333	47.92916667	0.04	-0.02	2.26	1.69
650	12.77255	51.006826	0.24	-0.02	27.44	28.24
651	13.15416667	51.07916667	0.42	-0.02	3.91	3.85
652	13.76603	50.843051	0.2	-0.02	0.82	0.76

653	14.8	51.27	0.19	-0.02	1.31	1.26
654	8.100414	51.10887	0.24	-0.02	7.89	7.63
655	8.8125	49.32083333	0.41	-0.02	2.13	2.08
656	8.904166667	50.5375	0.49	-0.02	4.42	4.21
657	12.71	50.59	0.25	-0.03	7.27	6.95
658	13.44583333	48.6875	0.35	-0.03	17.03	15.33
659	13.52083333	48.72916667	0.2	-0.03	2.7	2.33
660	7.204166667	50.8375	0.47	-0.03	19.56	19.38
661	7.345833	49.70417	0.46	-0.03	14.06	13.33
662	7.4875	51.70417	0.39	-0.03	0.73	0.68
663	8.1375	48.57916667	0.47	-0.03	3.57	3.64
664	8.305563	49.038985	0.12	-0.03	985.66	995.15
665	13.02	50.89	0.27	-0.04	24.23	23.55
666	13.35416667	48.82083333	0.12	-0.04	4.7	8.15
667	13.504361	48.5824	0.08	-0.04	1090.87	1087.82
668	8.1375	48.2875	0.5	-0.04	18.53	18.78
669	8.22	48.31	0.61	-0.04	7.82	7.76
670	8.745833333	47.99583333	0.33	-0.04	3.6	3.65
671	9.11	50.49	0.5	-0.04	3.21	3.18
672	10.831	50.713	0.43	-0.05	0.69	0.7
673	12.060282	50.774489	0.33	-0.05	2.44	2.52
674	12.6875	48.67083333	0.08	-0.05	129.01	130.7
675	12.92	49.28	0.3	-0.05	0.92	0.89
676	7.845408	47.699494	0.53	-0.05	9.13	8.83
677	7.86	47.71	0.54	-0.05	8.8	8.95
678	7.8625	50.79583	0.5	-0.05	21.92	20.84
679	9.3125	52.92083	0.47	-0.05	0.71	0.72
680	11.25416667	47.42916667	0	-0.06	3.44	3.51
681	12.4875	50.4625	0.18	-0.06	2.48	2.22
682	6.995833333	51.07083333	0.47	-0.06	16.64	16.49
683	8.22	48.32	0.55	-0.06	6.25	5.44
684	9.22	49.26	0.44	-0.06	7.56	7.49
685	9.295833333	48.1125	0.24	-0.06	3.82	3.84
686	11.67083333	50.42916667	0.39	-0.07	0.95	0.96
687	12.2625	49.0875	0.43	-0.07	3.74	3.85
688	12.32	47.77	0.05	-0.07	3.05	3.17
689	12.78	51.01	0.2	-0.07	29.13	26.41
690	8.029166667	48.3875	0.45	-0.07	26.91	27.91
691	8.633705	49.156612	0.27	-0.07	1.14	1.04
692	9.54	47.67	0.3	-0.07	18.93	18.82
693	9.779166667	53.70416667	0.28	-0.07	0.71	0.71

694	10.4375	47.6125	0.18	-0.08	3.42	3.32
695	13.45416667	48.69583333	0.4	-0.08	24.31	25.1
696	8.34	48.52	0.35	-0.08	4.22	3.47
697	9.279166667	48.97083333	0.24	-0.08	1.04	1.07
698	12.50416667	50.7125	0.18	-0.09	17.96	15.08
699	7.729166667	47.92083333	0.47	-0.09	5.04	5.38
700	8.03	48.39	0.44	-0.09	28.51	29.34
701	10.013584	53.392938	0.23	-0.1	4.06	4.06
702	12.841559	49.309549	0.37	-0.1	3.89	3.77
703	11.57083	53.6625	0.51	-0.11	1.35	1.29
704	8.279166667	51.34583333	0.38	-0.11	3.29	3.33
705	9.86	47.68	0.35	-0.11	6.7	6.64
706	11.16	52.97	0.49	-0.12	6.5	6.46
707	7.154166667	50.79583333	0.46	-0.12	67.65	67.37
708	8.297143	48.818715	0.48	-0.12	18.49	18.89
709	10.27083333	47.42916667	0	-0.13	4.49	4.55
710	11.2625	50.6875	0.55	-0.13	6.8	6.85
711	11.74583	53.72083	0.47	-0.14	2.1	2.09
712	7.159168	50.797901	0.46	-0.14	72.88	68.75
713	8.354166667	48.62083333	0.48	-0.14	12.52	11.14
714	9.1875	49.0875	0.36	-0.14	0.87	0.92
715	10.1125	49.02083333	0.5	-0.15	5.89	5.34
716	10.14583333	48.92916667	0.61	-0.15	8.74	8.93
717	12.822811	47.681177	0.01	-0.15	37.52	38.2
718	7.920833333	47.65416667	0.58	-0.15	7.87	7.98
719	9.820833333	47.70416667	0.34	-0.15	10.57	9.95
720	12.8125	47.6875	0.01	-0.16	39.28	38
721	9.145833	51.3875	0.57	-0.16	1.07	1.05
722	9.4125	52.8875	0.37	-0.16	0.89	0.88
723	10.56	47.58	0.08	-0.17	5.5	5.66
724	13.23	49.02	0.28	-0.17	8.62	8.32
725	7.604167	53.34583	0.39	-0.17	0.79	0.81
726	12.73	48.53	0.24	-0.18	1.48	1.41
727	6.795833333	52.10416667	0.41	-0.18	4.78	4.85
728	9.5125	48.17083333	0.36	-0.18	1.41	1.36
729	11.02083	53.4875	0.58	-0.19	0.74	0.81
730	11.9354	47.71	0.07	-0.19	4.57	4.5
731	8.21	47.91	0.49	-0.19	11.18	11.44
732	8.595833	50.35417	0.62	-0.2	0.75	0.77
733	9.74	49.6	0.61	-0.2	12.93	13.33
734	6.829166667	50.77916667	0.38	-0.21	3.5	3.44

735	8.97	48.53	0.2	-0.21	1.1	1.18
736	9.904166667	49.22916667	0.35	-0.21	10.72	10.51
737	11.54583333	47.77916667	0.02	-0.22	28.47	28.33
738	8.854166667	48.27916667	0.37	-0.22	10.75	10.62
739	9.37	49.38	0.62	-0.22	16.03	17.19
740	9.604166667	47.62916667	0.25	-0.22	24.42	23.39
741	8.962187	48.526377	0.18	-0.23	1.24	1.23
742	9.24	49.19	0.36	-0.23	1.11	1.1
743	9.59	50.59	0.63	-0.23	16.73	16.88
744	7.995833333	48.12083333	0.06	-0.24	10.38	10.8
745	9.995833333	48.37083333	0.09	-0.24	57.25	59.38
746	8.3125	47.80416667	0.43	-0.25	7.82	7.83
747	9.4875	47.6625	0.36	-0.25	11.36	10.98
748	10.00416667	48.8875	0.57	-0.26	20.82	21.89
749	7.7625	50.70416667	0.53	-0.26	1.88	1.9
750	7.99	47.92	0.6	-0.26	9.29	9.27
751	12.542419	47.99121	0.12	-0.27	13.49	14.42
752	6.81	51.3	0.14	-0.27	1.26	1.22
753	6.95	52.61	0.58	-0.27	19.84	20.34
754	7.9	47.99	0.5	-0.27	2.46	2.42
755	9.354166667	48.7125	0.5	-0.27	14.86	15.3
756	10.069745	53.576022	0.31	-0.28	0.96	0.94
757	10.89583333	51.25416667	0.47	-0.28	1.98	1.98
758	12.432654	51.376276	0.27	-0.28	1.73	1.71
759	8.2	49.14	0.5	-0.28	0.97	0.95
760	11.0649	47.4843	0.02	-0.29	7.53	7.38
761	12.64583333	47.82916667	0.16	-0.29	12.36	13.18
762	7.2125	50.85416667	0.56	-0.29	22.97	23.14
763	10.30417	52.67083	0.69	-0.3	0.85	0.81
764	10.888009	48.406968	0.03	-0.3	104.66	103.47
765	13.41	48.94	0.42	-0.3	9.52	9.3
766	11.66	49.91	0.58	-0.31	0.88	0.91
767	12.55	47.99	0.09	-0.31	15.14	13.78
768	12.715356	50.59635	0.24	-0.31	8.14	8.64
769	10.30416667	47.35416667	0	-0.32	1.09	1.04
770	6.445833	50.3625	0.55	-0.32	0.86	0.85
771	8.55	48.15	0.64	-0.32	24.88	25.39
772	8.954167	50.5125	0.51	-0.32	0.56	0.57
773	9.39	48.89	0.18	-0.32	0.99	0.92
774	9.83	48.63	0.61	-0.33	25.9	26.11
775	13.000042	49.042548	0.36	-0.34	17.02	16.51

776	8.74	48.72	0.44	-0.34	3.7	3.59
777	10.464	53.15	0.22	-0.35	9.56	9.27
778	12.62	52.27	0.29	-0.35	1.65	1.72
779	12.94583333	47.6125	0	-0.35	4.94	4.9
780	13.9875	53.52083	0.41	-0.35	5.59	5.51
781	9.26	48.96	0.39	-0.36	2.7	2.74
782	11.1125	47.47916667	0	-0.37	4.24	4.13
783	12.48	47.78	0.05	-0.37	38.69	39.33
784	9.145833333	48.52916667	0.48	-0.37	35.61	35.62
785	9.720833333	51.00416667	0.5	-0.37	11.16	11.24
786	8.470833333	49.94583333	0.29	-0.38	0.69	0.68
787	11.12916667	50.17083333	0.55	-0.4	44.82	46.58
788	11.57083333	47.67916667	0.02	-0.4	24.4	26
789	7.745833333	50.37916667	0.48	-0.4	29.16	30.36
790	8.979166667	48.9125	0.56	-0.4	48.03	50.69
791	10.316864	47.730411	0.06	-0.41	53.41	54.31
792	6.095833333	51.09583333	0.34	-0.41	25.1	24.99
793	8.0375	51.2875	0.53	-0.41	0.58	0.59
794	10.2375	47.4125	0.02	-0.42	9.41	9.49
795	9.129166667	49.34583333	0.52	-0.42	1.31	1.31
796	9.145833333	49.07916667	0.55	-0.42	116.7	116.62
797	10.32916667	47.50416667	0.02	-0.43	9.45	9.27
798	11.07916667	47.49583333	0.01	-0.44	12.52	12.41
799	12.9625	48.85416667	0.34	-0.44	0.96	0.92
800	10.87083333	48.0375	0	-0.45	80.22	79.19
801	14.92083	51.15417	0.19	-0.45	0.86	0.88
802	9.926	49.796	0.38	-0.45	149.96	150.99
803	11.759765	52.990829	0.38	-0.46	833.45	823.91
804	11.270803	47.443334	0.01	-0.47	12.16	12.26
805	7.54	49.19	0.43	-0.47	0.73	0.79
806	10.7875	51.8375	0.31	-0.49	0.7	0.75
807	10.88	53.23	0.29	-0.49	1204.3	1201.42
808	10.799034	47.696642	0	-0.5	75.57	73.42
809	13.8625	52.44583333	0.15	-0.5	35.15	34.82
810	11.14583333	47.8375	0.05	-0.51	21.54	17.1
811	11.9375	51.57916667	0.19	-0.51	121.68	121.64
812	13.01	47.64	0	-0.51	18.65	18.09
813	8.448717	49.323807	0.18	-0.52	1375.45	1372.84
814	9.579166667	48.25416667	0.34	-0.52	2.38	2.41
815	11.8125	51.92916667	0.15	-0.53	144.04	144.14
816	6.570833333	50.25416667	0.57	-0.53	0.74	0.76

817	8.46	49.94	0.32	-0.53	0.69	0.72
818	11.79583	51.7625	0.33	-0.54	2.03	2
819	9.895833333	53.75416667	0.31	-0.54	0.77	0.83
820	8.0125	47.94583333	0.49	-0.55	1.05	1.05
821	8.654167	53.3375	0.27	-0.55	0.75	0.71
822	8.66	49.07	0.31	-0.55	1.22	1.19
823	6.84	50.37	0.05	-0.56	4.48	4.49
824	9.145833333	50.4125	0.64	-0.56	0.97	0.98
825	10.94583333	51.57916667	0.26	-0.57	1.05	0.96
826	12.85	49.17	0.45	-0.57	1.7	1.77
827	13.5375	52.25417	0.29	-0.58	1.61	1.62
828	13.77	50.94	0.31	-0.59	0.88	0.92
829	11.09583333	47.5125	0	-0.62	18.27	18.47
830	8.295833333	48.29583333	0.32	-0.62	2.44	2.38
831	11.3625	47.67083333	0.01	-0.63	52.3	52.36
832	13.3875	48.85416667	0.13	-0.63	1.26	1.28
833	6.845833333	50.7625	0.22	-0.63	2.02	1.91
834	8.05	48.34	0.55	-0.63	1.17	1.15
835	6.879166667	51.87083333	0.52	-0.64	1.04	1.04
836	8.1	49.65	0.45	-0.66	1.21	1.2
837	6.9125	50.07083333	0.09	-0.67	0.96	0.93
838	8.6875	50.59583333	0.43	-0.67	1.25	1.24
839	13.8125	52.8375	0	-0.68	506.58	503.59
840	11.3875	52.5375	0.24	-0.69	0.72	0.68
841	9.72	49.37	0.34	-0.7	5.53	5.29
842	10.12917	52.89583	0.34	-0.71	1.06	1.07
843	10.65416667	47.8375	0.15	-0.73	4.44	4.48
844	12.22916667	48.0625	0	-0.73	413.59	415.89
845	10.55416667	51.85416667	0.29	-0.74	1.02	0.94
846	13.9625	51.84583333	0	-0.74	205.17	204.72
847	10.62	50.853	0.48	-0.76	0.71	0.75
848	8.370833333	48.60416667	0.53	-0.76	1.82	1.86
849	11.82916667	47.89583333	0.04	-0.77	4.55	4.62
850	6.84604	50.766075	0.26	-0.77	1.95	1.88
851	7.1375	51.37083	0.5	-0.77	0.98	0.95
852	10.04583333	49.4375	0.46	-0.78	0.9	0.87
853	8.37	48.26	0.38	-0.78	1.33	1.32
854	9.0375	51.35417	0.67	-0.78	1.31	1.25
855	10.3125	49.02916667	0.46	-0.8	0.88	0.87
856	6.845833333	50.7625	0.32	-0.8	1.76	1.77
857	10.811	51.469	0.52	-0.81	3.01	3.02

858	7.7625	50.82083333	0.48	-0.81	0.68	0.74
859	9.05	49.49	0.48	-0.81	0.73	0.71
860	6.945833333	51.15416667	0.23	-0.83	1.23	1.22
861	6.954166667	50.6625	0.38	-0.84	1.03	1.02
862	13.95417	52.4625	0.18	-0.86	0.8	0.75
863	10.02083333	49.09583333	0.35	-0.92	0.91	0.89
864	9.4125	48.7125	0.34	-0.92	23.24	22.29
865	10.84583333	54.15416667	0.3	-0.96	0.73	0.71
866	6.454166667	50.17916667	0.42	-0.97	0.79	0.73
867	9.320833333	54.12916667	0.26	-0.98	0.81	0.81
868	10.07083333	49.45416667	0.52	-0.99	1.82	1.8
869	8.93	47.84	0.12	-1	1.08	1.05
870	11.38	49.83	0.38	-1.01	0.89	0.87
871	8.71	48.87	0.2	-1.01	2.89	2.94
872	6.97	50.51	0.39	-1.02	0.94	0.95
873	9.079167	50.34583	0.39	-1.03	0.68	0.67
874	8.1625	48.15416667	0.31	-1.04	0.97	0.97
875	12.45416667	49.4875	0.48	-1.07	0.78	0.76
876	8.65	48.16	0.36	-1.09	3.76	3.82
877	6.77	50.76	0.23	-1.13	1.79	1.75
878	9.76	49.26	0.43	-1.15	6.24	6.28
879	7.7875	49.72917	0.37	-1.18	0.85	0.88
880	9.870833333	48.4125	0.26	-1.19	1.84	1.8
881	10.987	50.889	0.28	-1.21	1.38	1.49
882	13.86	51.94	0.47	-1.24	3.09	3.03
883	11.95417	52.24583	0.32	-1.25	0.92	0.94
884	13.3625	49.0125	0.09	-1.25	0.92	0.94
885	10.61	50.8	0.34	-1.3	0.94	0.97
886	7.26	52.13	0.35	-1.3	0.78	0.81
887	8.8	47.99	0.18	-1.31	1.28	1.26
888	8.920833	51.7125	0.49	-1.32	0.71	0.72
889	8.145833333	47.75416667	0.31	-1.34	1.74	1.74
890	12.02917	51.4375	0.13	-1.37	0.94	0.92
891	6.645833333	51.12916667	0.38	-1.4	7.67	7.72
892	11.85417	52.27917	0.38	-1.41	0.98	0.97
893	7.320833333	50.77916667	0.39	-1.44	12.13	12.19
894	11.43	50.62	0.44	-1.46	11.1	11.12
895	9.420833	51.49583	0.39	-1.51	0.82	0.84
896	13.8	52.42	0.39	-1.52	1.33	1.28
897	10.12	49.08	0.3	-1.53	0.96	0.9
898	11.8875	51.77917	0.07	-1.54	0.7	0.69

