Advancing Global Assessments of Freshwater Systems

Valerio Barbarossa

Advancing Global Assessments of Freshwater Systems

Valerio Barbarossa

Coverdesign: Ridderprint B.V. | www.ridderprint.nl
Lay-out & Printing: Ridderprint B.V. | www.ridderprint.nl

ISBN: 978-94-6375-697-6

© 2019, Valerio Barbarossa, the Netherlands

All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior written permission of the author.

Advancing Global Assessments of Freshwater Systems

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op 6 februari 2020,
om 16.30 uur precies

door

Valerio Barbarossa
geboren op 16 juni 1989
te Perugia, Italië

Promotoren:
Prof. dr. M.A.J. Huijbregts
Prof. dr. J.M.H. King (University of Surrey, Verenigd Koninkrijk)

Copromotor:
Dr. A.M. Schipper

Manuscriptcommissie:
Prof. dr. L.P.M. Lamers
Prof. dr. C. Kroeze (WUR)
Dr. M.T.H. van Vliet (UU)

Paranimfen:
Jelle Hilbers
Marco Zenoni

Dit onderzoek werd gefinancierd door de "European Union's Horizon 2020 research and innovation program" onder de Marie Sklodowska-Curie grant agreement No 641459 (RELIEF)

Dit onderzoek werd uitgevoerd onder auspiciën van de "Graduate Scholl for Socio-Economic and Natural Sciences of the Environment" (SENSE)

All models are wrong, but some are useful

| George E. P. Box

TABLE OF CONTENTS

1 Introduction

1.1 Freshwater systems

Freshwater systems are vital for natural and human communities. Their comparatively scattered and restricted distribution, totaling less than 1% of the globe's surface, belies the exceptionally high number of species they harbor and the essential services they provide to our society (Dudgeon et al., 2006; Strayer et al., 2010). According to the freshwater animal diversity assessment, 9.5% (or ~126,000) of the globally described number of animal species live in freshwater, including one third of all vertebrates and ~40% of all fish species (Balian et al., 2008; Nelson, 2006; Tedesco et al., 2017). The ecosystems comprising this remarkably diverse assemblage of species and their physical habitat are essential for humans. Freshwater ecosystems, indeed, provide water for consumptive and non-consumptive use (including natural filtration/treatment ensuring water quality), aquatic organisms for food and medicines, buffering of flood flows and erosion, and opportunities for recreational and touristic activities (Díaz et al., 2019; Millennium Ecosystem Assessment, 2005). Yet, the human-driven engineering and management of freshwater resources that maximize human access to water and the transformation of freshwater systems through urbanization, industrialization, and land-cover change directly modify and degrade freshwater ecosystems (Dudgeon et al., 2006; Strayer et al., 2010; Vörösmarty et al., 2010). Future climate change is expected to further threaten freshwater ecosystems worldwide (Knouft and Ficklin, 2017; Reid et al., 2019).

1.2 Integrated assessments

The complex nexus of multiple and interacting anthropogenic pressures on natural systems needs to be assessed in an integrated way to gather a synoptic and solution-focused outlook of the problem (Dowlatabadi, 1995). Through a transdisciplinary approach to interpret and quantify the drivers-pressures-state-impact-response pathways (DPSIR; **Figure 1.1**), integrated assessment models (IAMs) are designed to provide useful information to politicians and policy makers (Parson et al., 2007; Parson and Fisher-Vanden, 2002). IAMs allow for projections of human and natural systems based on future development scenarios of human energy use, industrial and urban development, agriculture and land-use/cover changes (Parson et al., 2007; Parson and Fisher-Vanden, 2002). For instance, scenario frameworks developed by the climate change research community such as the five narratives of the Shared Socioeconomic Pathways (SSPs) are implemented in IAMs to provide an integrated analysis of future (e.g., socioeconomic and environmental) climate impacts, vulnerabilities, and the potential of adaptation and mitigation measures (Riahi et al., 2017).

In IAMs, the status of the abiotic (i.e., physical) component of freshwater systems is generally modelled by land surface models that describe water resource variables aggregated over large spatial extents. For instance, LPJmL (used in IMAGE and REMIND-MAgPIE), and EPIC (in MESSAGE-GLOBIOM and WITCH-GLOBIOM) provide estimates of hydrological variables over ~55x55 km^2 grid cells (Fricko et al., 2017; Schaphoff et al., 2018; van Vuuren et al., 2017; Wang et al., 2013). A higher resolution would be beneficial to more accurately quantify human impacts on the water resources and identify parts of catchments that need to be prioritized for management, conservation and restoration measures (Bierkens et al., 2015; Knouft and Ficklin, 2017). The biotic component of ecosystems, on the other hand, remains under-represented in IAMs (Harfoot et al., 2014b). Further, among the biodiversity assessment models available, the representation of freshwater biodiversity lags behind compared to terrestrial or marine ecosystems, hampering a balanced quantification of impacts across all natural ecosystems (Di Marco et al., 2017; Merow et al., 2018). Therefore, it is highly needed to develop freshwater biodiversity assessment models that can be readily linked to IAMs.

1.3 Global modelling of streamflow: data-driven approaches to achieve higher spatial resolution

Large-scale hydrological models have long been developed to solve the lack of observational hydrological data and to project potential hydrological changes to forecast, for instance, droughts and floods. Measurements of water flow are indeed constrained by the scattered coverage of monitoring stations around the world, reflected by, for example, the spatially biased coverage of gauging stations from the recently released global streamflow indices and metadata archive (GSIM) (Do et al., 2018; Gudmundsson et al., 2018). The poor coverage in monitoring is due to the inaccessibility of most headwaters and a lack of financial and human resources, highlighted by the substantial decline in monitoring since the mid-1980s (Fekete and Vörösmarty, 2007; Hannah et al., 2011; Shiklomanov et al., 2002; Sivapalan, 2003). Modelling is therefore necessary to obtain a globally continuous and representative coverage of hydrological variables and predict future hydrological changes.

Process-driven global hydrological models (GHMs) and land surface models (LSMs) provide a gridded quantification of hydrological variables (Beck et al., 2016a; Haddeland et al., 2014, 2011; Schellekens et al., 2017; Schewe et al., 2014). Yet, the usage of such data for ecological modelling or water quality assessments is constrained by the relatively coarse spatial resolution characteristic of most GHMs/LSMs (~10 to 55 km), due to computational and input data constraints (Bierkens et al., 2015; Döll et al., 2016). Such coarse gridded data are unable to provide reliable estimates of hydrological variables for

small rivers (Strahler stream order < 5), which comprise 94.6 % of the total stream length and riparian interface on the planet (Downing et al., 2012). Small rivers and headwaters play a central role in understanding sediment and nutrient transport, ecology and connectivity of the overall water system (Wohl, 2017). Data-driven correlative models may provide a computationally efficient alternative for obtaining global hydrological data at high spatial resolution, which in turn can be used to downscale the output of GHMs/LSMs (Lin et al., 2019). Compared to GHMs/LSMs, regression-based and machine-learning approaches, like neural networks, are computationally efficient and relatively quick to parameterize at relatively high spatial resolutions (Verdin and Worstell, 2008). Recent studies have shown the feasibility of applying data-driven approaches to quantify streamflow, resulting in estimates that have a greater accuracy than the output of GHMs/LSMs (Beck et al., 2016a, 2015; Razavi and Coulibaly, 2013; Yaseen et al., 2015). Despite these encouraging results, consistent high-resolution global streamflow maps are not yet available.

1.4 Global freshwater biodiversity models: towards a species-based approach

Our knowledge of global biodiversity responses to anthropogenic threats is limited for freshwater ecosystems. Large-scale assessments typically focus on charismatic species in terrestrial systems, with less than 20% of recent papers dealing with aquatic species (Di Marco et al., 2017). For instance, effects of climate change on species distributions have been extensively modelled for terrestrial systems worldwide (Hof et al., 2011; Powers and Jetz, 2019; Visconti et al., 2016; Warren et al., 2018; Zurell et al., 2018), while freshwater species remain underrepresented in the literature. The bias towards terrestrial biodiversity is also reflected in the recent global assessment of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (Díaz et al., 2019; Merow et al., 2018). For the assessment, various global biodiversity assessments and models across different levels of biological organization (i.e., species, community and ecosystem) were available for terrestrial and marine systems, whereas only the community-based GLOBIO-Aquatic model (Janse et al., 2015) was available for freshwater biodiversity (Merow et al., 2018). The lack of comprehensive freshwater biodiversity assessments is in contrast with the high threat status of freshwater biodiversity. Between 1970 and 2014, populations of freshwater species experienced declines as high as 83% on average, which is considerably higher than the decline rates observed for marine or terrestrial species (WWF, 2018, 2016). About 30% of the ~28,000 freshwater-dependent species assessed so far by the International Union for Conservation of Nature are

threatened with extinction (IUCN, 2018). Large-scale models are needed to better assess current and future impacts of human activities on freshwater biodiversity.

Freshwater fish play a key role in the function and regulation of freshwater ecosystems and are of important economic and nutritional value as the livelihoods of many people, especially in developing countries, depend on fisheries (FAO, 2018; Holmlund and Hammer, 1999). The very few attempts to quantify human impacts on freshwater fish biodiversity at a global scale relied on relatively coarse-grain (i.e., low spatial resolution) assemblage-level metrics, by relating potential changes in the total number of species occurring over relatively large regions, to the degree of change in the underlying hydrology, e.g., streamflow or connectivity (Liermann et al., 2012; Tedesco et al., 2013; Xenopoulos et al., 2005). The assemblage-level approach employed in these type of models hampers the identification of specific species more at risk than others. In addition, the coarse spatial grain these studies relied on (e.g., freshwater ecoregions or gridded estimates at ~55 km) limits the spatial detail, essential for a reliable impact quantification given the dendritic nature of freshwater systems. Yet, advancements in availability of species occurrence data and environmental variables along with increased computational power make it now possible to model anthropogenic impacts at the level of individual fish species. Such refined approaches can provide a much higher level of spatial detail and accuracy, needed to delineate vulnerable areas and prioritize conservation efforts.

1.5 Aim and research questions

The overarching goal of this thesis is to improve the representation of freshwater systems in global integrated assessments. To that end, I aim to answer two main research questions:

1 How to improve the spatial representation of hydrological variables at the global scale?

2 How to model the direct (damming) and indirect (climate change) human impacts on freshwater fish biodiversity worldwide?

The first question addresses the need for improved hydrological data, while the second question begins to address the gap in global impact assessments for freshwater biodiversity.

1.6 Overview of the thesis

This thesis is divided into two main parts, each addressing one of the two research questions and covering a different aspect of the DPSIR chain (**Figure 1.1**). The first part (Chapters 2 and 3) describes the development and application of data-driven models to derive high-resolution gridded streamflow variables. Chapter 2 describes the development of a multiple regression model to predict mean annual streamflow at a global scale and compares its performance with a global hydrological model. Chapter 3 takes data-driven hydrological modeling a step further by employing a machine-learning approach to map yearly estimates of mean, minimum and maximum streamflow. This effort resulted in the FLO1K dataset, comprising gridded data at ~1 km spatial resolution for the three variables (mean, minimum and maximum streamflow) for each year from 1960 through to 2015. The second part of the thesis aims at quantifying the impacts of human activities on the geographic ranges of freshwater fish species worldwide. Chapter 4 quantifies changes in the connectivity of the geographic ranges of ~5,700 freshwater fish species due to dam construction, including ~40,000 existing dams as well as ~3,700 dams currently being constructed or planned. Chapter 5 quantifies potential future geographical range contractions driven by anthropogenic climate change for ~7,000 freshwater fish species. In particular, range contractions are predicted for different levels of global mean temperature increase, including 1.5°C, 2°C, 3.2°C and 4.5°C (compared to pre-industrial levels). Finally, Chapter 6 compares and integrates the results from the different chapters and provides general implications and recommendations.

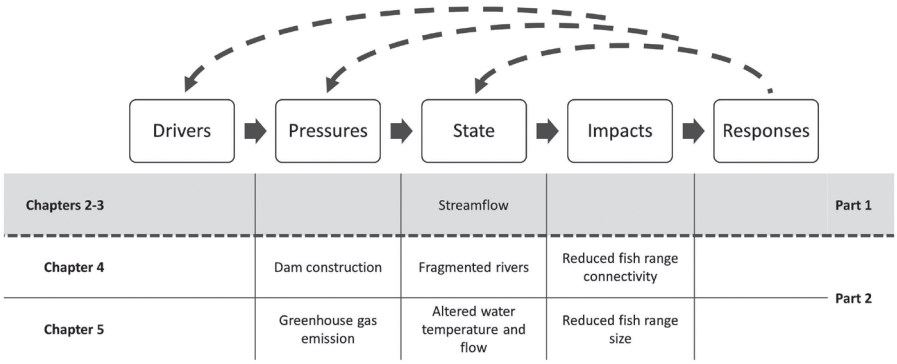
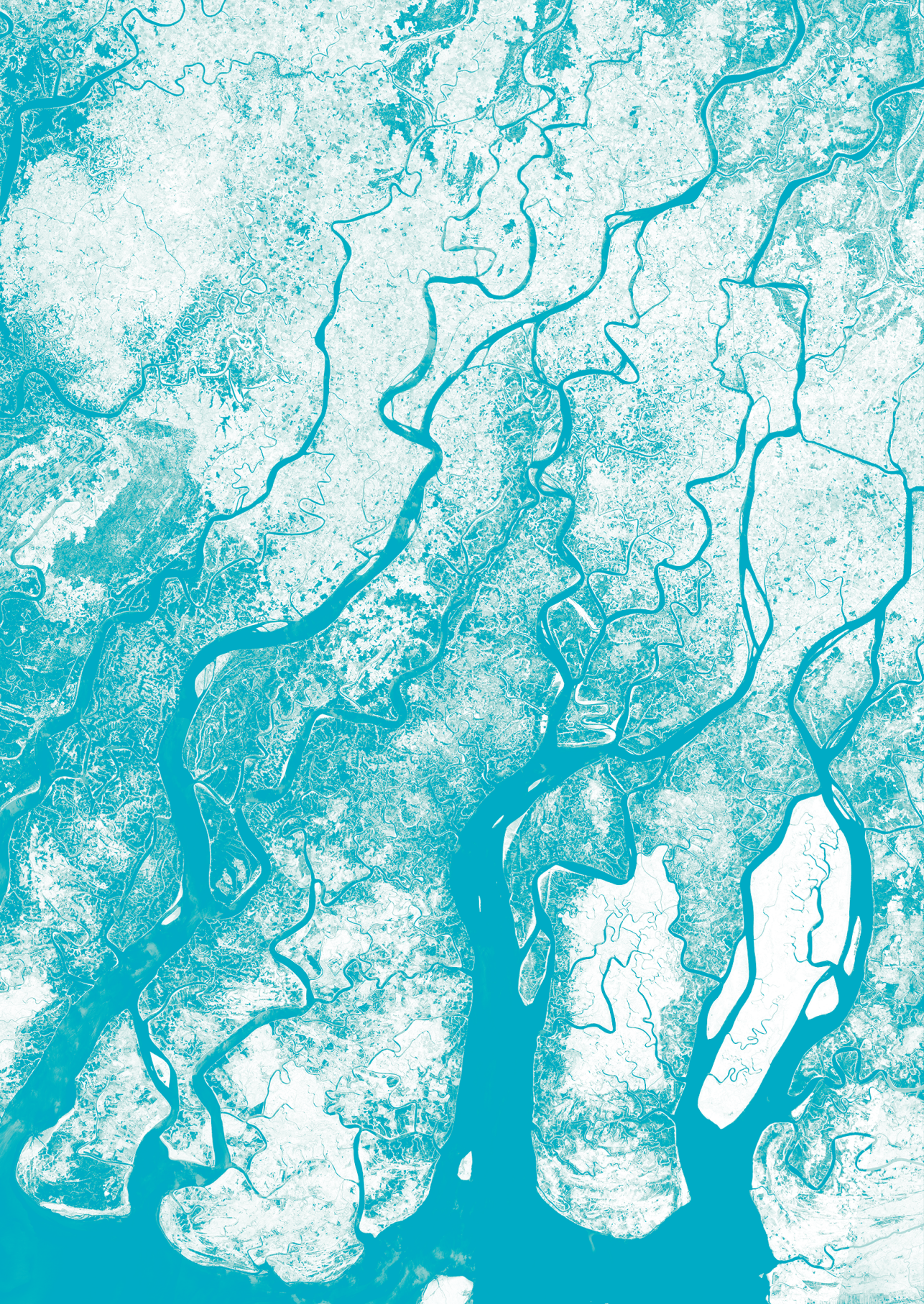


Figure 1.1 | Different aspects of the DPSIR chain covered by the studies of this thesis.

1



2

Developing and testing a global-scale regression model to predict mean annual streamflow

Valerio Barbarossa, Mark A. J. Huijbregts,
A. Jan Hendriks, Arthur H. W. Beusen,
Julie Clavreul, Henry King, Aafke M. Schipper

Published in Journal of Hydrology 544:479-487 (2017)

ABSTRACT

Quantifying mean annual flow of rivers (MAF) at ungauged sites is essential for assessments of global water supply, ecosystem integrity and water footprints. MAF can be quantified with spatially explicit process-based models, which might be overly time-consuming and data-intensive for this purpose, or with empirical regression models that predict MAF based on climate and catchment characteristics. Yet, regression models have mostly been developed at a regional scale and the extent to which they can be extrapolated to other regions is not known. In this study, we developed a global-scale regression model for MAF based on a dataset unprecedented in size, using observations of discharge and catchment characteristics from 1,885 catchments worldwide, measuring between 2 and 106 km2. In addition, we compared the performance of the regression model with the predictive ability of the spatially explicit global hydrological model PCR-GLOBWB by comparing results from both models to independent measurements. We obtained a regression model explaining 89% of the variance in MAF based on catchment area and catchment averaged mean annual precipitation and air temperature, slope and elevation. The regression model performed better than PCR-GLOBWB for the prediction of MAF, as root-mean-square error (RMSE) values were lower (0.29–0.38 compared to 0.49–0.57) and the modified index of agreement (d) was higher (0.80–0.83 compared to 0.72–0.75). Our regression model can be applied globally to estimate MAF at any point of the river network, thus providing a feasible alternative to spatially explicit process-based global hydrological models.

2.1 Introduction

Mean annual discharge or flow of rivers (hereafter abbreviated as MAF) is an important indicator of global water supply, with applications in irrigation supply assessment, climate change vulnerability assessment (Chang, 2003; Santini and di Paola, 2015), hydropower assessment (Hall et al., 2004), water footprinting (Hanafiah et al., 2011; Hoekstra et al., 2011; Jefferies et al., 2012; Pfister et al., 2009; Tendall et al., 2014), and for quantifying sediment fluxes (Syvitski et al., 2003). It also represents one of the most important factors determining the ecosystem integrity of freshwater biodiversity (Oberdorff et al., 2011, 1995; Poff and Zimmerman, 2010; Xenopoulos et al., 2005; Xenopoulos and Lodge, 2006). Despite its importance, streamflow data availability is limited, and monitoring is in rapid decline since the mid-1980s (Shiklomanov et al., 2002). Modelling approaches have long been used to estimate MAF at ungauged sites and are generally divided into two categories: spatially explicit process-based models and regression-based empirical models.

State-of-the-art spatially explicit numerical models for global-scale calculations of streamflow are Macroscale Hydrological Models (MHM) or Global Hydrology and Water Resources Models (GHWM) (Alcamo et al., 2003; Gosling and Arnell, 2011; Hanasaki et al., 2008; Van Beek and Bierkens, 2009; Van Der Knijff et al., 2010; Widén-Nilsson et al., 2007; Wisser et al., 2010). As these models account for the spatial variability of the physical processes involved within catchment hydrology and are capable of predicting streamflow even at the daily time scale, they are computationally and data intensive.

Regression-based approaches to calculate MAF are less time-consuming and computationally less intensive. Moreover, regression equations relating streamflow to explanatory catchment characteristics like upstream drainage area, precipitation and temperature may help to better understand general hydrological patterns and processes across different scales (Burgers et al., 2014; Farmer et al., 2015). However, to date, regression-based approaches relating mean annual streamflow to catchment characteristics have been mainly applied at a regional scale (Hortness and Berenbrock, 2001; Stuckey, 2006; Tran et al., 2015; Verdin and Worstell, 2008; Vogel et al., 1999) or to specific climate zones (Syvitski et al., 2003), and the extent to which these models can be extrapolated to other regions is not known. Regression relationships at the global scale have hardly been established so far. An exception is Burgers et al. (2014), who derived MAF relationships at a global scale using precipitation and catchment area as predictors. However, their model explained only 56% of the variance in MAF, which is low compared to the range of 77-99% achieved by regional regression models (e.g. Verdin and Worstell (2008)). Yet, the regional studies typically included a larger number of predictors, which suggests that the explanatory power of a global-scale regression model may increase

if relevant predictors are added. In addition, the applicability of global regression relationships for the prediction of mean annual streamflow has not yet been tested. Therefore, the aim of this study was twofold: 1) to establish an empirical regression model relating MAF to easily retrievable catchment characteristics at the global scale; 2) to test the predictive ability of the regression model in a backcasting analysis and compare its performance with the predictive performance of PCR-GLOBWB, a spatially explicit MHM (van Beek et al., 2011). To our knowledge, our study is the first to make an explicit comparison of the predictive abilities of a process-based and a regression-based global-scale model.

We based our regression model on measured long-term average MAF from 1,885 catchments worldwide, ranging from 2 km^2 to 10^6 km^2 in size. We used five predictor variables, including two climatic variables – mean annual precipitation and air temperature – and three geomorphologic variables – area, mean slope and mean elevation of the catchment. Drainage area, mean annual precipitation and mean annual temperature are often used as predictors of MAF in regional regression modelling studies (Verdin and Worstell, 2008; Vogel et al., 1999; Vogel and Sankarasubramanian, 2010). The dependence of MAF on drainage area is a well-accepted power relationship reflecting the self-similarity of river systems (Rodriguez-Iturbe and Rinaldo, 2001). Mean annual precipitation represents the potential runoff of the catchment, as it equals the amount of water supplied to the catchment (Thomas and Benson, 1970). We selected the mean annual temperature as a proxy for the potential evapotranspiration (PET), because temperature is a major determinant of evapotranspiration (Hamon, 1963; Lu et al., 2005; Thornthwaite, 1948). Furthermore, previous regression analyses of MAF have shown an increased explained variance when additional geomorphologic parameters were considered (Hortness and Berenbrock, 2001; Stuckey, 2006; Vogel et al., 1999). Therefore, we included average slope and elevation of the catchment as additional predictors in our study. Although elevation and slope alone may not directly influence MAF, they may serve as proxies for other factors causing inter-basin streamflow variation which are difficult to measure, e.g. radiation, wind, vegetation and basin ruggedness (Thomas and Benson, 1970).

2.2 Materials and methods

2.2.1 Mean annual discharge data

We retrieved worldwide MAF data from the Global Runoff Data Centre (GRDC) database, which provides daily or monthly observations of 9,213 gauging stations monitored from 1806 to 2015, with variable record length (GRDC, 2017). The GRDC has spent more than

25 years gathering river discharge data from the National Hydrological Services of all the World Meteorological Organization (WMO) state members, which has resulted in a discharge dataset unprecedented in size. For example, the SAGE Global River Discharge Database (http://nelson.wisc.edu/sage/data-and-models/riverdata/) and the RivDis database (Vorosmarty et al., 1998) provide discharge data for 3,500 and 1,018 stations, respectively. The accuracy of the discharge measurements included in the GRDC database is estimated to be about 10 – 20% (Syvitski, 2005). For our model development we selected discharge data for the period 1981 – 2010. We used a 30 year period because this is in accordance with the recommendations for climate analyses (World Meteorological Organization, 1992). We excluded years after 2010 because of a decrease in data availability for the most recent years. We averaged the daily discharge data over each year, using only those years where 100% of the daily observations were available. Next, we averaged the yearly observations for the period 1981 - 2010 in order to obtain a long-term mean annual discharge for each catchment. We selected monitoring stations with at least 15 yearly average discharge values in order to obtain representative long-term mean values (Kennard et al., 2009). This resulted in a dataset of 1,885 observations out of the 2,759 available GRDC gauged catchments with observations in the time range 1981 - 2010 (**Figure 2.1**).

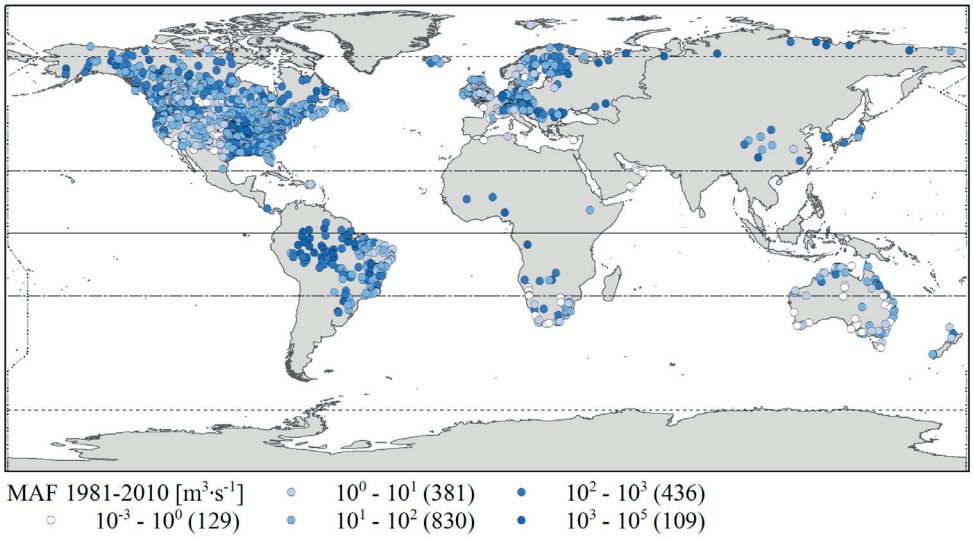


Figure 2.1 | Distribution of the 1885 GRDC gauging stations monitored for at least 15 years in the 1981–2010 period. The stations are grouped based on the mean annual flow (MAF) recorded at each station. Next to each MAF category, the number of observations is provided in brackets.

2.2.2. Catchment characteristics

We retrieved catchment-specific values for catchment area (A), average altitude (H), average slope (S), 30-years average temperature (T) and 30-years average precipitation (P) from a combination of data sources (**Table 2.1**). Catchment area was retrieved from the GRDC database, which includes a georeferenced map of the upstream catchment corresponding with each gauging station (GRDC, 2011). Catchment boundaries in this map have been established based on the HydroSHEDS drainage network, a 15 arc-seconds hydrological map derived from 3 arc-seconds elevation data of the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM), extended with the hydro1k hydrological network for latitudes above 60N, which are not covered by the SRTM data (GRDC, 2011; Lehner et al., 2008).

We derived altitude from the WorldClim digital elevation model (DEM), which is a 30 arc-seconds DEM based upon the SRTM elevation data extended with GTOPO30 elevation data for latitudes above 60N (Hijmans et al., 2005). We derived a raster slope map from the WorldClim DEM using the 'average maximum technique' in ArcGIS, similarly to Hortness and Berenbrock (2001). For precipitation and temperature, we averaged 30 arc-minutes resolution monthly raster maps from the Climatic Research Unit time series (CRU TS) 3.23 to annual values, and consequently averaged over the period 1981-2010 (Harris et al., 2014).

We resampled the raster maps obtained for precipitation, temperature, altitude and slope in order to match the 15 arc-seconds resolution of the HydroSHEDS drainage network, upon which the watershed boundaries were established. We then calculated a single mean value of each variable for each catchment corresponding with a GRDC gauging station, as described by Syvitski et al. (2003). Resampling and averaging were performed in ArcGIS 10.3. An overview of the summary statistics of the variables is available in **Table 2.1**.

Table 2.1 | Summary statistics of the mean annual streamflow (MAF) and catchment characteristics of 1,885 gauging stations in the period 1981-2010

Variable	Symbol	Unit	Mean	Median	SD[a]	Min[a]	Max[a]	γ_1[a]	Source database
MAF	Q	$m^3 \cdot s^{-1}$	$3.72 \cdot 10^2$	$3.38 \cdot 10^1$	$1.99 \cdot 10^3$	$3.47 \cdot 10^{-3}$	$4.73 \cdot 10^4$	14.2	GRDC[b]
Catchment area	A	m^2	$4.58 \cdot 10^{10}$	$4.61 \cdot 10^9$	$2.05 \cdot 10^{11}$	$2.00 \cdot 10^6$	$3.63 \cdot 10^{12}$	9.7	GRDC[c]
Altitude	H	m	$6.84 \cdot 10^2$	$4.52 \cdot 10^2$	$6.05 \cdot 10^2$	$1.35 \cdot 10^1$	$4.76 \cdot 10^3$	1.6	WorldClim[d]
Slope	S	(°)	$2.31 \cdot 10^0$	$1.27 \cdot 10^0$	$2.65 \cdot 10^0$	$4.71 \cdot 10^{-2}$	$1.56 \cdot 10^1$	2.0	WorldClim[d]
Precipitation	P	$m \cdot s^{-1}$	$2.78 \cdot 10^{-8}$	$2.46 \cdot 10^{-8}$	$1.46 \cdot 10^{-8}$	$3.53 \cdot 10^{-9}$	$1.09 \cdot 10^{-7}$	1.3	CRU TS 3.23[e]
Temperature	T	°C	$8.96 \cdot 10^0$	$8.24 \cdot 10^0$	$7.51 \cdot 10^0$	$-1.67 \cdot 10^1$	$2.76 \cdot 10^1$	0.4	CRU TS 3.23[e]

[a]SD = standard deviation; Min = minimum; Max = maximum; γ_1 = skewness, [b](GRDC, 2015), [c](GRDC, 2011), [d](Hijmans et al., 2005), [e](Harris et al., 2014)

2.2.3 Model fitting

Methods available for correlative modelling range from parametric and non-parametric regression-based approaches to machine-learning techniques (Chen et al., 2015; Danandeh Mehr et al., 2013; Fan et al., 2015; Okkan and Serbes, 2012; Wang et al., 2015; Wu et al., 2009). For the present study we selected ordinary least squares (OLS) regression because it results in an explicit equation, which facilitates interpretation and comparison with other studies. All the variables except for temperature were log-transformed to avoid heteroscedasticity as they revealed a right-skewed distribution (**Table 2.1**), in agreement with the choices made in previous studies for similar variables (Burgers et al., 2014; Hendriks et al., 2012; Syvitski et al., 2003; Verdin and Worstell, 2008; Vogel et al., 1999). This resulted in the following linear regression equation:

$$log_{10}Q = \beta_0 + \beta_A \cdot log_{10}A + \beta_P \cdot log_{10}P + \beta_T \cdot T + \beta_S \cdot log_{10}S + \beta_H \cdot log_{10}H + \varepsilon \quad \textbf{(2.1)}$$

where β_0 is the intercept, $<\beta_A, \beta_P, \beta_T, \beta_S, \beta_H>$ is the vector of the regression coefficients associated with the predictor variables (symbols are described in **Table 2.1**) and ε is the error term. The back-transformed form to real scale of **Equation 2.1** yields the nonlinear formulation:

$$Q = 10^{\beta_0} \cdot A^{\beta_A} \cdot P^{\beta_P} \cdot 10^{\beta_T \cdot T} \cdot S^{\beta_S} \cdot H^{\beta_H} \cdot 10^{\varepsilon} \quad \textbf{(2.2)}$$

Prior to performing the regression analysis, we assessed multi-collinearity among the predictors using Variance Inflation Factors (VIFs), employing the function "vif" of the package "HH" (Heiberger, 2015) in the R environment (r-project.org). We preferred VIFs over bivariate correlation analysis because pairwise correlation coefficients do not reveal more subtle forms of multicollinearity (Field, 2009). The maximum VIF was 2.8, well below the standard threshold of 5 (Zuur et al., 2009). We then fitted OLS regression models for an increasing number of predictors. To identify the best regression model for each of the five sets of predictors (first set with one predictor variable, second set with two predictor variables, etc.) as well as the best overall model, we employed the function "dredge" of the package "MuMIn" in the R environment (Barton, 2015). Within each set, the algorithm analyzes all possible combinations of predictor variables and ranks the regression models based on a user-defined criterion. To identify the most parsimonious model for each set of predictors, we used the Akaike Information Criterion (AIC) as well as the Bayesian Information Criterion (BIC), which employs a larger penalty term for additional predictor variables. Further, we used the Cooks D influence statistic in order to identify observations that may have biased the coefficients of the regression (Cook and Weisberg, 1982).

In order to assess potential bias in the regression coefficients induced by spatial autocorrelation resulting from the nested structure of the catchments, we compared the regression coefficients with the coefficients of a spatial error (SE) model. Since spatial autocorrelation can exist within either the residuals (spatial error) or the response variable (spatial lag), we performed a preliminary test for spatial autocorrelation based on the Lagrange Multiplier test (LM test) using the R package 'spdep' (Anselin, 1988; Bivand et al., 2013; Bivand and Piras, 2015). We preferred the LM test to the more commonly employed Moran's I test, for the LM test has a higher power to discriminate among either spatial error autocorrelation or spatial lag (Anselin and Rey, 1991). The LM test showed significantly higher autocorrelation in the error term (LM test value of 465, robust LM test value of 375) than in the response variable (with values of 115 and 25, respectively). Therefore we fitted an SE model that accounts for spatial autocorrelation in the residuals, expressing the error term of **Equation 2.1** as: $\varepsilon = \lambda W\varepsilon + \mu$, where λ is the coefficient in the spatial autoregressive structure, W is a weight matrix defined by the inverse distance between observations, and μ is the vector of identically distributed random errors (Ord, 1975). We calculated the distances between the GRDC stations across the stream network thereby considering as neighbors only those stations belonging to the same encompassing hydrologic basin. As we employed HydroSHEDS as rivers network for the calculation of W, we fitted and compared the OLS and SE regression coefficients based on a subset of observations within 58S - 60N latitude (n = 1,748).

2.2.4 Comparison with PCR-GLOBWB: backcasting analysis

We compared the predictive performance of our best regression model with the global hydrological model PCR-GLOBWB (van Beek et al., 2011; Van Beek and Bierkens, 2009). Defined as a "leaky bucket" type of model, PCR-GLOBWB calculates changes in water storage between two different soil layers, groundwater reservoir and atmosphere, forced by CRU TS 2.1 data, on a cell-by-cell basis at 30 arc-minutes resolution, for daily time steps. PCR-GLOBWB has been widely employed for assessments of global surface water and groundwater availability, nutrient transport modelling and biodiversity impact calculations (Beusen et al., 2016; Gleeson et al., 2012; Janse et al., 2015; Wada et al., 2011; Wanders and Wada, 2015). Compared to other GHMs, PCR-GLOBWB is a purely process-based model, as opposed to for example WaterGAP which is partially calibrated (Alcamo et al., 2003; Döll et al., 2003). Therefore, we considered PCR-GLOBWB a more suitable benchmark for the comparison.

We considered monitoring data of GRDC stations continuously monitored from 1971 to 1980 as an independent and common basis for the comparison between the regression model and PCR-GLOBWB. From the 2,219 GRDC stations used for the testing of PCR-

GLOBWB (van Beek et al., 2011), we selected the 543 stations that were continuously monitored from 1971 through 1980 (**Figure A.2**). We derived mean annual values of temperature and precipitation from the CRU TS 3.23 for each catchment, according to the approach described in section 2.2, and calculated mean annual streamflow for each year in the time span 1971-1980, and as a 10 years average.

We evaluated and compared the performances of the regression model and PCR-GLOBWB employing root mean square error (*RMSE*) and modified index of agreement (*d*). Thus, we employed an absolute as well as a relative error measure, following the recommendations for hydrological model evaluation as provided by Legates and McCabe (1999). We adopted the index of agreement d^2 in the modified form *d* to avoid inflation of errors by squared values (Legates and McCabe, 1999). In addition, the *d* represents an improvement over the coefficient of determination (R^2) (Legates and McCabe, 1999).

Given the six orders of magnitude covered by the data, we log transformed the mean annual streamflow values. The *RMSE* is calculated as:

$$RMSE = \sqrt{\frac{1}{n \cdot m} \cdot \sum_{x=1}^{m} \sum_{t=1}^{n} \left(\log_{10} O_{x,t} - \log_{10} P_{x,t}\right)^2} \tag{2.3}$$

where $n \cdot m$ are the dimensions of the matrix of observations of the *m* GRDC stations over the *n* years of the backcasting period, $O_{x,t}$ is the observed value for the station *x* at time *t*, and $P_{x,t}$ is the predicted value for the station *x* at time *t*.