899	12.49	53.16	0.22	-1.55	1.16	1.12
900	11.7875	52.57917	0.33	-1.58	0.85	0.85
901	9.770833333	49.00416667	0.45	-1.58	3.97	4
902	9.94	48.8	0.31	-1.6	0.89	0.85
903	9.495833333	49.37083333	0.25	-1.64	1.02	0.97
904	7.695833333	49.8875	0.5	-1.68	0.93	0.98
905	12.06	50.77	0.2	-1.76	2.98	2.87
906	10.5375	53.4375	0.36	-1.77	0.93	0.93
907	10.17916667	50.04583333	0.47	-1.8	1.08	1.11
908	11.979419	48.413618	0.15	-1.83	4.93	4.84
909	9.145833	51.9125	0.69	-1.83	2.44	2.45
910	12.32917	52.27083	0.13	-1.92	0.92	0.98
911	12.1125	51.62083	0.19	-1.93	1.05	0.96
912	7.9375	47.90416667	0.27	-1.94	1.23	1.21
913	9.9	48.42	0.42	-1.94	2	1.98
914	13.997	52.368	0.23	-2.07	26.41	26.69
915	8.704166667	48.8875	0.26	-2.09	10.02	9.96
916	11.432152	50.619021	0.06	-2.1	12.74	12.79
917	11.98	48.42	0.14	-2.16	3.51	3.52
918	12.47083	50.62083	0.18	-2.2	0.8	0.8
919	11.00417	51.44583	0.64	-2.21	3.94	3.91
920	6.2625	50.9625	0.15	-2.21	1.11	1.13
921	6.6375	51.2125	0.1	-2.22	0.68	0.66
922	9.04	48.56	0.01	-2.28	1.14	1.13
923	12.0375	53.3875	0.4	-2.3	0.9	0.84
924	10.65417	53.0125	0.55	-2.42	1.37	1.42
925	10.2625	48.90416667	0.3	-2.48	0.94	0.87
926	10.77083333	50.6875	0.13	-2.49	0.98	0.95
927	9.604166667	54.00416667	0.21	-2.49	0.73	0.69
928	11.6125	51.22083	0.08	-2.51	0.97	0.93
929	9.895833	53.27083	0.43	-2.53	1.04	0.99
930	12.2625	51.92083	0.28	-2.56	0.88	0.88
931	8.895833333	48.72083333	0.18	-2.58	2	2.03
932	12.27083333	49.0625	0.43	-2.64	1.19	1.19
933	10.37083333	49.3875	0.31	-2.72	0.93	0.9
934	9.779166667	53.27916667	0.23	-2.89	0.95	0.93
935	12.34	52.22	0.07	-2.92	0.98	1.03
936	8.020833333	49.77916667	0.38	-2.92	1.08	1.12
937	6.279167	51.35417	0.44	-2.95	1.41	1.4
938	10.47083333	52.90416667	0.15	-3.05	0.93	0.97
939	8.58	48.29	0.29	-3.13	5.96	6.09

940	9.757545	51.366474	0.37	-3.26	0.77	0.77
941	12.52083333	50.42916667	0.18	-3.27	0.93	0.9
942	8.79	50.31	0.47	-3.45	3.4	3.41
943	8.6	50.46	0.37	-3.48	0.94	0.94
944	8.51	48.57	0.45	-3.49	1.3	1.28
945	6.270833	51.1625	0.25	-3.58	1.03	1.03
946	8.8875	51.2375	0.37	-3.62	0.85	0.86
947	12.72917	52.17083	0.35	-3.66	1.44	1.4
948	11.8125	53.5625	0.42	-3.67	1.24	1.2
949	12.09583	52.00417	0.28	-3.75	1.34	1.35
950	12.35416667	53.3125	0.3	-3.82	0.91	0.88
951	10.085826	50.83704	0.54	-3.9	11.49	11.31
952	9.795833333	49.42916667	0.32	-4.1	1.02	1.01
953	8.945833333	49.30416667	0.43	-4.24	1.49	1.48
954	8.18	49.47	0.08	-4.29	1.16	1.15
955	8.395833333	48.1125	0.48	-4.3	1.12	1.11
956	9.76	51.37	0.39	-4.54	1.06	1.03
957	11.55416667	50.07083333	0.59	-4.57	2.4	2.4
958	8.954166667	48.0375	0.32	-4.6	11.92	11.75
959	14.3375	52.5875	0.19	-4.7	1.3	1.19
960	11.7375	51.6125	0.03	-4.79	1.14	1.16
961	14.4375	51.27917	0.06	-4.8	0.76	0.69
962	12.23	50.51	0.4	-4.89	1	0.98
963	9.46	49.13	0.38	-4.99	0.88	0.89
964	11.6625	52.57917	0.26	-5.01	0.73	0.78
965	10.08	49.14	0.23	-5.39	1	0.99
966	9.9625	50.2125	0.49	-5.46	1.15	1.09
967	13.65416667	50.77916667	0.24	-5.53	1.07	1.03
968	9.58	48.9	0.38	-5.9	1.08	1.1
969	10.94583	53.52083	0.24	-5.92	1.56	1.57
970	9.332153	49.234982	0.26	-6.02	11.25	11.42
971	13.62916667	51.4625	0.41	-6.16	7.59	7.6
972	12.8375	49.30416667	0.47	-6.33	2.3	2.3
973	10.37916667	48.49583333	0.2	-6.56	81.16	81.35
974	10.2625	48.4625	0.16	-7.35	65.35	66.9
975	10.3625	48.4875	0.21	-7.35	77.2	76.41
976	12.77083	51.85417	0.24	-7.59	1.45	1.39
977	6.320833333	51.52916667	0.52	-7.96	6.95	6.71
978	9.3375	48.6375	0.31	-8.01	22.83	23.13
979	8.14	49.07	0.21	-8.72	0.99	0.94
980	11.753	50.52	0.42	-8.78	4.07	4

981	9.129166667	48.5375	0.19	-9.02	19.2	18.88
982	9.71	49.58	0.46	-9.45	5.93	5.76
983	7.645833	49.67917	0.31	-9.57	7.98	8.02
984	10.0125	48.3375	0.13	-9.78	18.76	18.76
985	7.7875	50.0375	0.26	-10.04	0.95	0.97
986	6.554166667	51.22083333	0.13	-10.09	0.83	0.82
987	14.22917	52.05417	0.07	-10.14	0.82	0.86
988	6.270833	51.27083	0.13	-11.1	0.96	0.98
989	6.2875	51.1375	0.14	-11.58	0.81	0.81
990	10.49	52.34	0.41	-12.19	4.21	4.21
991	11.542917	50.769252	0.18	-13.82	21.47	21.6
992	9.06	48.94	0.18	-13.9	14.89	14.14
993	10.10416667	49.0125	0.37	-13.93	2.51	2.48
994	9.2875	49.25416667	0.36	-15.69	13.34	12.78
995	12.63	52.06	0.06	-17.36	1.11	1.14
996	12.92	47.77	0.09	-17.74	25.6	25.57
997	7.2125	50.77916667	0.29	-22.01	17.83	18.1
998	13.4375	51.87083	0.13	-22.82	1	0.99
999	10.90417	51.8625	0.25	-24.43	1.64	1.66
1000	11.55416667	47.5875	0.05	-27.87	8.5	8.59
1001	9.57	47.75	0.14	-33.74	6.92	6.88
1002	6.179167	51.00417	0.03	-34.1	3.18	3.14
1003	11.55416667	47.7625	0.06	-34.16	13.37	13.42
1004	12.4875	52.37916667	0.36	-39.03	49.06	49.65
1005	6.270833333	50.70416667	0.31	-40.68	1.17	1.2
1006	9.37	48.68	0.22	-41.09	24.98	24.2
1007	8.74	49.12	0.09	-41.11	1.74	1.7
1008	7.604166667	49.62916667	0.04	-45.22	2.78	2.77
1009	12.5125	52.3875	0.35	-67.99	86.51	87.47
1010	10.45416667	51.84583333	0.28	-74.71	1.68	1.59
1011	6.929167	51.3625	0.32	-123.1	46.21	45.56
1012	7.045833	51.39583	0.31	-161.06	44.42	44.55
1013	12.1625	47.70416667	0.01	-389.48	95.88	96.66

Table S2.3: GLM residual deviance and AIC values from both observed and simulated models. Full species list (n=34) and no of occurrences within each KGE threshold modelled.

species	kge_threshold	no_presences	no_of_sites	sim_res_deviance	sim_AIC	sim_df_res	obs_res_deviance	obs_AIC	obs_df_res
Anabolia nervosa	low	42	108	140.52	150.52	103	132.56	142.56	103
Anabolia nervosa	mid	42	116	145.06	155.06	111	142.92	152.92	111
Anabolia nervosa	high	50	103	139.40	149.40	98	139.51	149.51	98
Asellus aquaticus	low	63	108	130.18	140.18	103	130.59	140.59	103
Asellus aquaticus	mid	69	116	146.50	156.50	111	154.61	164.61	111
Asellus aquaticus	high	61	103	136.58	146.58	98	132.76	142.76	98
Athripsodes albifrons	mid	25	116	115.84	125.84	111	114.71	124.72	111
Athripsodes cinereus	mid	28	116	127.13	137.13	111	126.42	136.42	111
Athripsodes cinereus	high	22	103	101.82	111.82	98	101.44	111.44	98
Baetis fuscatus	low	26	108	115.93	125.93	103	107.93	117.93	103
Baetis fuscatus	mid	48	116	142.65	152.65	111	146.13	156.13	111
Baetis fuscatus	high	36	103	126.93	136.93	98	123.94	133.94	98
Baetis lutheri	mid	34	116	128.31	138.31	111	125.69	135.69	111
Baetis lutheri	high	23	103	102.80	112.80	98	106.45	116.45	98

Baetis muticus	mid	22	116	94.66	104.66	111	100.39	110.39	111
Baetis rhodani	low	86	108	100.96	110.96	103	102.60	112.60	103
Baetis rhodani	mid	90	116	96.49	106.49	111	99.07	109.07	111
Baetis rhodani	high	74	103	100.96	110.96	98	101.82	111.82	98
Baetis scambus	low	27	108	88.94	98.94	103	111.54	121.54	103
Baetis scambus	mid	38	116	123.65	133.65	111	134.45	144.45	111
Baetis scambus	high	35	103	118.13	128.13	98	116.83	126.83	98
Baetis vernus	low	49	108	133.05	143.05	103	120.77	130.77	103
Baetis vernus	mid	56	116	150.10	160.10	111	151.04	161.04	111
Baetis vernus	high	50	103	130.25	140.25	98	130.38	140.38	98
Brachycentrus subnubilus	mid	24	116	117.20	127.20	111	115.59	125.59	111
Brachycentrus subnubilus	high	26	103	101.49	111.49	98	115.01	125.01	98
Caenis luctuosa	mid	23	116	110.88	120.88	111	105.53	115.53	111
Caenis luctuosa	high	22	103	98.62	108.62	98	101.88	111.88	98
Centroptilum luteolum	mid	33	116	133.93	143.93	111	132.77	142.77	111
Centroptilum luteolum	high	26	103	115.81	125.81	98	111.27	121.27	98
Ceraclea dissimilis	mid	21	116	105.79	115.79	111	107.27	117.27	111
Ceraclea dissimilis	high	22	103	98.69	108.69	98	100.54	110.54	98

Ephemera danica	low	55	108	144.51	154.51	103	144.07	154.07	103
Ephemera danica	mid	72	116	141.38	151.38	111	146.06	156.06	111
Ephemera danica	high	54	103	127.50	137.50	98	128.73	138.73	98
Gammarus pulex	low	62	108	134.32	144.32	103	138.23	148.23	103
Gammarus pulex	mid	73	116	145.73	155.73	111	138.94	148.94	111
Gammarus pulex	high	73	103	107.34	117.34	98	107.42	117.42	98
Goera pilosa	mid	26	116	117.87	127.87	111	117.10	127.10	111
Goera pilosa	high	23	103	105.32	115.32	98	98.76	108.76	98
Habroleptoides confusa	low	23	108	107.03	117.03	103	94.66	104.66	103
Habroleptoides confusa	mid	21	116	97.84	107.84	111	101.27	111.27	111
Habrophlebia lauta	mid	25	116	96.34	106.34	111	110.31	120.31	111
Halesus radiatus	low	24	108	100.44	110.44	103	110.67	120.67	103
Halesus radiatus	mid	24	116	110.72	120.72	111	113.08	123.08	111
Heptagenia sulphurea	low	24	108	113.39	123.39	103	92.15	102.15	103
Heptagenia sulphurea	mid	24	116	114.79	124.79	111	114.75	124.75	111
Heptagenia sulphurea	high	31	103	120.44	130.44	98	115.86	125.86	98

Hydropsyche incognita	mid	30	116	120.67	130.67	111	122.86	132.86	111
Hydropsyche instabilis	low	25	108	112.33	122.33	103	109.28	119.28	103
Hydropsyche pellucidula	low	36	108	126.31	136.31	103	112.18	122.18	103
Hydropsyche pellucidula	mid	39	116	142.98	152.98	111	140.18	150.18	111
Hydropsyche pellucidula	high	44	103	137.84	147.84	98	139.49	149.49	98
Hydropsyche siltalai	low	64	108	141.37	151.37	103	138.49	148.49	103
Hydropsyche siltalai	mid	80	116	115.13	125.13	111	120.71	130.71	111
Hydropsyche siltalai	high	61	103	118.43	128.43	98	117.95	127.95	98
Lepidostoma basale	mid	24	116	113.64	123.64	111	115.01	125.01	111
Lepidostoma basale	high	25	103	97.49	107.49	98	100.94	110.94	98
Lepidostoma hirtum	low	32	108	120.78	130.78	103	123.33	133.33	103
Lepidostoma hirtum	mid	53	116	140.68	150.68	111	149.06	159.06	111
Lepidostoma hirtum	high	39	103	126.97	136.97	98	123.04	133.04	98
Limnephilus lunatus	low	27	108	107.55	117.55	103	106.05	116.05	103
Limnephilus lunatus	mid	21	116	104.56	114.56	111	104.18	114.18	111
Limnephilus lunatus	high	21	103	100.01	110.01	98	95.91	105.91	98
Mystacides azurea	low	22	108	99.11	109.11	103	99.37	109.37	103

Mystacides azurea	mid	38	116	146.31	156.31	111	142.66	152.66	111
Mystacides azurea	high	34	103	119.51	129.51	98	123.28	133.28	98
Polycentropus flavomaculatus flavomaculatus	low	31	108	120.32	130.32	103	119.11	129.11	103
Polycentropus flavomaculatus flavomaculatus	mid	48	116	151.97	161.97	111	148.81	158.81	111
Polycentropus flavomaculatus flavomaculatus	high	31	103	109.70	119.70	98	117.45	127.45	98
Prodiamesa olivacea	low	44	108	131.00	141.00	103	137.65	147.65	103
Prodiamesa olivacea	mid	55	116	152.74	162.74	111	149.44	159.44	111
Prodiamesa olivacea	high	65	103	130.71	140.71	98	127.44	137.44	98
Psychomyia pusilla	mid	38	116	142.68	152.68	111	140.29	150.29	111
Psychomyia pusilla	high	34	103	121.80	131.80	98	126.16	136.16	98
Serratella ignita	low	48	108	141.31	151.31	103	144.05	154.05	103
Serratella ignita	mid	62	116	141.79	151.79	111	147.74	157.74	111
Serratella ignita	high	51	103	140.53	150.53	98	132.64	142.64	98

Torleya major	high	22	103	79.95	89.95	98	82.81	92.81	98

Appendix B: Supporting information for Chapter 3

Table S3.1: Number of species per taxonomic group and relative percentage of each order over the entire community.

Order*	No. of Species	% of community
Bivalvia	5	7.46
Chironomidae	3	4.48
Coleoptera	5	7.46
Crustacea	4	5.97
Ephemeroptera	10	14.93
Gastropoda	6	8.96
Heteroptera	3	4.48
Hirudinae	4	5.97
Odonata	1	1.49
Oligochaeta	5	7.46
Plecoptera	3	4.48
Tricoptera	16	23.88
Turbellaria	2	2.99
Total	67	100

Table S3.2: All species (n=67) traits: stream zonation, current preference and feeding type.

				stream zonation preference										current preference							feeding type									
Species	Order	Occurences	% PREVALENCE	eucrenal - spring region	hypocrenal - spring-brook	epirhithral - upper-trout region	metarhitral - lower-trout region	hyporhithral - grayling region	epipotamal - barbel region	metapotamal - bream region	hypopotamal - brackish water region	littoral - lake and stream shorelines, lentic sites, ponds etc.	profundal - bottom of stratified lakes	limnobiont - occurring only in standig waters	limnophil - standing waters; avoids current; rarely in slowly flowing streams	limno to rheophil - standing waters but regularly occurring in slowly flowing streams	rheo ro limnophil - slowly flowing streams and lentic zones; also found in standing waters	rheophil - streams; prefers zones with moderate to high current	rheobiont - streams; bound to zones with high current	indifferent - no preference	grazers/ scrapers	miners	xylophagous taxa	shredders -	gatherers/ collectors	active filter feeders -	passive filter feeders	predators	parasites	other feeding types
Anabolia nervosa	Tricoptera	533	42.30	0	0	0	0	0	3	3	0	4	0		yes						2	0	0	5	1	0	0	2	0	0
Anacaena globulus	Coleoptera	47	3.73	2	5	0	0	0	0	0	0	3	0				yes				3	0	0	2	3	0	0	2	0	0
Anodonta anatina	Bivalvia	34	2.70	0	0	0	1	2	2	2	0	3	0				yes				0	0	0	0	0	10	0	0	0	0
Apsectrotanypus trifascipennis	Chironomidae	67	5.32	0	1	1	2	2	1	0	0	2	1					yes			0	0	0	0	1	0	0	9	0	0
Athripsodes cinereus	Tricoptera	194	15.40	0	0	0	0	3	3	0	0	3	1			yes					0	0	0	3	3	0	0	4	0	0
Baetis lutheri	Ephemeroptera	99	7.86	0	0	1	5	3	1	0	0	0	0					yes			5	0	0	0	5	0	0	0	0	0
Baetis rhodani	Ephemeroptera	1238	98.25	0	1	2	3	2	1	1	0	0	0					yes			5	0	0	0	5	0	0	0	0	0
Bithynia tentaculata	Gastropoda	381	30.24	0	0	0	0	1	2	2	2	3	0							yes	3	0	0	0	2	5	0	0	0	0
Brachycercus harrisella	Ephemeroptera	23	1.83	0	0	0	0	0	5	5	0	0	0				yes				0	0	0	0	10	0	0	0	0	0
Brachyptera seticornis	Plecoptera	58	4.60	0	2	5	3	0	0	0	0	0	0					yes			7	0	0	0	3	0	0	0	0	0
Caenis horaria	Ephemeroptera	143	11.35	0	1	0	0	0	1	1	1	5	1		yes						0	0	0	0	10	0	0	0	0	0
Clinotanypus nervosus	Chironomidae	29	2.30	0	0	0	0	0	0	0	0	10	0		yes						0	0	0	0	1	0	0	9	0	0
Cloeon dipterum	Ephemeroptera	104	8.25	1	1	1	1	1	1	1	1	2	0		yes						5	0	0	0	5	0	0	0	0	0
Corbicula fluminea	Bivalvia	21	1.67	0	0	0	0	0	3	3	2	2	0								0	0	0	0	0	0	0	0	0	0
Dugesia gonocephala	Turbellaria	512	40.63	0	0	3	4	2	1	0	0	0	0					yes			0	0	0	0	0	0	0	10	0	0
Dugesia tigrina	Turbellaria	36	2.86	0	0	0	2	1	0	0	0	7	0			yes					0	0	0	0	0	0	0	10	0	0
Elmis aenea	Coleoptera	326	25.87	0	0	3	6	1	0	0	0	0	0					yes			9	0	0	0	1	0	0	0	0	0
Ephemerella mucronata	Ephemeroptera	220	17.46	0	0	4	4	2	0	0	0	0	0				yes				5	0	0	0	5	0	0	0	0	0

Erpobdella vilnensis	Hirudinae	380	30.16	0	0	3	3	3	0	0	0	1	0								0	0	0	0	0	0	0	10	0	0
Galba truncatula	Gastropoda	42	3.33	1	2	1	1	1	1	0	0	3	0			yes					3	0	0	3	2	0	0	0	0	2
Gammarus fossarum	Crustacea	558	44.29	1	1	2	2	2	1	0	0	1	0					yes			1	0	0	7	2	0	0	0	0	0
Gammarus roeselii	Crustacea	485	38.49	0	1	1	1	2	2	1	0	2	0				yes				1	0	0	5	3	0	0	1	0	0
Gammarus tigrinus	Crustacea	82	6.51	0	0	0	0	1	2	3	4	0	0				yes				0	0	0	7	3	0	0	0	0	0
Gerris lacustris	Heteroptera	43	3.41	0	0	0	0	0	0	0	0	10	0			yes					0	0	0	0	0	0	0	10	0	0
Glyphotaelius pellucidus	Tricoptera	38	3.02	1	1	1	0	0	1	0	0	6	0		yes						1	0	0	6	0	0	0	3	0	0
Graptodytes pictus	Coleoptera	28	2.22	0	0	0	0	0	0	0	0	10	0			yes					0	0	0	0	0	0	0	10	0	0
Gyraulus albus	Gastropoda	136	10.79	0	1	1	1	1	1	1	1	3	0			yes					6	0	0	2	0	0	0	0	0	2
Habroleptoides confusa	Ephemeroptera	216	17.14	0	2	2	3	2	1	0	0	0	0				yes				0	0	0	0	10	0	0	0	0	0
Habrophlebia fusca	Ephemeroptera	117	9.29	1	2	2	2	2	0	0	0	1	0			yes					2	0	0	0	8	0	0	0	0	0
Haemopis sanguisuga	Hirudinae	57	4.52	1	1	1	1	1	1	1	1	1	1			yes					0	0	0	0	0	0	0	10	0	0
Halesus digitatus digitatus	Tricoptera	68	5.40	0	0	0	2	4	4	0	0	0	0			yes					1	0	0	7	0	0	0	2	0	0
Halesus tessellatus	Tricoptera	49	3.89	0	0	0	3	3	4	0	0	0	0			yes					1	0	0	7	0	0	0	2	0	0
Helobdella stagnalis	Hirudinae	323	25.63	0	0	0	1	2	2	2	1	2	0							yes	0	0	0	0	0	0	0	10	0	0
Hemiclepsis marginata	Hirudinae	31	2.46	0	0	0	0	2	3	2	1	2	0			yes					0	0	0	0	0	0	0	0	10	0
Hydraena gracilis	Coleoptera	238	18.89	0	1	1	3	3	2	0	0	0	0					yes			10	0	0	0	0	0	0	0	0	0
Ilyocoris cimicoides cimicoides	Heteroptera	24	1.90	0	0	0	0	0	0	2	0	8	0		yes						0	0	0	0	0	0	0	10	0	0
Isoperla grammatica	Plecoptera	47	3.73	0	0	3	3	3	1	0	0	0	0					yes			1	0	0	1	1	0	0	7	0	0
Lepidostoma basale	Tricoptera	242	19.21	0	0	0	5	5	0	0	0	0	0				yes				5	0	3	2	0	0	0	0	0	0
Limnephilus lunatus	Tricoptera	38	3.02	0	0	0	0	1	2	1	0	5	1		yes						2	0	0	5	0	0	0	3	0	0
Limnephilus extricatus	Tricoptera	298	23.65	2	0	0	0	0	2	2	0	4	0		yes						2	0	0	5	0	0	0	3	0	0
Limnius volckmari	Coleoptera	469	37.22	0	0	1	3	5	1	0	0	0	0					yes			8	0	0	1	1	0	0	0	0	0
Limnodrilus claparedeanus	Oligochaeta	53	4.21	0	0	1	1	1	1	1	1	2	2				yes				0	0	0	0	10	0	0	0	0	0
Limnodrilus hoffmeisteri	Oligochaeta	266	21.11	0	0	1	1	1	1	1	1	2	2				yes				0	0	0	0	10	0	0	0	0	0
Lithax obscurus	Tricoptera	21	1.67	2	4	4	0	0	0	0	0	0	0					yes			9	0	0	0	1	0	0	0	0	0
Lymnaea stagnalis	Gastropoda	93	7.38	0	0	0	0	1	1	0	0	8	0			yes					4	0	0	4	0	0	0	0	0	2
Lype reducta	Tricoptera	83	6.59	0	1	4	2	1	1	0	0	1	0							yes	8	0	2	0	0	0	0	0	0	0

Molanna angustata	Tricoptera	68	5.40	0	0	0	0	1	3	3	0	3	0		yes						0	0	0	0	3	0	0	7	0	0
Musculium lacustre	Bivalvia	46	3.65	0	1	0	0	0	2	2	0	5	0			yes					0	0	0	0	0	10	0	0	0	0
Nais elinguis	Oligochaeta	52	4.13	1	1	1	1	1	1	1	1	1	1				yes				5	0	0	0	5	0	0	0	0	0
Nemoura cinerea cinerea	Plecoptera	49	3.89	3	2	1	1	1	1	0	0	1	0		yes						1	0	0	4	5	0	0	0	0	0
Neureclipsis bimaculata	Tricoptera	25	1.98	0	0	0	0	0	5	5	0	0	0				yes				0	0	0	0	0	0	1	9	0	0
Paratendipes albimanus	Chironomidae	34	2.70	0	0	1	1	1	1	1	1	3	1							yes	1	0	0	0	8	1	0	0	0	0
Pisidium nitidum	Bivalvia	41	3.25	0	0	0	2	2	2	2	0	2	0				yes				0	0	0	0	0	10	0	0	0	0
Pisidium subtruncatum	Bivalvia	82	6.51	0	0	1	1	2	2	2	0	1	1				yes				0	0	0	0	0	10	0	0	0	0
Planorbarius corneus	Gastropoda	70	5.56	0	0	0	0	1	2	2	2	3	0			yes					4	0	0	2	2	0	0	0	0	2
Planorbis planorbis	Gastropoda	72	5.71	0	0	0	0	0	2	2	2	4	0			yes					6	0	0	2	0	0	0	0	0	2
Platycnemis pennipes	Odonata	131	10.40	0	0	0	0	0	2	4	0	4	0			yes					0	0	0	0	0	0	0	10	0	0
Plea minutissima minutissima	Heteroptera	25	1.98	0	0	0	0	0	0	0	0	10	0		yes						0	0	0	0	0	0	0	10	0	0
Polycentropus flavomaculatus flavomaculatus	Tricoptera	440	34.92	0	0	0	2	2	2	2	2	0	0			yes					0	0	0	0	0	0	1	9	0	0
Potamophylax rotundipennis	Tricoptera	82	6.51	0	0	1	3	3	3	0	0	0	0				yes				2	0	0	6	0	0	0	2	0	0
Potamothrix hammoniensis	Oligochaeta	92	7.30	0	0	0	0	1	2	2	1	2	2			yes					0	0	0	0	10	0	0	0	0	0
Potamothrix moldaviensis	Oligochaeta	24	1.90	0	0	0	0	0	3	3	0	3	1				yes				0	0	0	0	10	0	0	0	0	0
Proasellus coxalis	Crustacea	268	21.27	0	2	2	1	1	1	1	2	0	0						yes		1	0	0	8	1	0	0	0	0	0
Procloeon bifidum	Ephemeroptera	73	5.79	0	0	1	1	2	2	2	0	2	0				yes				0	0	0	0	10	0	0	0	0	0
Psychomyia pusilla	Tricoptera	133	10.56	0	0	0	1	4	4	1	0	0	0				yes				6	0	0	0	2	0	1	1	0	0
Rhithrogena semicolorata	Ephemeroptera	44	3.49	0	1	2	3	3	1	0	0	0	0					yes			10	0	0	0	0	0	0	0	0	0
Rhyacophila nubila	Tricoptera	256	20.32	0	2	5	3	0	0	0	0	0	0					yes		0	0	0	0	0	0	0	10	0	0	0

Table S3.3: All predictors (n=82), description, category and source, used in BRT variable selection process. Uniform predictor set and all custom predictor sets per species (n=67). Cat (predictor category), BIO (bioclimate), Hydro (hydrology), Topo (topography)

Name	Description	Cat	Source	Uniform	*Anabolia nervosa*	*Anaccaena globulus*	*Anodonta anatina*	*Apsectrotanypus trifascipennis*	*Athripsodes cinereus*	*Baetis lutheri*	*Baetis rhodani*	*Bithynia tentaculata*	*Brachycercus harrisella*	*Brachyptera seticornis*	*Caenis horaria*	*Clinotanypus nervosus*	*Cloeon dipterum*	*Corbicula fluminea*	*Dugesia gonocephala*	*Dugesia tigrina*	*Elmis aenea*	*Ephemerella mucronata*	*Erpobdella vilnensis*	*Galba truncatula*	*Gammarus fossarum*	*Gammarus roeseli*	*Gammarus tigrinus*	*Gerris lacustris*	*Glyphotaelius pellucidus*	*Graptodytes pictus*	*Gyraulus albus*	*Habroleptoides confusa*	*Habrophlebia fusca*	*Haemopis sanguisuga*	*Halesus digitatus*
Agriculture	Cultivated and managed vegetation	Land Use	Domish et al 2015		X	X	X			X	X		X	X		X						X				X	X						X	X	
Barren Land	Barren lands/sparse vegetation	Land Use	Domish et al 2015	X		X	X	X		X	X		X	X		X	X	X			X	X	X		X		X		X	X	X		X		X
Shrubs	Shrubs Herbaceous vegetation Regularly flooded shrub/herbaceous vegetation	Land Use	Domish et al 2015																															X	
Snow	Snow/ice	Land Use	Domish et al 2015																																
Forest	evergreen/deciduous needle leaf trees evergreen broadleaf trees deciduous broadleaf trees Mixed/other trees	Land Use	Domish et al 2015	X				X	X			X			X		X	X	X	X	X		X	X	X			X	X	X	X	X			X
Urban	Urban/built-up	Land Use	Domish et al 2015		X				X	X		X			X				X	X				X		X		X				X			
Water	Open water	Land Use	Domish et al 2015																																
Bio01	Annual Mean Upstream Temperature	BIO	Domish et al 2015					X	X											X								X							

Bio02	Mean Upstream Diurnal Range (Mean of monthly (max temp - min temp))	BIO	Domish et al 2015																																
Bio03	Upstream Isothermality (hydro_02 / hydro_07) (* 100)	BIO	Domish et al 2015										X					X																	
Bio04	Upstream Temperature Seasonality (standard deviation *100)	BIO	Domish et al 2015	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Bio05	Maximum Upstream Temperature of Warmest Month	BIO	Domish et al 2015		X							X		X	X		X		X				X												X
Bio06	Minimum Upstream Temperature of Coldest Month	BIO	Domish et al 2015																																
Bio07	Upstream Temperature Annual Range - hydro_06)	BIO	Domish et al 2015													X																			
Bio08	Mean Upstream Temperature of Quarter	BIO	Domish et al 2015	X		X		X			X		X		X			X		X		X		X		X				X	X				X
Bio09	Mean Upstream Temperature of Quarter	BIO	Domish et al 2015	X	X		X	X	X	X	X	X		X		X	X	X	X		X				X	X	X	X	X	X			X	X	
Bio10	Mean Upstream Temperature of Quarter	BIO	Domish et al 2015																							X									
Bio11	Mean Upstream Temperature of Quarter	BIO	Domish et al 2015																																
Bio12	Annual Upstream Precipitation	BIO	Domish et al 2015	X		X	X				X	X	X	X	X		X	X		X				X	X	X						X		X	X
Bio13	Upstream Precipitation of Wettest Month	BIO	Domish et al 2015						X							X			X		X										X				
Bio14	Upstream Precipitation of Driest Month	BIO	Domish et al 2015																				X										X		

Bio15	Upstream Precipitation Seasonality (Coefficient Variation)	BIO	Domish et al 2015				X						X		X							X											X		X
Bio16	Upstream Precipitation of Wettest Quarter	BIO	Domish et al 2015																																
Bio17	Upstream Precipitation of Driest Quarter	BIO	Domish et al 2015																																
Bio18	Upstream Precipitation of Warmest Quarter	BIO	Domish et al 2015																																
Bio19	Upstream Precipitation of Coldest Quarter	BIO	Domish et al 2015																																
DH1	Annual maxima of 1 day means of daily discharge	Hydro	Irving et al 2018		X																														
DH2	Annual maxima of 3 day means of daily discharge	Hydro	Irving et al 2018																																
DH3	Annual maxima of 7 day means of daily discharge	Hydro	Irving et al 2018																																
DH4	Annual maxima of 30 day means of daily discharge	Hydro	Irving et al 2018																																
DH5	Annual maxima of 90 day means of daily discharge	Hydro	Irving et al 2018																																
DL1	Annual minima of 1 day means of daily discharge	Hydro	Irving et al 2018																																
DL2	Annual minima of 3 day means of daily discharge	Hydro	Irving et al 2018																																
DL3	Annual minima of 7 day means of daily discharge	Hydro	Irving et al 2018																																
DL4	Annual minima of 30 day means of daily discharge	Hydro	Irving et al 2018																																

DL5	Annual minima of 90 day means of daily discharge	Hydro	Irving et al 2018																																
MA1	Mean daily flows	Hydro	Irving et al 2018																																
MA2	Median daily flows	Hydro	Irving et al 2018																																
MA12	Mean monthly flows	Hydro	Irving et al 2018					X																				X	X	X					
MA13	Mean monthly flows	Hydro	Irving et al 2018																																
MA14	Mean monthly flows	Hydro	Irving et al 2018																																
MA15	Mean monthly flows	Hydro	Irving et al 2018																																
MA16	Mean monthly flows	Hydro	Irving et al 2018																																
MA17	Mean monthly flows	Hydro	Irving et al 2018																																
MA18	Mean monthly flows	Hydro	Irving et al 2018																																
MA19	Mean monthly flows	Hydro	Irving et al 2018																																
MA20	Mean monthly flows	Hydro	Irving et al 2018																																
MA21	Mean monthly flows	Hydro	Irving et al 2018																																
MA22	Mean monthly flows	Hydro	Irving et al 2018																																
MA23	Mean monthly flows	Hydro	Irving et al 2018																																
MH1	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH2	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH3	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH4	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH5	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH6	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH7	Mean maximum monthly flows	Hydro	Irving et al 2018																																