The *d* is formulated as:

$$d = \frac{\sum_{x=1}^{m} \sum_{t=1}^{n} \left|\log_{10} O_{x,t} - \log_{10} P_{x,t}\right|}{\sum_{x=1}^{m} \sum_{t=1}^{n} \left(\left|\log_{10} P_{x,t} - \log_{10} \bar{O}\right| + \left|\log_{10} O_{x,t} - \log_{10} \bar{O}\right|\right)} \tag{2.4}$$

where $\bar{O}$ stands for the overall average of the observed streamflow of all the stations across the 10 years period. The index of agreement varies between 0 and 1, with higher values indicating a better fit. When the mean annual streamflow averaged over the 10 years is considered, *n* becomes 1 in **Equations 2.3** and **2.4**.

2.3 Results

2.3.1 Model fitting

The OLS regression analysis revealed that the model with the full set of predictors was the most parsimonious (i.e., lowest AIC; **Table A.1**). Nearly 90% of the variation in MAF could be explained by the five catchment characteristics (**Table 2.2**), indicating that the most relevant predictors for MAF were covered by the regression model. According to the standardized regression coefficients, which can be compared across explanatory variables to assess their relative importance (Bring, 1994), catchment area was the most important predictor of MAF, followed by precipitation, temperature, slope and elevation.

Table 2.2 | Coefficients (raw and standardized), goodness of fit (R^2) and number of underlying observations (m) of the most parsimonious regression model $Q = 10^{\beta 0} \cdot A^{\beta A} \cdot P^{\beta P} \cdot 10^{\beta T \cdot T} \cdot H^{\beta H} \cdot S^{\beta S}$. CI = confidence interval.

Coefficient	Value (95% CI)	Std. value	R^2	m
β_0	9.066 (8.503 - 9.630)	0	0.89	1,885
β_A	1.018 (1.001 - 1.035)	0.961		
β_P	2.070 (1.991 - 2.148)	0.486		
β_T	-0.038 (-0.040 - -0.035)	-0.290		
β_H	-0.509 (-0.565 - -0.454)	-0.212		
β_S	0.464 (0.421 - 0.507)	0.237		

The model performed better for higher MAF values (**Figure 2.2**). Furthermore, residual errors were slightly larger for catchments with lower precipitation values and at higher altitudes (**Figure A.1**). Residuals tended to be randomly distributed in relation to catchment area, precipitation or slope (**Figure A.1**). Only about 1.3% of the predicted values showed errors greater than one order of magnitude (**Figure 2.2**). For the Cooks D statistic, a maximum value of 0.03 was found, well below the threshold of 1 (Cook and Weisberg, 1982), meaning that none of the observations biased the regression coefficients.

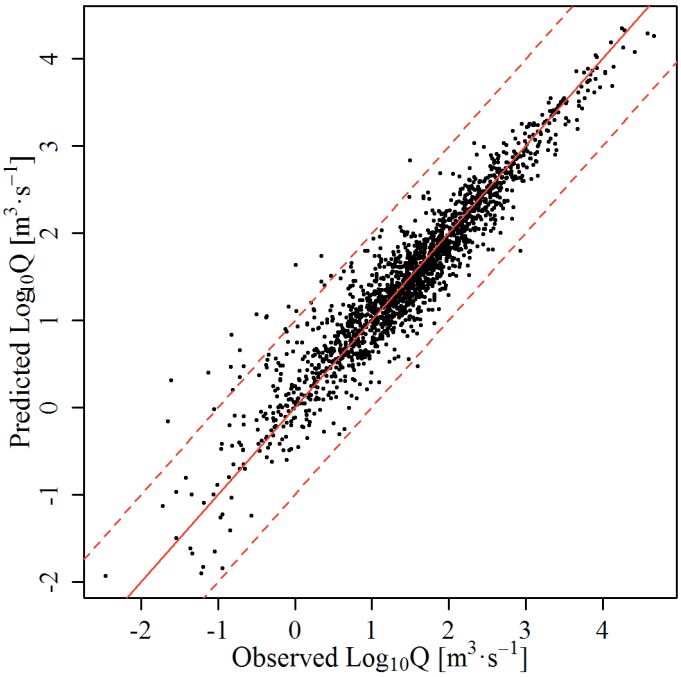


Figure 2.2 | Predicted versus observed MAF values of the 1885 GRDC stations employed in the regression analysis. The solid line represents perfect model fit (1:1 line) and the dashed lines represent a difference of one order of magnitude.

The comparison of the regression coefficients between the OLS and SE regression models revealed a large overlap of the confidence intervals of the coefficients (49-86% CI overlap; **Table A.2**). This indicated that the OLS regression coefficients were not significantly influenced by spatial autocorrelation (type I error).

2.3.2 Performance testing on independent data and comparison with PCR-GLOBWB

The testing of the regression model on independent data in the time period 1971 - 1980 (backcasting analysis), for both single year and 10-years average MAF, revealed that the predictions of the regression model were characterized by lower RMSE values and higher *d* values than the predictions of the global hydrological model PCR-GLOBWB (**Figure 2.3**).

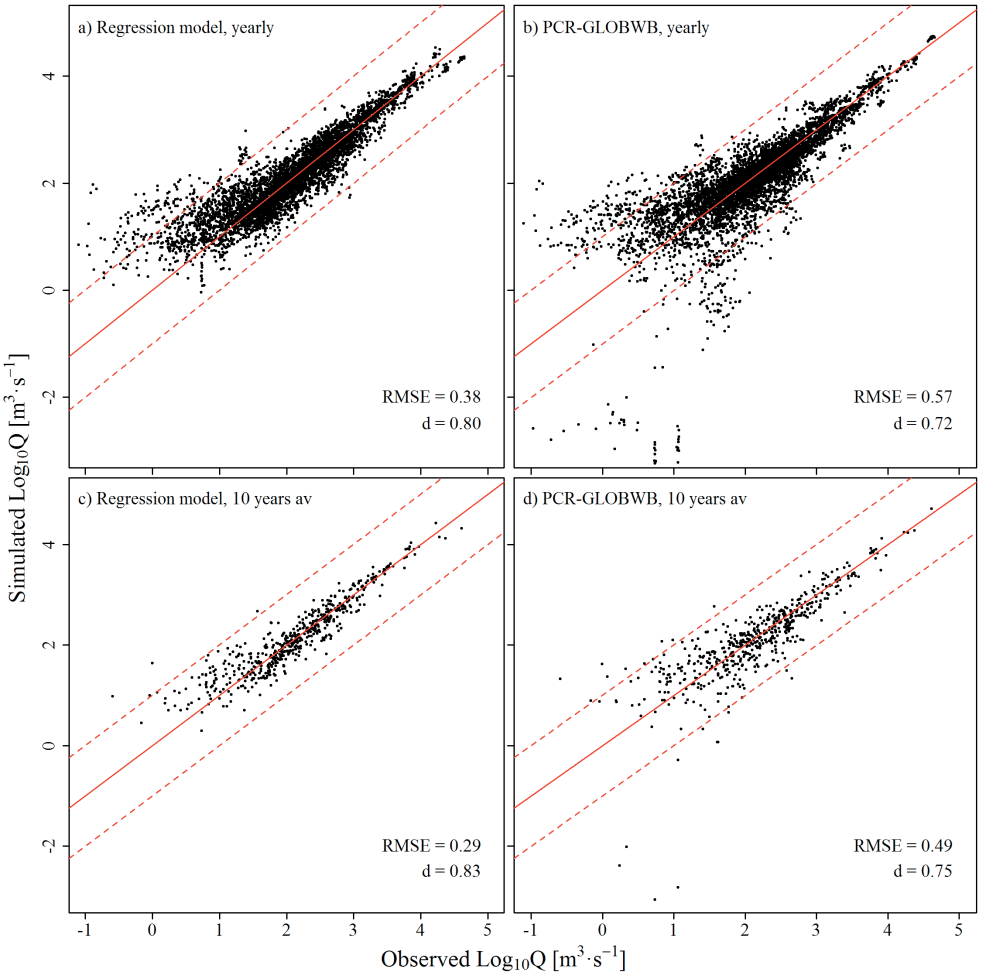


Figure 2.3 | Results of the backcasting analysis for the period 1971–1980, showing predicted versus observed MAF for the OLS regression model (left) and the GHM PCR-GLOBWB (right), based on yearly values (top) as well as 10-year average values (bottom). Within each chart, the solid line represents perfect model fit (1:1 line) and the dashed lines define a range of accuracy of plus/minus one order of magnitude. RMSE = root mean square error; d = modified index of agreement.

The PCR-GLOBWB simulation resulted in a greater number of outliers, in agreement with the higher RMSE values (**Figure 2.3**). For the OLS model, about 3% of the observations had residuals greater than one order of magnitude, whereas about 7% of the PCR-GLOBWB results deviated more than one order of magnitude from the measurements. PCR-GLOBWB performed slightly better than the OLS model for the highest MAF values (> 10,000 m^3/s; **Figure 2.3**). Both models performed poorly for MAF values lower than

10 m^3/s. In general, residuals of PCR-GLOBWB revealed a cone-shaped distribution, with larger errors at lower discharge values. In contrast, the OLS model showed a tendency to overestimate low discharge values.

Both models performed better when backcasting the 10-years average of MAF than MAF for single years. For the 10-years average MAF, the residuals greater than 1 order of magnitude reduced to about 1% (4 observations) and 4% (22 observations) for the regression model and PCR-GLOBWB, respectively.

2.4. DISCUSSION

2.4.1 Regression coefficients interpretation

We developed a global-scale multiple regression model for predicting mean annual flow of rivers based on easily retrievable input parameters. We calibrated the model on 1,885 catchments worldwide based on long-term average discharge data (1981-2010), resulting in a model explaining 89% of the variance in MAF based on catchment area and catchment-averaged precipitation, temperature, slope and altitude. The analysis revealed the catchment area to be the most important predictor of MAF (**Table 2.2**). Indeed, a single regression analysis based on catchment area alone already explains 61% of the variation in MAF (**Table A.1**), in agreement with scaling relationships reflecting self-similarity across catchments. Nevertheless, additional predictors considerably increased the explained variance (**Table A.1**). This suggests that multiple regression improves the interpretation of the spatial scaling of MAF at a global level, in agreement with recent findings that point toward multiple regression (called "multiscaling") to improve the interpretation of scaling behavior for daily streamflow of the Southeast United States (Farmer et al., 2015).

The MAF scaled to catchment area with an exponent of about 1, implying a linear relationship between MAF and drainage area. This is in agreement with the coefficients reported by a number of regional studies using multiple regression analysis (**Table 2.3**), and close to the value of 0.86 reported by the global study of Burgers et al. (2014). Further, MAF scaled to precipitation with an exponent of about two. This is highly similar to the findings reported by Tran et al. (2015) (**Table 2.3**), who conducted a regional study covering 533 catchments. On the other hand, the coefficient for precipitation found in this study is in contrast with the linear relationship observed in Burgers et al. (2014). However, in the current study we employed about three times more observations and our regression model explained about 33% more of the variance. In general, coefficients reported for precipitation tend to be greater than 1 and converging towards 2 (**Table 2.3**), reflecting nonlinearity in the physical process responsible for the runoff generation (see

e.g. (Yu et al., 2015)). Therefore, given the global coverage, the heterogeneity of the input variables values employed to calibrate the coefficients and the large explained variance, the value of about 2 is considered to be a reasonable estimate of the exponent for precipitation.

For temperature a negative coefficient was obtained, reflecting decreasing MAF with increasing temperature. Indeed, increases in temperature would lead to an increased evapotranspiration, which eventually implies that less water is routed through the drainage network and poured at the discharge point. The negative relationship is in agreement with the results of regional studies, but as these studies were based on log-transformed temperature data, the values of the coefficients are not directly comparable with ours (Tran et al., 2015; Verdin and Worstell, 2008; Vogel et al., 1999). The relationship between MAF and altitude is rather complex. The exponent resulting from the regression analysis was negative, which can be explained by the fact that at higher altitudes the solar radiation and wind are more intense, therefore enhancing the evapotranspiration process (Tran et al., 2015). In contrast, some studies reported positive exponents for altitude (**Table 2.3**). A possible explanation could be that these studies did not include temperature as a predictor, which implies that the positive exponent for the altitude term may reflect an effect of temperature. At high altitudes, the lower temperature likely results in less evapotranspiration hence an increase in MAF. However, if the model does not include temperature as predictor, the larger MAF at higher altitudes may result in a positive regression coefficient for altitude instead. The positive exponent of about ½ obtained for slope (S) compares well with the values reported in regional studies and reflects that in catchments with steeper slopes, the runoff and consequently the MAF is enhanced.

Table 2.3 | Regression coefficients found in this study compared with coefficients reported in regional and global studies available from the literature. n represents the number of catchments employed to calibrate the regression coefficients; A range is the range of the catchment areas employed in the respective study.

	This study	Burgers et al. (2013)	Hortness and Berenbrock (2001)	Stuckey (2006)	Tran et al. (2015)	Vogel et al. (1999)	Gyawali et al. (2015)
β_A	1.02	0.86	0.83 - 1.10	1.01	1.01	0.58 - 1.14	0.87
β_P	2.07	1.01	1.64 - 2.70	1.80	2.04	1.21 - 6.42	3.68
β_T	-0.04	-	-	-	-0.49[a]	-7.66 - -0.51[a]	-
β_S	0.46	-	-3.44 - 7.52[b]	-	-	0.33 - 0.51	0.30[b]
β_H	-0.51	-	-2.36 - 2.30	0.13	-	1.66[c]	-
n	1,885	663	200	195	533	1,553	93
A range [km^2]	$2.0 \cdot 10^0$ - $3.6 \cdot 10^6$	$7.3 \cdot 10^3$ - $4.6 \cdot 10^6$	$8.0 \cdot 10^0$ - $3.5 \cdot 10^4$	$5.6 \cdot 10^0$ - $4.5 \cdot 10^3$	$1.3 \cdot 10^1$ – $1.7 \cdot 10^2$	-	$6.2 \cdot 10^0$ - $4.4 \cdot 10^3$
Extent	Global	Global	Idaho, USA	Pennsylvania, USA	Upper Mississippi, USA	USA	western Great Lakes, USA

[a]values refer to the log-transformed form and are therefore not directly comparable with the coefficient obtained in this study
[b]values refer to slope in percent instead of degrees and are therefore not directly comparable with the coefficient obtained in this study
[c]used only for region 16, "Great Basin" in Vogel et al. (1999)

2.4.2 Performance comparison with PCR-GLOBWB

Overall, our regression model performed better than PCR-GLOBWB when applied to an independent test dataset. Differences were apparent in particular for smaller catchments (**Figure 2.3**). The residuals (absolute values) of PCR-GLOBWB for the 10-years average MAF revealed a significant negative trend in relation to area (p-value < 0.01, R^2 = 0.03), while for the regression model no significant correlation was found. However, new global hydrological models with greater spatial resolution than the 30 arc-minutes version of PCR-GLOBWB employed in this study may achieve better results, especially for smaller catchments (e.g. see the list of models provided in Bierkens (2015)). Yet, such refined models are more demanding in terms of computational costs (Bierkens, 2015), and might therefore be more suitable when monthly or daily discharge values are needed.

Both the regression model and PCR-GLOBWB performed worse for water-scarce regions, as revealed by larger errors at higher dryness ratio values (**Figure 2.4**). The dryness ratio reflects water losses due to evapotranspiration relative to the amount of precipitation. It is defined as the actual annual evapotranspiration divided by the total annual precipitation, where the actual evapotranspiration is calculated as the annual precipitation minus the unit discharge, in turn obtained by dividing the discharge by the area of the catchment (Vogel et al., 1999). The larger errors with higher dryness ratios are likely due to a combination of higher uncertainty in the precipitation values for water-scarce regions and hydrological processes that are particularly relevant in dry regions yet not described by the models (Döll et al., 2003; van Beek et al., 2011; Vogel et al., 1999). Examples of such processes include the almost instantaneous evaporation from many ephemeral post-rainfall ponds and relatively large losses from the river channel to groundwater (Döll et al., 2003). In addition, water abstraction by human activities is likely to affect the natural flow in water scarce regions more than in wet regions, thus providing an additional possible explanation for the overestimation of MAF in dry regions.

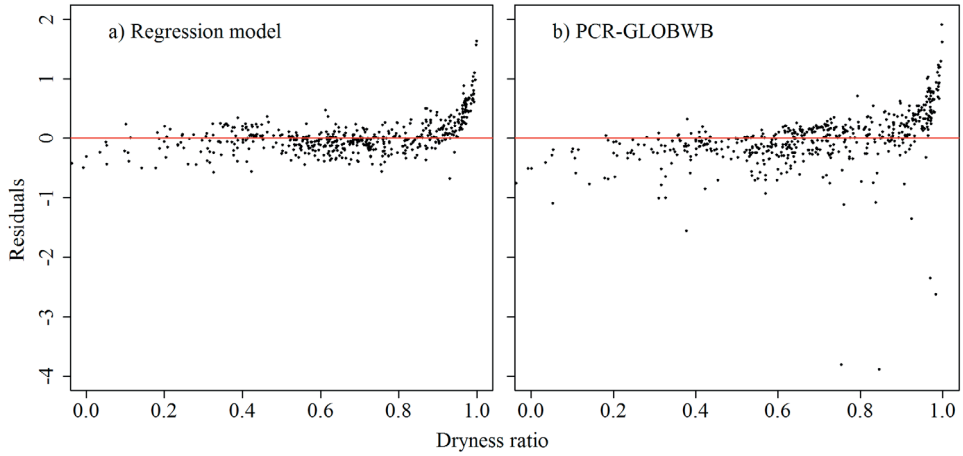


Figure 2.4 | Residuals plot of the backcasting of the 10-year average MAF vs the dryness ratio for (a) the regression model and (b) PCR-GLOBWB. Note that 18 values (about 3% of total data) with dryness ratios smaller than zero were excluded for clarity of representation.

2.4.3 Applicability of the regression model

The residuals of our regression model were not related to area (**Figure A.1-b**), suggesting that the model is area-independent and maintains similar performance across catchments ranging across at least six orders of magnitude in size (2 to 10^6 km^2). Although the distribution of the monitoring stations employed for the calibration of

the regression coefficients was skewed towards America and Europe (**Figure 2.1**), the residuals of the model-application (yearly and 10 years average MAF) were consistent across different continents (**Figure A.3**). This is supported by the wide range of latitudes covered by the monitoring stations, reflected by a large range in precipitation and temperature values (**Table 2.1**). This indicates that the model developed by this study can be applied to predict MAF at any point of any river network globally, taking into account the weaker prediction power for water scarce regions (see section 4.2). As such, the model is most suitable for assessments of water availability and ecological integrity in relation to changing future climatic conditions.

The model requires only a small number of input parameters, namely catchment area, catchment-averaged precipitation, catchment-averaged temperature, catchment-averaged altitude and catchment-averaged slope. Catchment area as well as slope and altitude can be easily derived from a flow direction raster map and a digital elevation model with standard Geographic Information System (GIS) tools. Annual mean precipitation and temperature can be obtained from observations covering a given period of interest. If predictor variable values are within the range of values employed in the calibration phase (**Table 2.1**), the uncertainty of the predictions is known. Application of the model to predictor values outside the calibration domain results in MAF values with unknown uncertainty. In addition to this, given the fact that the model was calibrated on climatic data different from the CRU TS 3.23 (Harris et al., 2014), with a river network different from HydroSHEDS 15-sec (Lehner et al., 2008) or with another DEM else than the one provided by the WorldClim database (Hijmans et al., 2005), we acknowledge that the model is not valid with other input sources.

The regression model performed worse for extreme MAF values when applied at finer temporal scales, as exemplified by the decreased performance of the model on a year-by-year basis (**Figure 2.3**). This is due to the fact that the regression coefficients have been calibrated on 30 years averaged data and therefore when the model is applied at a finer temporal scale it would underestimate high MAF values and overestimate low MAF values. Yet, the bias due to temporal downscaling was relatively low at high MAF values (**Figure 2.3**), suggesting that the overestimation of MAF at low values is related to the decreased model performance in water-scarce regions (**Figure 2.4**) rather than the temporal downscaling. Implementing catchment aridity as predictor in a regression model is, however, a non-trivial problem. Attempts to include an interaction term of precipitation and temperature to describe catchment aridity within this study were inconclusive due to the scarce increase in the explained variance and multicollinearity issues. Furthermore, including complex multi-variable predictors would go beyond the scope of developing a conceptually simple and easily applicable model. In addition,

process-based models like PCR-GLOBWB which implement the evapotranspiration process in a more mechanistic way reveal the same limitations when predicting MAF in arid catchments. Therefore, the scarcity and uncertainty of monitored precipitation and discharge values in such regions represent a major source of bias. Hence, further research should focus on improving models for predicting mean annual flow in water scarce regions, which is particularly relevant in assessments of water availability and water footprints. In addition, future work would benefit from an increase in the quality and quantity of monitoring data, for precipitation and discharge in particular.

2.5 CONCLUSIONS

We presented a conceptually simple model for predicting mean annual flow of rivers globally at any point of the river network. The model explained 89% of the variance in MAF based on observations retrieved from 1,885 catchments worldwide. The regression coefficients obtained were within the ranges reported by previous regional-scale studies and indicated that MAF scales linearly to catchment area, while it scales nonlinearly to precipitation with an exponent of about 2. Temperature, slope and altitude, which have not been used before in global regression models for MAF, further improve the explained variance. Our model can be applied to estimate MAF at any ungauged site in the river network globally. However, it should be noted that the model is valid only for input parameters within the range of the calibration variables and therefore, outside the applicability domain the uncertainty in the estimation of MAF is unknown. Application of the model to predict 1-year and 10-year average MAF based on independent test data revealed that only 3% and 1% of the simulated MAF values deviated more than one order of magnitude from the measurements, respectively. In addition to this, our model performed slightly better than the widely employed macro hydrological model PCR-GLOBWB, particularly for smaller catchments. Both the model developed in this study and PCR-GLOBWB performed worse for water-scarce regions by overestimating the MAF, due to the increased uncertainty in rainfall and discharge observations and to the difficulty in describing the catchment hydrology in such regions. This implies that in dry regions, our model should be applied very cautiously.

It is recommended that forthcoming studies on global models for the prediction of MAF concentrate on including a better description of water-scarce regions within the model. Nonetheless, streamflow prediction models would in general benefit from the improvement of quality and quantity of monitored precipitation and discharge data especially in arid regions. In addition, future research should focus on whether such a simplified framework for describing catchment hydrology worldwide can be derived for finer temporal resolutions (daily or monthly).

2.6 ACKNOWLEDGMENTS

This project has received funding from the Europeans Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 641459. The authors would like to thank the Global Runoff Data Center for providing the data of daily rivers discharge from worldwide gauging stations, and Rens van Beek from Utrecht University for providing global monthly gridded discharge values from the model PCR-GLOBWB. The data used in this work are listed in the references and provided in the supporting information.

3

FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015

Valerio Barbarossa, Mark A. J. Huijbregts, Arthur H. W. Beusen, Hylke E. Beck, Henry King, Aafke M. Schipper

Published in Scientific Data 5:180052 (2018)

ABSTRACT

Streamflow data is highly relevant for a variety of socio-economic as well as ecological analyses or applications, but a high-resolution global streamflow dataset is yet lacking. We created FLO1K, a consistent streamflow dataset at a resolution of 30 arc seconds (~1 km) and global coverage. FLO1K comprises mean, maximum and minimum annual flow for each year in the period 1960–2015, provided as spatially continuous gridded layers. We mapped streamflow by means of artificial neural networks (ANNs) regression. An ensemble of ANNs were fitted on monthly streamflow observations from 6,600 monitoring stations worldwide, i.e., minimum and maximum annual flows represent the lowest and highest mean monthly flows for a given year. As covariates we used the upstream-catchment physiography (area, surface slope, elevation) and year-specific climatic variables (precipitation, temperature, potential evapotranspiration, aridity index and seasonality indices). Confronting the maps with independent data indicated good agreement (R^2 values up to 91%). FLO1K delivers essential data for freshwater ecology and water resources analyses at a global scale and yet high spatial resolution.

3.1 Background & Summary

Quantifying streamflow is critical to a variety of socio-economic and ecological analyses and applications (Bakker, 2012; Vörösmarty et al., 2010; WWAP (United Nations World Water Assessment Programme), 2015). Examples include the study of freshwater biodiversity patterns (Iwasaki et al., 2012; Oberdorff et al., 2011; Poff and Zimmerman, 2010; Xenopoulos and Lodge, 2006), assessments of global water resources (Dai, 2016; Dai et al., 2009), for example irrigation supply, hydropower or water footprinting (Hanafiah et al., 2011; Hoekstra et al., 2011; Tendall et al., 2014), analyses of the fate of pollutants (Grill et al., 2016) and quantification of sediment fluxes (Syvitski et al., 2003; Syvitski, 2005). Most of the stream reaches in the world are poorly or not monitored at all (Fekete and Vörösmarty, 2007; Sivapalan, 2003), due to the inaccessibility of most headwaters and a lack of financial and human resources (Shiklomanov et al., 2002), highlighted by a substantial decline in monitoring since the mid-1980s (Fekete and Vörösmarty, 2007; Hannah et al., 2011; Shiklomanov et al., 2002). Streamflow is commonly quantified with process-driven global hydrological models (GHMs) and land surface models (LSMs) (Beck et al., 2016a; Bierkens, 2015; Haddeland et al., 2014, 2011; Schellekens et al., 2017). GHMs/LSMs are typically run at coarse spatial resolutions (~10 to 50 km), due to computational constraints, and consequently are unable to provide reasonable streamflow estimates for small rivers (defined here by Strahler stream order < 5), which comprise 94.6 % of the total stream length and riparian interface on the planet (Downing et al., 2012). Streamflow data at higher spatial resolution would be highly beneficial for ecological applications and water resources assessment, for example understanding/modelling freshwater species distributions or modelling the fate and effects of pollutants in the aquatic environment (Domisch et al., 2015a, 2015b; Grill et al., 2016; Labay et al., 2015; Thorp et al., 2006).

Compared to process-based models, data-driven models like regression equations and neural networks are more suited for generating high-resolution streamflow data with large spatial extent, thanks to their computational efficiency and relatively quick parameterization (Verdin and Worstell, 2008). Data-driven models typically quantify streamflow based on upstream catchment characteristics related to topography, climate, land cover, and soils (Barbarossa et al., 2017; Beck et al., 2015, 2013; Verdin and Worstell, 2008). Data-driven approaches have been mostly employed at a local scale (Razavi and Coulibaly, 2013). Recent studies demonstrated, however, the feasibility of applying a data-driven approach at a global scale, resulting in streamflow estimates that may have greater accuracy than the output of GHMs/LSMs (Barbarossa et al., 2017; Beck et al., 2015). Despite these encouraging results, consistent high-resolution global streamflow maps are not yet available.

Here we present FLO1K: a consistent dataset of global annual streamflow maps at 1 km resolution for each year in the period 1960-2015. Annual flow (AF) metrics include mean annual flow as well as minimum and maximum monthly flow for a given year. We produced the maps with feed-forward Artificial Neural Networks (ANNs) trained on yearly AF metric values from 6,600 monitoring stations worldwide, using catchment-averaged covariates representing topography and climate. We delineated the upstream catchments based on the 1-km HydroSHEDS (https://www.hydrosheds.org) hydrography (Lehner et al., 2008), extended with Hydro1k (https://lta.cr.usgs.gov/HYDRO1K) for latitudes above 60°N not covered by HydroSHEDS, thereby achieving a global coverage (excluding Antarctica). For the training of the ANNs, we used 10 yearly values of mean, minimum and maximum AF per monitoring station and climate covariates for the corresponding years. We then constructed the AF metric maps by first computing for each year and each 30 arc seconds grid cell the upstream catchment-averaged covariates (which varied from year to year for climate), and then applying the trained ANNs. The streamflow is calculated for each terrain grid cell, i.e., it represents the potential in-channel discharge that would occur in the presence of a natural watercourse. The flow maps have a resolution 10 to 50 times higher than those typically produced using state-of-the-art GHMs/LSMs (Verzano et al., 2012; Wada et al., 2014) and global data-driven approaches (Beck et al., 2015). For each of the three AF metrics, 56 yearly layers (1960-2015) are available packed in the NetCDF-4 format CF-compliant. In addition, we provide the FLO1K layers upscaled to 5 and 30 arc minutes resolutions for coarser-grain applications, including comparisons with GHMs/LSMs outputs. The FLO1K database can be downloaded from http://geoservice.pbl.nl/download/opendata/FLO1K and figshare (Data Citation 1: figshare https://dx.doi.org/10.6084/m9.figshare.c.3890224).

3.2 Methods

3.2.1 General approach and streamflow network

The procedure to generate the maps consisted of (i) model fitting, including observed streamflow data preparation, extraction of covariates, and training of the ANNs, and (ii) application of the ANNs to generate the global AF maps. **Figure 3.1** provides a general outline of the procedure. We used the 30 arc seconds (~1 km) version of HydroSHEDS35 extended with Hydro1k for latitudes above 60°N to retrieve the drainage direction network and delineate the upstream catchment of each grid cell (Wu et al., 2012, 2011). The HydroSHEDS hydrography is based on the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) (Farr et al., 2007), which covers the entire terrestrial land surface from

latitudes 56°S to 60°N. To achieve a global spatial coverage, we extended HydroSHEDS with Hydro1k (Wu et al., 2012, 2011), the latter being a United States Geological Survey (USGS) product derived from the GTOPO30 Digital Elevation Model (DEM) (https://lta.cr.usgs.gov/GTOPO30). The resulting drainage direction network is available at http://files.ntsg.umt.edu/data/DRT/.

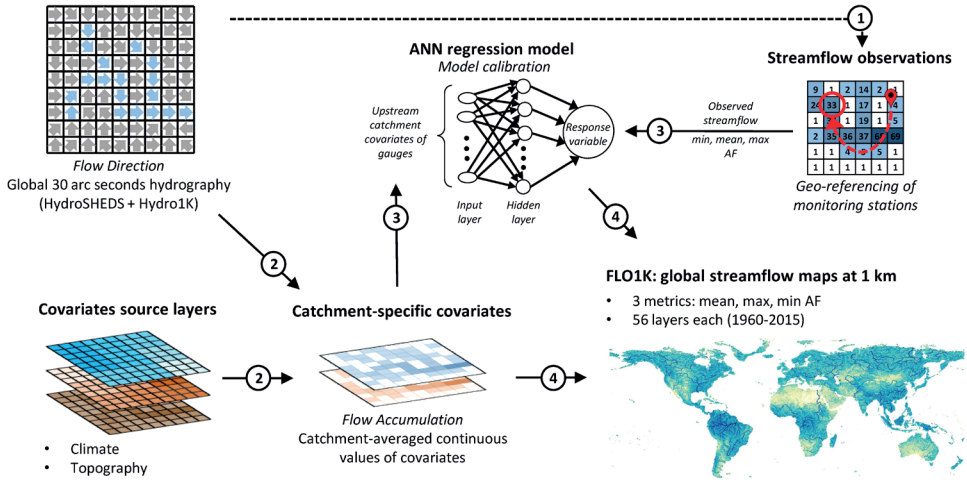


Figure 3.1 | Schematic overview of the streamflow mapping procedure. The procedure consisted of four main steps: 1) monitoring stations (gauges) are geo-referenced based on the global hydrography, 2) catchment-specific covariates are compiled by aggregating climatic and physiographic variables over the upstream catchment of each cell, 3) ANNs are trained on monitoring data of AF metrics and covariates of the corresponding upstream catchment, 4) the trained ANNs are applied to the spatially-continuous covariates to create the global streamflow maps.

3.2.2 Streamflow observations

We derived mean, maximum and minimum AF values from flow records in the Global Runoff Data Centre (GRDC) database (www.bafg.de/GRDC) (GRDC, 2017). The GRDC comprises daily and monthly streamflow records from 9,252 monitoring stations worldwide. The GRDC monitoring stations are not directly referenced on the hydrography employed in this study. This means that mismatched monitoring stations might encompass the wrong upstream catchment basin, which in turn may lead to errors when training the ANNs. As the GRDC dataset includes the estimated catchment area upstream of each monitoring station, we geo-referenced each station in order to

match the most similar upstream area on the 30 arc seconds stream network, following the procedure previously used to allocate GRDC stations on the HydroSHEDS 15 arc seconds hydrography (GRDC, 2011). For each station, a new location is selected that minimizes discrepancies in catchment area and distance from the original location, within a 5 grid cells (~5 km) search radius. Out of the original 9,252 monitoring stations, 285 were excluded as they did not report coordinates. Of the remaining 8,967, 746 (~8%) were excluded because there was no matching catchment area within the search radius (based on a threshold of maximum 50% difference (GRDC, 2011)). Out of the remaining 8,221, 65% reported an area difference smaller than 5%, 15% had an area difference between 5% and 10%, and 20% had an area difference between 10% and 50%.

We used the monthly records provided by the GRDC to calculate AF metrics for the period 1960-2015. We computed the mean AF for each year by averaging the 12 monthly values, and retrieved maximum and minimum AF by selecting the highest and lowest monthly values for each year, respectively. We considered only those years with a complete 12 months record and selected monitoring stations with at least 10 years of data from 1960 through 2015. The remaining set of stations totaled 6,600 and were globally distributed as shown in **Figure 3.2**.

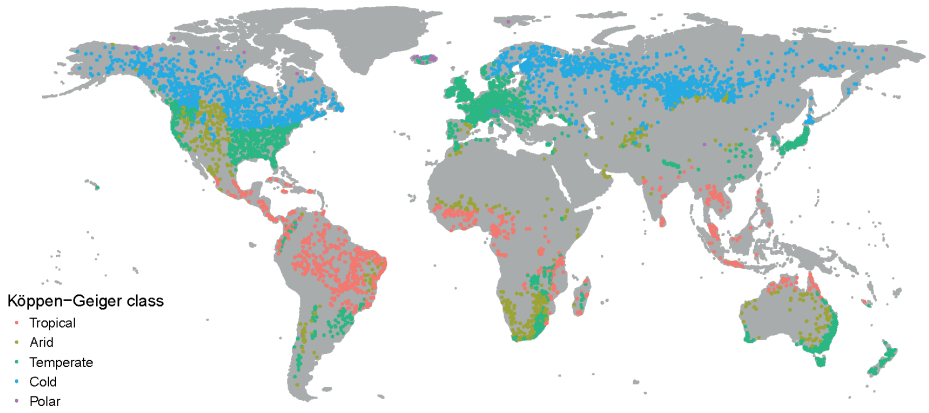


Figure 3.2 | Distribution of the 6,600 GRDC stations monitored for at least 10 years in the period 1960-2015. Stations are colored according to the Köppen-Geiger climate classification (Kottek et al., 2006).

3.2.3 Catchment-specific covariates

As covariates of the flow metrics we used topography and climate, which we retrieved from publicly available spatially explicit sources and then aggregated to the upstream catchment of each grid cell. The choice of the covariates set and source data was based on previous studies (Barbarossa et al., 2017; Beck et al., 2015, 2013; Farmer and Vogel, 2013; Razavi and Coulibaly, 2013; Verdin and Worstell, 2008; Vogel et al., 1999), expert knowledge and data availability. A list of the covariates and related source databases is provided in **Table 3.1**.

Table 3.1 | Description of the predictor variables used as input for the modelling of AF. The spatial resolution refers to the source data; for the analysis all variables were resampled to ~1 km.

Category	Variable description	Source	Annual metric	Unit	No. layers	Spatial resolution	Temporal coverage
Topography	Upstream catchment area	This study	-	km^2	1	~ 1 km	-
	Elevation	SRTM + GTOPO30	-	m	1	~ 1 km	-
	Surface slope	USGS	-	°	1	~ 1 km	-
Climate	Precipitation	MSWEP + GPCC	Mean	mm/ month	56 x 4	~ 25 km	1960 - 2015
			Minimum	mm/ month			
			Maximum	mm/ month			
			Seasonality index	-			
	Air temperature	CRU TS 3.24.01	Mean	K	56 x 3	~ 50 km	1960 - 2015
			Minimum	K			
			Maximum	K			
	Potential evapo-transpiration	This study	Mean	mm/ month	56 x 2	~ 50 km	1960 - 2015
			Seasonality index	-			
	Aridity index	This study	-	-	56	~ 25 km	1960 - 2015

We calculated the area of the upstream catchment of each cell by summing the areas of the upstream grid cells. We derived the upstream catchment-averaged elevation from the SRTM DEM40 resampled at 30 arc seconds as provided by HydroSHEDS35, supplemented with the GTOPO30 DEM for areas lacking SRTM coverage, i.e., latitudes above 60°N. We transformed the elevation values by adding a constant value of 500 m to avoid negative values, the lowest being represented by the shores of the Dead Sea at 430 m below sea level. We employed the USGS slope map developed for the Prompt Assessment of Global Earthquakes for Response (PAGER) system (Verdin et al., 2007) to calculate upstream catchment-averaged surface slope values. This map is based on the same SRTM+GTOPO30 DEM and has been corrected for the discrepancy between ground units (arc degrees) and elevation units (meters; **Figure 3.3**) (Verdin et al., 2007).