MH8	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH9	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH10	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH11	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH12	Mean maximum monthly flows	Hydro	Irving et al 2018																																
MH21	High flow volume	Hydro	Irving et al 2018	X		X	X	X			X			X	X	X	X	X	X	X	X	X	X	X				X		X	X		X	X	X
ML1	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML2	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML3	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML4	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML5	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML6	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML7	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML8	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML9	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML10	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML11	Mean minimum monthly flows	Hydro	Irving et al 2018																																
ML12	Mean minimum monthly flows	Hydro	Irving et al 2018																																
RA1	Rise rate	Hydro	Irving et al 2018																																
RA3	Fall rate	Hydro	Irving et al 2018							X												X								X					
TA1	Constancy	Hydro	Irving et al 2018																																
TA2	Predictability of flow	Hydro	Irving et al 2018		X				X	X		X	X												X	X	X		X			X			

Slope	% rise	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Eastness	-1+1	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Northness	-1+1	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table S3.3: Continued. All predictors (n=82), description, category and source, used in BRT variable selection process. Uniform predictor set and all custom predictor sets per species (n=67). Cat (predictor category), BIO (bioclimate), Hydro (hydrology), Topo (topography)

Name	Description	Category	Source	*Halesus tesselatus*	*Helobdella stagnalis*	*Hemiclepsis marginata*	*Hydraena gracilis*	*Ilyocoris cimicoides*	*Isoperla grammatica*	*Lepidostoma basale*	*Limnephilus extricatus*	*Limnephilus lunatus*	*Limnius volckmari*	*Limnodrilus claparedeanus*	*Limnodrilus hoffmeisteri*	*Lithax obscurus*	*Lymnaea stagnalis*	*Lype reducta*	*Molanna angustata*	*Musculium lacustre*	*Nais elinguis*	*Nemoura cinerea*	*Neureclipsis bimaculata*	*Paratendipes albimanus*	*Pisidium nitidum*	*Pisidium subtruncatum*	*Planorbarius corneus*	*Planorbis planorbis*	*Platycnemis pennipes*	*Plea minutissima*	*Polycentropus flavomaculatus*	*Potamophylax rotundipennis*	*Potamothrix hammoniensis*	*Potamothrix moldaviensis*	*Proasellus coxalis*	*Procloeon bifidum*	*Psychomyia pusilla*	*Rhithrogena semicolorata*	*Rhyacophila nubila*
Agriculture	Cultivated and managed vegetation	Land Use	Domish et al 2015	X		X		X	X						X	X		X		X		X								X						X	X	X	
Barren Land	Barren lands/sparse vegetation	Land Use	Domish et al 2015	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X		X			X	X	X	X	X	X	X	X	X	X	X		X	X	X	X
Shrubs	Shrubs Herbaceous vegetation Regularly flooded shrub/herbaceous vegetation	Land Use	Domish et al 2015																	X												X							
Snow	Snow/ice	Land Use	Domish et al 2015																																				
Forest	evergreen/deciduous needle leaf trees evergreen broadleaf trees deciduous broadleaf trees Mixed/other trees	Land Use	Domish et al 2015		X		X			X	X	X	X	X			X		X		X		X	X	X	X	X	X	X		X	X	X	X	X				X
Urban	Urban/built-up	Land Use	Domish et al 2015			X			X													X	X												X				
Water	Open water	Land Use	Domish et al 2015																																				
Bio01	Annual Mean Upstream Temperature	Bio	Domish et al 2015	X									X		X																		X	X					X
Bio02	Mean Upstream Diurnal Range (Mean of monthly (max temp - min temp))	Bio	Domish et al 2015																											X									
Bio03	Upstream Isothermality (hydro_02 / hydro_07) (* 100)	Bio	Domish et al 2015								X										X																		
Bio04	Upstream Temperature Seasonality (standard deviation *100)	Bio	Domish et al 2015	X	X	X	X	X	X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X		X		X	X	X	X	X	X	X	X	X
Bio05	Maximum Upstream Temperature of Warmest Month	Bio	Domish et al 2015			X					X										X	X			X				X	X			X	X					
Bio06	Minimum Upstream Temperature of Coldest Month	Bio	Domish et al 2015																								X												

Bio07	Upstream Temperature Annual Range - hydro_06)	Bio	Domish et al 2015																																				
Bio08	Mean Upstream Temperature of Quarter	Bio	Domish et al 2015	X		X		X	X			X		X	X	X		X	X		X		X		X	X	X					X		X		X	X	X	
Bio09	Mean Upstream Temperature of Quarter	Bio	Domish et al 2015	X	X	X		X		X			X	X	X				X	X		X		X		X		X	X	X	X		X	X		X	X	X	
Bio10	Mean Upstream Temperature of Quarter	Bio	Domish et al 2015						X							X								X								X							
Bio11	Mean Upstream Temperature of Quarter	Bio	Domish et al 2015																																				
Bio12	Annual Upstream Precipitation	Bio	Domish et al 2015	X	X			X			X		X			X			X	X			X									X		X	X	X	X	X	
Bio13	Upstream Precipitation of Wettest Month	Bio	Domish et al 2015				X										X	X									X				X								X
Bio14	Upstream Precipitation of Driest Month	Bio	Domish et al 2015																					X															
Bio15	Upstream Precipitation Seasonality (Coefficient Variation)	Bio	Domish et al 2015				X															X	X	X													X		X
Bio16	Upstream Precipitation of Wettest Quarter	Bio	Domish et al 2015																																				
Bio17	Upstream Precipitation of Driest Quarter	Bio	Domish et al 2015																																				
Bio18	Upstream Precipitation of Warmest Quarter	Bio	Domish et al 2015																																				
Bio19	Upstream Precipitation of Coldest Quarter	Bio	Domish et al 2015																																				
DH1	Annual maxima of 1 day means of daily discharge	Hydro	Irving et al 2018																			X																	
DH2	Annual maxima of 3 day means of daily discharge	Hydro	Irving et al 2018																																				
DH3	Annual maxima of 7 day means of daily discharge	Hydro	Irving et al 2018																																				
DH4	Annual maxima of 30 day means of daily discharge	Hydro	Irving et al 2018																																				
DH5	Annual maxima of 90 day means of daily discharge	Hydro	Irving et al 2018																																				
DL1	Annual minima of 1 day means of daily discharge	Hydro	Irving et al 2018																																				
DL2	Annual minima of 3 day means of daily discharge	Hydro	Irving et al 2018																																				

DL3	Annual minima of 7 day means of daily discharge	Hydro	Irving et al 2018																																				
DL4	Annual minima of 30 day means of daily discharge	Hydro	Irving et al 2018																																				
DL5	Annual minima of 90 day means of daily discharge	Hydro	Irving et al 2018																																				
MA1	Mean daily flows	Hydro	Irving et al 2018																																				
MA2	Median daily flows	Hydro	Irving et al 2018																																				
MA12	Mean monthly flows	Hydro	Irving et al 2018									X									X				X	X				X									
MA13	Mean monthly flows	Hydro	Irving et al 2018																																				
MA14	Mean monthly flows	Hydro	Irving et al 2018																																				
MA15	Mean monthly flows	Hydro	Irving et al 2018																																				
MA16	Mean monthly flows	Hydro	Irving et al 2018																																				
MA17	Mean monthly flows	Hydro	Irving et al 2018						X																			X											
MA18	Mean monthly flows	Hydro	Irving et al 2018																																				
MA19	Mean monthly flows	Hydro	Irving et al 2018																																				
MA20	Mean monthly flows	Hydro	Irving et al 2018																																				
MA21	Mean monthly flows	Hydro	Irving et al 2018																																				
MA22	Mean monthly flows	Hydro	Irving et al 2018																																				
MA23	Mean monthly flows	Hydro	Irving et al 2018																																				
MH1	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH2	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH3	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH4	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH5	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH6	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH7	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH8	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH9	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH10	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH11	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH12	Mean maximum monthly flows	Hydro	Irving et al 2018																																				
MH21	High flow volume	Hydro	Irving et al 2018	X	X	X		X		X	X	X		X	X	X	X	X	X	X		X		X	X	X	X	X	X	X				X		X		X	X

ML1	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML2	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML3	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML4	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML5	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML6	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML7	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML8	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML9	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML10	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML11	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
ML12	Mean minimum monthly flows	Hydro	Irving et al 2018																																				
RA1	Rise rate	Hydro	Irving et al 2018																																				
RA3	Fall rate	Hydro	Irving et al 2018			X				X				X	X														X				X						
TA1	Constancy	Hydro	Irving et al 2018																																				
TA2	Predictability of flow	Hydro	Irving et al 2018				X		X				X								X		X								X	X	X		X		X		
Slope	% rise	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Eastness	-1+1	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Northness	-1+1	Topo	Amatulli et al 2015	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table S3.4: BRT Coefficients

	Climate Category					Hydrology Category					Land use Category					Customized Predictor Set					Uniform Predictor Set				
taxa	no_of_trees	dev_mean	cor_mean	discrim_mean	cv_thres	no_of_trees	dev_mean	cor_mean	discrim_mean	cv_thres	no_of_trees	dev_mean	cor_mean	discrim_mean	cv_thres	no_of_trees	dev_mean	cor_mean	discrim_mean	cv_thres	no_of_trees	dev_mean	cor_mean	discrim_mean	cv_thres
Anabolia nervosa	1000	1.25	0.10	0.56	0.32	800	1.25	0.01	0.50	0.32	1000	1.25	0.04	0.53	0.32	1000	1.25	0.08	0.55	0.32	1000	1.25	0.06	0.54	0.32
Anacaena globulus	4700	0.29	0.14	0.62	0.06	1750	0.30	0.06	0.52	0.04	2250	0.30	0.07	0.58	0.05	4600	0.29	0.12	0.63	0.06	3450	0.29	0.08	0.63	0.05
Anodonta anatina	5850	0.21	0.08	0.67	0.06	2900	0.22	0.06	0.63	0.04	1850	0.22	0.05	0.62	0.03	6700	0.21	0.12	0.66	0.06	6000	0.21	0.11	0.66	0.06
Apsectrotanypus trifascipennis	7500	0.31	0.14	0.67	0.09	1000	0.32	0.05	0.54	0.04	2400	0.32	0.06	0.56	0.05	3550	0.32	0.10	0.57	0.06	3600	0.32	0.09	0.55	0.06
Athripsodes cinereus	5950	0.70	0.16	0.64	0.16	1900	0.72	0.08	0.59	0.13	3550	0.72	0.10	0.60	0.13	6150	0.70	0.17	0.66	0.16	5350	0.70	0.15	0.65	0.15
Baetis lutheri	2600	0.52	0.08	0.58	0.09	3700	0.52	0.11	0.60	0.09	1100	0.53	0.06	0.56	0.08	4900	0.52	0.12	0.62	0.10	3600	0.52	0.10	0.60	0.09
Baetis rhodani	9500	1.18	0.27	0.66	0.63	1000	1.25	0.08	0.55	0.68	4150	1.23	0.16	0.60	0.66	9250	1.19	0.26	0.66	0.64	9100	1.19	0.26	0.67	0.64
Bithynia tentaculata	7250	0.99	0.20	0.64	0.24	1000	1.01	0.06	0.54	0.21	1000	1.01	0.08	0.57	0.21	7200	0.99	0.18	0.64	0.23	7000	0.99	0.19	0.64	0.23
Brachyptera seticornis	1000	0.35	0.03	0.54	0.05	1750	0.35	0.05	0.58	0.05	1000	0.35	0.04	0.54	0.05	2050	0.36	0.07	0.59	0.05	1550	0.36	0.06	0.54	0.05
Caenis horaria	9300	0.56	0.20	0.70	0.14	3950	0.58	0.11	0.61	0.11	1000	0.59	0.06	0.56	0.09	8350	0.57	0.17	0.68	0.13	7100	0.57	0.15	0.66	0.13
Cloeon dipterum	8200	0.45	0.16	0.68	0.11	3250	0.47	0.09	0.61	0.08	4000	0.47	0.09	0.62	0.09	5050	0.46	0.10	0.64	0.09	7150	0.46	0.13	0.66	0.10
Corbicula fluminea	10000	0.11	0.16	0.77	0.08	3250	0.12	0.04	0.60	0.03	1350	0.12	0.02	0.62	0.02	7350	0.12	0.10	0.70	0.05	4850	0.12	0.08	0.59	0.03
Dugesia gonocephala	6800	1.20	0.22	0.63	0.35	1000	1.24	0.03	0.51	0.31	3600	1.23	0.14	0.59	0.32	6150	1.21	0.21	0.62	0.34	6750	1.20	0.22	0.64	0.34
Dugesia tigrina	1000	0.24	0.03	0.60	0.03	950	0.24	0.02	0.56	0.03	1000	0.24	0.03	0.59	0.03	4250	0.24	0.06	0.70	0.05	1000	0.24	0.02	0.63	0.03
Elmis aenea	6200	0.95	0.20	0.64	0.22	1000	0.97	0.03	0.53	0.19	1400	0.97	0.09	0.57	0.20	6250	0.95	0.19	0.64	0.22	5350	0.95	0.18	0.63	0.21
Ephemerella mucronata	3050	0.80	0.12	0.60	0.15	1000	0.81	0.09	0.53	0.14	1000	0.81	0.05	0.56	0.14	1000	0.81	0.06	0.56	0.14	2000	0.80	0.10	0.59	0.15
Galba truncatula	1200	0.28	0.03	0.58	0.04	1050	0.28	0.03	0.59	0.04	1300	0.28	0.05	0.51	0.04	1000	0.28	0.02	0.53	0.04	1000	0.28	0.03	0.56	0.04

Gammarus fossarum	10000	1.18	0.33	0.71	0.36	1000	1.28	0.08	0.55	0.34	8250	1.23	0.24	0.65	0.36	10000	1.20	0.30	0.69	0.37	10000	1.20	0.31	0.69	0.37
Gammarus roeselii	3800	1.12	0.16	0.59	0.26	1000	1.14	0.07	0.54	0.26	1000	1.14	0.06	0.54	0.26	2750	1.13	0.15	0.57	0.26	1500	1.13	0.13	0.56	0.26
Gammarus tigrinus	5500	0.33	0.10	0.67	0.07	1000	0.34	0.04	0.60	0.05	2850	0.34	0.06	0.61	0.05	3000	0.34	0.07	0.59	0.06	3000	0.34	0.07	0.58	0.06
Gerris lacustris	6800	0.25	0.14	0.70	0.09	4600	0.26	0.11	0.64	0.07	5950	0.25	0.14	0.69	0.08	5100	0.26	0.12	0.72	0.08	5100	0.26	0.12	0.70	0.07
Glyphotaelius pellucidus	2400	0.24	0.05	0.64	0.04	3150	0.24	0.05	0.70	0.04	3250	0.24	0.05	0.60	0.05	3050	0.24	0.07	0.61	0.04	2350	0.24	0.06	0.59	0.04
Graptodytes pictus	3400	0.19	0.06	0.68	0.05	1000	0.19	0.01	0.63	0.03	2750	0.19	0.04	0.60	0.03	5250	0.18	0.08	0.70	0.06	4050	0.18	0.09	0.69	0.05
Gyraulus albus	8350	0.57	0.21	0.68	0.14	2000	0.60	0.09	0.53	0.10	6650	0.58	0.17	0.66	0.13	8350	0.57	0.21	0.68	0.13	7400	0.58	0.19	0.67	0.13
Habroleptoides confusa	2850	0.77	0.11	0.59	0.15	1000	0.78	0.06	0.53	0.13	1000	0.78	0.04	0.53	0.13	3400	0.77	0.12	0.59	0.15	2150	0.77	0.10	0.57	0.14
Habrophlebia fusca	4100	0.51	0.11	0.60	0.10	1250	0.51	0.07	0.60	0.08	950	0.52	0.04	0.54	0.08	3000	0.51	0.10	0.62	0.09	3750	0.51	0.09	0.62	0.09
Haemopis sanguisuga	3550	0.35	0.07	0.63	0.06	2750	0.35	0.07	0.62	0.06	9000	0.33	0.14	0.72	0.10	7500	0.33	0.16	0.67	0.08	5750	0.34	0.11	0.66	0.08
Halesus digitatus	3450	0.32	0.07	0.60	0.06	1800	0.32	0.04	0.58	0.05	3750	0.32	0.09	0.59	0.06	1050	0.33	0.04	0.59	0.05	2650	0.33	0.07	0.60	0.06
Halesus tesselatus	3600	0.29	0.09	0.63	0.06	1600	0.29	0.06	0.49	0.04	6700	0.28	0.12	0.67	0.07	1900	0.29	0.05	0.61	0.05	2050	0.29	0.05	0.61	0.05
Helobdella stagnalis	3400	1.00	0.14	0.58	0.22	1000	1.01	0.05	0.54	0.21	1000	1.01	0.09	0.57	0.21	5100	1.00	0.16	0.62	0.22	4450	1.00	0.15	0.61	0.22
Hemiclepsis marginata	1250	0.22	0.05	0.57	0.03	1150	0.23	0.03	0.60	0.03	1500	0.22	0.04	0.61	0.03	1300	0.22	0.02	0.61	0.03	1800	0.22	0.04	0.63	0.03
Hydraena gracilis	3700	0.89	0.12	0.61	0.19	1000	0.90	0.02	0.51	0.17	1000	0.90	0.08	0.58	0.17	3600	0.89	0.13	0.61	0.19	2950	0.90	0.11	0.60	0.18
Ilyocoris cimicoides	4000	0.16	0.09	0.67	0.05	2900	0.16	0.06	0.63	0.04	2200	0.16	0.04	0.61	0.03	3800	0.16	0.11	0.60	0.05	3250	0.16	0.09	0.61	0.04
Isoperla grammatica	2700	0.20	0.08	0.60	0.04	8600	0.20	0.09	0.64	0.05	3650	0.20	0.07	0.64	0.04	750	0.20	0.02	0.56	0.03	1200	0.20	0.04	0.58	0.03
Lepidostoma basale	5200	0.85	0.15	0.63	0.18	1350	0.87	0.09	0.59	0.16	1000	0.87	0.07	0.56	0.16	3550	0.86	0.12	0.60	0.17	3000	0.86	0.12	0.59	0.17
Limnephilus extricatus	3800	0.21	0.07	0.63	0.05	2000	0.21	0.05	0.52	0.04	4750	0.21	0.15	0.59	0.06	2600	0.21	0.05	0.58	0.04	2250	0.21	0.05	0.58	0.04
Limnephilus lunatus	8750	0.89	0.24	0.68	0.23	2150	0.93	0.11	0.57	0.19	5800	0.91	0.17	0.62	0.21	8400	0.90	0.22	0.67	0.22	7200	0.91	0.21	0.66	0.21
Limnius volckmari	5400	1.21	0.18	0.62	0.33	1000	1.23	0.03	0.54	0.31	1350	1.23	0.11	0.57	0.31	4600	1.21	0.17	0.61	0.32	4450	1.21	0.17	0.60	0.32