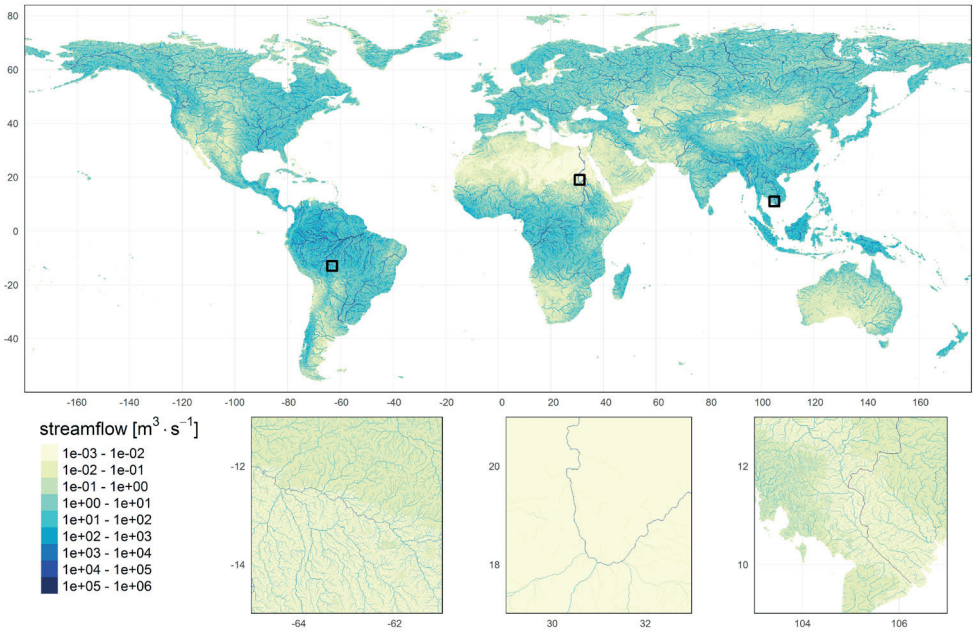


Figure 3.3 | Long-term mean annual flow map overview. The long-term mean annual flow represents the average of the year-specific FLO1K maps for mean AF over the period 1960-2015. The global map has been upscaled using maximum-value resampling by a factor of 20 for clarity of visualization. Insets show the original 1 km resolution. Location of each inset is marked on the global map by a black square.

We derived the upstream catchment-averaged values for annual mean, maximum and minimum air temperature (T_{air}) and precipitation (P), as well as potential

evapotranspiration (PET), aridity index (AI) and seasonality index for P and PET, for every year over the period 1960-2015. For air temperature, we employed the Climate Research Unit (CRU) Time Series (TS) dataset (Harris et al., 2014) (version 3.24.01; monthly temporal and 0.5° spatial resolution). For precipitation, we used the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset (Beck et al., 2017) (version 1.2; 3-hourly temporal and 0.25° spatial resolution; 1979-2015) supplemented with the Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis (Schneider et al., 2015) (version 7; monthly temporal and 0.5° spatial resolution) prior to 1979. MSWEP merges a wide range of gauge, satellite, and reanalysis datasets to achieve precipitation estimates with greater accuracy than any other global dataset (Beck et al., 2017). To combine the GPCC and MSWEP datasets, we rescaled the GPCC estimates such that the 1979-2013 mean of GPCC matched that of MSWEP. For each year and grid cell, we retrieved the mean annual value of T_{air} and P as the mean over the 12 monthly layers, and the minimum and maximum as the lowest and highest monthly values, respectively. We computed mean annual potential evapotranspiration from monthly T_{air} values following the temperature-based approach of Hargreaves et al., (1985) and employing the same CRU TS v. 3.24.01 source data for temperature. Similarly, we calculated seasonality index layers for P and PET as $X_{si} = X_{yr}^{-1} \sum |X_m - X_{yr}/12$, where *si*, *yr* and *m* stand for seasonality index, yearly and monthly values, respectively (Walsh and Lawler, 1981). We downscaled the raster layers for the climate-related covariates to match the 30 arc seconds resolution of the hydrography using nearest-neighbour resampling. In addition, we calculated the aridity index for each year as PET/P, using mean annual P and PET.

To calculate the upstream catchment-average values of the covariates, we employed the TauDEM software (Terrain Analysis Using Digital Elevation Models, http://hydrology.usu.edu/taudem). TauDEM is an open-source C++ software explicitly designed to implement the flow algebra for large datasets, employing a Message Passing Interface (MPI, http://mpi-forum.org) to implement highly parallelized processing algorithms (Tarboton, 2008; Tarboton et al., 2009; Tesfa et al., 2011). We extracted the covariates for the upstream catchment of each cell of the global hydrological network via the so-called flow accumulation technique ('AreaD8' in TauDEM). This technique considers each grid cell as a pour point and subsequently calculates the number of upstream grid cells or the sum of the attribute values of these upstream grid cells, using the flow direction map to delineate the watershed boundaries of the upstream catchment. To derive continuous upstream catchment-averaged values for the predictor variables, we divided the sum of the upstream covariate values by the total number of upstream grid cells at each pour point. To speed-up the calculations, we split the global flow direction layer into six continents (North America, South and Central America, Europe, Africa, Asia, Oceania). Adjacent continents (e.g., Europe and Asia) were separated along watershed boundaries.

3.2.4 Training of Artificial Neural Networks

We quantified the relationships between the flow metrics and the covariates using artificial neural networks (ANNs), which have been widely used for hydrological modelling from local (Yaseen et al., 2015) to global (Beck et al., 2015, 2013) scales. We employed the feed-forward ANN algorithm based on the multi-layer perceptron structure with one hidden layer (Bishop and M., 1995; Haykin, 1994) (**Figure 3.1**). We trained the ANNs based on year-specific values of mean, minimum and maximum AF, using the upstream-catchment topography and year-specific climate as covariates (**Table 3.1**). We applied a Box-Cox transformation to normalize the distributions of each variable (response and covariates) (Box and Cox, 1964). In addition, we standardized each distribution to zero mean and unit standard deviation, as required for the ANNs (Haykin, 1994). To avoid possible bias due to differences in monitoring intensity among the stations, we randomly picked 10 yearly values from those stations monitored at least 10 years across the 1960–2015 period. We then iterated the ANNs training 20 times, sampling different years from those stations having a record longer than 10 years. Prior to the training, we tuned the number of neurons of the hidden layer of the ANNs and the weights decay value to regularize the ANNs cost function, and therefore control for overfitting. To this end, we used 10-fold cross-validation (CV) whose folds were based on excluded monitoring stations, and identified the number of neurons and weights decay value that maximized the median coefficient of determination (R^2) and minimized the median Root Mean Square Error (RMSE) of the testing set. As a result, we employed 20 neurons for the ANNs hidden layer and a weights decay value of 0.01.

3.2.5 Generating mean, maximum and minimum AF global maps

We applied the ANNs model to produce 30 arc seconds maps with mean, maximum and minimum annual flow from 1960 through 2015 (Data Citation 1: figshare https://dx.doi.org/10.6084/m9.figshare.c.3890224). For each grid cell, we computed the AF metrics as the median across the outputs of 20 trained ANNs and back-transformed the values to $m^3 \cdot s^{-1}$.

We upscaled the 30 arc seconds layers to 5 and 30 arc minutes resolutions, in order to serve potential coarser-grain applications. We based the upscaled output on the 5 and 30 arc minutes flow direction grids produced by applying the dominant river tracing (DRT) algorithm to the same 30 arc seconds flow direction layer used in this study (Wu et al., 2012, 2011). The 5 and 30 arc minutes flow direction grids are freely available for download at http://files.ntsg.umt.edu/data/DRT/. We upscaled the 30 arc seconds streamflow values by choosing the value of the cell that minimized the differences in upstream-drainage area between the native 30 arc seconds and the coarser resolution

grid cell. For the 5 arc minutes grids it was necessary to employ a one-cell search radius to avoid losing connectivity.

3.2.6 Code availability

The code used to generate the covariate data, geo-reference the monitoring stations, train the ANNs and generate the flow maps (Data Citation 1: figshare https://dx.doi.org/10.6084/m9.figshare.c.3890224) was written and run in R version 3.3.2. TauDEM tools (Tarboton, 2008) were used to produce the catchment-specific covariate layers and GDAL library (GDAL Development Team, 2017) functions were employed to handle the analyses on large raster data. The scripts are available on request.

The ensemble of trained ANNs are available as R objects (.rds) and as Portable Model Markup Language (PMML) objects for cross-platform compatibility (.pmml, http://dmg.org). The parameters used for the Box-Cox transformation and standardization of the variables employed by the ANNs are also available in CSV format.

3.3 Data Records

The FLO1K dataset is a set of gridded layers packed as NetCDF-4 files freely available for download (Data Citation 1: figshare https://dx.doi.org/10.6084/m9.figshare.c.3890224). For each of the three AF metrics, 56 yearly layers are available from 1960 through 2015, yielding a total of 168 layers. Each non-null cell represents the potential streamflow in $m^3 \cdot s^{-1}$, stored as 32-bit floating point. Layers are in the WGS84 coordinate system with a cell size of 30 arc seconds (~ 1 km) and a global extent, including all continents except for Antarctica (90°N to 90°S latitude and 180°W to 180°E longitude). In addition, upscaled data are available at 5 and 30 arc minutes.

3.4 Technical Validation

To evaluate the quality of the FLO1K maps, we run a 10-fold cross-validation for each of the 20 ANN runs, such that each observation was included in the test set once and by splitting the folds by stations. We assessed the overall map quality with R^2 and RMSE calculated based on log-transformed values to evaluate the performance across the full spectrum of streamflow values (10^{-3}-10^5 $m^3 \cdot s^{-1}$). Cross-validation results showed high agreement between training (90%) and independent testing (10%) data, with negligible variation among the replicates (**Table 3.2**).

Table 3.2 | Model performance statistics. The R^2 and RMSE values represent medians (with standard deviations in brackets) of the 10-fold cross-validation of 200 replicates. Both R^2 and RMSE were calculated from log-transformed values, therefore the RMSE is unitless.

AF metric	R^2		RMSE	
	Training	Testing	Training	Testing
Mean	0.92 (0.001)	0.91 (0.002)	0.32 (0.002)	0.34 (0.004)
Maximum	0.91 (0.001)	0.90 (0.002)	0.33 (0.002)	0.34 (0.005)
Minimum	0.85 (0.001)	0.83 (0.003)	0.48 (0.002)	0.51 (0.005)

We assessed the uncertainty per grid cell resulting from the sub-sampling of the monitoring stations, by computing the coefficient of variation (CoV) over the 20 replicates. Uncertainty was very low (CoV < 0.5) for the main river stems globally and smaller reaches in wet regions (**Figure 3.4**). We found higher uncertainty (higher CoV values) for low streamflow values in dry areas, e.g., the upper basin of the Nile (central inset of **Figure 3.4**). These higher CoV values likely reflect the lower number of streamflow observations available for calibrating the ANNs in these areas. The highest CoV values (> 3.5) were found in grid cells with a low number of upstream grid cells (typically <5) in dry areas. In these grid cells, most of the ANN replicates yielded zero-flow values whereas one or few replicates yielded close-to-zero values, resulting in a low mean yet large CoV across the 20 replicates.

We checked for potential bias in streamflow estimates in the northern hemisphere due to snowmelt delays, e.g., the contributing effect of snowfall in November-December of the previous year on the streamflow in May-June. To this end, we generated streamflow maps based on the US water year (November-October) for stations north of 40N and compared their performance to the original (calendar year-based) FLO1K maps. We tuned the ANNs ensemble and computed the streamflow fields adopting the US water year for both the streamflow data and the climate input variables. Differences in R^2 between models based on calendar versus US water year were smaller than 0.01 and therefore considered negligible (**Table 3.3**).

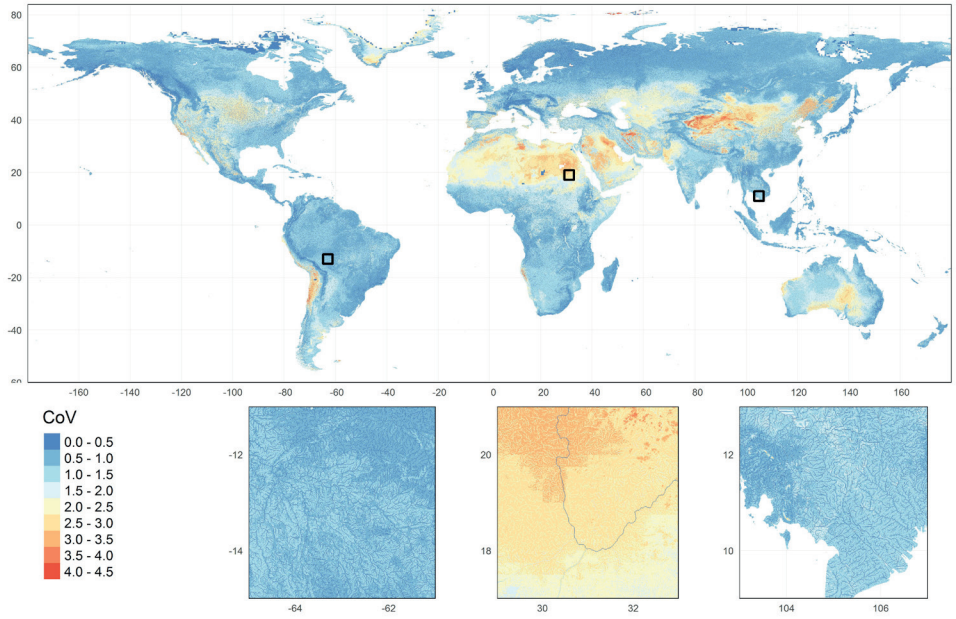


Figure 3.4 | Uncertainty in mean AF due to differences in sub-sampling among the 20 ANN replicates. Uncertainty is expressed as coefficient of variation (CoV) averaged across the overall period 1960-2015. The global map has been upscaled using mode resampling by a factor of 5 for clarity of visualization. Insets show the original 1 km resolution. Location of each inset is marked on the global map by a black square.

Table 3.3 | Comparison of R2 values for streamflow metrics based on calendar vs US water year. The comparison is based on 2,484 stations north of 40N latitude, monitored for at least 30 years in the period 1960-2015. The R2 was calculated from log-transformed values. LT: long term; YR: yearly.

	AF metric					
	Mean (LT)	Mean (YR)	Max (LT)	Max (YR)	Min (LT)	Min (YR)
Calendar year	0.975	0.953	0.969	0.936	0.933	0.885
US water year	0.975	0.955	0.969	0.942	0.928	0.879

3.5 Usage notes

The FLO1K dataset reports the potential streamflow in $m^3 \cdot s^{-1}$ in each grid cell, i.e., the discharge that would occur if there were a natural watercourse. To avoid confusion, we emphasize that the estimates represent volumetric streamflow rather than specific runoff. As such, the estimates cannot directly be compared with outputs from climate or land surface models without a streamflow routing component.

We refrained from filtering the output to the actual stream network because there are multiple methods for stream network delineation (Avcioglu et al., 2017; Hancock, 2005; Istanbulluoglu et al., 2002; Russell et al., 2015; Sangireddy et al., 2016; Tarboton et al., 1992, 1991; Thompson et al., 2001), which users of FLO1k may want to select or refine according to their needs. For global-scale analyses one might adopt an arbitrary upstream catchment area threshold in order to delineate the network (e.g., 25 upstream grid cells as in Hydrosheds (Lehner et al., 2008)), as to our knowledge more refined methods have not yet been developed/tested.

The estimated maximum and minimum flow values for a given year reflect the highest and the lowest monthly values of that year. This does not give an indication about which months of the year belong to the maximum or minimum flow. The corresponding months might change from year to year based on the yearly distribution of the precipitation.

Users of the upscaled streamflow grids should keep in mind that these are contingent on the respective DRT flow direction layers (Wu et al., 2012, 2011). Further, the accuracy of the upscaled grids has not been evaluated.

3.6 ACKNOWLEDGMENTS

The authors would like to thank the Global Runoff Data Center for providing the streamflow data. The authors would also like to thank Huan Wu and Kristine L. Verdin for providing the global hydrography base layer and the global surface slope data (PAGER), respectively. This project has received funding from the Europeans Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 641459.

4

Worldwide impacts of existing and future dams on the connectivity of freshwater fish ranges

Valerio Barbarossa, Rafael Schmitt,
Mark A. J. Huijbregts, Christiane Zarfl,
Henry King, Aafke M. Schipper

Submitted

ABSTRACT

Dams have a high socio-economic value (e.g., for water security, energy provisioning and flood protection), but also fragment freshwater habitats. Yet, a global species-level quantification of dam-induced fragmentation is lacking. Here, we assessed the degree of fragmentation of the occurrence ranges of ~5,700 lotic fish species worldwide due to ~40,000 existing and ~3,700 additional future dams. Per river basin, we quantified a connectivity index (CI) for each fish species by combining its occurrence range with a high-resolution hydrography and the locations of the dams. Ranges of non-diadromous fish species were more fragmented (CI=69 ± 30%; mean ± standard deviation) than ranges of diadromous species (CI=82 ± 22%). Fragmentation due to future dams was especially high in the tropics, with declines in mean CI of ~20-30% in the Amazon, Niger, Congo and Mekong basins. Our assessment can guide river management at multiple scales and in various domains, including strategic hydropower planning, identification of species at risk, and prioritization of restoration measures, such as dam removal and bypass construction.

4.1 Introduction

Freshwater habitats cover only about 0.8% of the Earth's surface, yet they host a disproportionately high diversity of species. One third of the described vertebrates, including ~40% of the fish species, are found in freshwater environments (Nelson, 2006). Freshwater biodiversity is also disproportionately threatened, with decline rates higher than observed for marine or terrestrial biodiversity (WWF, 2018). Damming of rivers is one of the main threats to freshwater biodiversity (Reid et al., 2019; Vörösmarty et al., 2010). While dams provide direct economic benefits (e.g. water security, flood protection and renewable energy), they affect freshwater ecosystems by inundation, hydrologic alteration and fragmentation, for example (Grill et al., 2019; Herbert and Gelwick, 2006). Fragmentation of the freshwater environment has major implications for freshwater fish (Carvajal-Quintero et al., 2019) as dams obstruct migration routes, essential for spawning or feeding, and limit dispersal (Fuller et al., 2015; van Puijenbroek et al., 2019). The near-future expansion of hydropower facilities will further threaten freshwater fish biodiversity (Reid et al., 2019). While an estimated ~50% of the river volume is currently altered by either flow regulation or fragmentation, the pending construction of ~3,700 major hydropower dams is expected to increase this percentage to 93% (Grill et al., 2015; Zarfl et al., 2015).

Efforts to quantify the effects of dams on the connectivity of freshwater fish habitat have mainly been carried out at local scales (e.g., 9, 12, 13). Global-scale assessments are key to identify and prioritize conservation needs, but thus far have focused on mapping river connectivity without quantifying impacts on freshwater biodiversity (Grill et al., 2019, 2015; Nilsson et al., 2005). An exception is Liermann et al. (Liermann et al., 2012) relating the degree of present-day river fragmentation within freshwater ecoregions to their overall freshwater fish diversity. However, freshwater ecoregions cover large extents (average area = 311,605 km^2, n = 426) (Abell et al., 2008) and do not account for the actual geographical range of species, which can be limited to smaller areas and hence be subject to different degrees of fragmentation. In addition, by encompassing and cutting through multiple watersheds (Abell et al., 2008), ecoregions do not account for the spatial connectivity of rivers, which defines the spatial template for aquatic biodiversity and species migration. Thus, a spatially resolved and species-specific assessment of fragmentation effects, accounting for the actual global drainage network, is missing.

Here, we assessed impacts of current and future dams on the geographic range connectivity for ~5,700 lotic (i.e., that live partially or exclusively in flowing freshwater bodies) fish species worldwide. We included ~40,000 existing dams (Lehner et al., 2011; Mulligan et al., 2009) and ~3,700 dams that are currently under construction or planned

(Zarfl et al., 2015). We employed a species-specific modelling approach to quantify connectivity for each of the ~5,700 fish species based on their actual occurrence range. We based our analysis on a high-resolution hydrological network comprising ~1M sub-basin units with an average size of ~100 km^2 (Lehner et al., 2008). We adopted connectivity measures specific to non-diadromous and diadromous fish species, respectively, as fragmentation impacts of dams might differ between fish that migrate between freshwater and marine environments and fish that complete their lifecycle in freshwater (Cote et al., 2009). The species-specific modelling approach combined with a detailed hydrography allowed us to identify species groups most at risk as well as geographic hotspots of impact.

4.2 Results

Based on the ~40,000 existing dams we found substantially lower range connectivity for non-diadromous species (CI=69±30%, mean ± standard deviation) compared to diadromous species (CI=82±22%, **Figure 4.1**). The completion of ~3,700 dams that are currently under construction or planned will further lower the connectivity of non-diadromous species' ranges (CI=60±31%, **Figure 4.1**). For diadromous species the decrease in CI due to future dams construction was smaller (CI=78±23%, **Figure 4.1**), but still locally relevant (**Figure 4.2**). In general, connectivity was lower and projected to further decrease in larger rather than smaller basins (**Figure 4.2**).

Our results revealed the lowest CI values for species occurring in North America, Europe, India and China. These values did not substantially decrease in the future (**Figure 4.2**). In contrast, we found the largest differences between present and future dams for species occurring in South America, Africa and South East Asia (**Figure 4.2**). Decreases in connectivity due to future dams were particularly high for non-diadromous species in large tropical and sub-tropical rivers, e.g., Amazon, Congo, Niger, Salween, and Mekong (**Figure 4.2**, **2.3a** and **B.4**). For instance, the mean CI across the species of the Amazon basin dropped by ~30 percent points in the future, and in the Mekong, Congo and Niger the drop was ~20 percent points (**Figure 4.3a**). In contrast, the connectivity of diadromous species' ranges declined most in small basins along the coastline of Central-Eastern Africa, Western Africa and the Malay Archipelago (**Figure 4.2**). For instance, in the Comoe in Western Africa and the Purari basin in Papua New Guinea, the construction of a few mainstream dams resulted in a strong decrease in connectivity. For instance, the mean connectivity dropped from ~100% to ~20% in the Purari basin and to ~50% in the Comoe basin (**Figure 4.3b**). Yet, high CI reductions for diadromous species were observed also in larger basins like the Danube, Niger and Salween, where connectivity drops were as high as ~30 percent points (**Figure 4.2**, **4.3b** and **B.4**).

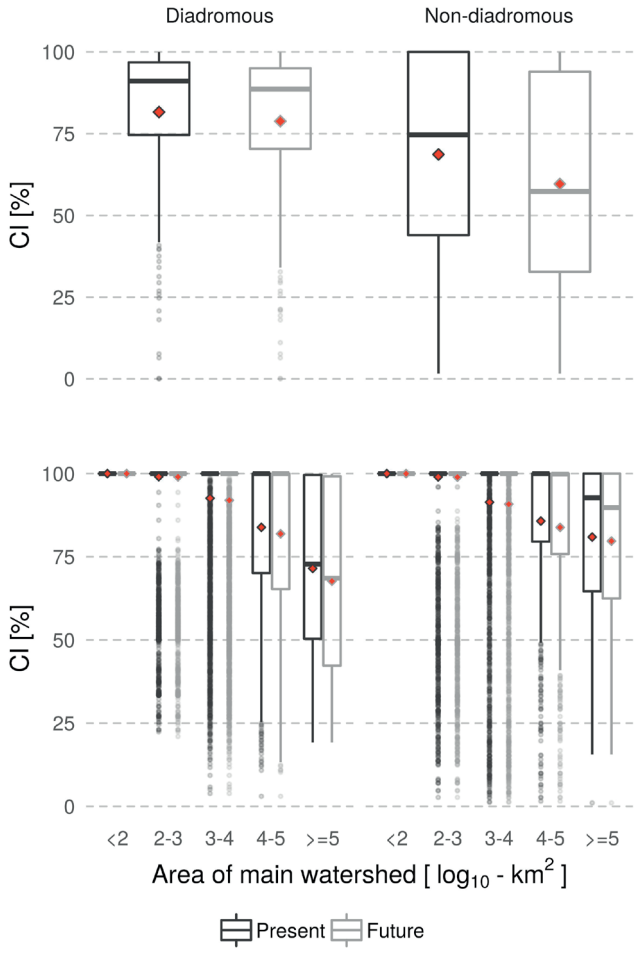


Figure 4.1 | Connectivity Index (CI) across species (top) and main river basins (bottom). Values are shown for present dams (black) as well as present and future dams together (gray). For species occurring in multiple watersheds, the area-weighted mean of the basin-specific CI values was calculated (top). The basin-level CI (bottom) represents the mean of the CI values across the species occurring within the basin. Boxes represent the interquartile range and the median, and whiskers the 95% interval. Red diamonds represent the mean.

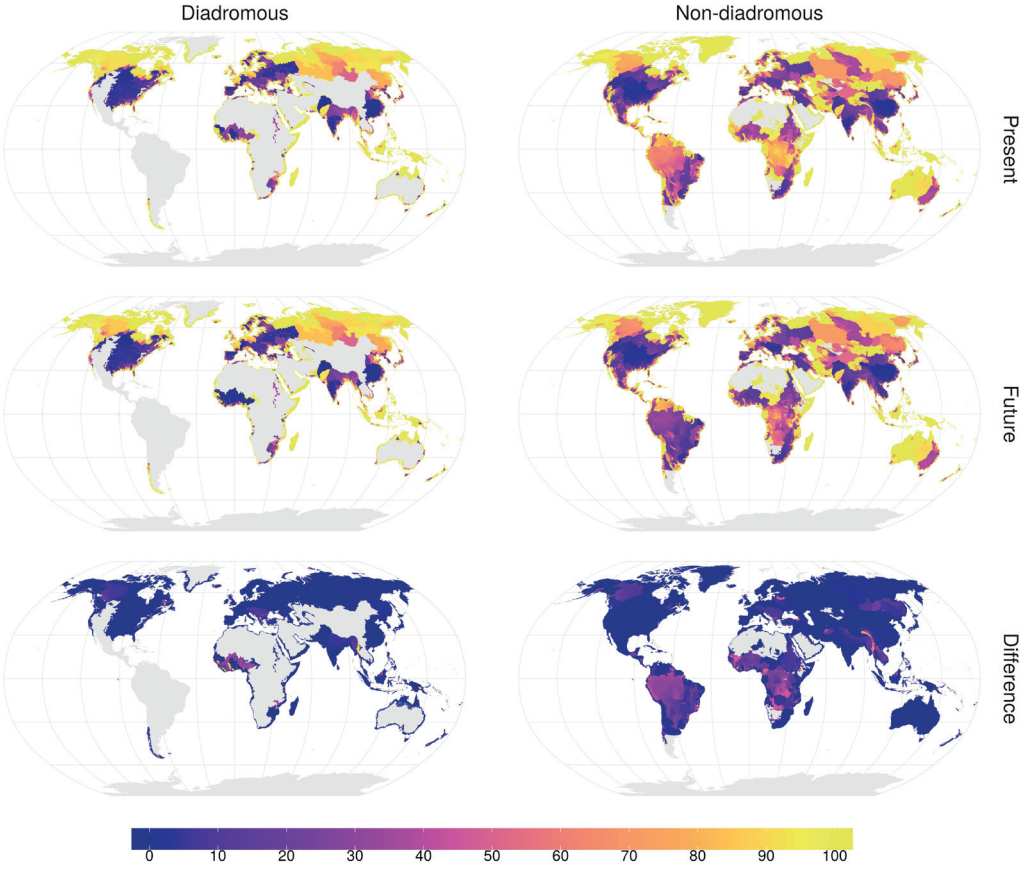


Figure 4.2 | Mean connectivity index (%) per sub-basin (~1M units) for present situation (top), future projection (center) and the difference between them (bottom) for diadromous (left) and non-diadromous (right) fish species. Grey represents areas without species range data.

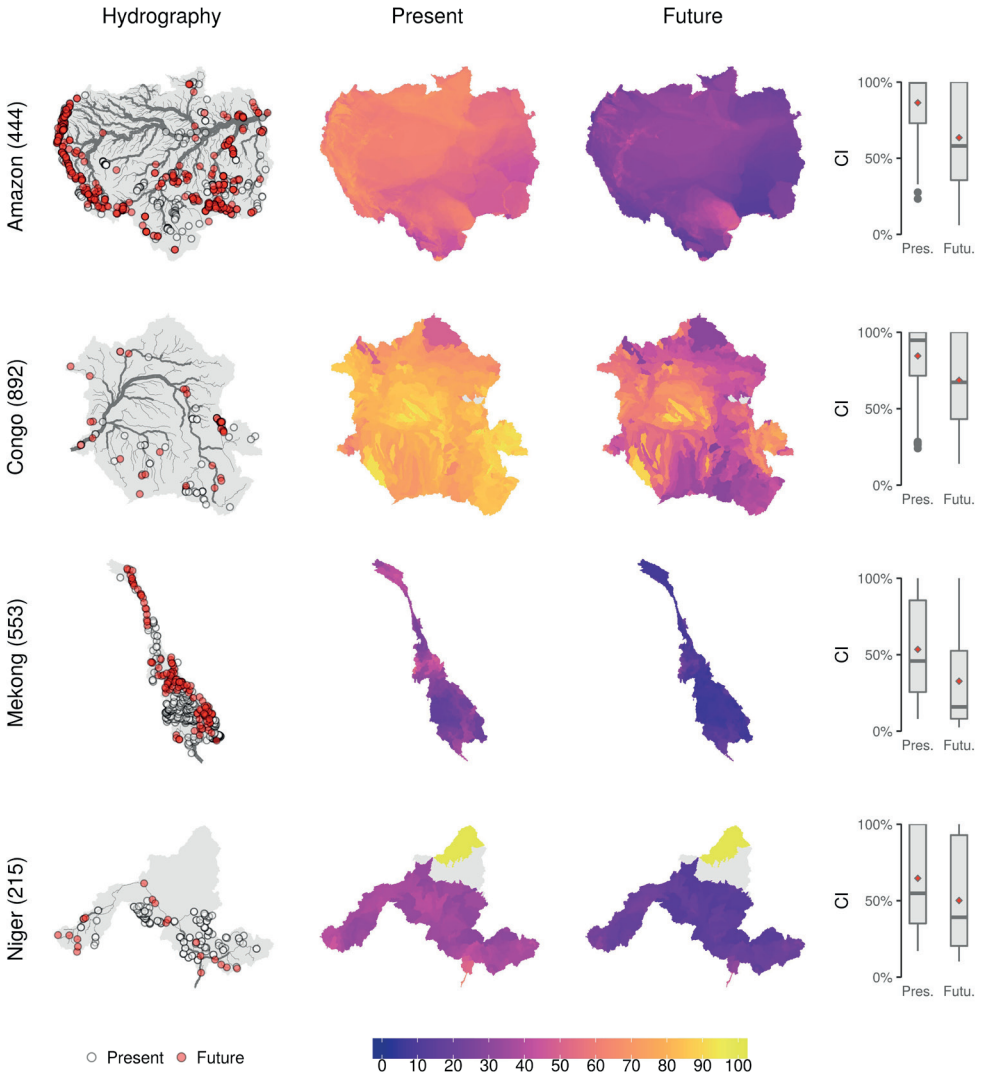


Figure 4.3a. Mean connectivity Index (CI) across non-diadromous species in four main hydrologic basins. The maps show the basin hydrography with the location of dams (on the left) and the connectivity index at the sub-basin level for the present situation and future projections (center). On the right, the species-specific CI values are summarized as boxplots, with diamonds representing the mean. Location of the selected basins is shown in **Figure B.6**.

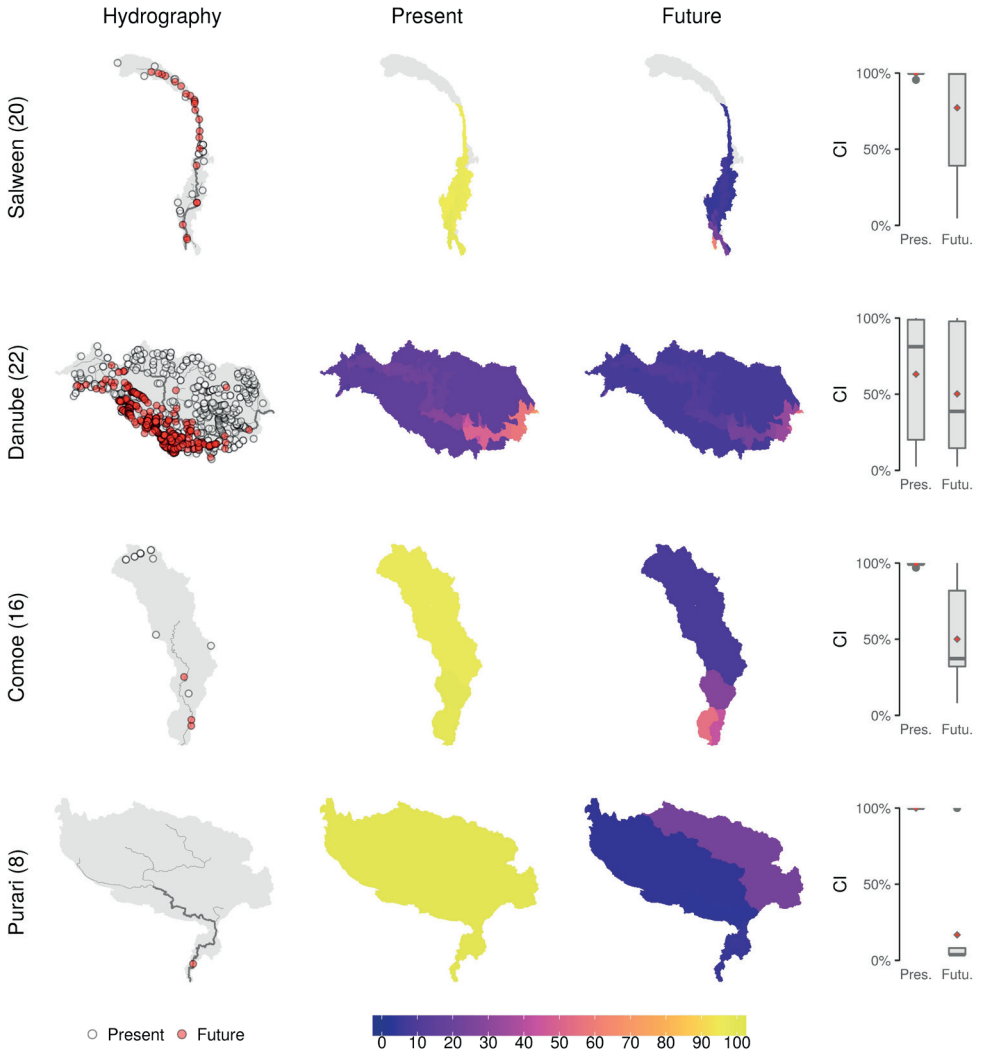


Figure 4.3b. Mean connectivity Index (CI) across diadromous species in four main hydrologic basins. The maps show the basin hydrography with the location of dams (on the left) and the connectivity index at the sub-basin level for the present situation and future projections (center). On the right, the species-specific CI values are summarized as boxplots, with diamonds representing the mean. Location of the selected basins is available in **Figure B.6**.