Limnodrilus claparedeanus	10000	0.30	0.16	0.76	0.10	1350	0.32	0.05	0.58	0.05	1950	0.32	0.05	0.60	0.05	8450	0.30	0.16	0.74	0.09	8100	0.31	0.15	0.73	0.08
Limnodrilus hoffmeisteri	10000	0.78	0.29	0.74	0.24	1000	0.86	0.04	0.55	0.16	6900	0.83	0.18	0.65	0.20	10000	0.78	0.30	0.74	0.23	10000	0.79	0.30	0.73	0.23
Lithax obscurus	4600	0.10	0.10	0.75	0.05	8150	0.09	0.19	0.68	0.06	1500	0.10	0.02	0.59	0.02	7300	0.09	0.19	0.69	0.04	5900	0.09	0.16	0.68	0.04
Lymnaea stagnalis	9700	0.45	0.18	0.72	0.13	1250	0.48	0.07	0.50	0.07	7700	0.46	0.17	0.69	0.12	8600	0.46	0.17	0.70	0.11	9150	0.46	0.17	0.71	0.11
Lype reducta	3550	0.42	0.09	0.60	0.07	1150	0.42	0.07	0.51	0.06	3100	0.42	0.08	0.57	0.07	2500	0.42	0.06	0.58	0.07	1000	0.42	0.04	0.58	0.06
Molanna angustata	10000	0.34	0.18	0.75	0.11	4650	0.35	0.15	0.54	0.09	4300	0.36	0.12	0.61	0.08	5650	0.35	0.12	0.67	0.08	5750	0.35	0.13	0.68	0.08
Musculium lacustre	1900	0.26	0.07	0.50	0.04	2050	0.26	0.05	0.51	0.05	1000	0.27	0.04	0.41	0.03	1000	0.26	0.02	0.58	0.04	1500	0.26	0.04	0.57	0.04
Nais elinguis	7800	0.28	0.16	0.74	0.09	200	0.30	0.01	0.49	0.04	4550	0.29	0.09	0.67	0.06	5750	0.29	0.10	0.69	0.06	8050	0.28	0.15	0.72	0.08
Nemoura cinerea	7300	0.30	0.15	0.70	0.08	2300	0.31	0.05	0.58	0.05	4300	0.31	0.08	0.63	0.06	6950	0.30	0.11	0.71	0.07	6850	0.30	0.11	0.70	0.08
Neureclipsis bimaculata	1700	0.16	0.05	0.58	0.03	5150	0.16	0.09	0.67	0.05	1000	0.16	0.02	0.62	0.02	2550	0.16	0.03	0.65	0.03	2700	0.16	0.04	0.64	0.03
Paratendipes albimanus	3300	0.20	0.07	0.58	0.04	4800	0.20	0.10	0.67	0.06	1000	0.21	0.02	0.54	0.03	4450	0.20	0.09	0.63	0.05	5700	0.20	0.11	0.66	0.06
Pisidium nitidum	9200	0.20	0.11	0.70	0.09	2900	0.21	0.06	0.64	0.05	9150	0.20	0.15	0.75	0.08	5050	0.21	0.09	0.70	0.06	3600	0.21	0.10	0.68	0.05
Pisidium subtruncatum	2900	0.34	0.06	0.62	0.06	600	0.34	0.01	0.50	0.04	4900	0.34	0.12	0.64	0.07	2300	0.34	0.06	0.60	0.06	2350	0.34	0.06	0.60	0.06
Planorbarius corneus	2600	0.37	0.06	0.62	0.07	350	0.37	0.01	0.48	0.05	2100	0.37	0.05	0.60	0.06	4400	0.36	0.10	0.66	0.07	4550	0.36	0.11	0.64	0.07
Planorbis planorbis	7600	0.35	0.17	0.67	0.10	3550	0.36	0.10	0.58	0.07	3450	0.36	0.07	0.62	0.07	5650	0.35	0.11	0.67	0.08	4800	0.36	0.10	0.63	0.08
Platycnemis pennipes	10000	0.51	0.20	0.72	0.14	2000	0.54	0.06	0.58	0.09	4800	0.53	0.12	0.63	0.11	8300	0.52	0.16	0.68	0.12	8150	0.52	0.16	0.68	0.12
Plea minutissima	3300	0.16	0.08	0.60	0.04	1300	0.16	0.03	0.52	0.02	2900	0.16	0.04	0.64	0.03	3100	0.16	0.09	0.61	0.04	2950	0.16	0.06	0.63	0.04
Polycentropus flavomaculatus	1000	1.17	0.08	0.54	0.28	1000	1.17	0.05	0.53	0.28	1000	1.17	0.08	0.54	0.28	950	1.17	0.02	0.52	0.28	1000	1.17	0.05	0.53	0.28

Potamothrix hammoniensis	9050	0.41	0.18	0.71	0.11	850	0.43	0.02	0.53	0.06	4200	0.43	0.10	0.63	0.08	10000	0.41	0.15	0.72	0.11	10000	0.41	0.15	0.72	0.11
Potamothrix moldaviensis	1000	0.14	0.01	0.63	0.02	1150	0.14	0.04	0.65	0.02	5950	0.13	0.14	0.68	0.08	4250	0.13	0.11	0.63	0.05	4550	0.13	0.14	0.64	0.05
Proasellus coxalis	6600	0.82	0.17	0.64	0.19	1000	0.85	0.06	0.56	0.16	3050	0.84	0.11	0.59	0.17	5600	0.83	0.15	0.63	0.19	6200	0.82	0.16	0.64	0.19
Procloeon bifidum	4700	0.33	0.11	0.61	0.07	1000	0.34	0.01	0.56	0.05	1950	0.34	0.06	0.53	0.05	3800	0.33	0.08	0.57	0.06	1000	0.34	0.03	0.56	0.05
Psychomyia pusilla	3150	0.59	0.09	0.61	0.11	1000	0.59	0.05	0.55	0.09	1250	0.59	0.07	0.57	0.09	2300	0.59	0.07	0.62	0.10	4500	0.58	0.10	0.63	0.11
Rhithrogena semicolorata	6250	0.24	0.13	0.72	0.08	3000	0.25	0.06	0.64	0.05	3150	0.25	0.05	0.65	0.04	4050	0.24	0.07	0.64	0.05	5300	0.24	0.12	0.67	0.06
Rhyacophila nubila	6750	0.87	0.18	0.65	0.21	1000	0.90	0.06	0.56	0.17	2850	0.89	0.11	0.59	0.18	5600	0.88	0.16	0.64	0.20	5250	0.89	0.16	0.63	0.20
Riolus subviolaceus	4050	0.30	0.09	0.66	0.06	1000	0.30	0.03	0.58	0.04	2900	0.30	0.07	0.59	0.05	3150	0.30	0.08	0.59	0.06	4100	0.30	0.09	0.60	0.06
Stictotarsus duodecimpustulatus	4550	0.43	0.09	0.63	0.09	1100	0.44	0.06	0.51	0.06	1900	0.44	0.07	0.58	0.07	1000	0.44	0.05	0.58	0.06	1000	0.44	0.03	0.59	0.06
Torleya major	3750	0.85	0.13	0.62	0.18	900	0.86	0.02	0.52	0.16	1250	0.86	0.10	0.56	0.16	2850	0.85	0.11	0.61	0.17	2850	0.85	0.12	0.61	0.17
Valvata piscinalis	2200	0.17	0.06	0.58	0.03	500	0.17	0.02	0.55	0.02	3150	0.17	0.07	0.58	0.04	3200	0.17	0.06	0.64	0.04	3850	0.17	0.07	0.68	0.04

Table S3.5: TSS value from SDMs for all species (n=92)

species	Uniform Weight TSS	Uniform mean TSS	Uniform Standard Error	Custom Weighted TSS	Custom Mean TSS	Custom standard error
Anabolia_nervosa	0.34	0.32	0.01	0.33	0.30	0.01
Anacaena_globulus	0.48	0.44	0.02	0.43	0.40	0.02
Anodonta_anatina	0.63	0.59	0.02	0.65	0.60	0.02
Apsectrotanypus_trifascipennis	0.48	0.45	0.02	0.54	0.51	0.02
Athripsodes_cinereus	0.42	0.39	0.01	0.46	0.45	0.01
Baetis_lutheri	0.59	0.59	0.01	0.55	0.54	0.01
Baetis_rhodani	0.40	0.39	0.01	0.41	0.40	0.01
Bithynia_tentaculata	0.46	0.45	0.01	0.45	0.45	0.01
Brachycercus_harrisella	0.63	0.56	0.03	0.73	0.67	0.03
Brachyptera_seticornis	0.70	0.69	0.01	0.64	0.62	0.02
Caenis_horaria	0.52	0.50	0.01	0.51	0.50	0.01
Clinotanypus_nervosus	0.57	0.51	0.03	0.62	0.50	0.04
Cloeon_dipterum	0.46	0.43	0.02	0.52	0.49	0.02
Corbicula_fluminea	0.62	0.56	0.03	0.70	0.63	0.03

Dugesia_gonocephala	0.46	0.46	0.01	0.47	0.46	0.01
Dugesia_tigrina	0.60	0.55	0.02	0.67	0.60	0.03
Elmis_aenea	0.39	0.36	0.01	0.35	0.33	0.01
Ephemerella_mucronata	0.62	0.61	0.01	0.63	0.63	0.01
Erpobdella_vilnensis	0.39	0.38	0.01	0.33	0.33	0.01
Galba_truncatula	0.37	0.32	0.02	0.39	0.35	0.02
Gammarus_fossarum	0.47	0.47	0.01	0.44	0.43	0.01
Gammarus_roeselii	0.44	0.38	0.02	0.45	0.40	0.02
Gammarus_tigrinus	0.53	0.44	0.03	0.55	0.53	0.02
Gerris_lacustris	0.52	0.47	0.02	0.55	0.49	0.03
Glyphotaelius_pellucidus	0.46	0.41	0.02	0.48	0.39	0.03
Graptodytes_pictus	0.57	0.50	0.03	0.53	0.46	0.02
Gyraulus_albus	0.40	0.37	0.02	0.41	0.38	0.01
Habroleptoides_confusa	0.59	0.58	0.01	0.57	0.57	0.01
Habrophlebia_fusca	0.50	0.48	0.01	0.61	0.59	0.02
Haemopis_sanguisuga	0.51	0.48	0.02	0.45	0.42	0.02
Halesus_digitatus	0.44	0.42	0.01	0.41	0.37	0.02
Halesus_tesselatus	0.60	0.58	0.02	0.60	0.57	0.02

Helobdella_stagnalis	0.33	0.30	0.01	0.33	0.30	0.01
Hemiclepsis_marginata	0.56	0.53	0.02	0.62	0.56	0.03
Hydraena_gracilis	0.56	0.56	0.01	0.57	0.57	0.01
Ilyocoris_cimicoides	0.67	0.60	0.03	0.56	0.49	0.03
Isoperla_grammatica	0.60	0.54	0.03	0.56	0.50	0.02
Lepidostoma_basale	0.47	0.47	0.01	0.46	0.45	0.01
Limnephilus_extricatus	0.59	0.53	0.03	0.63	0.58	0.02
Limnephilus_lunatus	0.40	0.37	0.02	0.40	0.38	0.01
Limnius_volckmari	0.34	0.33	0.01	0.35	0.34	0.01
Limnodrilus_claparedeanus	0.68	0.66	0.02	0.65	0.63	0.02
Limnodrilus_claparedeianus	0.55	0.52	0.02	0.55	0.50	0.02
Limnodrilus_hoffmeisteri	0.52	0.51	0.01	0.50	0.50	0.01
Lithax_obscurus	0.82	0.75	0.03	0.71	0.61	0.04
Lymnaea_stagnalis	0.57	0.55	0.02	0.49	0.47	0.01
Lype_reducta	0.36	0.32	0.02	0.33	0.30	0.01
Molanna_angustata	0.62	0.59	0.02	0.61	0.58	0.02
Musculium_lacustre	0.45	0.36	0.03	0.38	0.33	0.02
Nais_elinguis	0.73	0.70	0.02	0.76	0.74	0.02

Nemoura_cinerea	0.50	0.45	0.02	0.49	0.44	0.02
Neureclipsis_bimaculata	0.57	0.51	0.03	0.68	0.60	0.03
Paratendipes_albimanus	0.60	0.57	0.02	0.61	0.57	0.02
Pisidium_nitidum	0.67	0.62	0.02	0.72	0.68	0.02
Pisidium_subtruncatum	0.56	0.53	0.02	0.58	0.55	0.02
Planorbarius_corneus	0.57	0.53	0.02	0.53	0.51	0.01
Planorbis_planorbis	0.51	0.47	0.02	0.49	0.45	0.02
Platycnemis_pennipes	0.52	0.51	0.01	0.55	0.54	0.01
Plea_minutissima	0.60	0.49	0.03	0.59	0.54	0.02
Polycentropus_flavomaculatus	0.42	0.41	0.01	0.39	0.38	0.01
Potamophylax_rotundipennis	0.56	0.52	0.02	0.52	0.49	0.01
Potamothrix_hammoniensis	0.49	0.46	0.02	0.59	0.57	0.02
Potamothrix_moldaviensis	0.57	0.49	0.03	0.75	0.65	0.04
Proasellus_coxalis	0.34	0.31	0.01	0.34	0.33	0.01
Procloeon_bifidum	0.43	0.38	0.02	0.47	0.44	0.02
Psychomyia_pusilla	0.48	0.46	0.01	0.48	0.47	0.01
Rhithrogena_semicolorata	0.63	0.59	0.02	0.65	0.61	0.02
Rhyacophila_nubila	0.59	0.59	0.01	0.59	0.59	0.01

Appendix C: Supporting information for Chapter 4

Table S4.1: correlation matrix of all predictors from all model configurations

	dh1	mh21	Bio 08 (hC)	Bio 09 (hC)	Bio 12 (hC)	Bio 15 (bC)	Bio 02 (bC)	Bio 04 (bC)	Bio 08 (bC)	Bio 09 (bC)
dh1	1	0.04110397	0.04299391	0.12507105	0.05306071	0.04631987	-0.030846	-0.2104459	0.07266893	0.07059209
mh21	0.04110397	1	-0.1811403	0.04500541	-0.1460599	-0.0497881	0.11133953	0.01669001	-0.0178457	0.02258609
Bio 08 (hC)	0.04299391	-0.1811403	1	-0.1186342	0.04297953	0.11957353	0.18033317	-0.0168116	0.37173693	-0.1118949
Bio 09 (hC)	0.12507105	0.04500541	-0.1186342	1	0.03579034	-0.1671069	-0.3614894	-0.5589005	-0.1557763	0.30354887
Bio 12 (hC)	0.05306071	-0.1460599	0.04297953	0.03579034	1	0.0690387	-0.1302382	-0.0842928	0.1245857	0.04414832
Bio 15 (bC)	0.04631987	-0.0497881	0.11957353	-0.1671069	0.0690387	1	0.21512485	-0.0551016	0.31857108	-0.0173852
Bio 02 (bC)	-0.030846	0.11133953	0.18033317	-0.3614894	-0.1302382	0.21512485	1	0.28669032	0.31285451	-0.1640398

Bio 04 (bC)	-0.2104459	0.01669001	-0.0168116	-0.5589005	-0.0842928	-0.0551016	0.28669032	1	-0.0420521	-0.3169023
Bio 08 (bC)	0.07266893	-0.0178457	0.37173693	-0.1557763	0.1245857	0.31857108	0.31285451	-0.0420521	1	-0.1183476
Bio 09 (bC)	0.07059209	0.02258609	-0.1118949	0.30354887	0.04414832	-0.0173852	-0.1640398	-0.3169023	-0.1183476	1

Table S4.2: Coefficients for each predictor category in 1st run and all predictors in 2nd run.