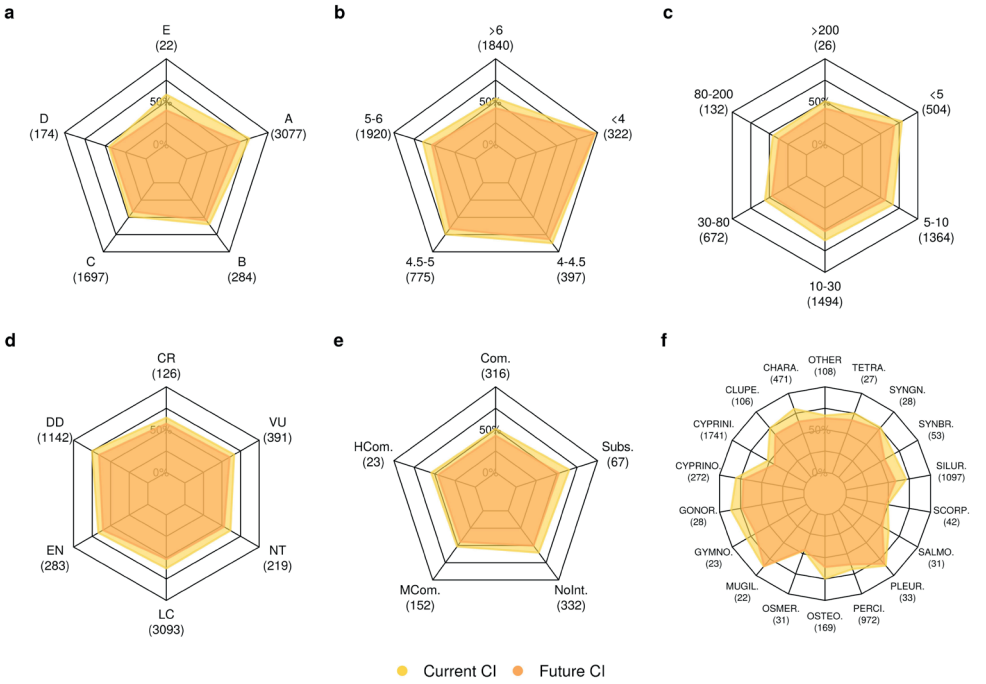


Figure 4.4 | Mean of species-specific CI values by different traits and categories for present and future, including a) Köppen-Geiger climate zones, where A=Equatorial, B=Arid, C=Warm temperate, D=Snow, E=Polar, b) initial geographical range area (log (base 10)-transformed km^2), c) body length (cm), d) IUCN threat status, where CR=Critically endangered, VU=Vulnerable, NT=Near Threatened, LC=Least Concern, EN=Endangered, DD=Data Deficient, e) commercial relevance , where Com. = commercial, HCom. = highly commercial, MCom. = minor commercial, NonInt. = of no interest and Subs. = subsistence fisheries, and f) species order, where groups with less than 20 species are grouped together in the "other" category (the full list of order names is provided in **Table B.1**). For species occurring in multiple watersheds, an area-weighted mean of the basin-specific CI values was calculated before we averaging across species. Each panel shows a subset of the analyzed 5,638 species in this study (underneath each category the number of species is reported in brackets), as metadata for species traits and categories was not available for all the species.

In the current situation the ranges of species in equatorial climates (n = 3,077) were the most connected, but the connectivity dropped considerably in response to the future dams (mean CI from ~80% to ~70%; **Figure 4.4a**). CI values for species with larger occurrence ranges were lower and projected to further decrease in the future compared to species occurring over smaller areas (<~10,000 km^2, **Figure 4.1** and **4.4b**). The CI was lower for large-bodied species (maximum body length >30cm, mean CI≈50%) than for very small species (< 5cm, mean CI≈75%; **Figure 4.4c**). Species classified as "least concern"

or "not threatened" by the IUCN, revealed CI values equal to or lower than threatened species (**Figure 4.4d**). Furthermore, we found overall low mean CI values for species of commercial importance (~50%), while projected future declines in connectivity were relatively high for species of no commercial interest and species subject to subsistence fisheries (**Figure 4.4e**). Taxonomic groups with the lowest present and future CI values were Cypriniformes, Osmeriformes, Salmoniformes and Scorpaeniformes (**Figure 4.4f**).

4.3 Discussion

In this study, we assessed impacts of present and future dams on geographical range connectivity for ~5,700 lotic fish species. Based on the largest consistent global compilation of existing dams, we found that the ranges of non-diadromous species are, on average, considerably more fragmented than the ranges of diadromous species (**Figure 4.1**). The marked difference is likely due to the different spatial distribution of the two species groups with respect to dams' location (**Figure B.2** and **B.3**). Diadromous species are concentrated along the coastlines of the world's continents, with hotspots of species richness in small basins of the Malay Archipelago and African coastline (**Figure B.2**), which are less fragmented (**Figure 4.2**) due to relatively low numbers of both present and future dams (**Figure B.3**). In contrast, the connectivity reduction of ~9 percent points for non-diadromous species (**Figure 4.1**) suggests that the relatively large number of small upstream dams planned for hydropower generation (**Figure B.5**) (Zarfl et al., 2015) will affect non-diadromous fish disproportionately. According to our results, ranges of many non-diadromous species in the world's largest tropical basins will become highly fragmented after the completion of dams currently under construction or planned (**Figure 4.2** and **4.3a**). This result is in line with the general expectation that the biodiverse tropical watersheds such as the Amazon, Mekong and Congo will experience large ecological consequences from the expected boom in hydropower in these regions (Latrubesse et al., 2017; Winemiller et al., 2016).

We acknowledge that the database of dams considered in our study is not exhaustive, as small barriers are not considered. Even the ~40,000 dams covered by the GRanD and GOODD databases represent a minority of all dams worldwide (McCully, 2001; World Comission on Dams, 2000), highlighting the need for increased efforts to geo-reference dams and barriers. From this perspective, the connectivity index values reported here most likely represent an optimistic estimate (i.e., ranges might be more fragmented in reality). In addition, present-day species ranges may have been contracted compared to a pristine scenario due to existing dams (van Puijenbroek et al., 2019), potentially leading to a further underestimation of the fragmentation impacts in our assessment. On the contrary, we could not include fish passes in our modelling approach hence

assumed full impassability of the barriers, potentially leading to an overestimation of the fragmentation in some river basins. Yet, while efforts have been carried out to include fish passes in dams to allow fish to migrate across the barrier, evidence suggests that fish passes are not very effective and sometimes even harmful, with nature-like bypasses as the only exception (Birnie-Gauvin et al., 2018; Silva et al., 2018). We further acknowledge that the species covered by the IUCN represent roughly half of the known freshwater fish species (Tedesco et al., 2017), with a spatial bias in terms of coverage. For instance, the IUCN assessed only about one fourth of the species actually occurring in the Amazon basin (Tedesco et al., 2017). This implies that the basin-average connectivity index might be lower or higher than estimated here, depending on the locations of the ranges of the missing species.

Higher levels of fragmentation will likely resolve in reduced fish populations (Alò and Turner, 2005; Dias et al., 2017; Ziv et al., 2012). By disconnecting the continuum of the river network, dams isolate fish populations, reduce access to feeding areas and disrupt access to spawning sites (Fuller et al., 2015; van Puijenbroek et al., 2019). Indirectly, dams also exert additional upstream and downstream pressures on the aquatic habitats, e.g., by impoundment and altering flow and thermal regime and sediment and nutrient dynamics (Grill et al., 2019; Herbert and Gelwick, 2006; Poff and Schmidt, 2016). The tradeoffs among ecosystem services and hydropower gains are often overlooked and might lead to high socio-economic costs (Intralawan et al., 2018). This holds especially where communities are highly reliant on inland fisheries as source of proteins and household income, e.g., in basins such as the Amazon, Congo, Niger, Mekong, Irrawaddy and Salween (FAO, 2018; Intralawan et al., 2018; McIntyre et al., 2016). On the one hand, implementation of aquaculture fish production can compensate for reduced inland fisheries but it is associated with environmental impacts like introduction of non-native fishes and degradation of water quality (Edwards, 2015; Jia et al., 2015). Further, increased fragmentation could potentially exacerbate the ecological effects of size-biased harvesting of freshwater fish species. Our findings revealed that larger species have, on average, a more fragmented range than smaller species (**Figure 4.4c**), meaning that damming could potentially increase the effects of direct human exploitation, that are larger on the bigger fish species (Olden et al., 2007).

To our knowledge, our assessment is the first to quantify losses of connectivity of freshwater fish species ranges globally. Our species-specific and high-resolution (watershed units of ~100 km^2) method enables further understanding of potential ecological effects of existing dams and future dam installations that are often neglected in hydropower planning (Latrubesse et al., 2017). For instance, the case of the Purari basin in Papua New Guinea showed that the potential completion of one downstream dam

could strongly reduce the connectivity of the basin for diadromous species (**Figure 4.3b**). By only considering the topologic connectivity of the river network, without accounting for the actual geographical ranges and migratory behavior of the species, such patterns would not emerge. Our assessment also showed that fish species in Western Europe, US, India and China retain the lowest connectivity values (**Figure 4.2**). In Europe and US, efforts to remove ecologically impactful dams and restore longitudinal connectivity are currently underway. For instance, ~1,500 dams have already been removed in the US (American Rivers, 2019; O'Connor et al., 2015) and ~2,500 in Europe (data from UK, Sweden, Spain, Finland, www.damremoval.eu), although potential ecological tradeoffs related to the removal of large dams are still not well understood (Foley et al., 2017; Ishiyama et al., 2018). Our study can aid designing optimization strategies to prioritize sites for hydropower expansion (Schmitt et al., 2018) as well as river restoration, e.g. through dam removal (Opperman et al., 2011) or nature-like bypasses (Silva et al., 2018), that minimize or reduce ecological impacts and maximize dams benefits.

4.4 Methods

4.4.1 Species occurrence data

We retrieved data on species' occurrence from the IUCN Red List of Threatened Species (IUCN, 2018). The IUCN database provides spatially explicit occurrence range data for 7,242 freshwater fish species, about half of the known freshwater fish species (Tedesco et al., 2017), compiled from occurrence records and expert knowledge. This database represents the best available data on the spatial occurrence of freshwater fish species worldwide (McManamay et al., 2018). Since our analysis focuses on lotic species (that are found in flowing water bodies), we excluded lentic species that occur exclusively in stagnant water bodies. In addition, we considered only geographical ranges of species reported as extant. In total, we employed geographical ranges for 5,638 fish species (**Figure B.2**). We defined species that according to fishbase.org (Froese and Pauly, 2018) migrate to/from the marine environment as "diadromous" (n = 221), while the remaining were classified as "non-diadromous" (n = 5,417). More details on the classification of species are provided in Appendix B.

4.4.2 Present and future dam locations data

We retrieved data on the locations of 39,912 dams worldwide (**Figure B.3**), including 7,320 major dams (> 15 m high) from the Global Reservoir and Dam (GRanD) database (Lehner et al., 2011) and 32,592 additional dams from the GlObal geOreferenced Database of Dams (GOODD), comprising dams visible on global remote-sensing imagery (Mulligan

et al., 2009). To date, these two databases represent the most comprehensive global source for georeferenced data on dams (Grill et al., 2019). We employed the Future Hydropower Reservoirs and Dams (FHReD) database for future dam locations (Zarfl et al., 2015), which includes 3,681 dams (from here on called "future dams") of which 574 are "under construction" while the others are planned. This collection of dams is limited to hydropower dams above 1 MW capacity with available data on location and capacity and which have, at least, passed the feasibility evaluation stage (Zarfl et al., 2015).

4.4.3 Connectivity index calculation

We calculated a habitat connectivity index for each species following the approach of Cote et al (2009), differentiating between diadromous and non-diadromous species. In our study, we applied the approach developed for potamodromous species to the non-diadromous category, which includes both resident migratory (potamodromous) and resident non-migratory lotic species. Connectivity indices are commonly calculated based on vectorized river networks. However, the IUCN database reports species' occurrence as geographical ranges, areas covering a portion of a main hydrologic basin (which is defined as a basin with an outlet to the sea/internal sink). To apply the connectivity indices to sub-basin (or watershed, from here onwards used interchangeably) areas, we converted watershed area to river length using a well-proven power law $l = \beta a^{\alpha}$ (Hack et al., 1957; Sassolas-Serrayet et al., 2018). β is dependent on the shape of the watershed, while α is a constant ranging between 0.5 and 0.6 (Sassolas-Serrayet et al., 2018). When substituting for l in the CI equations from (Cote et al., 2009), β cancels out (see Appendix B). Hence, we calculated the connectivity index for each non-diadromous fish species s in the main hydrologic basin b ($CI^{N}_{s,b}$) as:

$$CI^{N}_{s,b} = \frac{\sum_{i=1}^{n}(\sum_{j=1}^{m} a_{j,i,s,b}{}^{\alpha})^2}{(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{j,i,s,b}{}^{\alpha})^2} \cdot 100 \quad (4.1)$$

where $a_{j,i,s,b}$ represents the area of watershed j belonging to the isolated patch i (due to a dam) within a main hydrologic basin b and hosting species s. We used a value of 0.55 for α, i.e., the central value within the 0.5-0.6 range proposed by (Sassolas-Serrayet et al., 2018). The isolated patches are counted from the most downstream to the most upstream patch n, and are defined as the area upstream a dam or outlet/sink connecting zones of the species geographical range until the next upstream dam or the main basin boundary. The connectivity index of diadromous fish species s in basin b ($CI^{D}_{s,b}$) was calculated as:

$$CI^{D}_{s,b} = \frac{\sum_{j=1}^{m} a_{j,1,s,b}{}^{\alpha}}{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{j,i,s,b}{}^{\alpha}} \cdot 100 \quad (4.2)$$

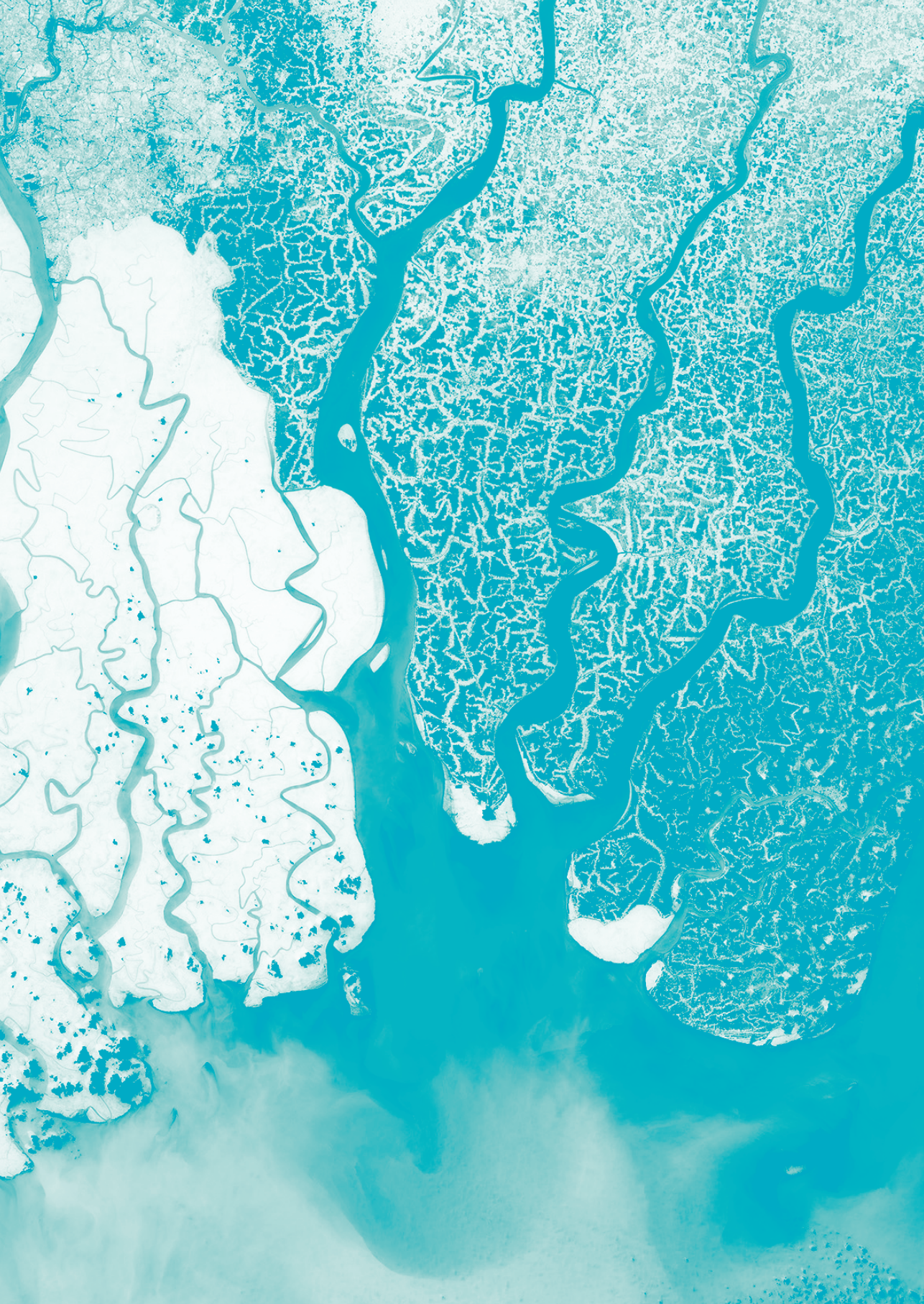
In this case, the CI for diadromous species is solely dependent on the sum of watershed areas $a_{j,1,s,b}$ belonging to the most downstream patch connected to the ocean. This means that the most downstream dam will affect connectivity for diadromous species to a much higher extent than non-diadromous species (Cote et al., 2009). To calculate the areas of the watersheds containing a certain species within a main basin, we allocated the IUCN geographical ranges of each species to the ~1M overlapping HydroBASINS watersheds (Pfafstetter level 12, median area = 137 km^2). Thus we determined in which watersheds a species occurs and in turn which species occur in any given watershed. The derivation of the equations along with an example sketch as well as additional details on choice of the underlying hydrography are provided in the Appendix B.

4.4.4 Aggregation of CI values across species

Based on the species-specific CI values, we calculated a global mean value of CI across all diadromous and non-diadromous species, respectively, for both the present and future scenario. For species occurring in multiple basins (see previous section for a definition of main hydrologic basin vs watershed), we calculated a mean CI value weighted by the occurrence range area of the species within the different main basins. We further calculated basin-specific CI values as the mean CI across all species occurring in the basin. Finally, we calculated the mean connectivity index for species groups characterized by different traits, again using an area-weighted mean for species occurring in multiple basins. We differentiated species according to main climate zone by overlaying the occurrence ranges with the Köppen-Geiger climate categories (Kottek et al., 2006). Species falling into multiple climate zones were assigned the climate zone with the largest overlap. We retrieved information on threat status and taxonomic group (Order) directly from the IUCN metadata (IUCN, 2018). Further, we retrieved data from FishBase (Froese and Pauly, 2018) on maximum body length and commercial importance (species of high-low commercial relevance or used in subsistence fisheries). We performed a synonyms check for the binomial nomenclature provided in the IUCN database to maximize the overlap with the FishBase database (Froese and Pauly, 2018).

4.5 ACKNOWLEDGMENTS

This project has received funding from the Europeans Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 641459. AMS was supported by the GLOBIO project (www.globio.info).



5

Implications of global warming for the world's freshwater fishes

Valerio Barbarossa, Joyce Bosmans,
Mark A. J. Huijbregts, Niko Wanders,
Marc F. P. Bierkens, Henry King, Aafke M. Schipper

Submitted

ABSTRACT

The Intergovernmental Panel on Climate Change recently emphasized the importance to limit global warming to 1.5°C rather than 2°C to reduce, amongst others, impacts on biodiversity and ecosystems. Yet, freshwater fish species have been ignored in global climate-change impact assessments. Here, we provide the first global quantification of potential geographical range contractions of freshwater fish species worldwide due to changes in water temperature and availability. We estimated that 20% of the freshwater fish species will lose >50% of their current geographical range in a 2°C warmer world. Intensifying mitigation efforts to limit warming to 1.5°C lowers this percentage to 13%. In comparison, 44% of the species will lose >50% of their range in a "current pledges" scenario (3.2°C). Projected range losses were largest in tropical and sub-arid regions. Further, range losses were larger than recent estimates for terrestrial species, highlighting the need to intensify (inter) national commitments to limit global warming, particularly to safeguard freshwater biodiversity and ecosystems.

5.1 Introduction

Freshwater habitats are disproportionally biodiverse. While they cover only 0.8% of the Earth's surface, they host ~15,000 fish species, corresponding to approximately half of the global known fish diversity (Dudgeon et al., 2006; Tedesco et al., 2017). As decline rates of global biodiversity have been unusually high over the last four decades (Butchart et al., 2010), freshwater biodiversity has experienced the highest plunge (Dudgeon et al., 2006). Amid human pressures on freshwater ecosystems (including water abstraction, diversion, damming, pollution), anthropogenic climate change is expected to become increasingly important in the future (Reid et al., 2019; Urban, 2015). Rising air temperatures and changing precipitation patterns modify water temperature and flow regimes worldwide, thus affecting two key habitat factors for freshwater species (Knouft and Ficklin, 2017). Being ectotherms, fish are directly influenced by water temperature, while the hydrologic regime determines the structure and dynamics of the freshwater habitat (Comte and Olden, 2017; Poff, 2018). Geographical range shifts or contractions are among the most prominent biotic responses to climate change (Daufresne et al., 2009), whereby the strongest responses are expected in ectotherms (Deutsch et al., 2008). Impacts are exacerbated in freshwater systems, as the insular nature of many freshwater habitats hampers compensatory movements to cooler locations, especially for fully aquatic organisms like fishes (Dudgeon et al., 2006). About 50% of the known freshwater fish species are expected to be affected by climate change (Darwall and Freyhof, 2016). Yet, warming-induced range changes of freshwater fish species have not yet been systematically quantified at the global scale, in sharp contrast with the many studies on species in terrestrial systems (Hof et al., 2011; Powers and Jetz, 2019; Warren et al., 2018; Zurell et al., 2018).

Here, we project changes in the geographical ranges of freshwater fish species worldwide (n = 6,924) for different global warming scenarios. Following the latest IPCC report (Ove Hoegh-Guldberg et al., 2018), we include scenarios that limit global mean temperature increases to 1.5ºC and 2.0ºC. For comparison purposes, we include two additional scenarios: a "current pledges" scenario set at 3.2ºC warming, where no emission cuts are performed after the first Nations Determined Contribution in 2030, and a "no-policy" scenario (no mitigation) set at 4.5ºC warming (all temperatures relative to pre-industrial) (Warren et al., 2018). We compiled species-specific thresholds for three key habitat factors determining the distribution of freshwater fish species (Eaton and Scheller, 1996), namely maximum weekly water temperature, minimum weekly flow and the number of zero flow weeks. Thresholds for these three factors were defined by extreme flow and water temperature conditions in the species' geographical ranges (see Methods). We validated the species-specific water temperature thresholds

with critical thermal maxima reported from laboratory tests (Comte and Olden, 2017) and found them in good agreement (mean standard error = 0.09 [-]; **Figure C.1**). We then used the thresholds to quantify reductions in geographical ranges due to future changes in flow and thermal regimes. We calculated the present and future weekly flow and water temperature values at a spatial resolution of 5 arc-minutes (~10 km) using a Global Hydrological Model (GHM) coupled to a dynamic water temperature model (Sutanudjaja et al., 2018; Wanders et al., 2019). The GHM was forced with the meteorological input from five different Global Climate Models (GCMs) combined with four Representative Concentration Pathway (RCP) future scenarios.

5.2 Results

5.2.1 Global patterns of geographical range contractions

The scenario without climate-change mitigation policy (+ 4.5ºC) resulted in projected range contractions larger than 50% for 69% (±7%) of the freshwater fish species (the value in brackets represents the standard deviation of the GCM-RCP combinations ensemble at each warming level, see **Figure C.2**). This number was projected to decrease to 44% (±9%), 20% (±5%) and 13% (±4%) for warming levels of 3.2ºC, 2ºC and 1.5ºC, respectively (inset of **Figure 5.1a**), suggesting substantial benefits of limiting global warming. We found larger range losses than recently estimated for terrestrial vertebrates, where 6%, 12%, 37% and 57% of the species were projected to lose >50% of their current range for the same four warming targets of 1.5ºC., 2ºC, 3.2ºC and 4.5ºC, respectively (Warren et al., 2018). Further, previous studies have reported that ectotherms are likely to be less resilient to climate change compared to endotherms, due to a lower buffering capacity to increased experienced temperature (Buckley et al., 2012; Deutsch et al., 2008). Our range loss estimates for freshwater fish species were closer to losses estimated for amphibians than for endothermic terrestrial vertebrates (Warren et al., 2018). This contributes to the evidence that climate change is likely to have a higher impact on ectothermic organisms.

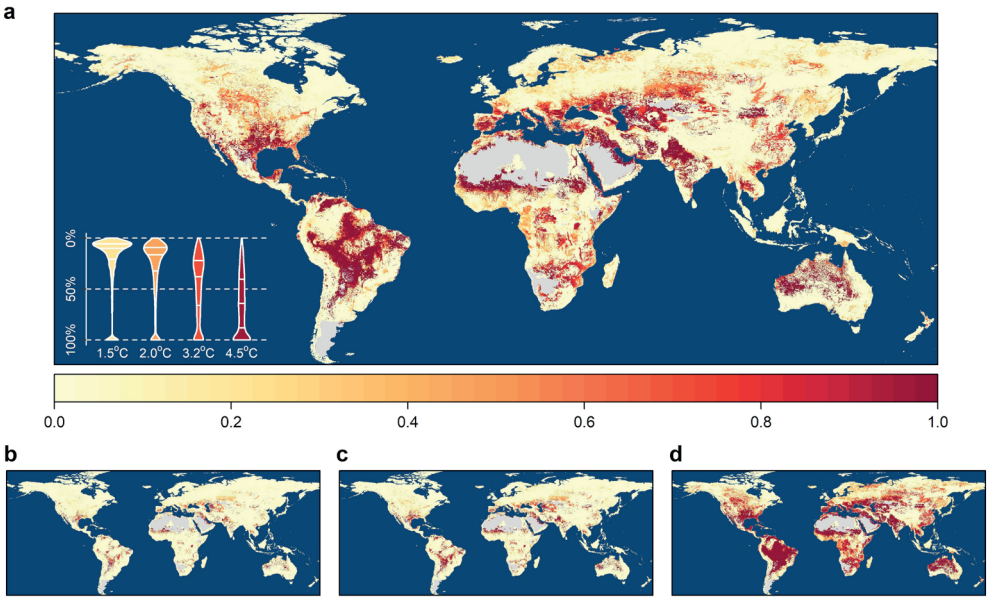


Figure 5.1 | Geographical range contractions of freshwater fish species for different global warming levels. The maps show the local cumulative range loss (see Methods) at five arc minutes (~ 10km) resolution for global warming levels of 3.2°C (a), 1.5°C (b), 2.0°C (c) and 4.5°C (d). The inset shows the distribution of geographical range contractions across the species, including the median and interquartile range (horizontal lines). Gray denotes no data areas (no species occurring or no data available).

Overall, hotspots of range contractions were found in tropical, sub-arid and Mediterranean regions (**Figure 5.1**). At higher latitudes, substantial range contractions became apparent only at higher warming levels (3.2°C, 4.5°C). As a result, the largest local relative species losses were recorded in tropical watersheds such as the Amazon, Niger, Congo and Mekong (**Figure 5.2**; see **Figure C.10** for a more exhaustive overview), in accordance with previous studies suggesting large climate-change induced habitat degradation in these watersheds (Béné et al., 2012; Frederico et al., 2016; Pokhrel et al., 2018). Tropical species are indeed expected to be highly affected by climate change (Tewksbury et al., 2008), and our results confirm this for freshwater fishes (**Figure 5.3a**). We note that many of the tropical watersheds host low-income food-deficit countries where local communities are highly dependent on fishery as a primary food source. Indeed, up to 50% of household incomes in countries along the Mekong, Zambezi and Brazilian Amazon depend on fishing (FAO, 2018). Hence, range contractions of freshwater fish species, particularly high in these regions, are likely to have socio-economic repercussions (Allison et al., 2009).

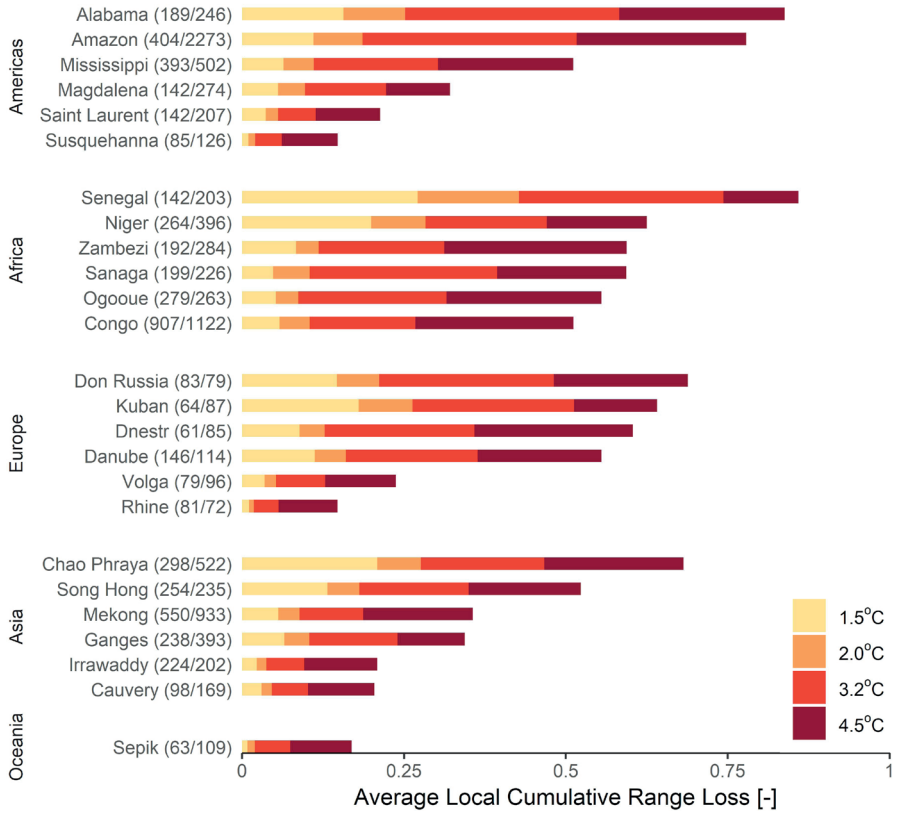


Figure 5.2 | Average local cumulative range loss at different warming levels for 25 large watersheds. For each continent defined according to the World Bank Development Indicators (www.worldbank.org), we selected the six watersheds with the largest numbers of species and covering at least 50% of the known species richness (according to Tedesco et al., (2017)). We made an exception for the Amazon, where our data covered less than 50% of the known species but still a considerable number (404 species). A more exhaustive overview of 200 watersheds is available in **Figure C.10**. Numbers in brackets represent the ratio of the number of species covered in our study over the total number of species known to occur in the watershed according to Tedesco et al., (2017). Our numbers may exceed those of Tedesco due to differences in sampling and catchment delineation.

5.2.2 Drivers of geographical range contractions

Our results show that the projected increase in water temperature has a much larger impact on freshwater fish species than more severe low flow conditions (**Figure C.3**). Further, we found only limited spatial overlap of contractions due to low flow and high water temperature (**Figure C.3**), reflecting the uneven spatial distribution of the two stressors (**Figure C.4**, **I-III**). Indeed, climate change results in reduced flows (including

low flows) only in specific regions, e.g., the Mediterranean, Sub-Saharan Africa, parts of the US, South America and Australia (Wiel et al., 2019), while water temperature will rise almost everywhere (**Figure C.4-III**) (Wanders et al., 2019). Changes in water temperature are expected to become the dominant stressor for salmonid populations (Isaak et al., 2012). In the face of climate change, there is a need to strengthen the understanding of the effects of water temperature changes also on other freshwater fish species.

5.2.3 Comparison across species groups and traits

According to our results, species with smaller initial occurrence range (up to 10^5 km^2) and smaller body size (up to 30 cm of maximum body length) are the most threatened by climate change (**Figure 5.3b-c**). This is in line with previous studies highlighting that smaller species are at higher risk of habitat degradation, in contrast to larger fishes that are mostly threatened by direct exploitation (Olden et al., 2007; Ripple et al., 2017). The largest geographical range contractions were projected for species in lakes (i.e., lentic species; **Figure 5.3d**). At 1.5ºC warming, the proportion of species losing more than half of their habitat was ~47% for lentic species as opposed to only ~6% for lotic species (including species found in both flowing and stagnant water bodies; **Figure 5.3d** and **Figure C.5**). It should be noted, however, that our range loss estimates for species living in lakes with large bathymetric ranges are likely to be overestimated. In fact, since our hydrological model provides only average water temperature estimates in lakes (Wanders et al., 2019), we could not account for microclimatic refugia offered by water stratification (Morrissey-McCaffrey et al., 2018). Further, species that spend part of their lives in the marine environment were predicted to have the smallest range reductions (that is, in the freshwater environment, thus without considering potential altered habitat conditions in the ocean; Figure 5.3d). A similar pattern was observed when looking at the migratory behavior for a subset of species (**Figure 5.3e**). We found species migrating between the freshwater and marine environment (i.e., diadromous species) to be the most resilient to climate change and non-migratory the most affected (**Figure 5.3e**). In line with this, relatively small range contractions were found for orders mostly comprising diadromous species, such as Mugiliformes (mullets), Osmeriformes (smelts), Syngnathiformes (e.g. pipefish), Tetraodontiformes and Pleuronectiformes (**Figure 5.3i**). In contrast, commercial categories, trophic groups and current IUCN threat categories did not serve as a good predictor for future threats posed by climate change (**Figure 5.3f-h**). Species currently belonging to a low threat category (e.g. "near threatened" or "least concern") might become threatened in the future (**Figure 5.3f** and **Figure C.6**). Furthermore, we notice that species currently marked as "data deficient" (~20% of the species analyzed in our study), will on average have larger range contractions than "least concern" or "near threatened" species.

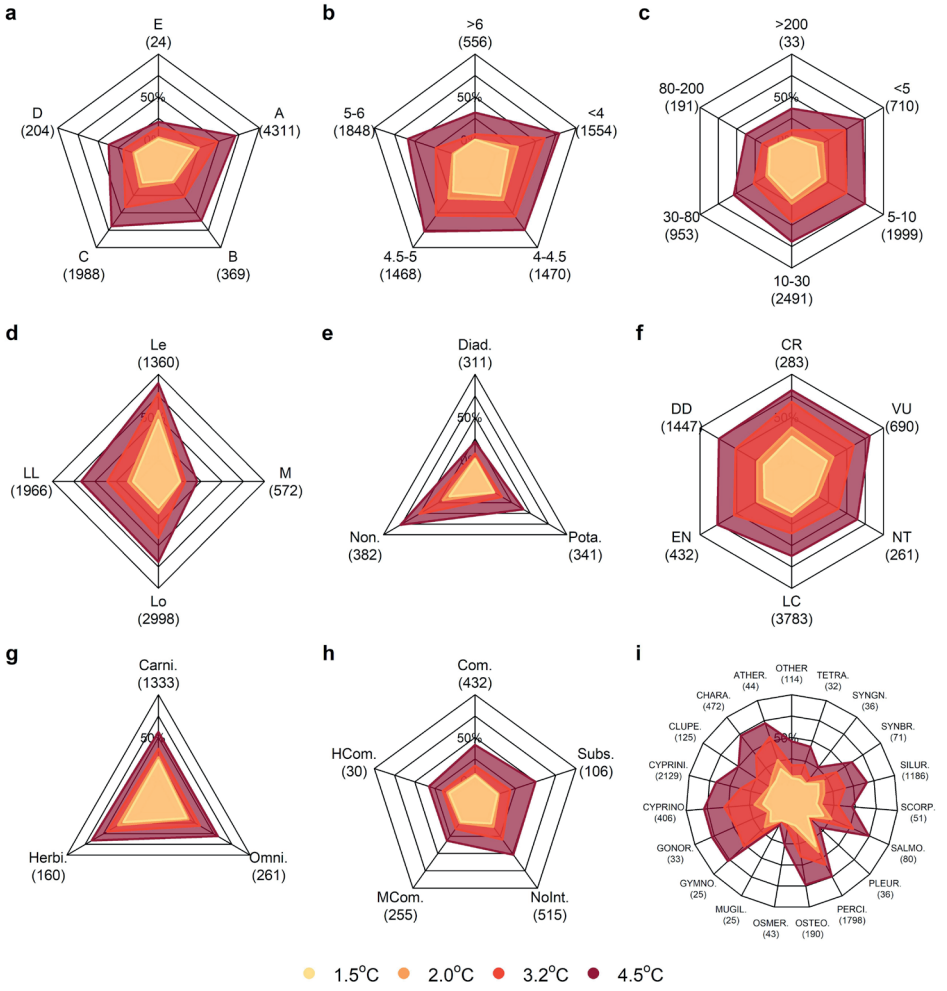


Figure 5.3 | Implications of different warming levels, expressed as the percentage of species losing more than 50% of their range, for various species groups. Köppen-Geiger climate zones (a), where A=Equatorial, B=Arid, C=Warm temperate, D=Snow, E=Polar. Initial geographical range area (b), in log (base 10)-transformed km^2, and body length (c) in cm. Habitat type (d), Le=Lentic, Lo=Lotic, LL=Lentic and lotic, M=Marine and freshwater. Three main migratory behaviors (e), diadromous (migrate to and from the marine environment), potamodromous (migrate within freshwater) and non-migratory. IUCN threat status (f), CR=Critically endangered, VU=Vulnerable, NT=Near Threatened, LC=Least Concern, EN=Endangered, DD=Data Deficient. Trophic level (g), categorized in Carnivorous, Omnivorous and Herbivorous species. Commercial relevance (h), commercial, "highly" and "minor" commercial, of no interest, of potential interest and "subsistence fisheries"(Froese and Pauly, 2018). Species order (i), where groups with less than 20 species represented in our database are grouped together in the "other" category for simplicity of representation (the full list of order names is provided in **Table C.2**). The outer line represents 100% and each inner line is drawn with a 25% interval.