1st run BRTs															2nd run BRTs				
Hydroclimate					Hydrology					Bioclimate					All predictors				
no_of_ trees	dev_ mean	cor_ mean	Discrim mean	cv_thres	no_of_ trees	dev_ mean	cor_mean	discrim mean	cv_thres	no_of_ trees	dev_ mean	cor_ mean	Discrim mean	cv_thres	no_of_ trees	dev_ mean	cor_ mean	Discrim mean	cv_thres
2300	0.1937637	0.03165908	0.67598	0.03830936	1900	0.19428371	0.05404798	0.59629	0.03170907	7350	0.18441022	0.12830968	0.69641	0.0676502	5750	0.18652672	0.12011354	0.70351	0.0581091
1000	1.25315613	0.08094596	0.55178	0.32424652	950	1.25476986	0.02568042	0.52875	0.3221976	2050	1.24902636	0.1214032	0.56919	0.32326098	1600	1.25016257	0.12464871	0.57844	0.32349401
3900	0.29219755	0.10516271	0.597	0.05683603	2150	0.29627332	0.07377855	0.55359	0.04563816	4300	0.29188986	0.11909036	0.60003	0.05991084	5000	0.28964948	0.13159019	0.63636	0.06583351
1000	1.26624139	0.10782563	0.56973	0.33374099	1000	1.2673596	0.07567551	0.5582	0.33221297	1000	1.26683226	0.08617803	0.55769	0.33308734	1000	1.2649108	0.1012037	0.56882	0.33200551
8550	0.46671325	0.17590127	0.68107	0.11088658	1000	0.48617864	0.04307111	0.52518	0.07180062	7350	0.46769535	0.14874212	0.68577	0.11008312	5100	0.47513322	0.11950853	0.64959	0.09373251
6800	0.21129914	0.14390221	0.64855	0.06276316	4600	0.21345727	0.13108658	0.5256	0.05120029	3450	0.21647912	0.07150302	0.60419	0.0435151	5300	0.21281996	0.10187682	0.62452	0.05312693
5100	0.31733486	0.10357068	0.61339	0.06741931	1000	0.32391668	0.02363363	0.47236	0.0435876	6450	0.31335035	0.11050955	0.68912	0.08400376	6850	0.31278158	0.12770394	0.65352	0.07875275
1000	0.972131	0.05810074	0.52868	0.19258685	1000	0.97307397	0.02535141	0.51918	0.19315048	2150	0.96833388	0.09688076	0.58951	0.20003746	1000	0.97449399	0.07552963	0.57726	0.1949268
5400	0.70392614	0.14613499	0.6481	0.15252678	2850	0.7163072	0.09791766	0.60275	0.13323407	5950	0.7036952	0.15353576	0.65107	0.14968799	6900	0.69679414	0.16740644	0.65362	0.15490218
4100	0.51973064	0.11124054	0.61841	0.1005876	1000	0.52698223	0.05673255	0.55006	0.07917108	3600	0.52098521	0.10368145	0.61558	0.09905672	3950	0.51960736	0.0992249	0.62582	0.09808779
9200	1.18923783	0.25534439	0.65411	0.63336375	1000	1.24646848	0.06303068	0.52791	0.68354758	10000	1.17381811	0.29200076	0.67698	0.63581327	8900	1.18332972	0.271841	0.66407	0.63317806
6900	0.98979462	0.18699228	0.62362	0.23404093	1000	1.01364735	0.08335809	0.56223	0.20859108	7400	0.98522865	0.19739832	0.64789	0.24603726	6900	0.98800483	0.1885931	0.64138	0.23664088
10000	0.55801811	0.21132477	0.69594	0.14489597	4700	0.58091638	0.11956676	0.6236	0.11254861	6800	0.57243486	0.15776527	0.66771	0.12373908	7950	0.56683971	0.17025079	0.66666	0.13083726
4500	0.61354874	0.10760254	0.61721	0.11651633	1550	0.62077208	0.07164686	0.57749	0.1013942	3500	0.61595387	0.0974182	0.60813	0.11369397	5300	0.61006964	0.11624877	0.63755	0.1232073
700	0.50332579	0.0154709	0.51373	0.07312298	1000	0.50289058	0.03830439	0.48185	0.07415479	1000	0.50304531	0.02766079	0.55409	0.07477235	1000	0.50169	0.03309828	0.57409	0.07443081
850	0.35936224	0.02162931	0.5023	0.04991044	2850	0.35659846	0.07142315	0.59603	0.06066708	2750	0.35526905	0.0994785	0.51759	0.06644689	3050	0.35448416	0.065043	0.61453	0.06510285
8350	0.45666404	0.14079677	0.66188	0.10680004	1550	0.47224743	0.06035681	0.61289	0.07292957	7900	0.45387179	0.14778611	0.69406	0.10707395	7400	0.45578305	0.14673978	0.6715	0.09917412
10000	0.10863266	0.18110769	0.77106	0.08538344	3400	0.11993549	0.07127804	0.52947	0.03501758	2900	0.1213404	0.03493757	0.64111	0.02621418	3300	0.12053525	0.04646298	0.6847	0.02955502
6350	1.20734126	0.20917943	0.6222	0.33533612	1000	1.23970415	0.04265829	0.52012	0.31309919	6700	1.20434519	0.21298253	0.63028	0.34070308	6450	1.20361841	0.21242585	0.63295	0.33889403
5450	0.17612103	0.10748157	0.68829	0.06010374	4000	0.17936354	0.06054844	0.68621	0.0451251	2700	0.17992565	0.06742343	0.67166	0.04198846	4250	0.18375806	0.068618	0.70016	0.05242042
7100	0.94300551	0.19189642	0.64592	0.22678454	1000	0.9697819	0.07130525	0.52136	0.19188376	6050	0.94928437	0.17268436	0.63228	0.22327083	6800	0.9454221	0.18798886	0.64333	0.2217386
8900	0.33950855	0.19142707	0.67747	0.09068936	4950	0.3523303	0.09473527	0.62718	0.07720208	10000	0.33489812	0.17032866	0.74349	0.10848268	8250	0.34015746	0.15076398	0.69985	0.08997417
1450	0.80471719	0.09417233	0.5871	0.14500022	1000	0.80638829	0.07978769	0.5454	0.14224659	4000	0.79679498	0.14015473	0.59807	0.15171745	3700	0.79947251	0.12752967	0.60724	0.15005716
3450	0.47729582	0.0809657	0.62661	0.08532334	1000	0.48211905	0.02822801	0.52478	0.06978815	4450	0.47513352	0.0949845	0.64429	0.08900044	3250	0.47651647	0.0830952	0.61	0.08378231
1600	0.27553479	0.0556004	0.55904	0.04063612	1000	0.2761634	0.03339038	0.50962	0.03704158	1200	0.2760707	0.03405051	0.5484	0.03853897	3900	0.27703635	0.08190026	0.60266	0.05930378
10000	1.19144072	0.31324023	0.69544	0.36891258	1000	1.27466969	0.10194021	0.55911	0.33824553	10000	1.17972548	0.33136783	0.70841	0.3785064	10000	1.17874339	0.33787498	0.70885	0.38128059
1000	1.02816214	0.01843957	0.51563	0.78793745	1000	1.02738292	0.06097359	0.53082	0.78715078	1000	1.02711148	0.06340699	0.55668	0.78798622	1000	1.02787868	0.04661074	0.55088	0.78739104
2950	1.12790984	0.13630616	0.57217	0.26135286	1000	1.13531588	0.06997227	0.5459	0.25813785	1000	1.13528626	0.07160625	0.55597	0.25854781	1000	1.13436829	0.10484991	0.56462	0.25810764
4600	0.33215561	0.09552712	0.64608	0.06941239	1000	0.3390397	0.05029379	0.61228	0.04518873	5150	0.33166863	0.07857821	0.6693	0.07257761	4600	0.3309513	0.09715902	0.68923	0.07577864
5750	0.25397501	0.13732206	0.71781	0.09623958	4450	0.25592923	0.13739362	0.61839	0.07858397	4550	0.25863633	0.11130295	0.68814	0.06368173	7450	0.25126891	0.17001845	0.76537	0.0948298
1200	0.51448089	0.04865502	0.60352	0.07892815	850	0.51562648	0.03049305	0.48925	0.075276	4050	0.50960245	0.08204424	0.63336	0.09417406	2650	0.51118237	0.07147029	0.61302	0.08582372
2250	0.24130439	0.04035451	0.60492	0.04022721	1700	0.2419915	0.04192852	0.61683	0.03504001	3450	0.2386788	0.08082682	0.66366	0.05316365	2150	0.24064986	0.05042826	0.62651	0.04331998
2800	0.28488761	0.05615149	0.60379	0.0506888	2550	0.28545385	0.05867086	0.5561	0.04519348	2000	0.28596958	0.03786472	0.60386	0.04541022	4850	0.2814608	0.07341479	0.65615	0.05850006

4250	0.18345867	0.09347746	0.67519	0.05936731	1600	0.1881348	0.03580658	0.57983	0.03043654	1850	0.18804795	0.02985553	0.62964	0.03067177	6750	0.17797319	0.11865092	0.74718	0.06489723
8300	0.57348709	0.18894456	0.67578	0.1337499	2300	0.59739494	0.11960301	0.50288	0.09602426	8400	0.57293621	0.18712501	0.68334	0.13484951	8000	0.57351798	0.20129917	0.66172	0.1238317
4750	0.50772714	0.10811515	0.60364	0.10209239	4650	0.50619591	0.11148512	0.65547	0.09921738	1700	0.51346927	0.06962603	0.58136	0.0817001	4950	0.50889005	0.11505411	0.62957	0.10211663
2400	0.34716234	0.06021262	0.61501	0.05741722	5050	0.34145021	0.10397965	0.60708	0.07642303	1100	0.34878905	0.0402579	0.59189	0.04910508	3400	0.34454015	0.07177687	0.63514	0.06391145
4150	0.31847358	0.09634044	0.59638	0.06175988	1000	0.32423297	0.02162689	0.52326	0.0435183	1750	0.32289586	0.04198085	0.61505	0.05135001	4800	0.32047194	0.10398232	0.65955	0.07312384
2550	0.88176952	0.09910198	0.59674	0.17475598	1000	0.88619586	0.056148	0.54122	0.16621986	1000	0.88571222	0.07081725	0.58066	0.16649315	2050	0.88388774	0.09188811	0.58266	0.17425482
3950	0.28674741	0.1087954	0.61645	0.05876479	1550	0.29170329	0.05443327	0.44536	0.04195372	2000	0.29114678	0.0479282	0.63619	0.04871699	2800	0.28949222	0.05149474	0.63265	0.0543576
4150	1.00056473	0.14286411	0.59827	0.22551207	1000	1.01245284	0.05306634	0.53334	0.20768439	5100	0.99341907	0.16810746	0.61018	0.23564021	4650	0.99517233	0.1577411	0.61281	0.22959132
2050	0.22416416	0.04239536	0.57064	0.03357007	1900	0.22450452	0.03330944	0.62233	0.03532921	3750	0.22199612	0.08702276	0.58369	0.04258114	2250	0.22333108	0.05847446	0.57588	0.03792959
2500	0.57726287	0.09969655	0.50754	0.09360334	1000	0.58205477	0.04544489	0.46593	0.08837318	3200	0.57564257	0.10342102	0.54919	0.10016635	2900	0.57507299	0.10762558	0.53135	0.09820575
1000	0.8942686	0.04761072	0.54962	0.16796638	1000	0.89396821	0.05914978	0.55271	0.16848919	4500	0.88264544	0.13798468	0.60624	0.18391757	2950	0.88931392	0.11762438	0.58348	0.17552021
1000	0.86995238	0.07113798	0.56636	0.16127416	1000	0.87085506	0.0343487	0.53212	0.16052224	2150	0.86676579	0.09718447	0.55022	0.16456799	1000	0.87215229	0.02225379	0.52148	0.16134701
3350	0.8196641	0.10700618	0.6093	0.16146804	1000	0.82654636	0.04817825	0.53354	0.14797718	4200	0.81514416	0.13602222	0.61529	0.16997317	3500	0.81935745	0.12076839	0.6048	0.16462908
1650	0.1755591	0.04442416	0.48013	0.02746385	1750	0.17533885	0.03935936	0.49174	0.0309434	1950	0.17495327	0.06757788	0.5801	0.0327375	1350	0.17532089	0.0293258	0.58678	0.02641901
4400	0.1572429	0.09068565	0.61385	0.05332133	3650	0.15977465	0.08817427	0.6157	0.04429875	3550	0.16053746	0.07559499	0.63563	0.0379922	3050	0.15968547	0.05974092	0.57017	0.03818589
2850	0.19905558	0.06350409	0.58225	0.0404157	4600	0.19855854	0.06676124	0.61218	0.04206872	2350	0.19956095	0.03978531	0.59581	0.03886025	3450	0.19725046	0.07195326	0.56922	0.04471173
7000	0.2872528	0.11447311	0.69217	0.07874036	1400	0.29710768	0.06092638	0.41733	0.04282977	2550	0.2961708	0.05074827	0.60272	0.04881582	3900	0.29293853	0.06766935	0.61323	0.05942145
5100	0.85473309	0.14536139	0.6171	0.18309413	1600	0.86559216	0.09012937	0.58285	0.16381528	4750	0.85585231	0.13246903	0.61273	0.18099397	6350	0.85002357	0.17353159	0.64665	0.18716293
8100	0.30677971	0.13256486	0.67069	0.07813193	1000	0.31898543	0.02814411	0.55038	0.04359179	10000	0.29576712	0.189397	0.75447	0.1004148	7100	0.30451838	0.15135451	0.69976	0.0825335
5150	0.48858164	0.10432795	0.63155	0.10067902	600	0.49928547	0.00843133	0.48179	0.070849	2750	0.49441698	0.08161584	0.61936	0.08511602	4800	0.48687727	0.11416927	0.65274	0.09740739
4200	0.20878403	0.09624252	0.58081	0.05013981	1000	0.21331925	0.02194767	0.56571	0.02850901	3350	0.21017507	0.08762058	0.62404	0.04247143	1500	0.21230625	0.044562	0.58878	0.03326838
1300	0.14290969	0.02441119	0.62415	0.02273327	3000	0.14116749	0.04891473	0.66147	0.04000942	3300	0.14020639	0.09099517	0.64307	0.03756173	5000	0.13794458	0.08135285	0.6705	0.0501783
8650	0.88971933	0.24448326	0.67126	0.22940354	2100	0.93025583	0.105157	0.57324	0.18655479	10000	0.87626015	0.26694169	0.69942	0.22876301	10000	0.88193663	0.27521501	0.69286	0.22755135
8700	0.35896391	0.18852812	0.72289	0.10707401	1550	0.38238	0.04664233	0.59171	0.05607483	9850	0.35082955	0.21552831	0.74471	0.10734991	10000	0.29317518	0.22042677	0.78768	0.10251533
10000	0.7812952	0.3061333	0.74248	0.23842116	1000	0.85962605	0.05443062	0.52763	0.15800856	10000	0.75401602	0.36730746	0.76991	0.24187064	10000	0.75150094	0.36862574	0.76906	0.24092196
1250	0.09998726	0.02730088	0.63936	0.01928321	7900	0.0920253	0.22179389	0.66936	0.05702291	3800	0.09480987	0.09304785	0.66517	0.05054209	5000	0.09468033	0.10321426	0.73269	0.05513078
9300	0.44945688	0.18765642	0.71502	0.12580223	1150	0.47704634	0.05870127	0.52414	0.07108627	8100	0.4568346	0.14144166	0.7095	0.11267048	6500	0.46383186	0.12991963	0.69034	0.10306529
2450	0.41810613	0.0707562	0.60018	0.06762889	2300	0.41809505	0.06746123	0.53074	0.06564898	1000	0.42032628	0.03240005	0.58331	0.06108304	2450	0.41705945	0.07108751	0.57124	0.07093441
3250	0.21684952	0.06271438	0.64081	0.04466156	1000	0.21936098	0.02077929	0.5811	0.02903026	4800	0.21386949	0.08448696	0.67278	0.05442072	3750	0.21521649	0.05640694	0.6671	0.04887778
9250	0.34549759	0.13907925	0.7229	0.10269793	4900	0.35344334	0.13781962	0.53305	0.08727192	8050	0.34467772	0.13062358	0.73907	0.10724271	7150	0.34675119	0.13430698	0.7136	0.0975402
2850	0.26293277	0.05748723	0.55528	0.04629155	1600	0.26447808	0.03845204	0.56172	0.04255383	2850	0.2632735	0.0534185	0.59919	0.04677167	5700	0.25777216	0.09721651	0.61363	0.06236011
7750	0.2798824	0.15128231	0.73917	0.08911839	1250	0.29728245	0.04393542	0.55044	0.04177197	7250	0.28455998	0.13815941	0.72487	0.0846699	6700	0.28515192	0.13667908	0.70078	0.07069381
7700	0.2982922	0.13276155	0.71399	0.08355574	3950	0.30943405	0.06370515	0.65895	0.05833971	1400	0.313097	0.04068088	0.59765	0.04452012	4400	0.30538074	0.0963458	0.67778	0.06592586
2100	0.37188854	0.06573749	0.58863	0.06065696	1000	0.37393826	0.01978681	0.5038	0.05289213	1900	0.37202881	0.07212589	0.57627	0.05742044	1200	0.3770664	0.04748939	0.58456	0.05424892
1950	0.16233571	0.0357477	0.5481	0.02920744	4500	0.15919117	0.06507533	0.67945	0.05008718	3050	0.161274	0.04296404	0.71834	0.04107022	850	0.16289184	0.03001803	0.55363	0.02183146
2100	0.32276773	0.05109326	0.60195	0.04987901	900	0.3242158	0.02474064	0.51346	0.04302782	4200	0.31689226	0.12240692	0.59943	0.06943114	1600	0.32210281	0.06189341	0.55417	0.05315501
4500	0.18256794	0.11235168	0.67863	0.05249917	2600	0.18625811	0.09135079	0.50843	0.03911564	4350	0.18073032	0.15748299	0.59986	0.06565237	6250	0.17661473	0.16822941	0.70578	0.07478203
3100	0.20352399	0.08539747	0.5947	0.04589198	4250	0.20145076	0.10631651	0.58032	0.05489573	3450	0.20472621	0.06216075	0.68996	0.04511625	4500	0.20077975	0.09583239	0.6744	0.05311938
5600	0.21831414	0.08091412	0.70029	0.06984601	1100	0.22499559	0.02903264	0.47765	0.03248131	4700	0.22029924	0.08559706	0.66376	0.05487733	5200	0.21964909	0.06615434	0.68185	0.05735247

7400	0.20325378	0.10691422	0.71946	0.08399839	3000	0.21066543	0.06870883	0.60945	0.04583219	3500	0.21029402	0.06668205	0.64995	0.04531753	3950	0.2084462	0.08220825	0.65208	0.0562408
2050	0.34253413	0.06395478	0.56635	0.05517892	1000	0.34396753	0.03402184	0.55171	0.04784213	1650	0.34335639	0.04241084	0.60862	0.05356352	1800	0.34219597	0.06536022	0.56455	0.05421728
6200	0.35060573	0.14461897	0.67346	0.08525745	2200	0.36197987	0.06252256	0.56265	0.05958634	1000	0.36387071	0.03079733	0.57906	0.0506557	2950	0.35965816	0.06881223	0.62219	0.06832792
8200	0.52106251	0.17257767	0.68216	0.12256582	1000	0.54317528	0.04927481	0.52307	0.08250794	8600	0.51839811	0.17538306	0.68982	0.12658774	8000	0.52081874	0.16471018	0.6893	0.11765578
2950	0.16119824	0.06150167	0.55189	0.03589426	2600	0.16107127	0.06637521	0.56448	0.03721665	8850	0.15302318	0.11617488	0.75084	0.06301739	4250	0.1578908	0.10576433	0.62419	0.04307184
1000	1.17341167	0.04906896	0.52113	0.27567672	1000	1.17391484	0.03015312	0.52502	0.27552409	1000	1.1721777	0.09231855	0.54697	0.27599344	1000	1.17251465	0.07273329	0.53281	0.2754163
1000	0.42928681	0.05215852	0.54419	0.06069828	1000	0.42966953	0.02833018	0.53735	0.06042288	6950	0.41685272	0.15532077	0.63737	0.09203756	3400	0.42403066	0.10531126	0.57575	0.07679977
2700	0.13496622	0.04303732	0.64193	0.0328626	1300	0.13621697	0.01403386	0.60828	0.02210713	1400	0.13607036	0.02231243	0.60967	0.02316323	3050	0.13372449	0.04819297	0.67965	0.03864518
7400	0.8189683	0.17977278	0.65566	0.1937911	1000	0.84620063	0.04112693	0.5483	0.15526369	7550	0.81773241	0.17951098	0.65594	0.19447182	5950	0.82377082	0.16551812	0.64531	0.18831433
4550	0.33322978	0.08490588	0.59173	0.07009525	1750	0.33841144	0.04726642	0.55012	0.05056722	4750	0.32991356	0.12878189	0.60434	0.0745274	2750	0.33565782	0.07285234	0.5939	0.06112798
4000	0.58478391	0.1238017	0.62494	0.11334917	1000	0.59270622	0.0536189	0.56785	0.09234345	3350	0.58731146	0.09910939	0.60317	0.10780333	4550	0.58230899	0.11612007	0.64697	0.11446139
2800	1.13127407	0.12963214	0.58629	0.27196708	1000	1.13957604	0.03962188	0.52159	0.25915377	2100	1.1342866	0.10989505	0.5811	0.26869202	1600	1.13752395	0.10323375	0.57035	0.26611563
7700	0.23317793	0.15760112	0.71813	0.08141804	3950	0.24367155	0.08811328	0.67001	0.05800582	4250	0.24313295	0.08632029	0.68525	0.05853554	7200	0.23676173	0.11945281	0.70731	0.06742312
4950	0.28265122	0.05984648	0.67427	0.06142352	1000	0.28747735	0.02845891	0.5056	0.03832393	1000	0.28723635	0.02245251	0.56831	0.03905607	1000	0.28660687	0.03417832	0.57275	0.03935263
4050	0.43333479	0.08398068	0.64083	0.08346897	1000	0.43839096	0.03604436	0.54636	0.06326497	750	0.43900166	0.00894233	0.54691	0.06101185	1000	0.43732557	0.03851347	0.60952	0.06313936
1000	0.96973026	0.07316307	0.54762	0.1913268	1000	0.97067678	0.02273037	0.51435	0.19154153	1000	0.96990785	0.05943413	0.54344	0.19275194	1000	0.96816282	0.06514032	0.54477	0.19126795
4050	0.84635615	0.12771363	0.61195	0.18072612	1000	0.85695854	0.0545318	0.53973	0.15779567	5150	0.8426416	0.13805536	0.62927	0.18105421	3850	0.84623171	0.11931027	0.62194	0.17211078
7800	0.15893864	0.13139818	0.70661	0.07423819	3450	0.16613823	0.08339707	0.63257	0.04877448	5250	0.16514976	0.06748255	0.69145	0.0559193	5100	0.16454292	0.09286566	0.65051	0.05161833
1900	0.2123116	0.04377545	0.57998	0.0366287	1800	0.21264666	0.0358918	0.58673	0.03304986	3100	0.21023535	0.07139409	0.6042	0.04471894	1650	0.21207737	0.04167373	0.6005	0.03485937
9850	0.08930872	0.0989168	0.87412	0.08335255	4700	0.09679052	0.09402248	0.70433	0.04885551	2450	0.09890522	0.03182937	0.7521	0.02988998	5100	0.09617347	0.06911153	0.77796	0.05631622

Table S4.3: Mean TSS and Sensivity (SENS) values for each species and all model configurations. Mean calculated from 50 models per species (5 algorithms , 10 repeats),w = weighted mean, m = non weighted mean, se = standard error

.																								
	Hydroclimate & Hydrology						Bioclimate & Hydrology						Hydroclimate & Bioclimate						Hydroclimate, Bioclimate & Hydrology					
species	**TSS (w)**	**Sensitivity (w)**	**TSS (m)**	**Sensitivity (m)**	**TSS (se)**	**Sensitivity (se)**	**TSS (w)**	**Sensitivity (w)**	**TSS (m)**	**Sensitivity (m)**	**TSS (se)**	**Sensitivity (se)**	**TSS (w)**	**Sensitivity (w)**	**TSS (m)**	**Sensitivity (m)**	**TSS (se)**	**Sensitivity (se)**	**TSS (w)**	**Sensitivity (w)**	**TSS (m)**	**Sensitivity (m)**	**TSS (se)**	**Sensitivity (se)**
Alainites muticus	0.71	0.89	0.69	0.88	0.01	0.01	0.77	0.91	0.75	0.90	0.01	0.01	0.81	0.92	0.8	0.91	0.01	0.01	0.77	0.90	0.76	0.89	0.02	0.02
Anabolia nervosa	0.53	0.75	0.48	0.71	0.02	0.02	0.36	0.63	0.34	0.63	0.01	0.01	0.34	0.65	0.31	0.65	0.01	0.02	0.37	0.68	0.34	0.67	0.01	0.01
Anacaena globulus	0.58	0.79	0.55	0.77	0.02	0.02	0.54	0.74	0.53	0.73	0.01	0.02	0.56	0.74	0.54	0.73	0.01	0.02	0.54	0.76	0.52	0.74	0.01	0.03
Ancylus fluviatilis	0.54	0.76	0.47	0.71	0.03	0.02	0.42	0.74	0.41	0.74	0.01	0.01	0.47	0.74	0.46	0.73	0.01	0.01	0.44	0.77	0.43	0.77	0.01	0.01
Anisus vortex	0.69	0.83	0.66	0.82	0.02	0.01	0.70	0.85	0.69	0.85	0.01	0.01	0.56	0.78	0.55	0.77	0.01	0.02	0.58	0.79	0.57	0.79	0.01	0.01
Anodonta anatina	0.74	0.91	0.73	0.91	0.01	0.01	0.80	0.93	0.79	0.92	0.01	0.01	0.68	0.79	0.65	0.76	0.02	0.02	0.74	0.87	0.73	0.86	0.01	0.02
Apsectrotanypus trifascipennis	0.62	0.81	0.59	0.80	0.02	0.02	0.67	0.86	0.65	0.85	0.02	0.02	0.47	0.74	0.45	0.72	0.01	0.03	0.64	0.83	0.63	0.82	0.02	0.02
Atherix ibis	0.59	0.80	0.55	0.78	0.02	0.01	0.63	0.82	0.60	0.80	0.02	0.02	0.51	0.75	0.5	0.76	0.01	0.01	0.63	0.81	0.59	0.79	0.02	0.02
Athripsodes cinereus	0.63	0.80	0.58	0.77	0.02	0.02	0.60	0.80	0.56	0.78	0.02	0.02	0.46	0.7	0.43	0.68	0.02	0.02	0.48	0.72	0.46	0.71	0.01	0.01
Baetis lutheri	0.58	0.77	0.54	0.75	0.02	0.02	0.70	0.85	0.67	0.85	0.02	0.01	0.61	0.88	0.6	0.87	0.01	0.01	0.60	0.84	0.60	0.84	0.01	0.01
Baetis rhodani	0.57	0.77	0.54	0.75	0.02	0.02	0.52	0.81	0.49	0.80	0.02	0.02	0.43	0.8	0.42	0.79	0.01	0.01	0.44	0.81	0.43	0.80	0.01	0.01
Bithynia tentaculata	0.62	0.80	0.59	0.78	0.02	0.02	0.60	0.78	0.56	0.75	0.02	0.02	0.48	0.7	0.46	0.68	0.01	0.02	0.48	0.73	0.45	0.72	0.01	0.02
Caenis horaria	0.61	0.81	0.59	0.80	0.02	0.01	0.71	0.84	0.69	0.84	0.01	0.01	0.58	0.78	0.57	0.78	0.01	0.01	0.56	0.76	0.54	0.75	0.01	0.01
Caenis luctuosa	0.62	0.76	0.57	0.74	0.02	0.02	0.68	0.82	0.66	0.82	0.02	0.01	0.52	0.74	0.47	0.72	0.02	0.02	0.52	0.77	0.50	0.76	0.02	0.01
Calopteryx virgo	0.56	0.76	0.52	0.73	0.02	0.02	0.63	0.80	0.60	0.79	0.02	0.02	0.45	0.72	0.41	0.7	0.02	0.02	0.49	0.72	0.47	0.72	0.01	0.02
Cheumatopsyche lepida	0.79	0.89	0.77	0.89	0.02	0.01	0.74	0.89	0.72	0.88	0.02	0.01	0.73	0.89	0.72	0.89	0.01	0.01	0.73	0.88	0.72	0.89	0.01	0.01

Cloeon dipterum	0.61	0.81	0.57	0.79	0.02	0.02	0.72	0.87	0.70	0.86	0.01	0.01	0.53	0.82	0.51	0.81	0.01	0.02	0.56	0.79	0.55	0.78	0.01	0.02
Corbicula fluminea	0.77	0.91	0.76	0.90	0.02	0.02	0.83	0.94	0.82	0.94	0.01	0.01	0.78	0.92	0.76	0.9	0.01	0.02	0.91	0.97	0.91	0.96	0.01	0.01
Dugesia gonocephala	0.58	0.76	0.55	0.73	0.02	0.02	0.61	0.83	0.57	0.81	0.02	0.02	0.5	0.82	0.5	0.82	0.01	0.01	0.50	0.80	0.50	0.79	0.01	0.01
Ecclisopteryx dalecarlica	0.68	0.85	0.65	0.83	0.02	0.02	0.82	0.93	0.81	0.93	0.01	0.01	0.64	0.79	0.62	0.79	0.02	0.02	0.66	0.82	0.63	0.81	0.02	0.02
Elmis aenea	0.58	0.76	0.53	0.73	0.02	0.02	0.59	0.77	0.55	0.74	0.02	0.02	0.43	0.64	0.39	0.61	0.02	0.02	0.45	0.69	0.43	0.68	0.02	0.02
Ephemera vulgata	0.58	0.81	0.56	0.80	0.01	0.02	0.76	0.86	0.76	0.85	0.01	0.01	0.63	0.81	0.62	0.8	0.01	0.01	0.64	0.82	0.63	0.81	0.01	0.01
Ephemerella mucronata	0.47	0.71	0.46	0.70	0.01	0.02	0.71	0.85	0.69	0.84	0.02	0.01	0.57	0.84	0.57	0.84	0.01	0.01	0.58	0.85	0.56	0.85	0.01	0.01
Erpobdella nigricollis	0.44	0.74	0.40	0.71	0.02	0.02	0.61	0.82	0.57	0.80	0.02	0.02	0.49	0.73	0.48	0.72	0.01	0.02	0.49	0.71	0.48	0.71	0.01	0.02
Galba truncatula	0.47	0.67	0.43	0.63	0.02	0.03	0.64	0.82	0.62	0.79	0.02	0.02	0.47	0.67	0.44	0.64	0.02	0.02	0.68	0.86	0.66	0.85	0.01	0.02
Gammarus fossarum	0.40	0.66	0.36	0.63	0.02	0.02	0.63	0.84	0.61	0.84	0.02	0.01	0.54	0.82	0.53	0.81	0.01	0.01	0.51	0.78	0.49	0.78	0.01	0.01
Gammarus pulex	0.30	0.67	0.25	0.64	0.02	0.02	0.44	0.76	0.39	0.74	0.02	0.02	0.34	0.75	0.3	0.74	0.02	0.01	0.46	0.80	0.41	0.78	0.02	0.02
Gammarus roeselii	0.41	0.68	0.31	0.65	0.02	0.02	0.60	0.85	0.56	0.85	0.02	0.01	0.49	0.76	0.44	0.73	0.02	0.02	0.62	0.80	0.60	0.79	0.02	0.02
Gammarus tigrinus	0.59	0.79	0.54	0.77	0.02	0.02	0.69	0.88	0.67	0.87	0.02	0.01	0.62	0.82	0.6	0.82	0.02	0.02	0.63	0.82	0.61	0.81	0.02	0.02
Gerris lacustris	0.53	0.70	0.51	0.68	0.02	0.02	0.66	0.82	0.65	0.81	0.01	0.02	0.56	0.73	0.54	0.71	0.02	0.03	0.60	0.75	0.58	0.74	0.01	0.02
Glossiphonia nebulosa	0.59	0.80	0.57	0.79	0.01	0.01	0.72	0.87	0.70	0.86	0.02	0.01	0.59	0.82	0.57	0.8	0.02	0.02	0.57	0.79	0.55	0.77	0.02	0.02
Glyphotaelius pellucidus	0.61	0.80	0.57	0.77	0.02	0.02	0.71	0.82	0.70	0.81	0.01	0.02	0.64	0.78	0.6	0.75	0.02	0.02	0.81	0.90	0.78	0.88	0.02	0.02
Gomphus vulgatissimus	0.54	0.80	0.51	0.79	0.02	0.02	0.75	0.87	0.74	0.86	0.01	0.01	0.59	0.83	0.57	0.82	0.01	0.02	0.60	0.82	0.58	0.80	0.02	0.02
Graptodytes pictus	0.52	0.70	0.49	0.66	0.02	0.03	0.75	0.92	0.74	0.92	0.01	0.01	0.64	0.8	0.62	0.79	0.01	0.02	0.65	0.83	0.62	0.80	0.02	0.02
Gyraulus albus	0.43	0.73	0.39	0.71	0.02	0.02	0.67	0.83	0.66	0.82	0.01	0.01	0.47	0.65	0.46	0.65	0.01	0.02	0.65	0.83	0.63	0.83	0.02	0.01
Habrophlebia fusca	0.53	0.74	0.50	0.73	0.02	0.02	0.67	0.85	0.64	0.84	0.02	0.01	0.56	0.73	0.52	0.7	0.02	0.02	0.69	0.82	0.66	0.80	0.02	0.02
Haemopis sanguisuga	0.53	0.77	0.50	0.75	0.02	0.02	0.68	0.82	0.66	0.80	0.02	0.02	0.54	0.75	0.52	0.73	0.02	0.02	0.55	0.75	0.52	0.72	0.02	0.02