5.3 Discussion

This study represents the first comprehensive quantification of geographical range contractions for freshwater fish, covering the entire globe and about half of the known freshwater fish species diversity. We showed that climate change impacts on freshwater fish species are expected to be particularly high in tropical areas, retaining ~60% of the fish species assessed in this study. Our results further highlight that limiting global warming will substantially decrease potential impacts.

The projected range changes presented in our study are worst-case estimates in the sense that we only account for range contractions. Projecting range expansion was considered unfeasible due to the many barriers (e.g., dams, weirs, culverts, sluices) and uncertainties and data gaps associated with the barriers' current and future locations (McManamay et al., 2018). Additionally, data needed to reliably estimate dispersal ability is still lacking for the majority of the species (Radinger et al., 2017). For terrestrial vertebrates, percentages of species losing more than half of their geographical ranges were reduced by 33 to 23% when dispersal was accounted for (Warren et al., 2018). Nonetheless, we observe that the most impacted species reported in our study are small-bodied species (**Figure 5.3e**), typically having weaker dispersal abilities (Radinger et al., 2017). Further, our results might overestimate range contractions if current geographical range boundaries are not primarily determined by climate-related factors, but also influenced by biogeographic dispersal barriers or anthropogenic pressures (Faurby and Araújo, 2018). This is an inherent caveat of bioclimatic envelope model projections in general, highlighting that these should be considered as baseline estimates of the magnitude and the distribution of climate-induced effects on biodiversity rather than predictions of the future distributions of individual species (Lawler et al., 2009).

The spatial patterns found in this study are contingent on the availability of species occurrence data (**Figure C.7**). For some basins in northern Asia, South America, and Australia, the fish species range maps available covered less than 20% of the known occurring species (Tedesco et al., 2017) (**Figure C.8**). Additionally, we recognize that a given increase in global mean temperature may lead to locally different range contractions depending on the GCM-RCP combination, as each is characterized by specific distributions of changes in water temperature and flow (Seneviratne et al., 2018). We notice a greater variability across GCMs than RCPs when looking at species-specific range contractions (**Figure C.2**), similar to previous findings for hydro-climatic variables (Greve et al., 2018). However, variability across the GCMs did not affect the species-specific thresholds, which were consistent across the models (**Figure C.9**).

Although reported range contractions in this study represent worst-case scenarios, our comparative analysis across the different warming levels and targets clearly showed an increase in impacts with increasing global warming. Limiting global warming to 2°C will reduce the percentage of freshwater fish species potentially losing more than half of their ranges by 55% compared to current pledges of governments (3.2°C). Restricting the global mean temperature rise to 1.5°C will lower the percentage by an additional 15% (or 35% compared to 2°C). While we acknowledge that the ecological realism of our model projections can be improved, these first comparative baseline estimates highlight the need to intensify (inter)national commitments to limit global warming if potentially severe disruption of freshwater biodiversity is to be prevented.

5.4 Methods

5.4.1 Species occurrence data

We used the IUCN Red List of Threatened Species occurrence range data to map the habitat of each freshwater fish species (IUCN, 2018). The IUCN database provides spatial occurrence data for 7,242 freshwater fish species (~56% of freshwater fish species (Nelson, 2006)), compiled from literature and expert knowledge. This database represents the best available spatial data on freshwater fish species geographical ranges (McManamay et al., 2018). We referenced the occurrence data of each species to the 5 arc-minutes (~10 km) hydrography used in this study by selecting grid cells with the majority of the area falling within the occurrence range of the species. Out of the initial 7,242 freshwater fish species, 318 (~4%) were excluded because the occurrence range was smaller than ~1,000 km^2 (i.e., 10 grid cells), resulting in 6,924 species employed in this study (**Figure C.2**).

5.4.2 Hydrological data

We employed the Global Hydrological Model (GHM) PCR-GLOBWB (Sutanudjaja et al., 2018) with a full dynamical two-way coupling to Dynamical Water temperate model (DynWAT) (Wanders et al., 2019) at 5 arc-minutes spatial resolution (~10 km at the equator), to retrieve weekly streamflow and water temperature worldwide (Sutanudjaja et al., 2018; Wanders et al., 2019). We selected this model combination for it allows a full representation of the hydrological cycle (considering also anthropogenic stressors, e.g. water use). It fully integrates water temperature and calculates the hydrological variables on a high-resolution hydrography. The choice of one hydrological model over an ensemble was motivated by the fact that very few GHMs or Land Surface Models (LSMs) calculate water temperature at the spatial resolution desired for this study (Sutanudjaja et al., 2018; Wanders et al., 2019). The PCR-GLOBWB model setup was

similar to Wanders et al.(Wanders et al., 2019), with the exception that flow and water temperature were aggregated at the weekly scale to capture the fish species' tolerance levels to extreme events (Eaton and Scheller, 1996).

5.4.3 Species-specific thresholds for extreme flow and water temperature

We quantified thresholds for minimum weekly flow, maximum number of zero flow weeks and maximum weekly water temperature for each species based on the present-day distribution of these characteristics within the occurrence range, similarly to (not exhaustive) Eaton and Scheller (1996), Azevedo et al., (2013), Leuven et al., (2010). To this end, we overlaid the species' range maps with the weekly flow and water temperature metrics calculated for each year from the output of the hydrological model and averaged over a long-term 30-years historical period, to conform to the standard for climate analyses (Rogelj et al., 2017) (1976-2005, for each GCM employed in the study). We calculated for each 5 arc-minutes grid cell the long-term minimum weekly flow (Q_{min}, **Equotation 5.1**), the long-term frequency of zero-flow weeks (Q_{zf}, **Equotation 5.2**) and the long-term maximum weekly temperature (Tw_{max}, **Equotation 5.3**), as follows:

$$Q_{min} = \frac{\sum_{i=1}^{N} \min(\mathrm{Q7}_i)}{N} \tag{5.1}$$

$$Q_{zf} = \frac{\sum_{i=1}^{N} \{j \in \{1, \dots, M\} : q7_j = 0\}_i}{N} \tag{5.2}$$

$$Tw_{max} = \frac{\sum_{i=1}^{N} \max(\mathrm{Tw7}_i)}{N} \tag{5.3}$$

where Q7 and Tw7 are the vectors of weekly streamflow and water temperature values for a given year *i*, respectively; q7 is the streamflow value for the week *j*; *N* is the number of years considered for the long-term (30 in this case) and *M* is the number of weeks in a year (~52). We then used the spatial distributions of these values within the range of each species to determine species-specific 'thresholds' for each of the three variables, defined as the 2.5 percentile of the minimum flow and the 97.5 percentile of the maximum water temperature and zero flow weeks values. We preferred these to using the absolute minimum and maximum values to reduce the influence of uncertainties and outliers in the threshold definition. An overview of the thresholds' distribution is available in **Figure C.5**.

5.4.4 Climate forcing and warming targets

We considered four main future scenarios based on increases of global mean air temperature equal to 1.5, 2.0, 3.2 and 4.5ºC. The global mean temperature increase refers to a 30-years average, in accordance with guidelines for climate analyses (Rogelj et al., 2017), and with pre-industrial reference set at 1850-1900 (Seneviratne et al., 2018). To obtain estimates of weekly water temperature and flow for each warming level, we forced the hydrological model with the output from an ensemble of five Global Climate Models (GCMs), each run for four Representative Concentration Pathway (RCP) scenarios, namely RCP 2.6, 4.5, 6.0 and 8.5 (see Supplementary Methods in Appendix C for details). Hence, each RCP-GCM combination would reach each warming level at a different point in time, with some of the RCP-GCM combinations not reaching certain warming levels. Consequently, the number of scenarios available differed among warming levels (an overview is provided in **Table C.1**). In total we modelled 42 scenarios (one scenario = one GCM-RCP combination at a certain point in the future), including 17 scenarios for 1.5ºC, 15 for 2.0ºC, 7 for 3.2ºC and 3 for 4.5ºC.

5.4.5 Patterns at the species level

We assessed the percentage of species losing more than half of their geographical range at each warming level, following Warren et al., (2018). To that end, we first calculated the percentage of species losing >50% of their range for each GCM-RCP scenario ensemble. Then, we calculated the mean and standard deviation at each warming level. In addition, we looked at the percentage of species losing more than half of their range according to different species traits and categories. We overlaid each species occurrence range with the historic Köppen-Geiger climate categories to obtain the main climate zone per species (i.e., capital letter of the climate classification) (Kottek et al., 2006). Species falling into multiple climate categories were assigned the climate zone with the largest overlap. We retrieved information on threat status and species order directly from the IUCN metadata (IUCN, 2018). We also gathered a list of potential habitats for each species from the same IUCN database. We classified species as lotic if they were associated with habitats containing at least one of the words “river”, “stream”, “creek”, “canal”, “channel”, “delta”, “estuaries”, and as lentic if the habitat descriptions contained at least one of the words “lake”, ”pool”, ”bog”, ”swamp”, ”pond”. Further, we retrieved data from Fishbase (www. fishbase.org) on migratory behavior, maximum body length and commercial importance (Froese and Pauly, 2018). From the same database we also retrieved trophic level values and aggregated them into Carnivore (trophic level >2.79), Omnivore (2.19 < trophic level ≤ 2.79) and Herbivore (trophic level ≤ 2.19) (Froese and Pauly, 2018). Prior to the comparison with Fishbase, we performed a synonym check for

the binomial nomenclature provided by in the IUCN database to maximize the overlap with the Fishbase database.

5.4.6 Spatial patterns of geographical range contractions

We stacked together the range contractions for the individual species to study spatial trends. To this end, we calculated the local cumulative range loss (LCRL) at each 5 arc-minutes (~10 km) grid-cell, globally. We calculated the LCRL as the median of the relative losses across the GCM-RCP scenarios, as follows:

$$LCRL_{i,w} = median\left(\frac{SN_{future,i,w,s} - SN_{present,i}}{SN_{present,i}}\right) \quad (5.4)$$

where *SN* represents the number of species with their range in grid-cell *i*, for the scenario *s* within the warming target *w*. We used the median rather than the mean because the data showed skewed distributions across the GCM-RCP combinations. In addition, we aggregated the LCRL for each watershed with an outlet to the ocean/sea or internal sink (e.g. lake), as follows:

$$LCRL_{x,w} = \frac{\sum_{i=1}^{I} LCRL_i}{I} \quad (5.5)$$

where *I* represents the number of grid cells within the watershed *x*.

5.5 ACKNOWLEDGEMENTS

This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. This project has received funding from the Europeans Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 641459. AMS was supported by the GLOBIO project (www.globio.info).

Synthesis 6

6.1 Introduction

The central aim of this thesis was to advance integrated global assessments of freshwater systems by 1) improving the accuracy and spatial detail of global hydrological data and 2) developing models to assess the impacts of dams and future climate change on the geographic ranges of freshwater fish species. Chapters 2 and 3 presented data-driven models to estimate streamflow at high spatial resolution. Impacts on freshwater fish geographical ranges due to dam-driven fragmentation and climate change were quantified in Chapters 4 and 5, respectively. This final chapter synthesizes the findings by evaluating the benefits, limitations and trade-offs of data-driven streamflow models and opportunities for further improvements (section 6.2). Next, section 6.3 discusses the added value of high-resolution species-specific routines to improve freshwater biodiversity assessments, evaluates combined impacts of damming and climate change on the fish species ranges, and lists a number of opportunities for future improvements of the species-based models. Finally, section 6.4 provides overall conclusions and recommendations.

6.2 Improving the spatial detail of global streamflow estimates with data-driven models

6.2.1 Pros and cons of data-driven hydrological models

An improved spatial resolution of hydrological data benefits various global water resource assessments needs (e.g., irrigation supply assessments, water footprinting) as well as the modeling of chemical fate, sediments transport and freshwater biodiversity (Bakker, 2012; Grill et al., 2019, 2016; Hanafiah et al., 2011; Hoekstra et al., 2011; Oberdorff et al., 2011; Poff and Zimmerman, 2010; Syvitski, 2005; Tendall et al., 2014). Whereas data availability and computational costs are a constraint for process-based hydrological models (Bierkens et al., 2015), data-driven models are a promising alternative to reach very high spatial resolution (≤ 1 km) while maintaining a global coverage. Chapters 2 and 3 of this thesis applied different data-driven approaches to quantify streamflow variables worldwide at high spatial resolution. The correlative model based on multiple linear regression (Chapter 2) yielded estimates of long-term average streamflow that were in good agreement with observed data. Further, streamflow estimates were relatively more accurate than those from a GHM, similarly to previous studies that compared runoff estimates from data-driven models with the output of GHMs/LSMs (Beck et al., 2016a, 2015). When estimating mean annual flow on a yearly basis, the regression model performance was diminished but it still outperformed the GHM. In turn, the machine-learning approach applied in Chapter 3 outperformed the regression

model, with R^2 values of ~0.91 even for yearly streamflow estimates (compared to an R^2 of 0.89 for long-term annual mean streamflow reported in Chapter 2). In addition, the more complex model structure and parameterization of artificial neural networks used to produce FLO1K, and the inclusion of additional climatological predictors (e.g. seasonality, aridity index) allowed the calculation of more yearly streamflow metrics, namely maximum and minimum annual streamflow (Chapter 3). The spatial resolution of FLO1K is ~10 to ~50 times higher than the output of global process-based hydrological models, allowing the quantification of streamflow for small reaches and headwaters (Chapter 3). Small upstream reaches provide unique habitat and refuges for riparian and aquatic biota and govern the overall connectivity of the watershed by connecting upstream sources to downstream flows (Wohl, 2017). Indeed, they are highly relevant for sediments, nutrients and chemicals transport, which are highly dependent upon upstream retention and release (Alexander et al., 2007; MacDonald and Coe, 2007; McClain and Naiman, 2008; Wohl, 2017). Therefore, modelling of such physical and biological processes can highly benefit from a high-resolution hydrology. An example of this is the application of FLO1K to develop a model to quantify exposure to pharmaceuticals in European surface waters, ePiE (Oldenkamp et al., 2018). The high spatial resolution of FLO1K was an essential prerequisite to correctly allocate the point sources of pharmaceuticals along the stream network and headwaters, and to route the compounds concentrations downstream the river network by considering average but also high and low flow conditions (Oldenkamp et al., 2018).

Nevertheless, the temporal resolution achieved by data-driven models at a global scale remains low compared to global process-based models (Beck et al., 2016, 2013; Razavi and Coulibaly, 2013; Schewe et al., 2014). Whereas GHMs/LSMs can achieve streamflow estimates at daily and sub-daily time steps (Schellekens et al., 2017), FLO1K is restricted to annual mean, maximum and minimum streamflow metrics (Chapter 3). The main factor limiting the temporal resolution of data-driven models is the relatively simple correlation-based structure as opposed to the mechanistic approach of GHMs/LSMs, which can account for a more holistic representation of hydrological processes and interactions therein. Further, even if more complete and detailed daily time series would be available for training machine-learning type of models, the amount of input data necessary to describe processes at the daily scale would likely decrease the computational performance advantage over mechanistic models. Therefore, mechanistic models such as GHMs/LSMs remain the preferred choice when applications require streamflow estimates with a high temporal resolution. Hybrid models constitute a recent development to obtaining hydrological variables at both high spatial and high temporal resolution by leveraging data-driven high-resolution estimates of hydrological variables to bias-correct and downscale the output of GHMs/LSMs (Lin et al., 2019). Such

approaches are very promising but still in their infancy. In any case, hybrid approaches still impose the computational burden of GHMs/LSMs, which have to be run prior to a downscaling routine applied by the hybrid approach.

Simple linear regression models remain a useful tool for generating or testing hypotheses. In Chapter 2, I considered a simple multiple regression framework to test variables' importance and contribution to the explained variance in estimating streamflow. The information of variables' importance and ranking derived from the results of Chapter 2, was then used to build the more complex ANNs-based FLO1K model (Chapter 3). Differently from linear regression, ANNs do not give any useful insights concerning variable importance and coefficients, as that information cannot be extracted from the complex constructs of ANNs (Haykin, 1994).

The potential for extrapolation and projection of the correlative streamflow models developed in this thesis remains to be tested. Therefore, GHMs/LSMs continue to be the preferred choice to derive specific hydrological variables for scenario analyses. Yet, when a small set of hydrological variables is needed at relatively low temporal resolutions (e.g., yearly) and for a relatively far future (e.g., the year 2100), running a mechanistic model might be inefficient from a computational point of view, as the model needs to run through every day up till the aimed point in time. In addition, the resulting spatial resolution achieved for future scenarios would still be bound to the GHM/LSM resolution. As discussed above, the spatial resolution limitation could be overcome by the employment of hybrid models, but so far these have only been applied for historical natural conditions (Lin et al., 2019), and their applicability for future scenarios remains to be tested.

In conclusion, different types of global models to estimate streamflow involve tradeoffs between spatial and temporal resolution as well as between complexity and computational costs (**Figure 6.1**). Correlative models are preferable when the aim is to achieve high spatial resolution. In contrast, daily or sub-daily time series can only be achieved by mechanistic models, yet at the expense of a lower spatial resolution. Hybrid models might offer a solution to achieve both high temporal and spatial resolution, but at high computational costs, as they involve both a process-based model and a downscaling routine (Lin et al., 2019). As of now, process-based models are the only tool available for obtaining streamflow data for scenario analysis as the effectiveness of regression approaches is yet to be tested for this type of application.

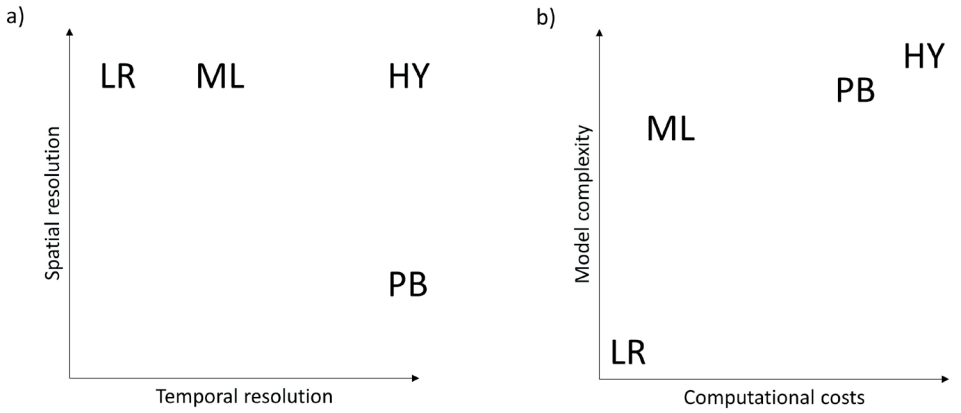


Figure 6.1 | Schematic representation of tradeoffs among different types of streamflow models with respect to a) spatial and temporal resolution and b) model complexity and computational costs. LR: linear regression, ML: machine learning, PB: process-based and HY: hybrid.

6.2.3 Outlook

Data-driven hydrological/streamflow models require observational data for training and testing. The spatial distribution of gauging stations with streamflow time series was highly scattered across the globe and biased towards western countries at the time of the development of the models presented in this thesis (Chapters 2 and 3). The recent release of a new streamflow monitoring dataset (Do et al., 2018; Gudmundsson et al., 2018), including about four times more monitoring stations than the GRDC data (GRDC, 2017) and with a sub-weekly temporal resolution, unlocks new opportunities to refine and validate global data-driven models, such as FLO1K. Higher temporal resolutions may be obtained if more complete and less spatially biased streamflow time series become available, by accounting for the time lag in rainfall-runoff processes and including human water use. Additionally, developments in remote sensing are expected to further boost data availability for covariates used in data-driven models. An example is the development of land cover time series (https://www.esa-landcover-cci.org/) or of the Global River Width from Landsat (GRWL) database (Allen and Pavelsky, 2018). Furthermore, the launch of the NASA Surface Water and Ocean Topography (SWOT) mission in the coming years will provide highly detailed altimetry data of surface water bodies that may further advance hydrological data-driven models (https://swot.jpl.nasa.gov/). This increased availability of data will make it possible to improve the accuracy of or develop new data-driven models of hydrological variables at a global scale. For instance, efforts have already been carried out in the past to model runoff (Beck et al., 2016b), groundwater (Fan et al., 2013), sediment fluxes (Syvitski et al., 2003;

Syvitski, 2005) or water temperature (Krider et al., 2013; Piccolroaz et al., 2016; Rabi et al., 2015; Toffolon and Piccolroaz, 2015) by means of regression-based models. Future studies may investigate how to develop or improve such models based on new remote-sensing data.

Data-driven streamflow models as presented in this thesis (Chapter 2 and 3) have potential for forecasting streamflow under changing future climatic conditions (Peel and Blöschl, 2011). Yet, additional uncertainty introduced by potential non-stationarity of model parameters under transient climatic conditions should be further investigated to better understand the reliability of such projections (Blöschl and Montanari, 2012; Lima and Lall, 2010). In addition, the promising hybrid approach that combines the output of mechanistic models with data-driven estimates to downscale streamflow to high spatial resolution (Lin et al., 2019) should also be explored, especially to apply such an approach to obtain high spatial and temporal resolution streamflow forecasts for scenario analyses.

6.3 Assessing the impacts of dams and climate change on freshwater fish biodiversity

6.3.1 A global species-based freshwater biodiversity modelling framework

Large-scale assessments of freshwater biodiversity were limited to coarse-grain assemblage-level estimates (Liermann et al., 2012; Winemiller et al., 2016; Xenopoulos et al., 2005). Chapters 4 and 5 started filling this gap by assessing the impacts of two key threats to freshwater fish, namely damming and climate change (Reid et al., 2019), on the geographic ranges of individual species. A species-based modelling approach was developed to leverage the availability of detailed information on the current geographical range of freshwater fish species (IUCN, 2018). This allowed both studies to quantify changes at the species level, detect inter-species differences in impacts and hence identify species groups most at risk, which in turn may help prioritizing conservation efforts. The spatial detail of the underlying hydrology used in both studies constitutes a considerable improvement compared to previous assemblage-level global freshwater biodiversity assessments, which typically employed large spatial units (e.g., freshwater ecoregions) or coarse spatial grids (e.g., 55x55 km^2) (Liermann et al., 2012; Xenopoulos et al., 2005). The species-specific connectivity index values (Chapter 4) were calculated at the level of sub-basins (~100 km^2), based on the 15 arc seconds (~500 m) HydroSHEDS hydrography (Lehner et al., 2008; Lehner and Grill, 2013). Chapter 5 employed the 5 arcminutes (~10 km) version of the GHM PCR-GLOBWB (Sutanudjaja et al., 2018; Wanders et al., 2019) to project water flow and temperature estimates driven by a combination of representative

concentration pathway scenarios and global climate models. A process-based rather than data-driven hydrological model was necessary to include water temperature and obtain the necessary level of temporal detail (Section 6.2.1).

Combining the results of the fragmentation and climate change impacts assessments highlights that a considerable number of species is threatened by both projected dam infrastructure building and climate-driven alteration in water temperature and flow. Specifically, 3% (n=96) of tropical and 2% (n=36) of warm temperate species may experience increases in impacts > 50% from both stressors (**Figure 6.2**). At 3.2ºC warming, projected impacts of climate change were much higher than impacts of ongoing and future dam building (**Figure 6.2**). Yet, at lower warming levels the difference between fragmentation and climate change impacts leveled, and at 1.5ºC warming the projected impacts of dams exceeded the climate change effects (**Figure D.1-3**). The disproportional combined impact of dam building and climate change in tropical regions, as shown in Figure 6.1, complements the findings of both Chapters 4 and 5, which showed that tropical species are expected to be mostly at risk from future fragmentation and climate change impacts. In the Amazon, Congo, Niger, Chao Phraya, Salween and Mekong basins, for example, large impacts were found for both stressors (Chapter 4 and 5). Therefore, species in these areas may experience disproportional impacts of future environmental change, in agreement with concerns expressed by previous qualitative surveys (Winemiller et al., 2016). Furthermore, both Chapters 4 and 5 highlighted that species that complete their life cycle in freshwater, which constitute the majority of the species, are more threatened than species that migrate between fresh water and the ocean (**Figure 4.1** and **5.3**). For instance, Chapter 4 reported a substantial difference in potential habitat fragmentation impacts between diadromous and non-diadromous species (current Connectivity Index = 82% for diadromous versus 69% for non-diadromous).

It should be noted that potential interactions between changes in flow/temperature and future dams were not included in the climate change projections (Chapter 5), as the location and number of dams was kept constant in PCR-GLOBWB. This can lead to potential overestimation of impacts in sub-arid or Mediterranean regions, where depletion of low flows is the leading driver of habitat loss (Chapter 5). Indeed, dams constructed in these regions can arguably contrast the effect of climate change by increasing downstream flows during the dry season (Pokhrel et al., 2018, 2012). Yet, Chapter 5 highlighted that water temperature is the main driver of climate-change induced fish habitat loss, indicating that potential overestimations of impacts are likely limited to small areas where increased droughts are the dominant driver of habitat loss.

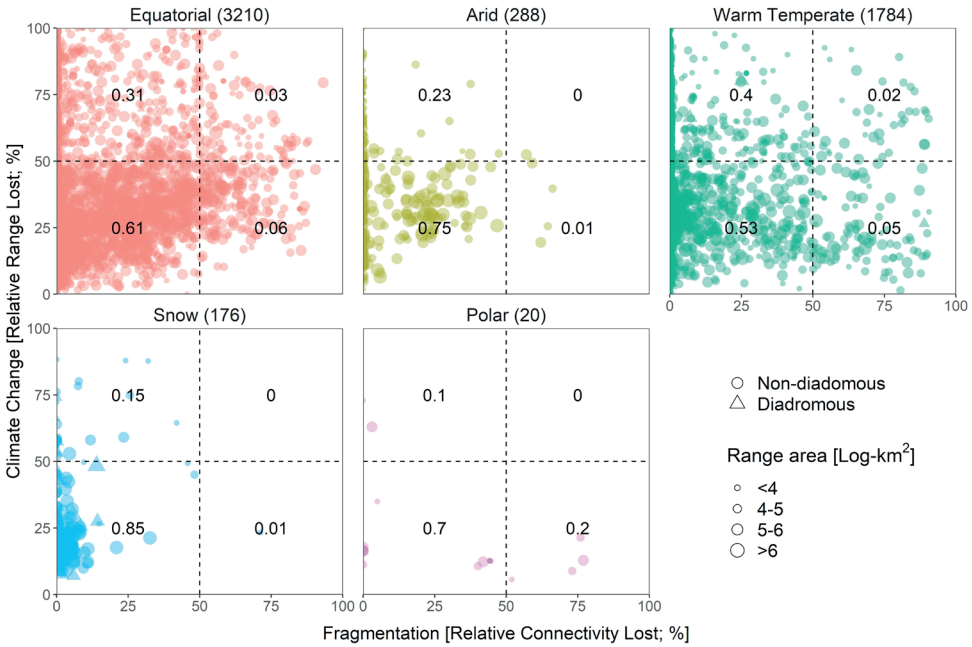


Figure 6.2 | Impacts from climate change (at 3.2ºC global warming) and fragmentation on the geographical ranges of 5,450 freshwater fish species per Köppen-Geiger climate zone (with number of species per zone in brackets). Each plotted point represents a species, whereby point size is proportional to the original range size. Numbers in the quadrants represent the proportions of species. Relative range/connectivity losses were calculated as the difference between current and future range area/ connectivity index divided by the current range area/connectivity index. Climate category was assigned based on the overlap of the range of each species with the Köppen-Geiger climate zones. Figures for different warming levels are shown in Appendix D.

6.3.3 Outlook

The studies presented in Chapters 4 and 5 constitute a first step in species-based assessments of the impact of human pressures on freshwater fish species. A need remains to express both fragmentation (Chapter 4) and geographical range contractions (Chapter 5) in the same impact metric to integrate/compare them. A solution would be to translate the loss in connectivity as presented in Chapter 4 to spatially resolved geographical range contractions. For instance, Carvajal-Quintero et al. (2017) related the range areas of fish species occurring in the Magdalena basin (South America) to their body size, to translate loss of longitudinal connectivity to loss of suitable habitat. To illustrate, Carvajal-Quintero et al. (2017) fitted a linear regression to the lower quantile of the species range area-body size relationship and used it to determine patches of the range that would become unsuitable due to dam construction. While I could reproduce

the range-body size relationship of this study for the Magdalena basin, I did not find significant relationships for the majority of other hydrologic basins (**Figure D.4**). This suggests that more refined relationships are needed, for instance bases on a larger set of environmental or physiological variables.

A number of aspects relevant to quantify species responses to human pressures are not yet included in the species-based modelling frameworks presented in Chapters 4 and 5. Dispersal of fish species in response to climate change was not considered as the literature is currently lacking models that allow quantifying dispersal ability across large numbers of fish species (McManamay et al., 2018; Radinger et al., 2017; Radinger and Wolter, 2015, 2014). Therefore, future studies should focus on developing strategies to incorporate dispersal, e.g., based on phylogenetic similarity (Comte and Olden, 2017). Further, future work is needed to more accurately assess and validate the species-specific flow and water temperature tolerance thresholds that were inferred from species ranges in Chapter 5. While the water temperature thresholds compared reasonably well with laboratory data (**Figure C.1**), the comparison was done for only ~200 species for which data was readily available. Thus, more comprehensive databases on fish environmental thresholds would be highly beneficial to more thoroughly validate the underlying assumptions of the model and findings presented in Chapter 5.

The inclusion of ~40,000 dams in Chapter 4 constitutes a large improvement compared to previous connectivity studies (Liermann et al., 2012), but this number is still far from representing the actual number of existing small and mid-size man-made barriers (McCully, 2001). For instance, it has been estimated that there are over 800,000 dams across the world (Wisser et al., 2010). There is a need to improve existing dam databases, for example through the development of new techniques to leverage remote-sensing data to detect dams using supervised machine learning techniques. Additionally, the analysis presented in Chapter 4 is based on the assumption that all dams are impassable. While fish ladders are indeed ineffective for most species (Birnie-Gauvin et al., 2018; King et al., 2017; Silva et al., 2018), future efforts should focus on integrating information about the presence of more natural or nature-like bypasses within dams databases. Such nature-like bypasses are the most effective solution to partially restore longitudinal connectivity of freshwater waterways, without the need to remove the barrier and compromise socio-economic benefits (Silva et al., 2018). Improving information on dam locations and associated (nature-like) bypasses, combined with improved information on species' dispersal ability, will considerably improve future connectivity assessments.

A number of other important stressors remain to be assessed to complement the freshwater fish impact assessments presented in this thesis (Dudgeon et al., 2006; Reid et al., 2019). For instance, the impacts of nutrients (e.g., harmful algal blooms leading

to depleted oxygen levels) on fish species are increasingly well understood but global species-based assessments are still lacking (Janssen et al., 2019). Other potentially relevant stressors include toxic compounds, nanomaterials and microplastics, and more research is needed to assess their impacts on freshwater biodiversity (Reid et al., 2019). Further, future studies should investigate interaction effects among multiple stressors and among species/food chains, which are not yet implemented in global freshwater biodiversity models (Harfoot et al., 2014a). Finally, a parallel development of more complex species distribution models (SDMs) for fish (Domisch et al., 2015; Labay et al., 2015; Radinger et al., 2017) at a global scale would help to validate the conclusions of the studies presented in this thesis.

6.4 Conclusions

Based on the work conducted in this thesis, the following main conclusions are drawn:

- Computationally efficient and spatially accurate gridded predictions of annual streamflow metrics can be achieved by means of data-driven models at a global scale
- Mechanistic global hydrological models retain a higher level of temporal accuracy and are preferred over data-driven models when daily time series are needed
- Future hydropower development will pose a higher threat to non-diadromous fish species than to diadromous fish species
- Percentages of freshwater fish species losing more than half of their current geographical range were projected be 13% at 1.5°C, 20% at 2°C, 44% at 3.2°C and 69% at 4.5°C of global warming (global temperature raises refer to pre-industrial levels)
- Fish species occurring in tropical and warm temperate basins might be particularly at risk due to the combined effects of climate change and future hydropower development

Recommendations for future work based on the findings of this thesis are:

- Future studies should aim to improve the temporal resolution of data-driven hydrological models by leveraging the increasing availability of remotely sensed covariates
- Future studies should test the applicability of global data-driven hydrological models for scenario projections and/or for downscaling the output of process-based models

- Future research should develop global fish dispersal models to be included in projections of geographical range changes
- More extensive datasets on environmental thresholds of freshwater fish species should be developed to more accurately assess future range changes
- More inclusive global datasets of dams and associated by-passes should be developed to improve freshwater fish impact assessments
- Interactions among multiple stressors and among species/food chains should be further investigated to improve the ecological realism of freshwater fish assessment models

Appendices

APPENDIX A

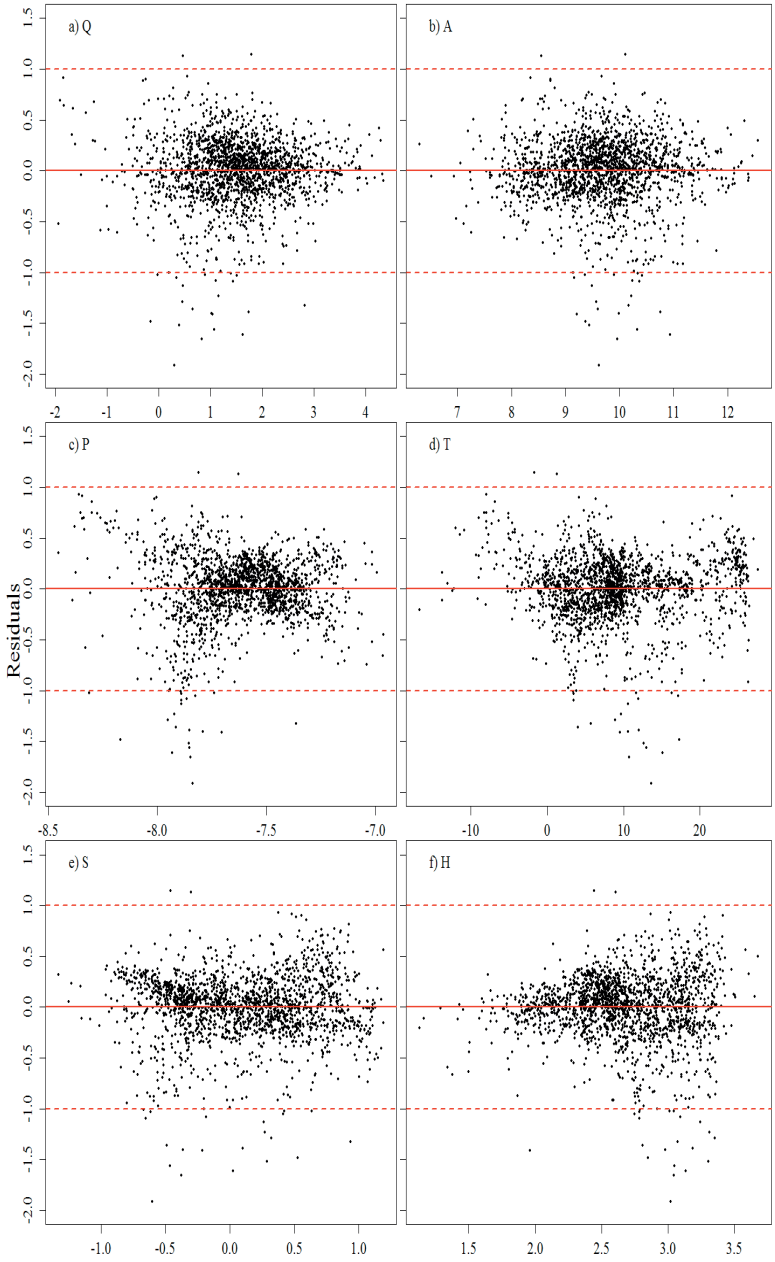


Figure A.1 | Residuals plot of the regression analysis (MAF 1981-2010) vs a) the response variable mean annual flow Q, b) area of the catchment A, c) 30-years mean precipitation P, d) 30-years mean temperature, e) average slope of the catchment and f) average altitude. The x axes are dimensionless as the variables have been log-transformed, except for temperature which is in °C. Dashed red lines delimit the 1 order magnitude deviation.

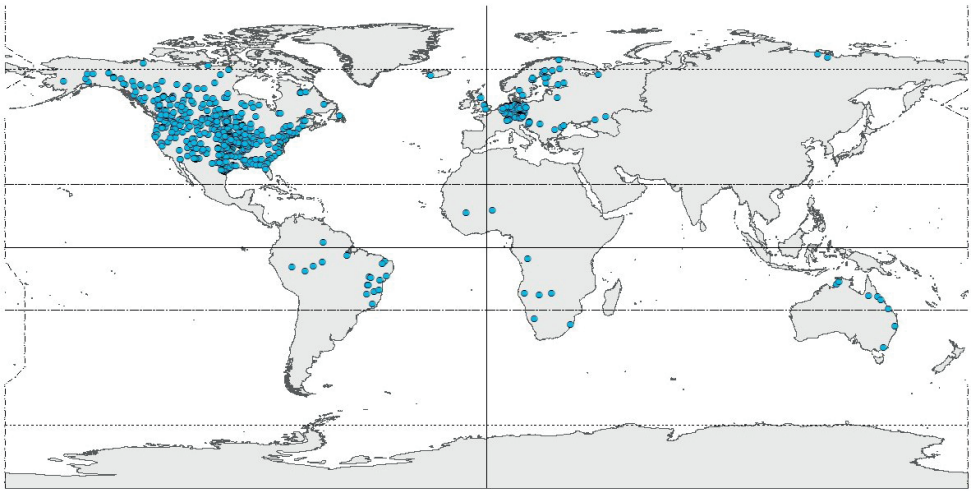


Figure A.2 | Distribution of the 543 GRDC gauging stations used for performance testing of the MAF model in the backcasting analysis.

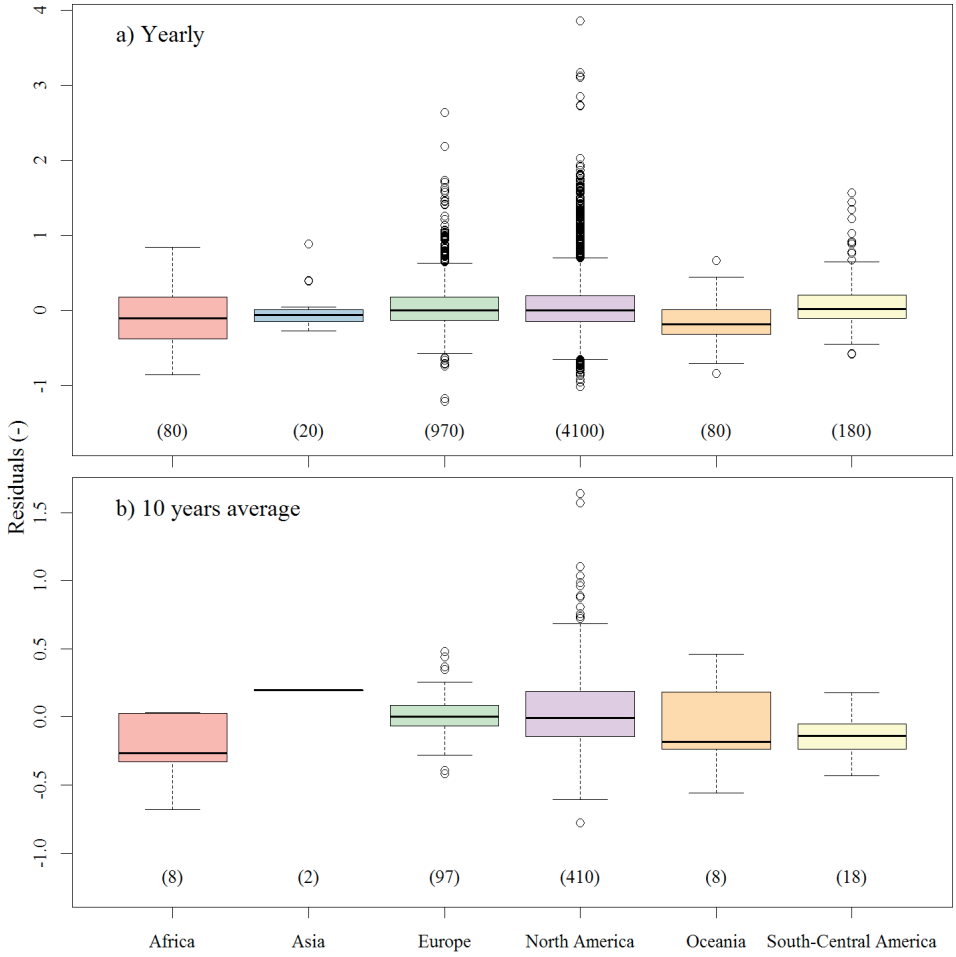


Figure A.3 | Model residuals per continent of the backcasting analysis for a) the yearly MAF and b) the 10-years averaged MAF values. Residuals are dimensionless as MAF is in base-10 log scale. In brackets the number of observations for each continent.

Table A.1 | Regression analysis for increasing number of predictors (MAF 1981-2010) based on $Q = 10^{\beta 0} \cdot A^{\beta A} \cdot P^{\beta P} \cdot 10^{\beta T \cdot T} \cdot H^{\beta H} \cdot S^{\beta S}$. R^2: coefficient of determination; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; Int.: intercept.

N. predictors	Int.	A	P	T	S	H	R^2	AIC	BIC
1	-6.505	0.828	-	-	-	-	0.611	3471	3488
2	6.565	0.933	1.850	-	-	-	0.790	2312	2335
3	10.956	0.957	2.411	-0.039	-	-	0.866	1464	1492
4	10.403	0.965	2.354	-0.037	0.184	-	0.874	1345	1378
5	9.066	1.018	2.070	-0.038	0.464	-0.509	0.893	1049	1088
Standardized coefficients									
1	0.000	0.782	-	-	-	-			
2	0.000	0.881	0.434	-	-	-			
3	0.000	0.903	0.566	-0.305	-	-			
4	0.000	0.911	0.553	-0.284	0.094	-			
5	0.000	0.961	0.486	-0.290	0.237	-0.212			

Table A.2 | Comparison of the coefficients from the spatial error (SE) and OLS regression models. The analysis is based on the subset of observations within 58S-60N latitudes. R^2: coefficient of determination; LM: Lagrange Multiplier test score; m: number of observations.

Variable	SE coeff. (95% CI)	OLS coeff. (95% CI)
λ	0.614 (0.573 -- 0.656)	-
Int.	9.900 (9.245 - 10.556)	10.365 (9.778 - 10.952)
A	0.982 (0.965 - 0.999)	0.999 (0.981 - 1.017)
P	2.184 (2.090 - 2.278)	2.268 (2.186 - 2.351)
T	-0.034 (-0.037 - -0.031)	-0.032 (-0.035 - -0.030)
H	-0.396 (-0.466 - -0.327)	-0.386 (-0.446 - -0.327)
S	0.368 (0.313 - 0.422)	0.396 (0.351 - 0.440)
R^2	0.93	0.90
LM	7	465
m	1,748	1,748

APPENDIX B

B.1 Supplementary Methods

B.1.1 Lotic/lentic species classification

We classified species as lotic, lentic or both lotic and lentic, using metadata from the IUCN Red List (IUCN, 2018). For each species, we retrieved a list of habitat types where the species was known to be found. We classified species as lotic if they were associated with habitats containing at least one of the words "river", "stream", "creek", "canal", "channel", "delta", "estuaries", and as lentic if the habitat descriptions contained at least one of the words "lake", "pool", "bog", "swamp", "pond".

B.1.2 Derivation of the connectivity index equations.

The equations proposed by Cote et al., (2009) allow to calculate the connectivity index for non-diadromous (N) and diadromous (D) fish species, assuming barriers are impassable, as follows:

$$CI_{s,b}^{N} = \frac{\sum_{i=1}^{n} {l_{i,s,b}}^{2}}{(\sum_{i=1}^{n} l_{i,s,b})^{2}} \cdot 100 \quad \text{(B.1)}$$

$$CI_{s,b}^{D} = \frac{l_{1,s,b}}{\sum_{i=1}^{n} l_{i,s,b}} \cdot 100 \quad \text{(B.2)}$$

where $l_{i,s,b}$ represents the length of stream segment i isolated due to a dam for a species s within a main basin b and n is the number of isolated segments due to $n-1$ dams within that basin. Hence, $CI_{s,b}$ expresses the habitat connectivity, with smaller values indicating less connectivity. The equation for diadromous species differs from the one for non-diadromous as the most downstream dam obstructing the passage to/from the marine environment is likely to have the highest impact (Cote et al., 2009). In **Equation B2**, $l_{1,s,b}$ is the length of the longest river segment that is connected to the ocean. While the measures of Cote et al., (2009) can in principle account for different passability of barriers, we assume here that the dams considered in this analysis are impassable.

The IUCN species occurrence locations are not reported per stream segment, but as geographical ranges occupying a portion of the hydrologic basin, while **Equations B1-B2** were developed for river segments. To make these equations applicable to the areal

range data from IUCN, we propose the following conversion between a watershed area and the length of the streams in that area based on Hack's law. According to Hack's law (Hack et al., 1957), $l = \beta a^{\alpha}$ i.e., the length of a stream (l) is proportional to its drainage area (a). Therefore, **Equations B1** and **B2** can be rewritten as:

$$CI^{N}_{s,b} = \frac{\sum_{i=1}^{n}(\sum_{j=1}^{m} \beta {a_{j,i,s,b}}^{\alpha})^2}{(\sum_{i=1}^{n}\sum_{j=1}^{m} \beta {a_{j,i,s,b}}^{\alpha})^2} \cdot 100 \quad \text{(B.3a)}$$

$$CI^{D}_{s,b} = \frac{\sum_{j=1}^{m} \beta {a_{j,1,s,b}}^{\alpha}}{\sum_{i=1}^{n}\sum_{j=1}^{m} \beta {a_{j,i,s,b}}^{\alpha}} \cdot 100 \quad \text{(B.4a)}$$

where, m are the watershed areas from j *to* m within the isolated patch i where the species s occurs. **Equations B3a** and **B4a** can in turn be rewritten as:

$$CI^{N}_{s,b} = \frac{\beta^2 \sum_{i=1}^{n}(\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha})^2}{\beta^2(\sum_{i=1}^{n}\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha})^2} \cdot 100 \quad \text{(B.3b)}$$

$$CI^{D}_{s,b} = \frac{\beta \sum_{j=1}^{m} {a_{j,1,s,b}}^{\alpha}}{\beta \sum_{i=1}^{n}\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha}} \cdot 100 \quad \text{(B.4b)}$$

Therefore, β can be eliminated and the final equations are as follows:

$$CI^{N}_{s,b} = \frac{\sum_{i=1}^{n}(\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha})^2}{(\sum_{i=1}^{n}\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha})^2} \cdot 100 \quad \text{(B.3c)}$$

$$CI^{D}_{s,b} = \frac{\sum_{j=1}^{m} {a_{j,1,s,b}}^{\alpha}}{\sum_{i=1}^{n}\sum_{j=1}^{m} {a_{j,i,s,b}}^{\alpha}} \cdot 100 \quad \text{(B.4c)}$$

We used the HydroBASINS (henceforth abbreviated as HB) watershed units (Pfafstetter level 12) to determine the watershed areas a (Lehner et al., 2008; Lehner and Grill, 2013). We allocated the IUCN geographical ranges of each species to the overlapping HB watersheds so that each watershed was assigned a list of species for which it provides habitat. In turn, we identified all HB watersheds that provide habitat for each species. HB divides the globe in 1,034,083 watersheds (area median = 135 km^2, interquartile range = 64 km^2) following the Pfafstetter coding scheme (Lehner and Grill, 2013) and based on the high-resolution 15 arc-seconds (~500m) HydroSHEDS hydrography (Lehner et al., 2008). We employed HB as most of the freshwater fish IUCN ranges are established

based on HB and with Pfaffstetter level 12 we used the highest level of spatial definition, i.e., the smallest watershed units. Each of the watersheds in HB carries information on the connectivity to the next downstream watershed, which allows to determine the total connected area within a basin. Dams falling within a HB unit were georeferenced to the downstream boundary of that HB unit so that isolated patches were a collection of HB watershed units. **Figure B.1** shows an example application of **Equations B.4a** and **B.4c**.

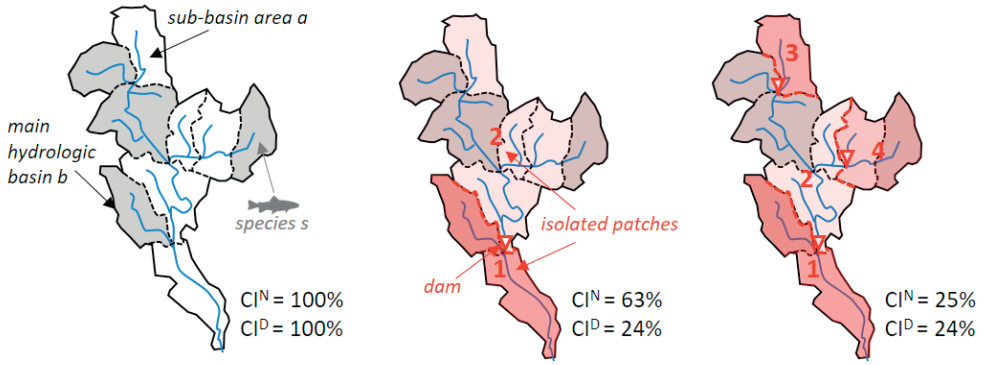


Figure B.1 | CI calculation for a hypothetical species *s* (occupying the gray areas) in a fictitious basin *b* (external solid line) partitioned in HB watersheds (internal boundary dashed lines). The addition of dams fragments the basin in isolated patches (red hues with numbers). For each configuration, the CI is given for species *s* being either diadromous or non-diadromous. Note that the CI would not change for a diadromous species between the center and the right panel, even though the right panel contains more dams, as the connectivity for diadromous species is controlled by the most downstream dam.

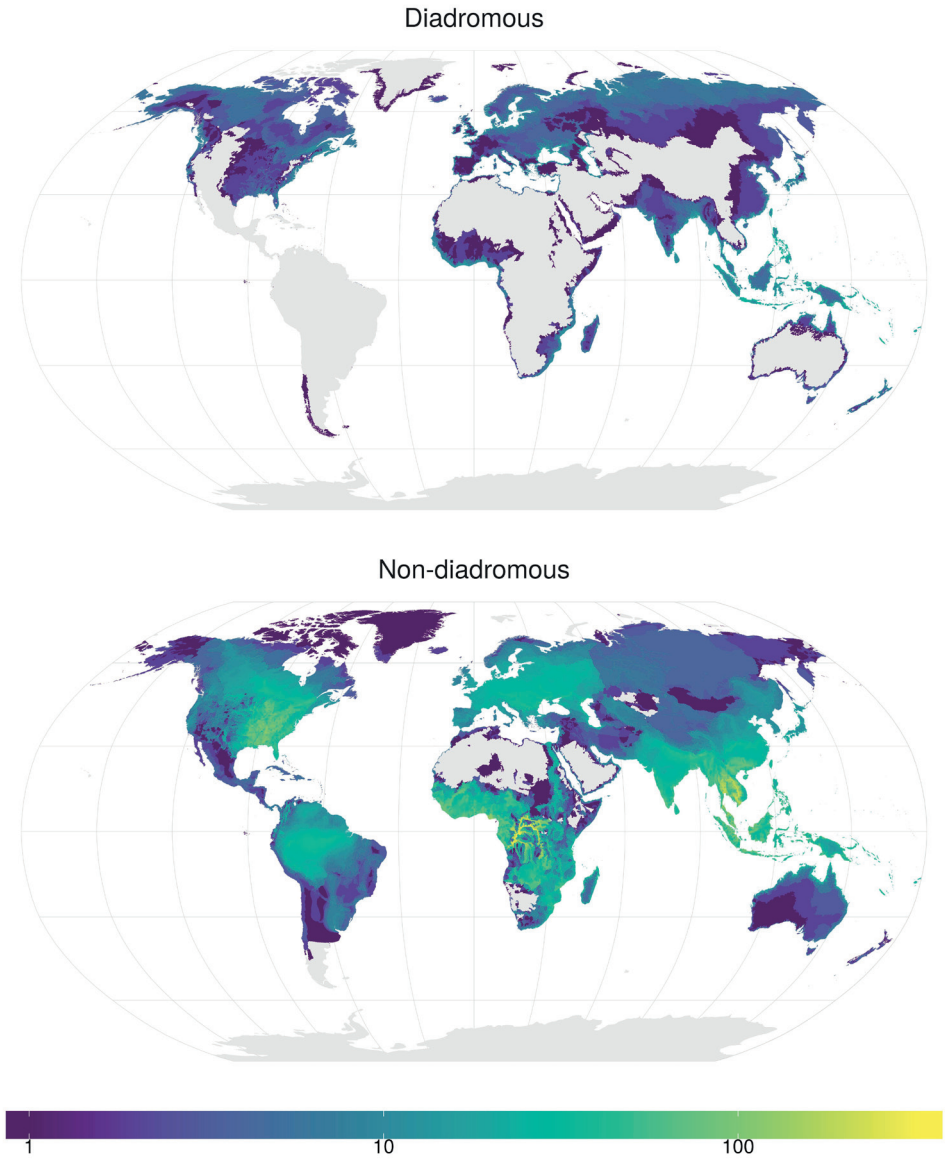


Figure B.2 | Occurrence data of 221 diadromous (top) and 5,417 non-diadromous (bottom) freshwater fish species included in this study, represented as the number of species in each of the ~1M sub-basins. Grey represents areas without species according to the IUCN database.

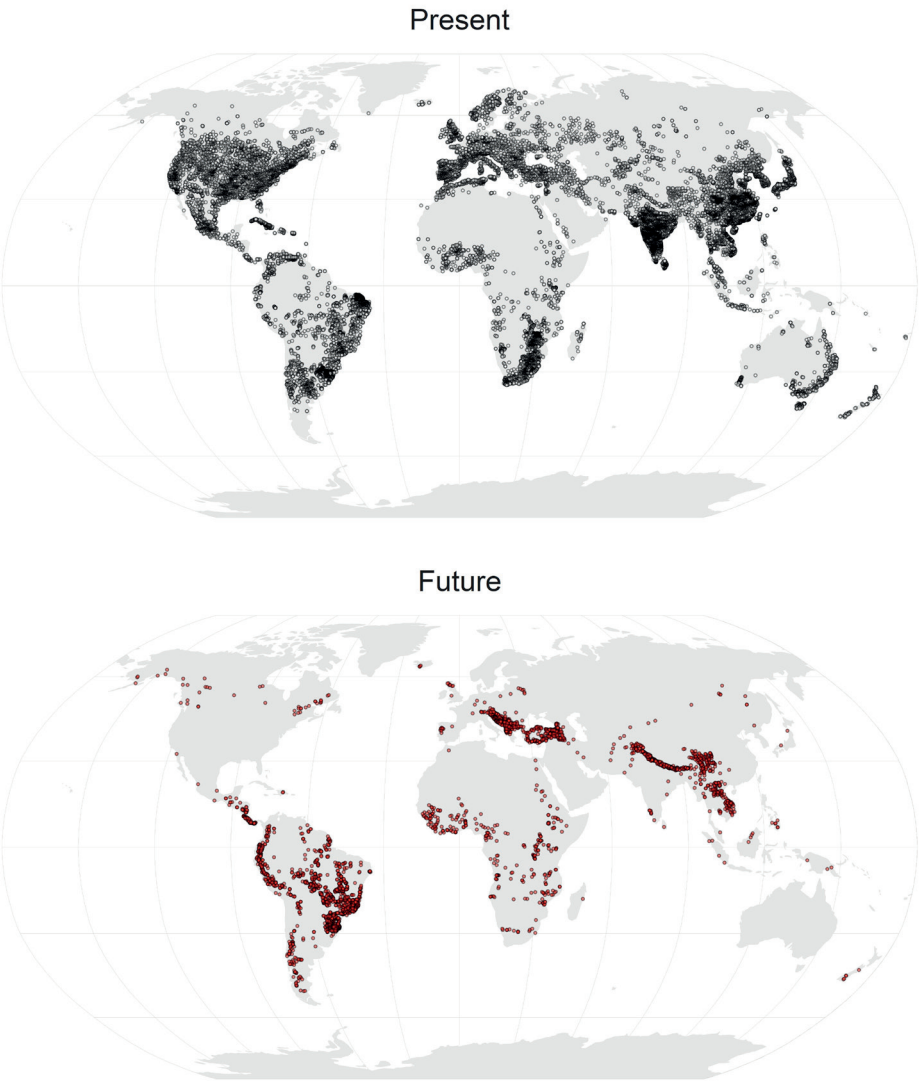


Figure B.3 | Location of dams included in this study for the present situation (39,912 dams; top) and for the future projection (3,681 dams; bottom).

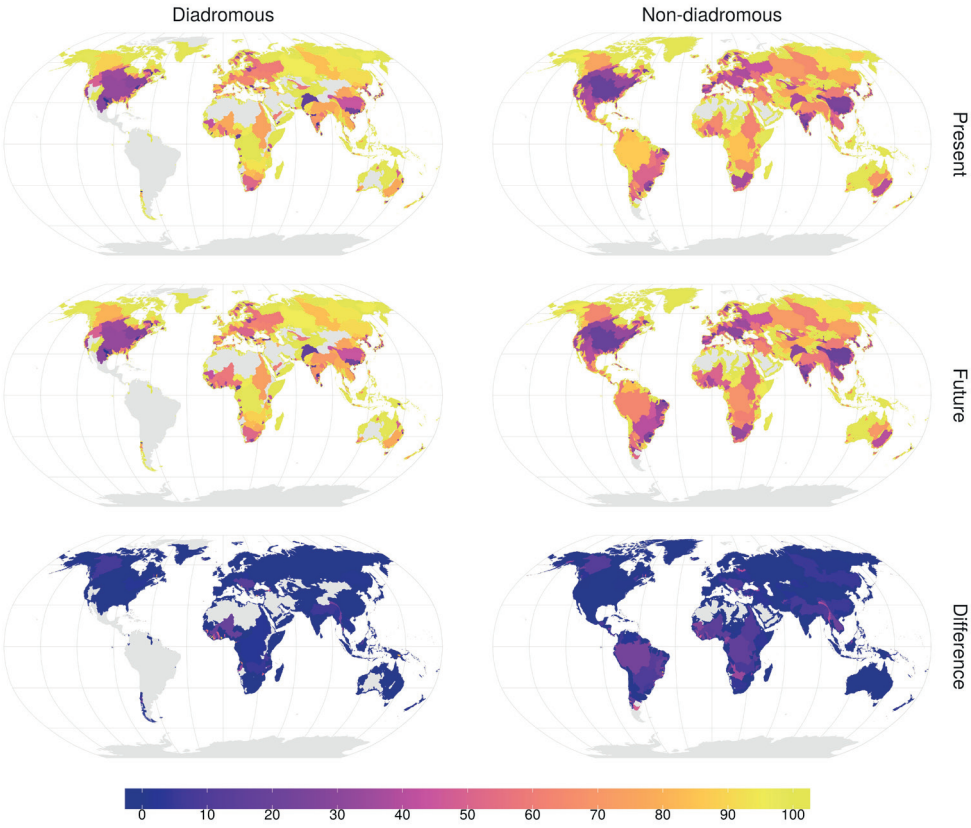


Figure B.4 | Connectivity index (%) for present and future scenarios aggregated for the world's main basins. The difference refers to the delta between present and future scenarios. For each of the ~58,000 basins, the mean of the CI value for the species occurring within that basin is reported. Grey represent areas without species according to the IUCN database.

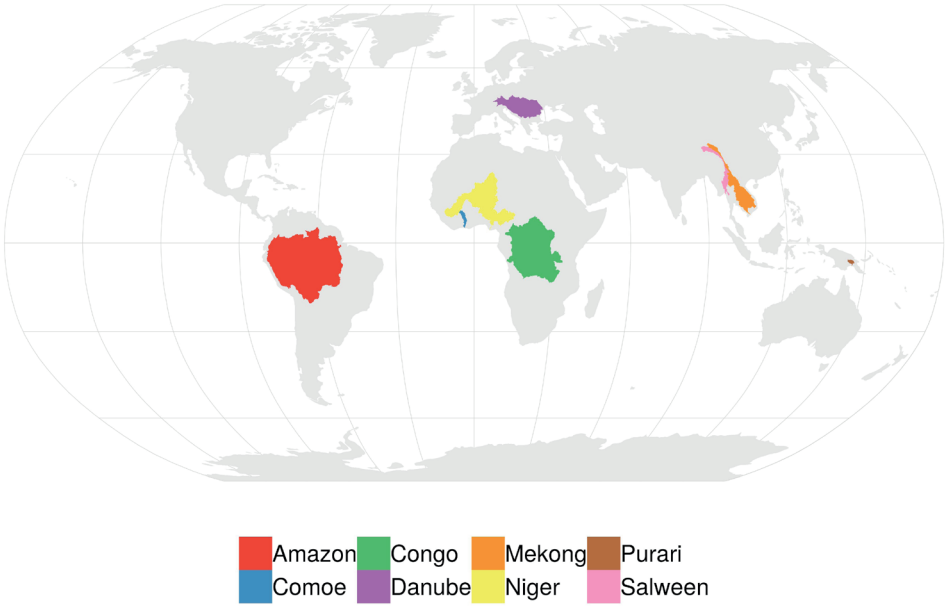


Figure B.5 | Location of example basins shown in **Figures 4a** and **4b**.

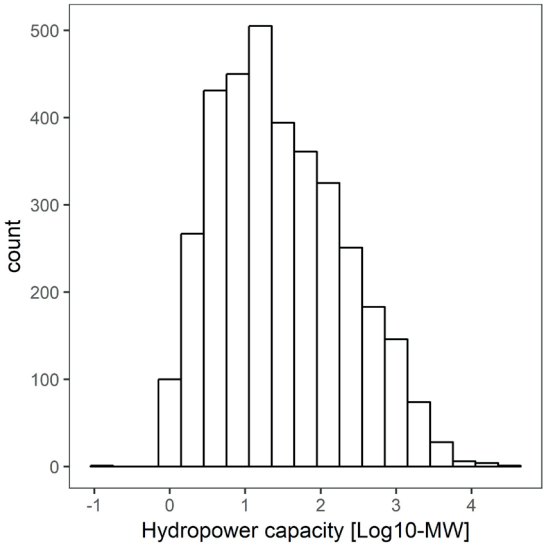


Figure B.6 | Hydropower capacity in log10-transformed MW based on the metadata available for future dams (Zarfl et al., 2015). Median = 23 MW, IQR = 109 MW, n = 3,527.

Table B.1 | Abbreviations used for the species order names of **Figure 4.4i** in the main text. "OTHER" groups together order names with less than 20 species available for our analysis.

Order name	Abbreviation used
ACIPENSERIFORMES	ACIPE.
ANGUILLIFORMES	ANGUI.
ATHERINIFORMES	ATHER.
BATRACHOIDIFORMES	OTHER
BELONIFORMES	BELON.
CHARACIFORMES	CHARA.
CLUPEIFORMES	CLUPE.
CYPRINIFORMES	CYPRINI.
CYPRINODONTIFORMES	CYPRINO.
ELOPIFORMES	OTHER
ESOCIFORMES	OTHER
GASTEROSTEIFORMES	GASTE.
GONORYNCHIFORMES	GONOR.
GYMNOTIFORMES	GYMNO.
LEPISOSTEIFORMES	OTHER
MUGILIFORMES	MUGIL.
OPHIDIIFORMES	OTHER
OSMERIFORMES	OSMER.
OSTEOGLOSSIFORMES	OSTEO.
PERCIFORMES	PERCI.
PERCOPSIFORMES	OTHER
PLEURONECTIFORMES	PLEUR.
POLYPTERIFORMES	OTHER
SALMONIFORMES	SALMO.
SCORPAENIFORMES	SCORP.
SILURIFORMES	SILUR.
SYNBRANCHIFORMES	SYNBR.
SYNGNATHIFORMES	SYNGN.
TETRAODONTIFORMES	TETRA.
GADIFORMES	OTHER

APPENDIX C

C.1 Supplementary Methods

C.1.1 Representativeness of species richness

We compared fish species richness, as derived from IUCN range maps, against a recent collection of freshwater fish species richness for the major world's watersheds (Tedesco et al., 2017) (**Figure C.8**).Tedesco et al., (2017) provide a list of 14,953 freshwater fish species occurring in 3,119 watersheds, collected from both literature and web-based sources. Using the same watershed boundaries, we sampled the species of the IUCN database (IUCN, 2018) occurring in each watershed.

C.1.2 Validation of the water temperature threshold

We compared the species-specific maximum weekly water temperature thresholds derived in this study to laboratory data of critical thermal maxima for 213 species (Comte and Olden, 2017). This allowed us to check whether the threshold inferred from spatial data was representative of the actual thermal tolerance of freshwater fish species. For those species for which multiple experiments to quantify the thermal limit were reported, we averaged over the reported outcomes. We then compared such data to the maximum weekly water temperature threshold calculated in our study. Since the threshold is calculated for each historical period forced by a different GCM, we considered the median of the five values and reported the range in terms of maximum and minimum values across the GCM ensemble.

C.1.3 Details on the GCM-RCP ensemble

For the climate forcing applied to the hydrological model PCR-GLOBWB, we employed the five global climate models (GCMs) available from the InterSectoral Impact Model Intercomparison Project (ISI-MIP). ISI-MIP provides bias-corrected and spatially down-scaled climate inputs standardized for impact models (Hempel et al., 2013; Schewe et al., 2014; Warszawski et al., 2014). The climate variables were used as input for the hydrological model PCR-GLOBWB and temperature model DynWAT and consisted of precipitation, temperature and radiation components.

C.2 Supplementary Discussion

C.2.1 Choice of ecologically relevant thresholds

A plethora of flow and water temperature metrics potentially affecting the spatial occurrence of freshwater species have been reported by previous studies (Arismendi et al., 2013; Carlisle et al., 2017; Dhungel et al., 2016; Döll and Zhang, 2010; Gao et al., 2009; Maheu et al., 2016; Olden and Poff, 2003; Poff et al., 2010; Poff and Zimmerman, 2010; Richter et al., 1996; Steel et al., 2017; van Vliet et al., 2013). Such metrics describe different features of the hydrograph and thermograph (magnitude, frequency, duration, timing and rate of change) and are usually selected depending on the study area, river type and species studied. Given the high number of metrics available and the redundancy among them (Arismendi et al., 2013; Olden and Poff, 2003), defining habitat suitability thresholds common to ~7,000 freshwater fish species is a non-trivial task. We adopted maximum water temperature (Tmax), minimum flow (Qmin), and the number of zero flow weeks (Qzf) as suitability thresholds, based on the weekly hydrograph and thermograph of the hydrological model. The maximum water temperature (Tmax) is the most important physiological threshold for fish species as mortality of ectothermic species occurs above lethal thresholds (Comte and Olden, 2017). Low flow (Qmin) and the number of zero flow events (Qzf) represent the range of low flow conditions that the species can withstand. Decreases in Qmin directly affect the riffle-pool systems and connectivity between viable habitat patches, leading to a rapid loss of biodiversity (Poff et al., 2010). Qzf gives an indication about the number of zero-flow events that a species can tolerate. Indeed, the increase in frequency of dry-spells directly correlates with reduction in diversity and biomass due to the loss of suitable aquatic habitat (Poff et al., 2010).

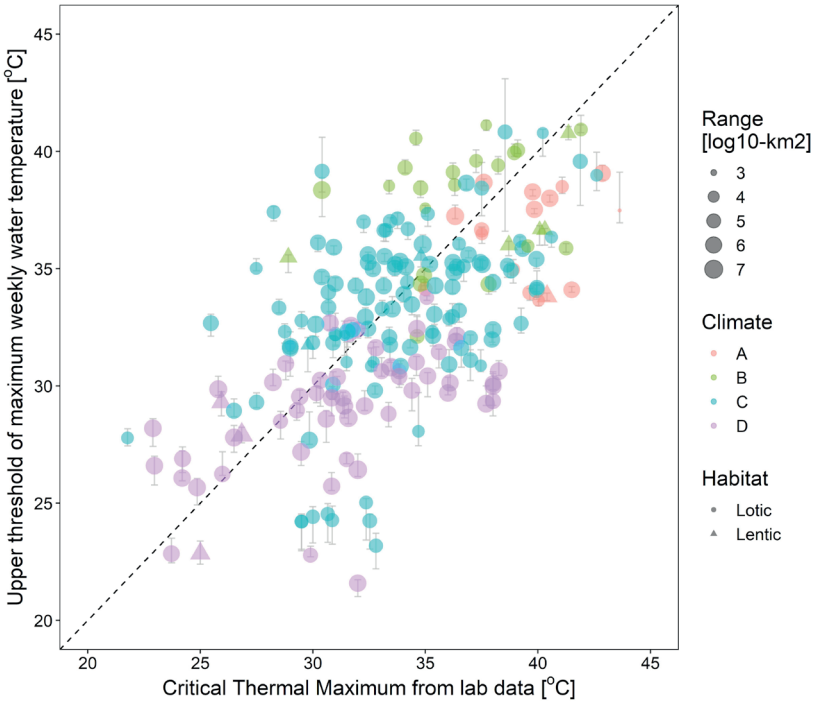


Figure C.1 | Validation of the upper threshold of maximum weekly water temperature used in this study against critical thermal maxima from laboratory experiments (n = 213)(Comte and Olden, 2017). The error bars represent full range of values from the five GCMs used in this study, while the dots show the median value. The dots size is proportional to the area of the IUCN occurrence range of each species and the values in the scale stands for a round of ±0.5. Circles represent lotic species while triangles stand for exclusively lentic species. Climate zoning is according to the historic Köppen-Geiger classification, A=Equatorial, B=Arid, C=Warm temperate, D=Snow. The 1:1 line is shown as a black dashed line. Mean standard error = 0.09; R^2 = 0.36.

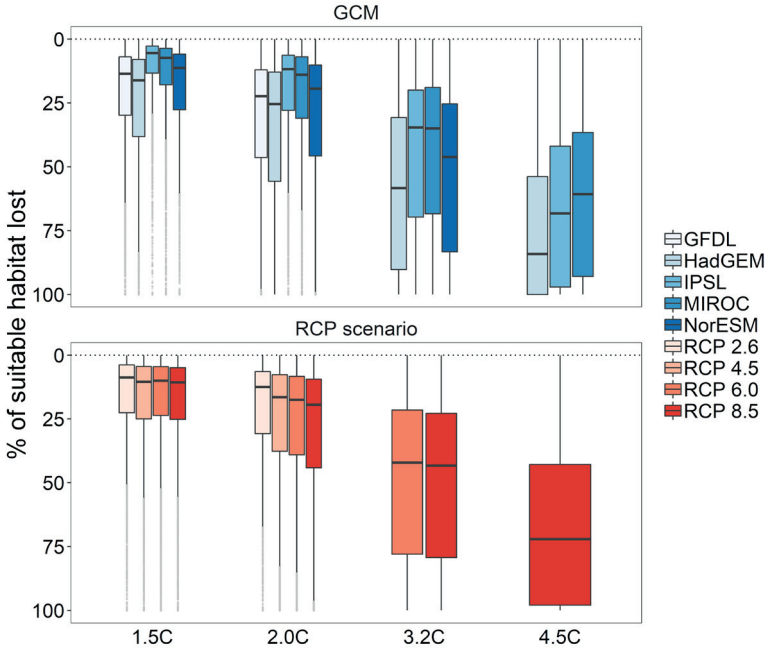


Figure C.2 | Output variability stemming from the different inputs of Global Climate Models and Representative Concentration Pathway scenarios. Each box delimits the interquartile range and shows the median, and whiskers stand for the 95% interval for the 6,924 fish species assessed in this study.