Halesus digitatus	0.45	0.74	0.43	0.71	0.02	0.03	0.66	0.83	0.65	0.82	0.01	0.02	0.51	0.79	0.5	0.79	0.01	0.02	0.73	0.85	0.71	0.84	0.01	0.02
Halesus radiatus	0.34	0.65	0.30	0.64	0.02	0.02	0.62	0.78	0.60	0.77	0.02	0.02	0.38	0.64	0.35	0.64	0.01	0.02	0.60	0.80	0.56	0.79	0.02	0.02
Halesus tesselatus	0.58	0.84	0.55	0.82	0.02	0.02	0.76	0.88	0.75	0.87	0.01	0.01	0.64	0.84	0.62	0.83	0.01	0.02	0.72	0.87	0.70	0.86	0.02	0.02
Helobdella stagnalis	0.33	0.64	0.31	0.63	0.01	0.02	0.59	0.77	0.55	0.75	0.02	0.02	0.36	0.63	0.34	0.61	0.01	0.02	0.62	0.77	0.59	0.75	0.02	0.02
Hemiclepsis marginata	0.57	0.79	0.54	0.75	0.02	0.03	0.76	0.87	0.75	0.86	0.01	0.01	0.64	0.82	0.63	0.81	0.01	0.02	0.68	0.87	0.66	0.86	0.01	0.02
Heptagenia sulphurea	0.40	0.72	0.35	0.68	0.02	0.02	0.66	0.81	0.63	0.80	0.02	0.01	0.47	0.76	0.45	0.77	0.01	0.02	0.67	0.83	0.65	0.83	0.02	0.01
Hydropsyche angustipennis	0.41	0.71	0.37	0.71	0.02	0.01	0.61	0.79	0.56	0.77	0.02	0.02	0.41	0.65	0.38	0.62	0.02	0.02	0.45	0.69	0.40	0.67	0.02	0.02
Hydropsyche instabilis	0.42	0.68	0.39	0.67	0.01	0.01	0.65	0.81	0.63	0.79	0.02	0.02	0.55	0.87	0.54	0.87	0.01	0.01	0.69	0.85	0.67	0.84	0.02	0.01
Hydropsyche pellucidula	0.34	0.66	0.30	0.63	0.02	0.02	0.56	0.76	0.52	0.73	0.02	0.02	0.37	0.66	0.34	0.64	0.02	0.02	0.59	0.79	0.56	0.78	0.02	0.01
Hyphydrus ovatus	0.61	0.80	0.57	0.76	0.02	0.03	0.81	0.94	0.80	0.94	0.01	0.01	0.58	0.82	0.56	0.8	0.02	0.03	0.68	0.87	0.65	0.85	0.02	0.02
Ilyocoris cimicoides	0.51	0.74	0.45	0.69	0.02	0.03	0.81	0.91	0.79	0.90	0.02	0.01	0.73	0.88	0.72	0.86	0.01	0.02	0.84	0.95	0.83	0.95	0.01	0.01
Isoperla grammatica	0.58	0.76	0.52	0.70	0.03	0.03	0.68	0.86	0.64	0.85	0.02	0.02	0.57	0.81	0.53	0.78	0.02	0.03	0.81	0.90	0.79	0.89	0.02	0.02
Laccophilus hyalinus	0.49	0.75	0.46	0.73	0.02	0.02	0.65	0.82	0.64	0.82	0.01	0.01	0.57	0.81	0.56	0.8	0.01	0.02	0.82	0.91	0.82	0.91	0.01	0.01
Lepidostoma basale	0.35	0.62	0.32	0.61	0.01	0.02	0.63	0.86	0.60	0.85	0.02	0.01	0.55	0.84	0.54	0.84	0.01	0.01	0.52	0.83	0.51	0.83	0.01	0.01
Leuctra fusca	0.62	0.86	0.59	0.83	0.02	0.02	0.82	0.92	0.82	0.92	0.01	0.01	0.76	0.85	0.76	0.84	0.01	0.01	0.87	0.92	0.87	0.92	0.01	0.01
Leuctra geniculata	0.51	0.77	0.49	0.77	0.01	0.02	0.74	0.87	0.73	0.86	0.01	0.01	0.63	0.78	0.61	0.77	0.01	0.01	0.73	0.87	0.72	0.87	0.01	0.01
Limnephilus extricatus	0.58	0.76	0.54	0.74	0.02	0.02	0.68	0.82	0.66	0.81	0.02	0.02	0.6	0.75	0.58	0.73	0.02	0.02	0.75	0.87	0.73	0.86	0.02	0.02
Limnephilus flavicornis	0.67	0.81	0.65	0.78	0.02	0.02	0.84	0.92	0.83	0.91	0.02	0.02	0.72	0.85	0.7	0.83	0.02	0.02	0.65	0.82	0.63	0.80	0.02	0.03
Limnephilus lunatus	0.42	0.74	0.38	0.74	0.02	0.01	0.59	0.77	0.54	0.74	0.02	0.02	0.44	0.72	0.41	0.72	0.02	0.01	0.44	0.69	0.41	0.69	0.02	0.02
Limnodrilus claparedeanus	0.73	0.86	0.72	0.85	0.01	0.01	0.77	0.87	0.76	0.86	0.01	0.01	0.76	0.87	0.75	0.87	0.01	0.01	0.84	0.91	0.83	0.91	0.01	0.01

Limnodrilus hoffmeisteri	0.60	0.79	0.56	0.76	0.02	0.02	0.66	0.82	0.64	0.81	0.02	0.01	0.55	0.69	0.54	0.69	0.01	0.01	0.69	0.83	0.67	0.82	0.02	0.01
Lithax obscurus	0.89	0.98	0.87	0.97	0.02	0.01	0.94	0.99	0.93	0.98	0.01	0.01	0.91	0.97	0.88	0.94	0.02	0.02	0.96	0.99	0.96	0.99	0.01	0.01
Lymnaea stagnalis	0.72	0.87	0.71	0.87	0.01	0.01	0.69	0.83	0.66	0.83	0.02	0.01	0.63	0.81	0.61	0.8	0.01	0.02	0.75	0.89	0.74	0.89	0.01	0.01
Lype reducta	0.57	0.84	0.52	0.82	0.02	0.02	0.59	0.82	0.56	0.81	0.02	0.02	0.43	0.71	0.4	0.67	0.02	0.03	0.40	0.66	0.38	0.65	0.01	0.03
Melampophylax mucoreus	0.62	0.85	0.60	0.83	0.02	0.02	0.77	0.92	0.76	0.92	0.01	0.01	0.66	0.88	0.65	0.87	0.01	0.02	0.88	0.95	0.87	0.95	0.01	0.01
Molanna angustata	0.74	0.91	0.70	0.88	0.03	0.02	0.82	0.92	0.81	0.92	0.01	0.01	0.65	0.87	0.64	0.86	0.01	0.02	0.86	0.94	0.86	0.94	0.01	0.01
Musculium lacustre	0.67	0.85	0.64	0.84	0.02	0.02	0.60	0.81	0.57	0.79	0.02	0.02	0.52	0.77	0.49	0.75	0.02	0.03	0.53	0.73	0.49	0.70	0.02	0.03
Nais elinguis	0.76	0.90	0.74	0.89	0.02	0.01	0.81	0.90	0.79	0.89	0.02	0.01	0.76	0.87	0.74	0.86	0.02	0.01	0.84	0.93	0.83	0.93	0.01	0.01
Nemoura cinerea	0.58	0.81	0.54	0.80	0.02	0.02	0.67	0.81	0.63	0.78	0.02	0.02	0.54	0.75	0.52	0.74	0.01	0.02	0.72	0.88	0.70	0.87	0.02	0.02
Nepa cinerea	0.59	0.79	0.56	0.77	0.02	0.02	0.59	0.79	0.54	0.78	0.02	0.02	0.47	0.77	0.45	0.76	0.01	0.02	0.47	0.77	0.44	0.74	0.02	0.02
Neureclipsis bimaculata	0.78	0.93	0.75	0.91	0.02	0.02	0.77	0.91	0.75	0.90	0.02	0.01	0.69	0.8	0.65	0.76	0.02	0.03	0.80	0.93	0.79	0.92	0.02	0.01
Notidobia ciliaris	0.58	0.83	0.53	0.80	0.02	0.02	0.67	0.83	0.63	0.80	0.02	0.02	0.53	0.73	0.49	0.7	0.02	0.02	0.66	0.84	0.65	0.83	0.02	0.02
Oecismus monedula	0.81	0.93	0.80	0.92	0.01	0.01	0.78	0.88	0.77	0.87	0.01	0.01	0.73	0.87	0.71	0.85	0.02	0.02	0.87	0.94	0.86	0.93	0.02	0.01
Paratendipes albimanus	0.72	0.87	0.69	0.85	0.02	0.02	0.70	0.84	0.68	0.82	0.02	0.02	0.63	0.8	0.6	0.79	0.02	0.02	0.78	0.91	0.77	0.90	0.01	0.01
Pisidium casertanum	0.70	0.89	0.67	0.88	0.02	0.01	0.77	0.89	0.75	0.88	0.02	0.02	0.6	0.81	0.57	0.78	0.02	0.03	0.61	0.82	0.59	0.80	0.02	0.03
Pisidium nitidum	0.78	0.89	0.77	0.88	0.02	0.01	0.82	0.91	0.81	0.90	0.01	0.01	0.75	0.83	0.74	0.83	0.01	0.01	0.87	0.93	0.86	0.92	0.01	0.01
Pisidium subtruncatum	0.74	0.85	0.72	0.84	0.01	0.01	0.71	0.89	0.69	0.88	0.01	0.01	0.56	0.8	0.55	0.79	0.01	0.02	0.81	0.90	0.79	0.89	0.02	0.01
Planorbis planorbis	0.64	0.77	0.63	0.77	0.01	0.02	0.70	0.85	0.69	0.84	0.01	0.01	0.56	0.78	0.55	0.77	0.01	0.02	0.58	0.76	0.57	0.75	0.01	0.02
Platycnemis pennipes	0.62	0.85	0.59	0.85	0.02	0.01	0.69	0.85	0.67	0.84	0.02	0.01	0.57	0.75	0.56	0.74	0.01	0.02	0.72	0.82	0.70	0.81	0.02	0.01
Plea minutissima	0.65	0.84	0.63	0.82	0.02	0.02	0.77	0.91	0.76	0.90	0.01	0.01	0.66	0.84	0.63	0.82	0.02	0.02	0.79	0.94	0.78	0.93	0.01	0.01
Polycentropus flavomaculatus	0.55	0.73	0.48	0.66	0.03	0.03	0.61	0.83	0.58	0.81	0.02	0.01	0.44	0.74	0.42	0.73	0.01	0.01	0.61	0.82	0.58	0.80	0.02	0.02

Potamophylax rotundipennis	0.67	0.85	0.65	0.85	0.02	0.01	0.59	0.79	0.56	0.78	0.02	0.02	0.59	0.76	0.57	0.75	0.01	0.01	0.70	0.80	0.68	0.79	0.02	0.02
Potamothrix moldaviensis	0.78	0.94	0.76	0.93	0.02	0.01	0.80	0.92	0.80	0.91	0.01	0.01	0.81	0.89	0.77	0.86	0.02	0.02	0.78	0.89	0.77	0.87	0.02	0.02
Proasellus coxalis	0.59	0.81	0.54	0.78	0.02	0.02	0.59	0.77	0.53	0.74	0.02	0.02	0.38	0.67	0.37	0.66	0.01	0.02	0.64	0.84	0.62	0.83	0.02	0.01
Procloeon bifidum	0.67	0.85	0.65	0.84	0.02	0.02	0.62	0.84	0.59	0.83	0.02	0.02	0.58	0.85	0.55	0.84	0.02	0.01	0.72	0.88	0.70	0.87	0.02	0.01
Psychomyia pusilla	0.62	0.83	0.58	0.82	0.02	0.01	0.63	0.80	0.60	0.79	0.02	0.02	0.52	0.7	0.5	0.69	0.01	0.01	0.69	0.83	0.66	0.82	0.02	0.02
Radix balthica	0.54	0.73	0.48	0.69	0.03	0.02	0.58	0.78	0.55	0.76	0.02	0.02	0.34	0.66	0.31	0.65	0.01	0.02	0.60	0.78	0.58	0.76	0.02	0.02
Rhithrogena semicolorata	0.66	0.84	0.62	0.81	0.02	0.02	0.78	0.93	0.77	0.92	0.01	0.01	0.69	0.88	0.68	0.87	0.01	0.02	0.67	0.83	0.65	0.81	0.02	0.02
Rhyacophila evoluta	0.68	0.85	0.65	0.83	0.02	0.02	0.77	0.91	0.76	0.91	0.01	0.01	0.71	0.84	0.69	0.83	0.02	0.02	0.80	0.92	0.79	0.92	0.01	0.01
Stictotarsus duodecimpustulatus	0.59	0.80	0.56	0.78	0.02	0.02	0.60	0.81	0.55	0.78	0.02	0.02	0.51	0.78	0.5	0.77	0.01	0.02	0.73	0.88	0.72	0.87	0.01	0.01
Stylodrilus heringianus	0.56	0.77	0.51	0.75	0.02	0.02	0.58	0.77	0.55	0.75	0.02	0.02	0.49	0.79	0.47	0.78	0.01	0.01	0.60	0.78	0.56	0.75	0.02	0.02
Torleya major	0.61	0.78	0.57	0.76	0.02	0.02	0.70	0.86	0.68	0.86	0.02	0.01	0.58	0.86	0.57	0.86	0.01	0.01	0.67	0.89	0.65	0.89	0.02	0.01
Valvata piscinalis	0.67	0.82	0.63	0.79	0.02	0.02	0.75	0.88	0.74	0.87	0.01	0.02	0.68	0.86	0.66	0.84	0.02	0.02	0.71	0.86	0.68	0.85	0.02	0.02
Velia caprai	0.58	0.78	0.54	0.76	0.02	0.02	0.70	0.83	0.69	0.82	0.01	0.01	0.52	0.73	0.47	0.67	0.02	0.03	0.71	0.84	0.67	0.82	0.02	0.02
Viviparus viviparus	0.83	0.96	0.82	0.95	0.01	0.01	0.82	0.94	0.81	0.93	0.02	0.01	0.87	0.94	0.84	0.92	0.02	0.02	0.94	1.00	0.93	1.00	0.01	0.00
Mean	0.59	0.80	0.56	0.78	0.02	0.02	0.68	0.85	0.66	0.83	0.02	0.02	0.57	0.79	0.55	0.77	0.01	0.02	0.66	0.83	0.64	0.82	0.01	0.02
max	0.89	0.98	0.87	0.97	0.03	0.03	0.94	0.99	0.93	0.98	0.02	0.02	0.91	0.97	0.88	0.94	0.02	0.03	0.96	1.00	0.96	1.00	0.02	0.03
min	0.30	0.62	0.25	0.61	0.01	0.01	0.36	0.63	0.34	0.63	0.01	0.01	0.34	0.63	0.30	0.61	0.01	0.01	0.37	0.66	0.34	0.65	0.01	0.00

Statement of academic integrity

I hereby certify that the submitted thesis "*Improvement of global change projections for riverine benthic macroinvertebrates*" is my own work, and that all published or other sources of material consulted in its preparation have been indicated. All collaboration that has taken place with other researchers is indicated and I have clearly stated my own personal share in those investigations in the Thesis Outline. I confirm that this work has not been submitted to any other university or examining body for a comparable academic award.

Berlin, 07 August 2019

Katherine Sarah Irving

Acknowledgments

This thesis would not have been possible without the support and encouragement of many devoted people in my academic and personal life. I am, and will always be, immensely grateful to all of them. I would especially like to thank the following people:

My main supervisor Dr. Sonja Jähnig for, first giving me the tremendous opportunity to join the working group and pursue this PhD, and second, all the support throughout the last 4 years. My co-supervisor Dr. Mathias Kuemmerlen, whose support and guidance has been exceptional. Thank you for your time, patience and imparted wisdom. I would also like to extend my gratitude to Prof. Dr. Klement Tockner, for support and reviewing the thesis.

Everyone at IGB, senior scientists, fellow PhDers and technicians, you have all been wonderful. Special appreciation goes to everyone in Adlershof; Dr. Markus Venohr, Dr. Alain Maasri, Dr. Andreas Gericke, Annett Wetzig, Vanessa Bremerich, Melissa Schulte and Dr. Simone Podshun for sharing this demanding journey. With special thanks to Judith Mahnkopf, for continual motivational talks, commendation and the fixing of data catastrophes. Also, Dr. Sami Domisch and Dr. Jens Kiesel for technical advice, supportive discussions and always having the time. My science brothers; Dr. Karan Kakouei, Dr. Fengzhi He and Martin Friedrichs, thank you all the helpful discussions, team conference trips, continual reassurance and fun times.

On a personal note, I would like to thank; Andrew LePage for never losing faith and tolerating my nonsense. Also, Oliver Dyson, Lucille Pearce, Karlie & Adam Wagner for reliable support, even from a distance. Extra thanks to Adam Wagner for extreme table formatting. My Berlin friends; Mike Savage for reality checks, perspective, and flow diagrams, and Leah Hinton for pep talks and creative englittering.

Last but not least, I would like to thank my parents: my father Keith Irving, who sadly did not survive to see me finish, I know you would be proud. And to my wonderful mother, Rosie Mathisen, you are my absolute pillar of strength, without you this thesis would not have been physically, or mentally possible.

List of publications relevant to the thesis

Irving K, Kuemmerlen M, Kiesel J, Kakouei K, Domisch S, Jähnig SC. 2018. A high-resolution streamflow and hydrological metrics dataset for ecological modeling using a regression model [Data Descriptor]. *Scientific Data.* 5:180224. doi: 10.1038/sdata.2018.224 (2018).

Kakouei, K**.**, Kiesel, J., Domisch, S., **Irving, K,** Jähnig, S. C., and Kail, J., 2018. Projected effects of climate-change-induced flow alterations on stream macroinvertebrate abundances. *Ecology and evolution*, 8(6), 3393-3409. doi; 10.1002/ece3.3907

Irving K, Jähnig SC. Kuemmerlen M, 2019. Identifying and applying an optimum set of environmental variables in species distribution models. *Inland Waters.* (Accepted). doi; 10.1080/20442041.2019.1653111

Irving K, Jähnig SC. Kuemmerlen M, 2019. Disentangling the influence of climatic and hydrological predictor variables on benthic macroinvertebrate distribution. To be submitted in August 2019