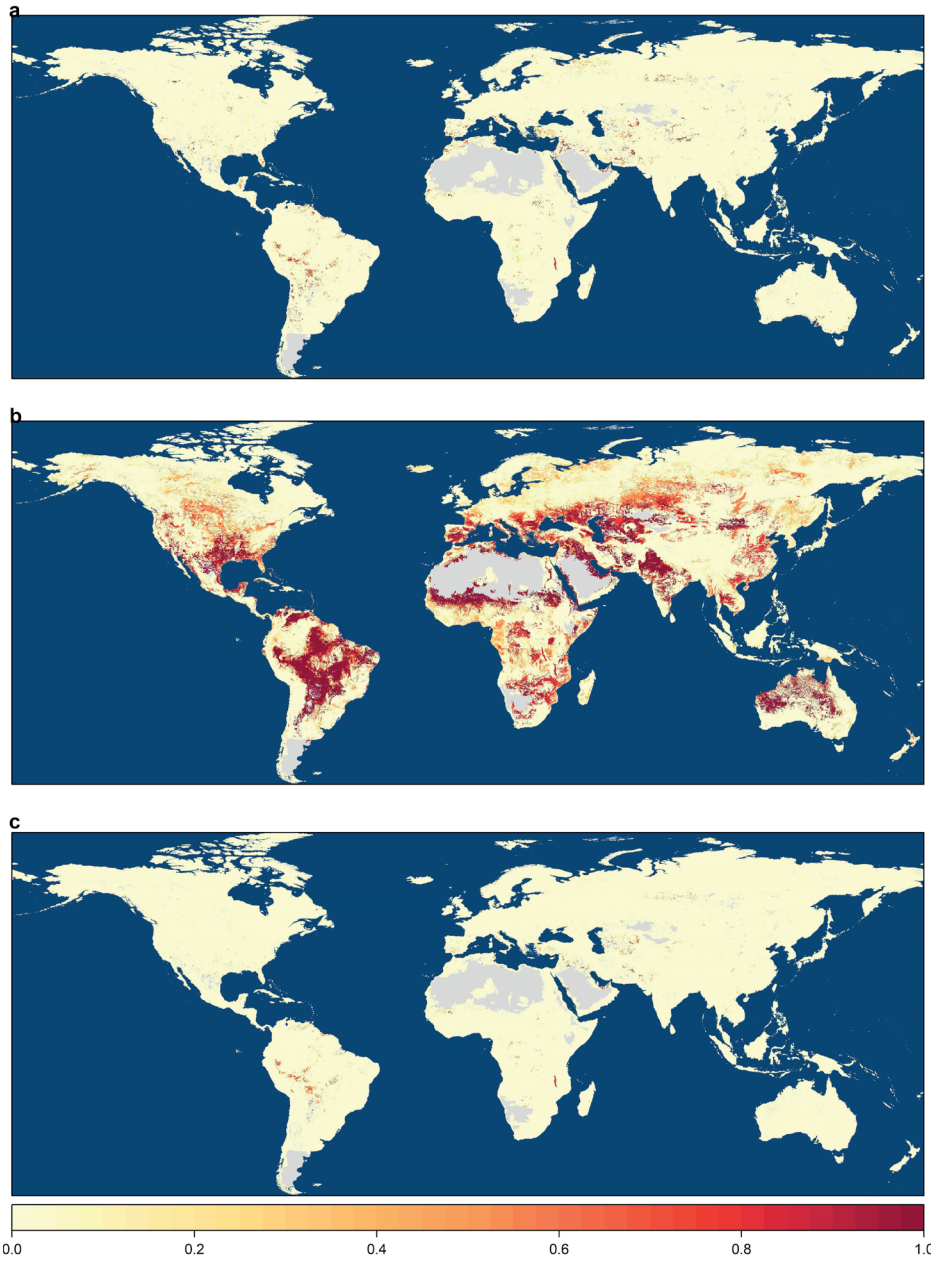


Figure C.3 | Stressors contribution to the relative loss of species at each five arc minutes grid-cell at 3.2ºC for flow (a) and water temperature (b). The overlap (relative loss of species due to both flow and water temperature) of the two stressors is also reported (c). The maps represent the median of the ensemble of the GCM-RCP combinations available for the 3.2ºC warming scenario. Grey denotes areas without information available on freshwater fish species.

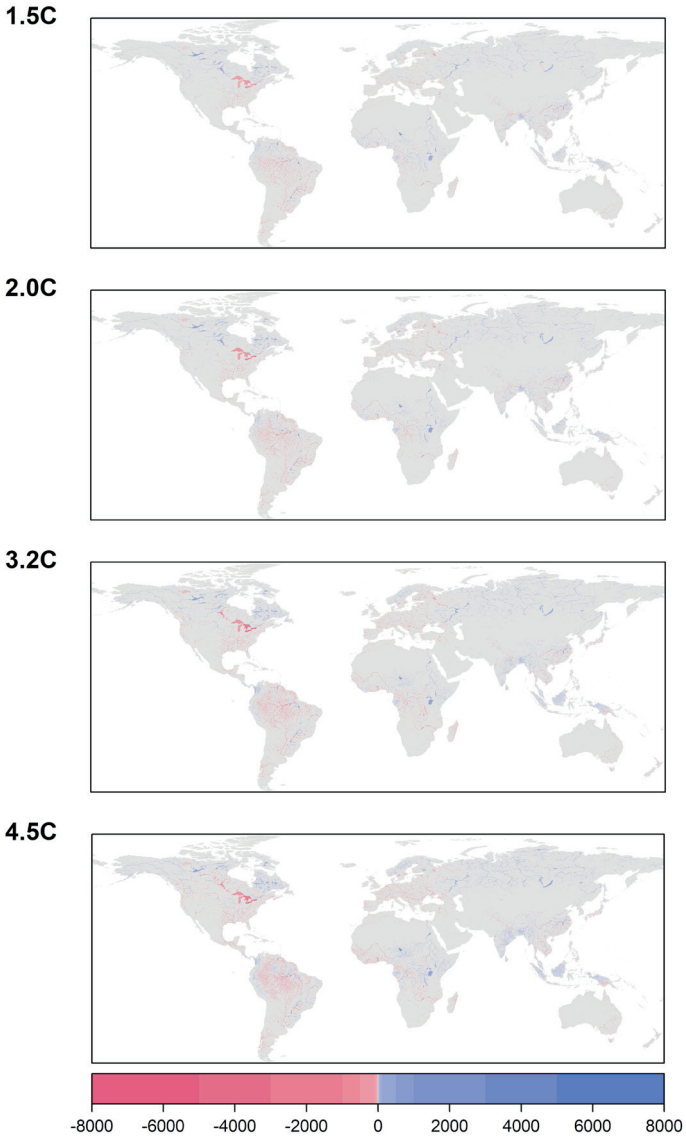


Figure C.4-I | Minimum weekly flow anomaly at the four different warming levels compared to the historical period (1976-2005). Each map represents the median of the GCM-RCP ensemble at each warming level. The difference is reported in units of m^3/s.

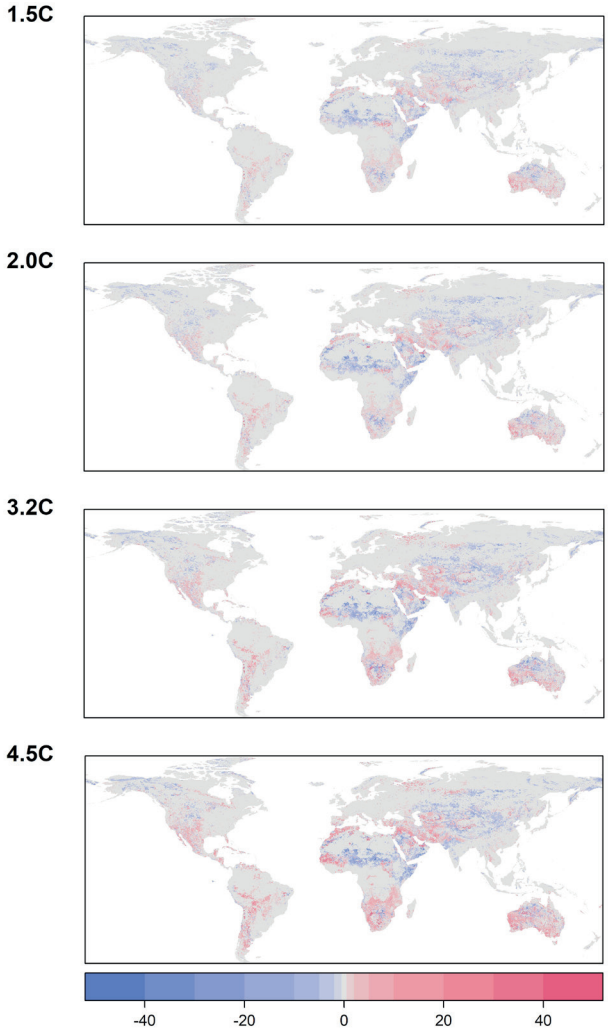


Figure C.4-II | Number of zero flow weeks anomaly at the four different warming levels compared to the historical period (1976-2005). Each map represents the median of the GCM-RCP ensemble at each warming level. The difference is number of weeks.

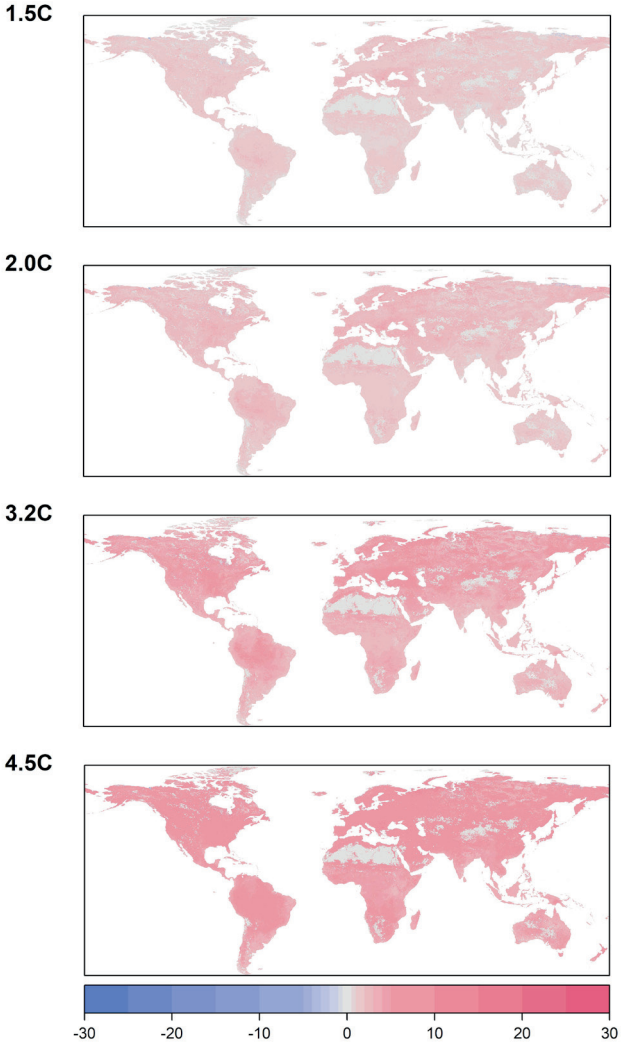


Figure C.4-III | Maximum weekly water temperature anomaly at the four different warming levels compared to the historical period (1976-2005). Each map represents the median of the GCM-RCP ensemble at each warming level. The difference is reported in units of ºC.

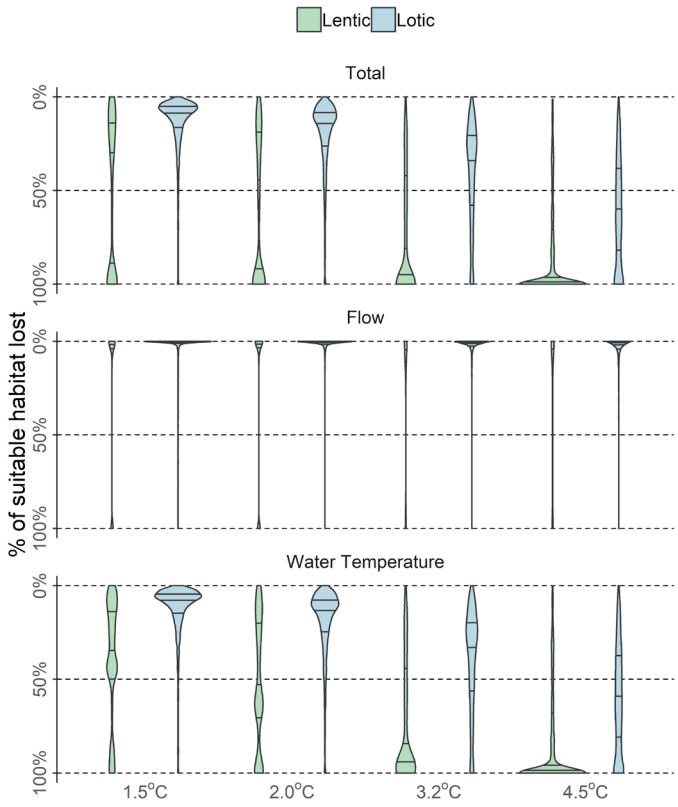


Figure C.5 | Breakdown of range contractions percentage for lotic vs (exclusively) lentic species at the four global warming levels. The solid horizontal lines of the violin plots represent the interquartile range and median. Contribution to the percentage of range contractions from flow and water temperature are reported in the middle and lower panels.

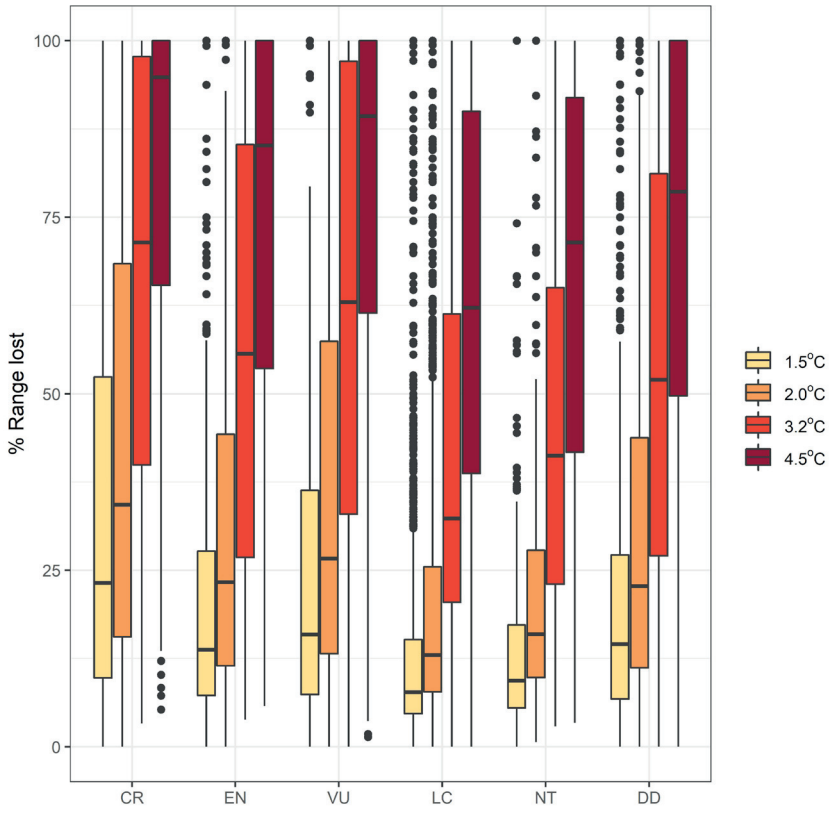


Figure C.6 | Percentage of range lost for the 6,924 species grouped by different IUCN threat categories at each warming level. For each species, the median of the ensemble of the GCM-RCP combinations is reported. CR=Critically endangered, EN=Endangered, VU=Vulnerable, LC=Least Concern, NT=Near Threatened, DD=Data Deficient.

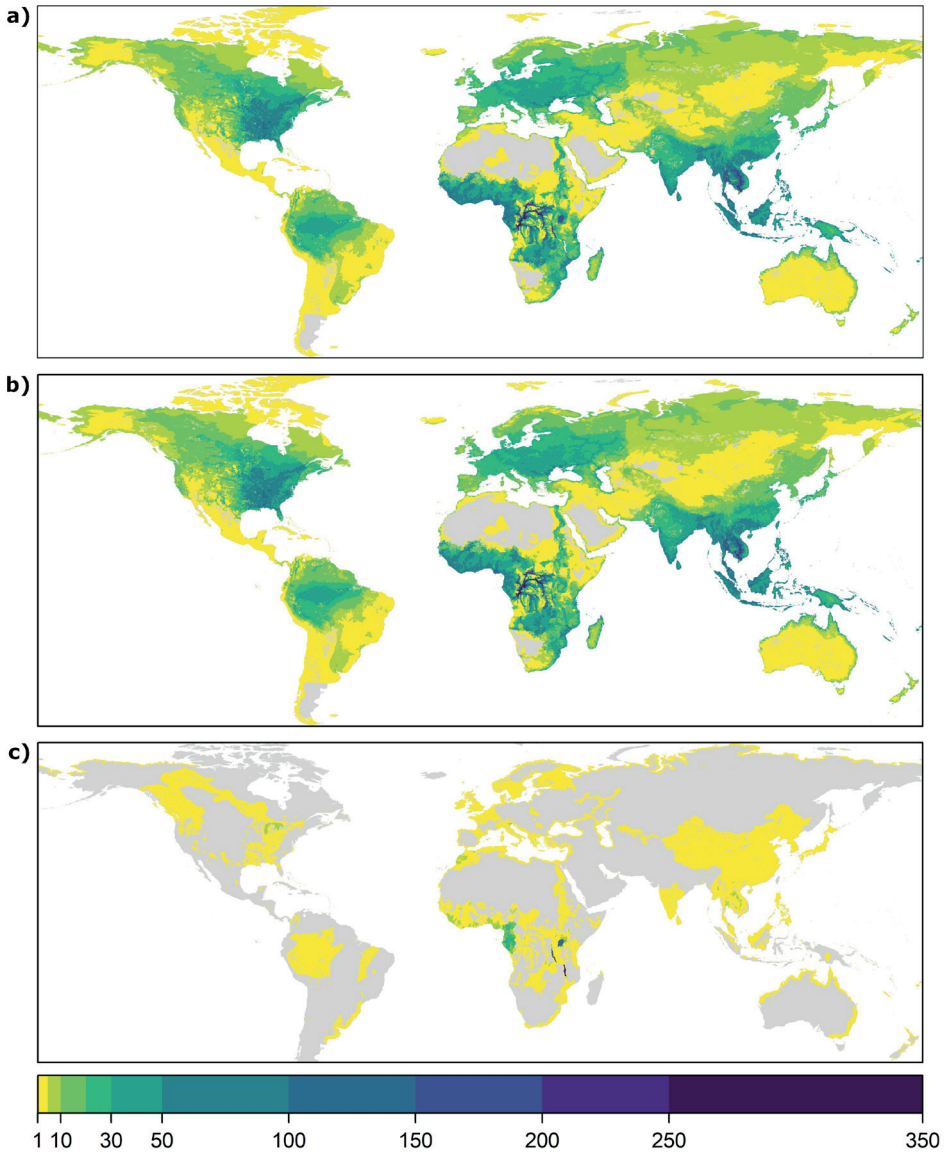


Figure C.7 | Current species richness at each five arc-minutes grid cell derived from the IUCN range maps for the 6,924 species used in this study (a). Distributions of lotic species that can live in flowing water bodies (b) and of exclusively lentic species that can live only in stagnant water bodies (c) are also reported (see Methods for further details of lentic and lotic categories). Gray denotes no data.

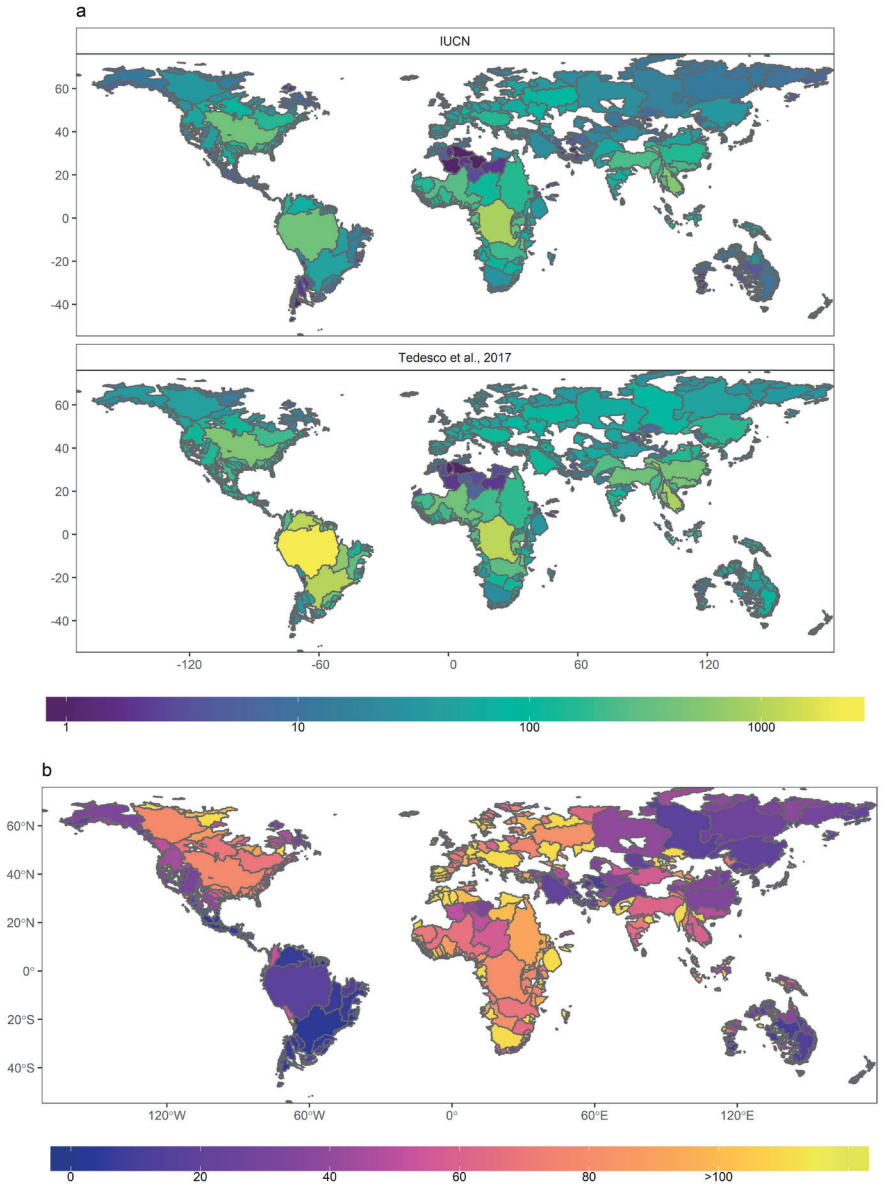


Figure C.8 | Number of freshwater fish species available for the major watersheds of the world for the IUCN data used in this study(IUCN, 2018) and according to Tedesco et al.(Tedesco et al., 2017) (a). The percentage of species covered in each watershed by this study (IUCN data)(IUCN, 2018) compared to Tedesco et al.(Tedesco et al., 2017) is also reported.

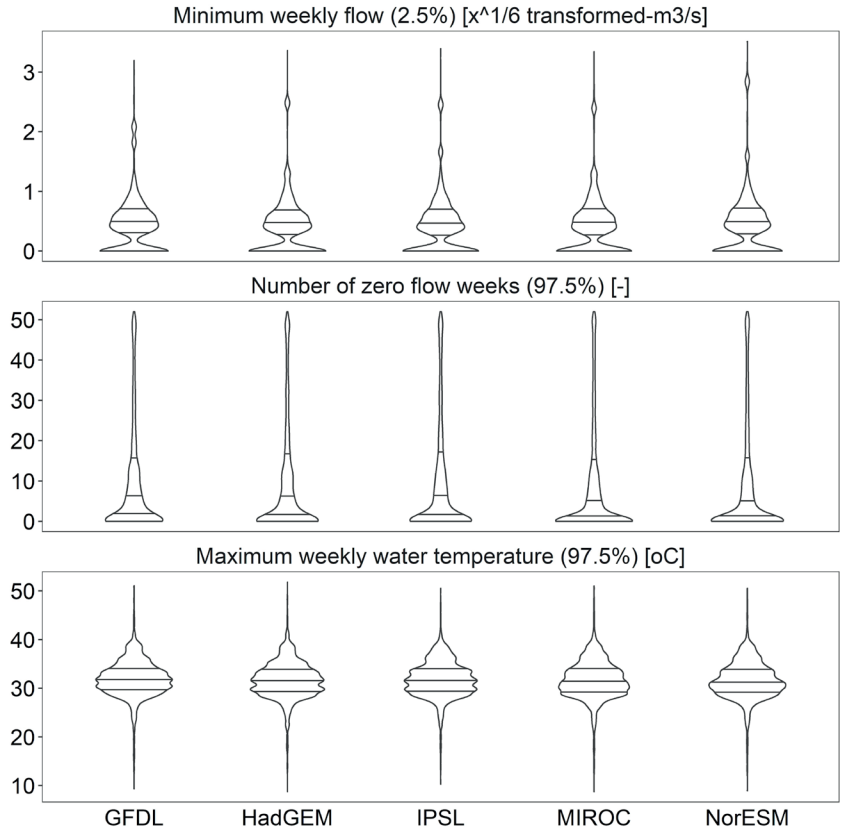
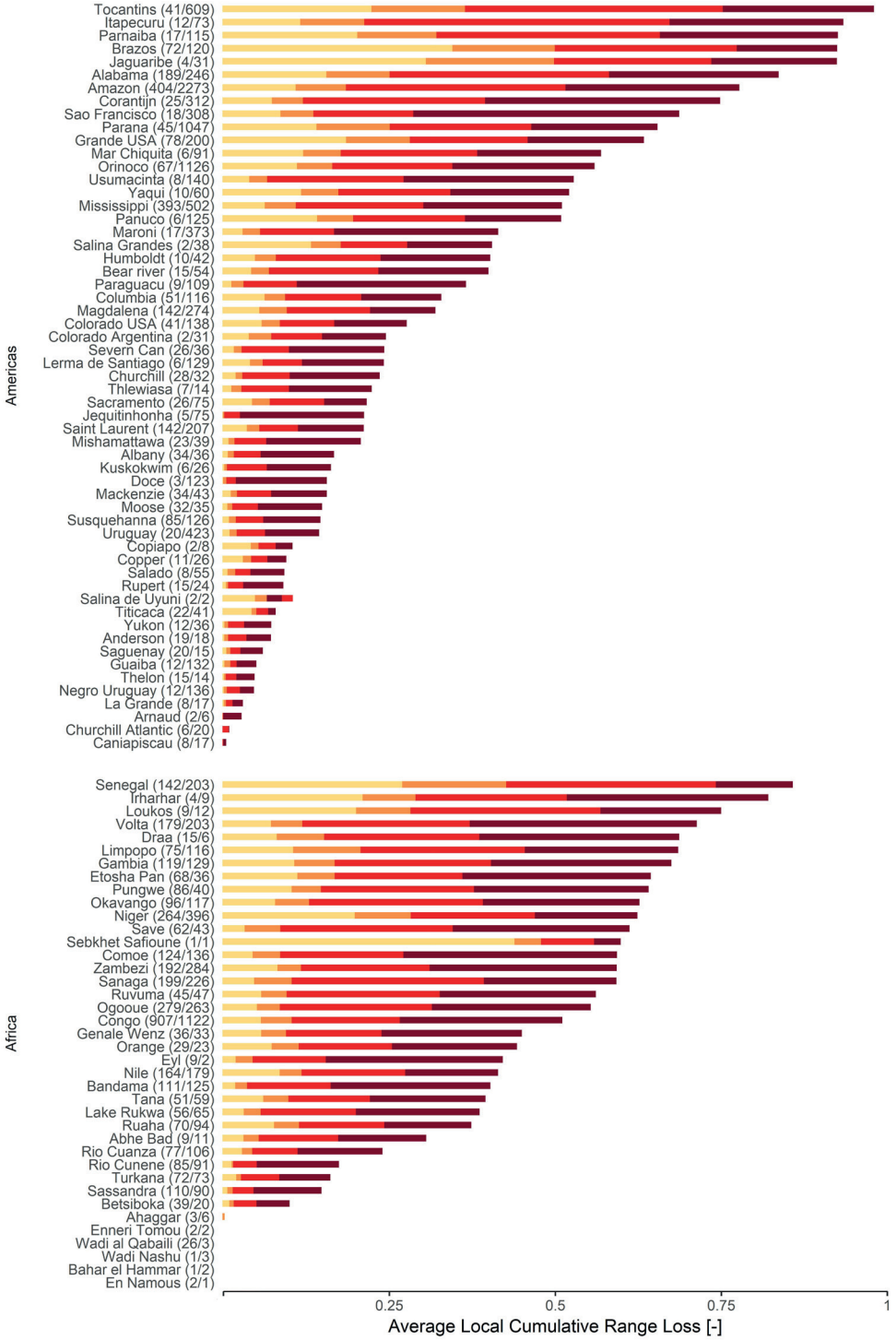
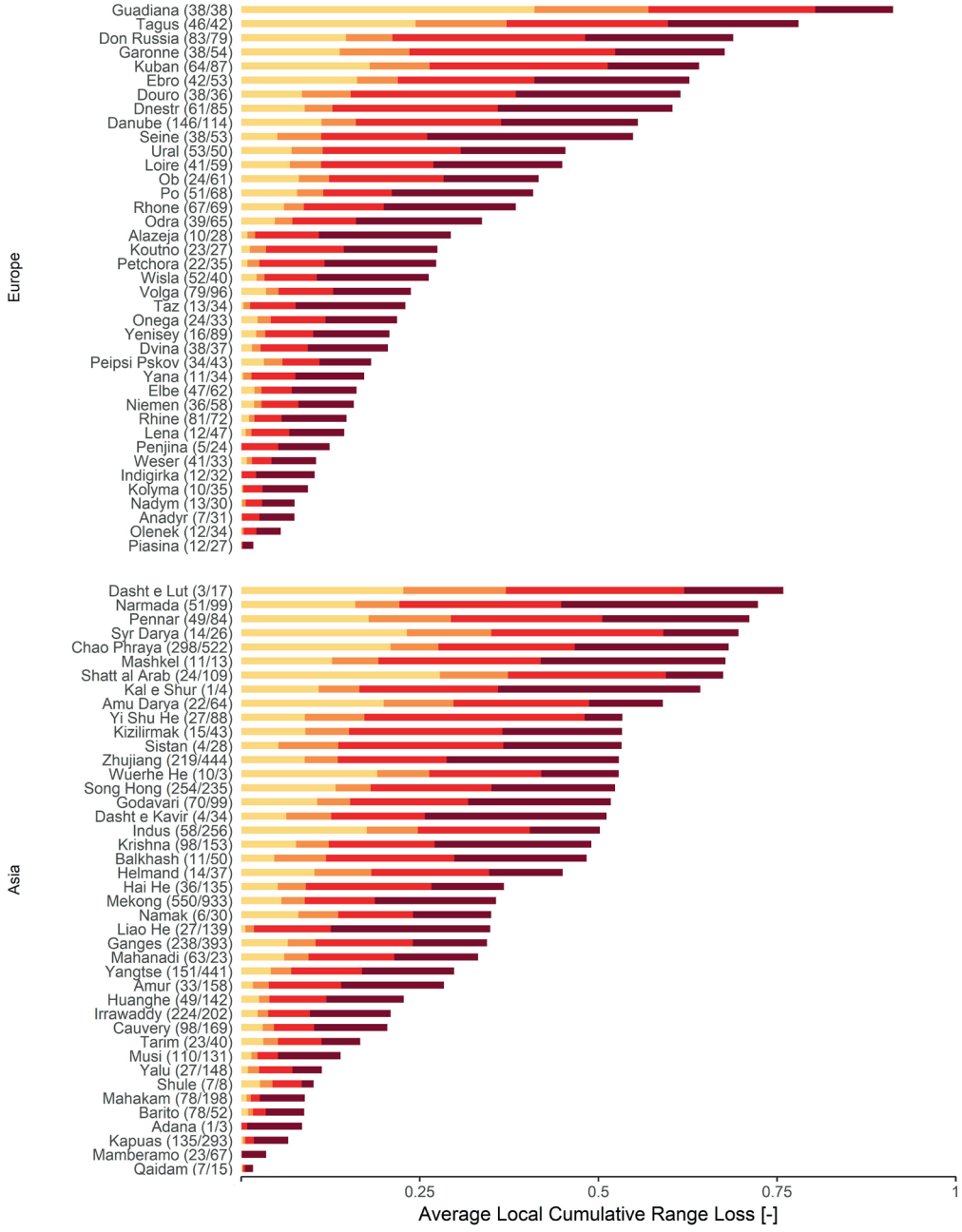


Figure C.9 | Statistical distribution of the flow and water temperature thresholds used in this study for the 6,924 species, represented as violin plots. For each variable, the threshold is calculated for each GCM historical period 1976-2005. In brackets, the percentile used for each threshold. Horizontal lines represent the interquartile range and median.


Americas
Tocantins (41/609)
Itapecuru (12/73)
Parnaiba (17/115)
Brazos (72/120)
Jaguaribe (4/31)
Alabama (189/246)
Amazon (404/2273)
Corantijn (25/312)
Sao Francisco (18/308)
Parana (45/1047)
Grande USA (78/200)
Mar Chiquita (6/91)
Orinoco (67/1126)
Usumacinta (8/140)
Yaqui (10/60)
Mississippi (393/502)
Panuco (6/125)
Maroni (17/373)
Salina Grandes (2/38)
Humboldt (10/42)
Bear river (15/54)
Paraguacu (9/109)
Columbia (51/116)
Magdalena (142/274)
Colorado USA (41/138)
Colorado Argentina (2/31)
Severn Can (26/36)
Lerma de Santiago (6/129)
Churchill (28/32)
Thlewiasa (7/14)
Sacramento (26/75)
Jequitinhonha (5/75)
Saint Laurent (142/207)
Mishamattawa (23/39)
Albany (34/36)
Kuskokwim (6/26)
Doce (3/123)
Mackenzie (34/43)
Moose (32/35)
Susquehanna (85/126)
Uruguay (20/423)
Copiapo (2/8)
Copper (11/26)
Salado (8/55)
Rupert (15/24)
Salina de Uyuni (2/2)
Titicaca (22/41)
Yukon (12/36)
Anderson (19/18)
Saguenay (20/15)
Guaiba (12/132)
Thelon (15/14)
Negro Uruguay (12/136)
La Grande (8/17)
Arnaud (2/6)
Churchill Atlantic (6/20)
Caniapiscau (8/17)
Africa
Senegal (142/203)
Irharhar (4/9)
Loukos (9/12)
Volta (179/203)
Draa (15/6)
Limpopo (75/116)
Gambia (119/129)
Etosha Pan (68/36)
Pungwe (86/40)
Okavango (96/117)
Niger (264/396)
Save (62/43)
Sebkhet Safioune (1/1)
Comoe (124/136)
Zambezi (192/284)
Sanaga (199/226)
Ruvuma (45/47)
Ogooue (279/263)
Congo (907/1122)
Genale Wenz (36/33)
Orange (29/23)
Eyl (9/2)
Nile (164/179)
Bandama (111/125)
Tana (51/59)
Lake Rukwa (56/65)
Ruaha (70/94)
Abhe Bad (9/11)
Rio Cuanza (77/106)
Rio Cunene (85/91)
Turkana (72/73)
Sassandra (110/90)
Betsiboka (39/20)
Ahaggar (3/6)
Enneri Tomou (2/2)
Wadi al Qabaili (26/3)
Wadi Nashu (1/3)
Bahar el Hammar (1/2)
En Namous (2/1)
0.25
0.5
0.75
1
Average Local Cumulative Range Loss [-]


Europe
Guadiana (38/38)
Tagus (46/42)
Don Russia (83/79)
Garonne (38/54)
Kuban (64/87)
Ebro (42/53)
Douro (38/36)
Dnestr (61/85)
Danube (146/114)
Seine (38/53)
Ural (53/50)
Loire (41/59)
Ob (24/61)
Po (51/68)
Rhone (67/69)
Odra (39/65)
Alazeja (10/28)
Koutno (23/27)
Petchora (22/35)
Wisla (52/40)
Volga (79/96)
Taz (13/34)
Onega (24/33)
Yenisey (16/89)
Dvina (38/37)
Peipsi Pskov (34/43)
Yana (11/34)
Elbe (47/62)
Niemen (36/58)
Rhine (81/72)
Lena (12/47)
Penjina (5/24)
Weser (41/33)
Indigirka (12/32)
Kolyma (10/35)
Nadym (13/30)
Anadyr (7/31)
Olenek (12/34)
Piasina (12/27)
Asia
Dasht e Lut (3/17)
Narmada (51/99)
Pennar (49/84)
Syr Darya (14/26)
Chao Phraya (298/522)
Mashkel (11/13)
Shatt al Arab (24/109)
Kal e Shur (1/4)
Amu Darya (22/64)
Yi Shu He (27/88)
Kizilirmak (15/43)
Sistan (4/28)
Zhujiang (219/444)
Wuerhe He (10/3)
Song Hong (254/235)
Godavari (70/99)
Dasht e Kavir (4/34)
Indus (58/256)
Krishna (98/153)
Balkhash (11/50)
Helmand (14/37)
Hai He (36/135)
Mekong (550/933)
Namak (6/30)
Liao He (27/139)
Ganges (238/393)
Mahanadi (63/23)
Yangtse (151/441)
Amur (33/158)
Huanghe (49/142)
Irrawaddy (224/202)
Cauvery (98/169)
Tarim (23/40)
Musi (110/131)
Yalu (27/148)
Shule (7/8)
Mahakam (78/198)
Barito (78/52)
Adana (1/3)
Kapuas (135/293)
Mamberamo (23/67)
Qaidam (7/15)
0.25
0.5
0.75
1
Average Local Cumulative Range Loss [-]

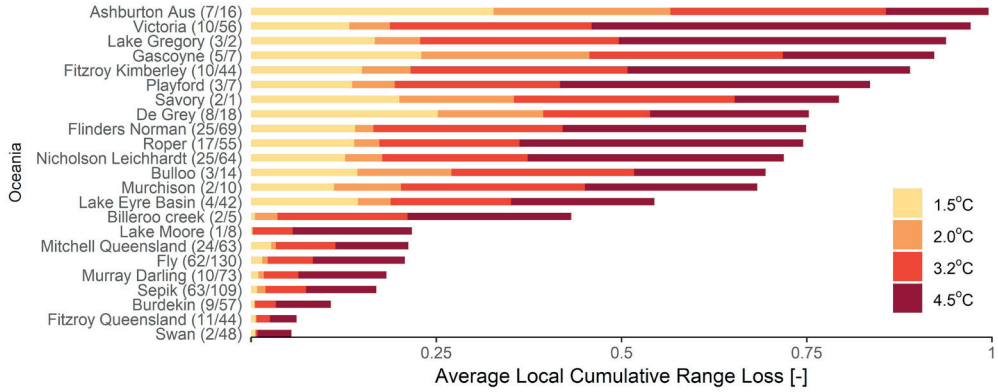


Figure C.10 | Local cumulative range loss averaged over the 200 largest watersheds of the world. Numbers in brackets represent the ratio of the number of species covered in our study over the total number of species known to occur in the watershed according to Tedesco et al.(Tedesco et al., 2017). Differences among the total number of species within each watershed are dependent upon sampling and differences in catchment delineation.

Table C.1 | Years at which each GCM-RCP scenario reaches one of the four global warming levels. The levels are defined by global mean temperature rises compared to pre-industrial (1850-1900). Years represent the central value of the 30-years moving average. Three scenarios do not report values as they do not reach even the 1.5°C warming level.

	1.5°C	2°C	3.2°C	4.5°C
gfdl_rcp2p6				
gfdl_rcp4p5	2044			
gfdl_rcp6p0	2052	2073		
gfdl_rcp8p5	2034	2051		
hadgem_rcp2p6	2028			
hadgem_rcp4p5	2029	2044		
hadgem_rcp6p0	2034	2048	2082	
hadgem_rcp8p5	2024	2036	2058	2080
ipsl_rcp2p6	2011	2033		
ipsl_rcp4p5	2012	2031		
ipsl_rcp6p0	2012	2032	2078	
ipsl_rcp8p5	2011	2026	2051	2074
miroc_rcp2p6	2018	2035		
miroc_rcp4p5	2021	2035		
miroc_rcp6p0	2023	2039	2073	
miroc_rcp8p5	2018	2030	2053	2074
noresm_rcp2p6				
noresm_rcp4p5				
noresm_rcp6p0	2050	2072		
noresm_rcp8p5	2033	2048	2078	

Table C.2 | Abbreviations used for the species order names of **Figure 5.3i** in the main text. "OTHER" groups together order names with less than 20 species available for our analysis.

Order name	Abbreviation used
ACIPENSERIFORMES	OTHER
ANGUILLIFORMES	OTHER
ATHERINIFORMES	ATHER.
AULOPIFORMES	OTHER
BATRACHOIDIFORMES	OTHER
BELONIFORMES	OTHER
CHARACIFORMES	CHARA.
CLUPEIFORMES	CLUPE.
CYPRINIFORMES	CYPRINI.
CYPRINODONTIFORMES	CYPRINO.
ELOPIFORMES	OTHER
ESOCIFORMES	OTHER
GASTEROSTEIFORMES	OTHER
GONORYNCHIFORMES	GONOR.
GYMNOTIFORMES	GYMNO.
LEPISOSTEIFORMES	OTHER
MUGILIFORMES	MUGIL.
OPHIDIIFORMES	OTHER
OSMERIFORMES	OSMER.
OSTEOGLOSSIFORMES	OSTEO.
PERCIFORMES	PERCI.
PERCOPSIFORMES	OTHER
PLEURONECTIFORMES	PLEUR.
POLYPTERIFORMES	OTHER
SALMONIFORMES	SALMO.
SCORPAENIFORMES	SCORP.
SILURIFORMES	SILUR.
SYNBRANCHIFORMES	SYNBR.
SYNGNATHIFORMES	SYNGN.
TETRAODONTIFORMES	TETRA.
GADIFORMES	OTHER

APPENDIX D

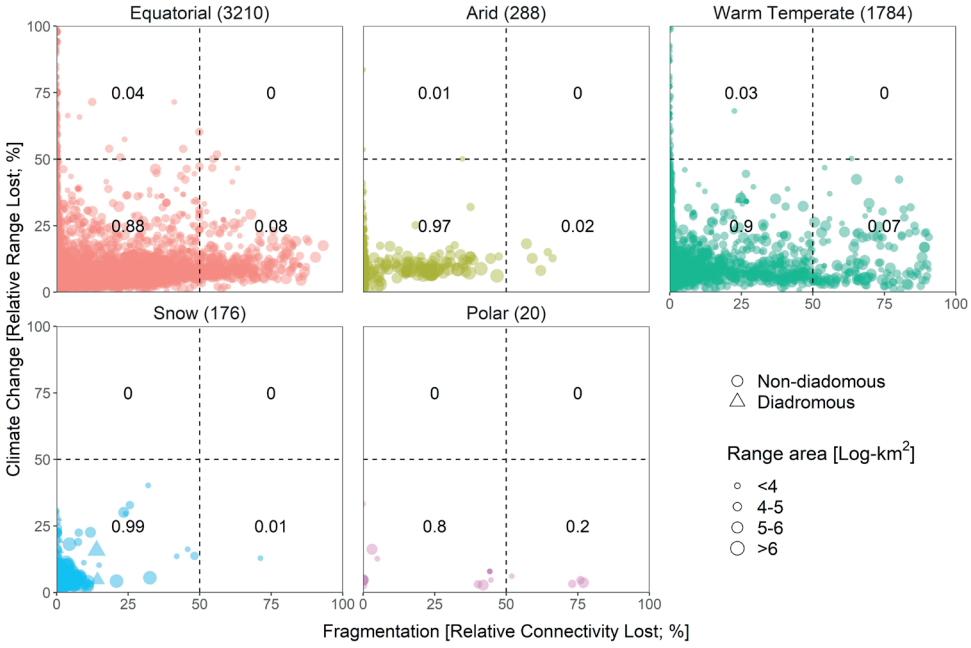


Figure D.1 | Impacts from climate change (at 1.5ºC global warming) and fragmentation on the geographical ranges of 5,450 freshwater fish species per Köppen-Geiger climate zone (with number of species per zone in brackets). Each plotted point represents a species, whereby point size is proportional to the original range size. Numbers in the quadrants represent the proportions of species. Relative range/connectivity losses were calculated as the difference between current and future range area/ connectivity index divided by the current range area/connectivity index. Climate category was assigned based on the overlap of the range of each species with the Köppen-Geiger climate zones.

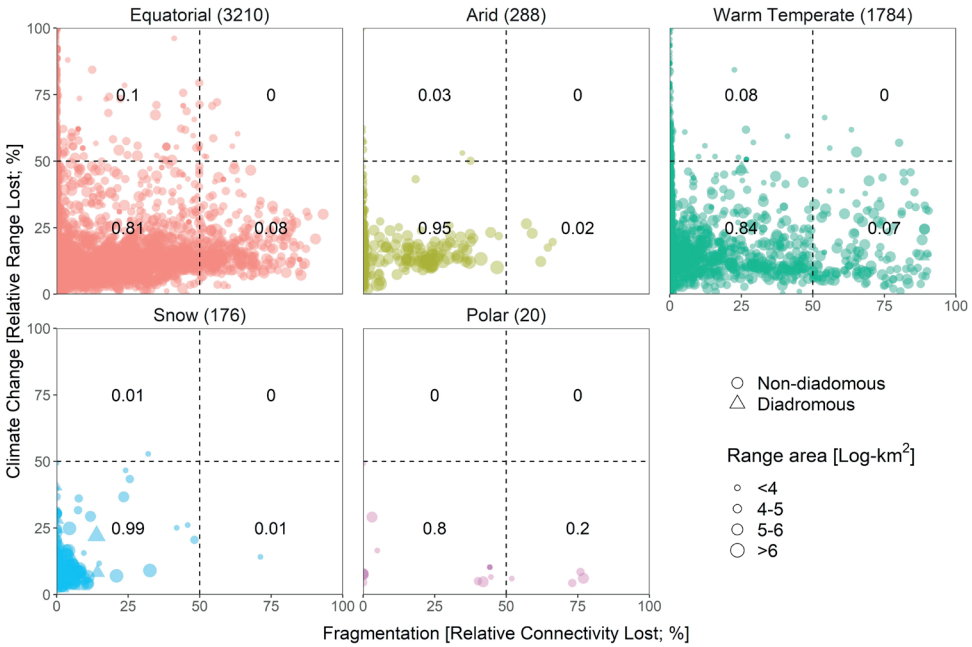


Figure D.2 | Impacts from climate change (at 2ºC global warming) and fragmentation on the geographical ranges of 5,450 freshwater fish species per Köppen-Geiger climate zone (with number of species per zone in brackets). Each plotted point represents a species, whereby point size is proportional to the original range size. Numbers in the quadrants represent the proportions of species. Relative range/connectivity losses were calculated as the difference between current and future range area/ connectivity index divided by the current range area/connectivity index. Climate category was assigned based on the overlap of the range of each species with the Köppen-Geiger climate zones.

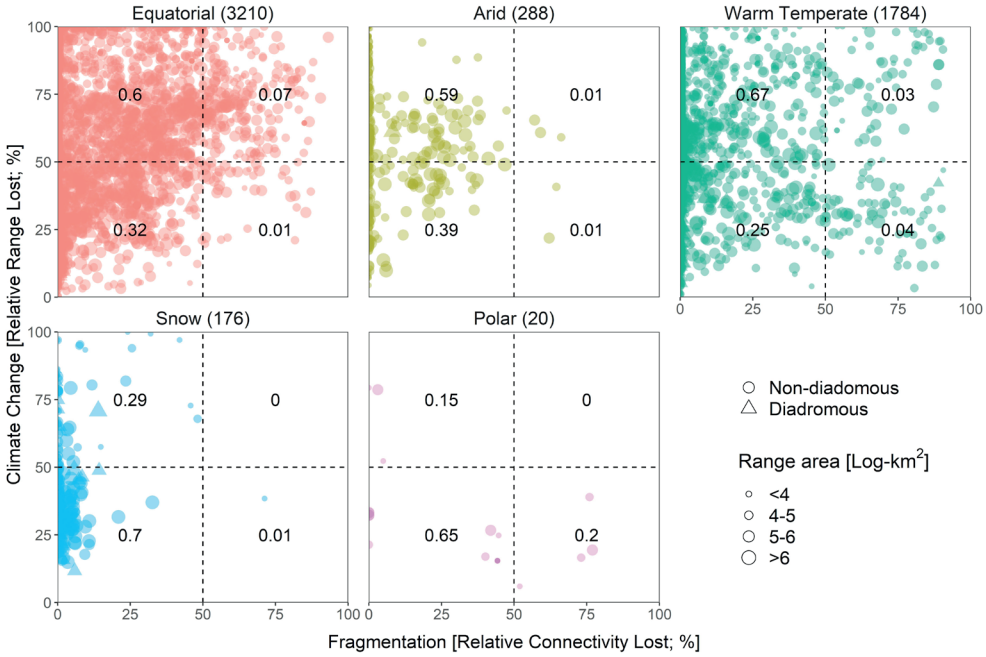
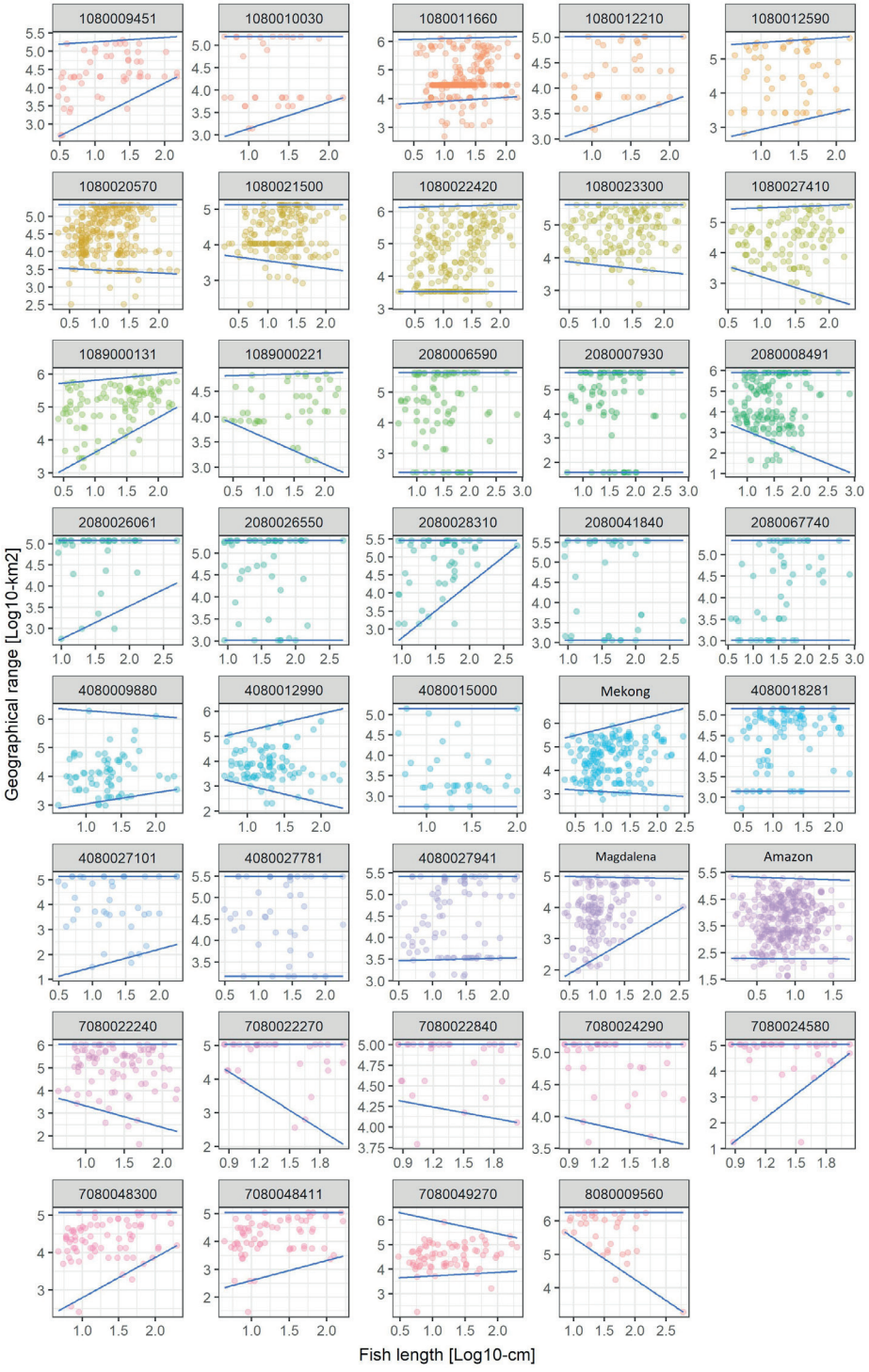


Figure D.3 | Impacts from climate change (at 4.5ºC global warming) and fragmentation on the geographical ranges of 5,450 freshwater fish species per Köppen-Geiger climate zone (with number of species per zone in brackets). Each plotted point represents a species, whereby point size is proportional to the original range size. Numbers in the quadrants represent the proportions of species. Relative range/connectivity losses were calculated as the difference between current and future range area/ connectivity index divided by the current range area/connectivity index. Climate category was assigned based on the overlap of the range of each species with the Köppen-Geiger climate zones.


1080009451
1080010030
1080011660
1080012210
1080012590
1080020570
1080021500
1080022420
1080023300
1080027410
1089000131
1089000221
2080006590
2080007930
2080008491
2080026061
2080026550
2080028310
2080041840
2080067740
4080009880
4080012990
4080015000
Mekong
4080018281
4080027101
4080027781
4080027941
Magdalena
Amazon
7080022240
7080022270
7080022840
7080024290
7080024580
7080048300
7080048411
7080049270
8080009560
Geographical range [Log10-km2]
Fish length [Log10-cm]

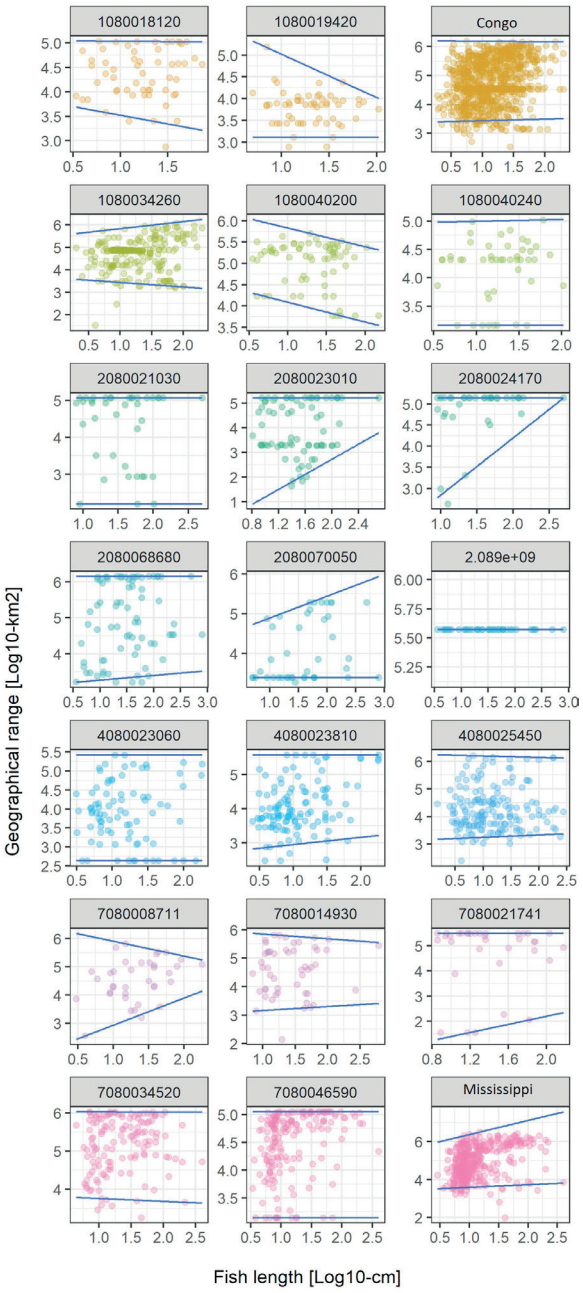


Figure D.4 | Geographical range area vs fish length and associated quantile regression models for the 0.05 and 0.95 quantiles for hydrologic basins with an area greater than 100,000 km2 and with at least 30 species according to the IUCN data (IUCN, 2018). The basin ID is according to the HydroBASINS database (Lehner and Grill, 2013), while the name of some basins are reported.

Literature cited

LITERATURE CITED

Abell, R., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S.C., Bussing, W., Stiassny, M.L.J., Skelton, P., Allen, G.R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J. V., Heibel, T.J., Wikramanayake, E., Olson, D., López, H.L., Reis, R.E., Lundberg, J.G., Sabaj Pérez, M.H., Petry, P., 2008. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. Bioscience 58, 403-414. doi:10.1641/B580507

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability. Hydrol. Sci. J. 48, 317-338. doi:10.1623/hysj.48.3.317.45290

Alexander, R.B., Boyer, E.W., Smith, R.A., Schwarz, G.E., Moore, R.B., 2007. The Role of Headwater Streams in Downstream Water Quality1. JAWRA J. Am. Water Resour. Assoc. 43, 41–59. doi:10.1111/j.1752-1688.2007.00005.x

Allen, G.H., Pavelsky, T.M., 2018. Global extent of rivers and streams. Science (80-.). 361, 585–588. doi:10.1126/science.aat0636

Allison, E.H., Perry, A.L., Badjeck, M.-C., Neil Adger, W., Brown, K., Conway, D., Halls, A.S., Pilling, G.M., Reynolds, J.D., Andrew, N.L., Dulvy, N.K., 2009. Vulnerability of national economies to the impacts of climate change on fisheries. Fish Fish. 10, 173–196. doi:10.1111/j.1467-2979.2008.00310.x

Alò, D., Turner, T.F., 2005. Effects of habitat fragmentation on effective population size in the endangered Rio Grande silvery minnow. Conserv. Biol. 19, 1138-1148. doi:10.1111/j.1523-1739.2005.00081.x

American Rivers, 2019. American Rivers Dam Removal Database [WWW Document]. doi:10.6084/m9.figshare.5234068.v5

Anselin, L., 1988. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geogr. Anal. 20, 1–17. doi:10.1111/j.1538-4632.1988.tb00159.x

Anselin, L., Rey, S., 1991. Properties of tests for spatial dependence in linear regression models. Geogr. Anal. 23, 112–131. doi:10.1111/j.1538-4632.1991.tb00228.x

Arismendi, I., Johnson, S.L., Dunham, J.B., Haggerty, R., 2013. Descriptors of natural thermal regimes in streams and their responsiveness to change in the Pacific Northwest of North America. Freshw. Biol. 58, 880–894. doi:10.1111/fwb.12094

Avcioglu, B., Anderson, C.J., Kalin, L., 2017. Evaluating the Slope-Area Method to Accurately Identify Stream Channel Heads in Three Physiographic Regions. JAWRA J. Am. Water Resour. Assoc. 53, 562–575. doi:10.1111/1752-1688.12512

Azevedo, L.B., van Zelm, R., Elshout, P.M.F., Hendriks, A.J., Leuven, R.S.E.W., Struijs, J., de Zwart, D., Huijbregts, M.A.J., 2013. Species richness-phosphorus relationships for lakes and streams worldwide. Glob. Ecol. Biogeogr. 22, 1304–1314. doi:10.1111/geb.12080

Bakker, K., 2012. Water Security: Research Challenges and Opportunities. Science (80-.). 337, 914–915. doi:10.1126/science.1226337

Balian, E. V., Lévêque, C., Segers, H., Martens, K. (Eds.), 2008. Freshwater Animal Diversity Assessment, Freshwater Animal Diversity Assessment. Springer Netherlands, Dordrecht. doi:10.1007/978-1-4020-8259-7

Barbarossa, V., Huijbregts, M.A.J., Hendriks, A.J., Beusen, A.H.W., Clavreul, J., King, H., Schipper, A.M., 2017. Developing and testing a global-scale regression model to quantify mean annual streamflow. J. Hydrol. 544, 479–487. doi:10.1016/j.jhydrol.2016.11.053

Barton, K., 2015. MuMIn: Multi-model inference. R package version 1.9.13. doi:citeulike:11961261

Beck, H.E., de Roo, A., van Dijk, A.I.J.M., 2015. Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*. J. Hydrometeorol. 16, 1478–1501. doi:10.1175/JHM-D-14-0155.1

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Dutra, E., Fink, G., Orth, R., Schellekens, J., 2016a. Global evaluation of runoff from ten state-of-the-art hydrological models. Hydrol. Earth Syst. Sci. Discuss. 1–33. doi:10.5194/hess-2016-124

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016b. Global-scale regionalization of hydrologic model parameters. Water Resour. Res. 52, 3599–3622. doi:10.1002/2015WR018247

Beck, H.E., Van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., De Roo, A., 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21, 589-615. doi:10.5194/hess-21-589-2017

Beck, H.E., Van Dijk, A.I.J.M., Miralles, D.G., De Jeu, R.A.M., Bruijnzeel, L.A., McVicar, T.R., Schellekens, J., 2013. Global patterns in base flow index and recession based on

streamflow observations from 3394 catchments. Water Resour. Res. 49, 7843–7863. doi:10.1002/2013WR013918

Béné, C., Sinaba, F., Lemoalle, J., Morand, P., Kodio, A., Andrew, N., 2012. Vulnerability and adaptation of African rural populations to hydro-climate change: experience from fishing communities in the Inner Niger Delta (Mali). Clim. Change 115, 463–483. doi:10.1007/s10584-012-0492-7

Benejam, L., Saura-Mas, S., Bardina, M., Solà, C., Munné, A., García-Berthou, E., 2016. Ecological impacts of small hydropower plants on headwater stream fish: From individual to community effects. Ecol. Freshw. Fish 25, 295-306. doi:10.1111/eff.12210

Beusen, A.H.W., Bouwman, A.F., Van Beek, L.P.H., Mogollón, J.M., Middelburg, J.J., 2016. Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. Biogeosciences 13, 2441-2451. doi:10.5194/bg-13-2441-2016

Bierkens, M.F.P., 2015. Global hydrology 2015: State, trends, and directions. Water Resour. Res. 51, 4923–4947. doi:10.1002/2015WR017173

Bierkens, M.F.P., Bell, V.A., Burek, P., Chaney, N., Condon, L.E., David, C.H., de Roo, A., Döll, P., Drost, N., Famiglietti, J.S., Flörke, M., Gochis, D.J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R.M., Reager, J.T., Samaniego, L., Sudicky, E., Sutanudjaja, E.H., van de Giesen, N., Winsemius, H., Wood, E.F., 2015. Hyper-resolution global hydrological modelling: what is next? Hydrol. Process. 29, 310–320. doi:10.1002/hyp.10391

Birnie-Gauvin, K., Franklin, P., Wilkes, M., Aarestrup, K., 2018. Moving beyond fitting fish into equations: Progressing the fish passage debate in the Anthropocene. Aquat. Conserv. Mar. Freshw. Ecosyst. 1–11. doi:10.1002/aqc.2946

Bishop, C.M., M., C., 1995. Neural networks for pattern recognition. Clarendon Press.

Bivand, R., Hauke, J., Kossowski, T., 2013. Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods. Geogr. Anal. 45, 150–179. doi:10.1111/gean.12008

Bivand, R., Piras, G., 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. J. Stat. Softw. 63. doi:10.18637/jss.v063.i18

Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Jackson, S.T., Ferrier, S., 2013. Space can substitute for time in predicting climate-change effects on biodiversity. Proc. Natl. Acad. Sci. 110, 9374–9379. doi:10.1073/pnas.1220228110

Blöschl, G., Montanari, A., 2012. Catchment scale controls the temporal connection of transpiration and diel fluctuations in streamflow. Hydrol. Process. 26, 1–16. doi:10.1002/hyp

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations, in: Journal of the Royal Statistical Society. Series B (Methodological). pp. 211--252.

Bring, J., 1994. How to Standardize Regression Coefficients. Am. Stat. 48.

Buckley, L.B., Hurlbert, A.H., Jetz, W., 2012. Broad-scale ecological implications of ectothermy and endothermy in changing environments. Glob. Ecol. Biogeogr. 21, 873–885. doi:10.1111/j.1466-8238.2011.00737.x

Burgers, H.E.R., Schipper, A.M., Jan Hendriks, A., 2014. Size relationships of water discharge in rivers: scaling of discharge with catchment area, main-stem length and precipitation. Hydrol. Process. 28, 5769–5775. doi:10.1002/hyp.10087

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vie, J.-C., Watson, R., 2010. Global Biodiversity: Indicators of Recent Declines. Science (80-.). 328, 1164–1168. doi:10.1126/science.1187512

Carlisle, D.M., Grantham, T.E., Eng, K., Wolock, D.M., 2017. Biological relevance of stream fl ow metrics : regional and national perspectives 36, 927–940. doi:10.1086/694913.

Carvajal-Quintero, J., Villalobos, F., Oberdorff, T., Grenouillet, G., Brosse, S., Hugueny, B., Jézéquel, C., Tedesco, P.A., 2019. Drainage network position and historical connectivity explain global patterns in freshwater fishes' range size. Proc. Natl. Acad. Sci. 116, 13434-13439. doi:10.1073/pnas.1902484116

Carvajal-Quintero, J.D., Januchowski-Hartley, S.R., Maldonado-Ocampo, J.A., Jézéquel, C., Delgado, J., Tedesco, P.A., 2017. Damming Fragments Species-Ranges and Heightens Extinction Risk. Conserv. Lett. 10, 708-716. doi:10.1111/conl.12336

Chang, H., 2003. Basin Hydrologic Response to Changes in Climate and Land Use: the Conestoga River Basin, Pennsylvania. Phys. Geogr. 24, 222–247. doi:10.2747/0272-3646.24.3.222

Chen, X.Y., Chau, K.W., Busari, A.O., 2015. A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. Eng. Appl. Artif. Intell. 46, 258–268. doi:10.1016/j.engappai.2015.09.010

Comte, L., Olden, J.D., 2017. Climatic vulnerability of the world's freshwater and marine fishes. Nat. Clim. Chang. 7, 718–722. doi:10.1038/nclimate3382

Cook, R.D., Weisberg, S., 1982. Residuals and influence in regression. New York: Chapman and Hall.

Cote, D., Kehler, D.G., Bourne, C., Wiersma, Y.F., 2009. A new measure of longitudinal connectivity for stream networks. Landsc. Ecol. 24, 101–113. doi:10.1007/s10980-008-9283-y

Dai, A., 2016. Historical and Future Changes in Streamflow and Continental Runoff, in: Terrestrial Water Cycle and Climate Change. John Wiley & Sons, Inc., pp. 17–37. doi:10.1002/9781118971772.ch2

Dai, A., Qian, T., Trenberth, K.E., Milliman, J.D., 2009. Changes in Continental Freshwater Discharge from 1948 to 2004. J. Clim. 22, 2773–2792. doi:10.1175/2008JCLI2592.1

Danandeh Mehr, A., Kahya, E., Olyaie, E., 2013. Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. J. Hydrol. 505, 240–249. doi:10.1016/j.jhydrol.2013.10.003

Darwall, W.R.T., Freyhof, J., 2016. Lost fishes, who is counting? The extent of the threat to freshwater fish biodiversity, in: Closs, G.P., Krkosek, M., Olden, J.D. (Eds.), Conservation of Freshwater Fishes. Cambridge University Press, Cambridge, UK, pp. 1–36.

Daufresne, M., Lengfellner, K., Sommer, U., 2009. Global warming benefits the small in aquatic ecosystems. Proc. Natl. Acad. Sci. 106, 12788–12793. doi:10.1073/pnas.0902080106

David L. Strayer, David Dudgeon, Strayer, D.L., Dudgeon, D., 2010. Freshwater biodiversity conservation: recent progress and future challenges. J. North Am. Benthol. Soc. 29, 344–358. doi:10.1899/08-171.1

Deutsch, C.A., Tewksbury, J.J., Huey, R.B., Sheldon, K.S., Ghalambor, C.K., Haak, D.C., Martin, P.R., 2008. Impacts of climate warming on terrestrial ectotherms across latitude. Proc. Natl. Acad. Sci. 105, 6668–6672. doi:10.1073/pnas.0709472105

Dhungel, S., Tarboton, D.G., Jin, J., Hawkins, C.P., 2016. Potential Effects of Climate Change on Ecologically Relevant Streamflow Regimes. River Res. Appl. 32, 1827–1840. doi:10.1002/rra.3029

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., Maina, J.M., Possingham, H.P., Rogalla von Bieberstein, K., Venter, O., Watson, J.E.M., 2017. Changing trends and persisting biases in three decades of conservation science. Glob. Ecol. Conserv. 10, 32–42. doi:10.1016/J.GECCO.2017.01.008

Dias, M.S., Tedesco, P.A., Hugueny, B., Jézéquel, C., Beauchard, O., Brosse, S., Oberdorff, T., 2017. Anthropogenic stressors and riverine fish extinctions. Ecol. Indic. 79, 37-46. doi:10.1016/j.ecolind.2017.03.053

Díaz, S., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Do, H.X., Gudmundsson, L., Leonard, M., Westra, S., 2018. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. Earth Syst. Sci. Data 10, 765–785. doi:10.5194/essd-10-765-2018

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., Wada, Y., 2016. Modelling Freshwater Resources at the Global Scale: Challenges and Prospects. Surv. Geophys. 37, 195-221. doi:10.1007/s10712-015-9343-1

Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. J. Hydrol. 270, 105–134. doi:10.1016/S0022-1694(02)00283-4

Döll, P., Zhang, J., 2010. Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations. Hydrol. Earth Syst. Sci. 14, 783–799. doi:10.5194/hess-14-783-2010

Domisch, S., Amatulli, G., Jetz, W., 2015a. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Sci. data 2, 150073. doi:10.1038/sdata.2015.73

Domisch, S., Jähnig, S.C., Simaika, J.P., Kuemmerlen, M., Stoll, S., 2015b. Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data. Fundam. Appl. Limnol 186, 1–2. doi:10.1127/fal/2015/0627

Dowlatabadi, H., 1995. Integrated Assessment Models of Climate Change Economics: An Imcomplete Overview. Energy Policy 23, 1–181. doi:10.1007/978-981-10-3945-4

Downing, J.A., Cole, J.J., Duarte, C.M., Middelburg, J.J., Melack, J.M., Prairie, Y.T., Kortelainen, P., Striegl, R.G., McDowell, W.H., Tranvik, L.J., 2012. Global abundance and size distribution of streams and rivers. Inl. Waters. doi:10.5268/IW-2.4.502

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. 81, 163. doi:10.1017/S1464793105006950

Eaton, J.G., Scheller, R.M., 1996. Effect of climate warming on fish therm habitat in streams. Limnol. Oceanogr. 41, 1109–1115.

Edwards, P., 2015. Aquaculture environment interactions: Past, present and likely future trends. Aquaculture 447, 2–14. doi:10.1016/j.aquaculture.2015.02.001

Fan, Y., Li, H., Miguez-Macho, G., 2013. Global Patterns of Groundwater Table Depth. Science (80-.). 339, 940–943. doi:10.1126/science.1229881

Fan, Y.R., Huang, W., Huang, G.H., Li, Z., Li, Y.P., Wang, X.Q., Cheng, G.H., Jin, L., 2015. A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections. Stoch. Environ. Res. Risk Assess. 29, 1557–1569. doi:10.1007/s00477-015-1048-y

FAO, 2018. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome. doi:issn 10

Farmer, W.H., Over, T.M., Vogel, R.M., 2015. Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States. Water Resour. Res. 51, 1775–1796. doi:10.1002/2014WR015924

Farmer, W.H., Vogel, R.M., 2013. Performance-weighted methods for estimating monthly streamflow at ungauged sites. J. Hydrol. 477, 240–250. doi:10.1016/j.jhydrol.2012.11.032

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Rev. Geophys. 45, RG2004. doi:10.1029/2005RG000183

Faurby, S., Araújo, M.B., 2018. Anthropogenic range contractions bias species climate change forecasts. Nat. Clim. Chang. 8, 252-256. doi:10.1038/s41558-018-0089-x

Fekete, B.M., Vörösmarty, C.J., 2007. The current status of global river discharge monitoring and potential new technologies complementing traditional discharge

measurements. Predict. Ungauged Basins PUB Kick-off (Proceedings PUB Kick-off Meet. held Bras. Novemb. 2002), IAHS Publ. no. 309 309, 129–136.

Field, A., 2009. Discovering statistics using SPSS. Sage publications.

Foley, M.M., Magilligan, F.J., Torgersen, C.E., Major, J.J., Anderson, C.W., Connolly, P.J., Wieferich, D., Shafroth, P.B., Evans, J.E., Infante, D., Craig, L.S., 2017. Landscape context and the biophysical response of rivers to dam removal in the United States. PLoS One 12, e0180107. doi:10.1371/journal.pone.0180107

Frederico, R.G., Olden, J.D., Zuanon, J., 2016. Climate change sensitivity of threatened, and largely unprotected, Amazonian fishes. Aquat. Conserv. Mar. Freshw. Ecosyst. 26, 91–102. doi:10.1002/aqc.2658

Fricko, O., Havlik, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., Ermolieva, T., Forsell, N., Herrero, M., Heyes, C., Kindermann, G., Krey, V., McCollum, D.L., Obersteiner, M., Pachauri, S., Rao, S., Schmid, E., Schoepp, W., Riahi, K., 2017. The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. Glob. Environ. Chang. 42, 251–267. doi:10.1016/j.gloenvcha.2016.06.004

Froese, R., Pauly, D., 2018. FishBase [WWW Document]. World Wide Web Electron. Publ. URL www.fishbase.org (accessed 2.11.18).

Fuller, M.R., Doyle, M.W., Strayer, D.L., 2015. Causes and consequences of habitat fragmentation in river networks. Ann. N. Y. Acad. Sci. 1355, 31–51. doi:10.1111/nyas.12853

Gao, Y., Vogel, R.M., Kroll, C.N., Poff, N.L., Olden, J.D., 2009. Development of representative indicators of hydrologic alteration. J. Hydrol. 374, 136–147. doi:10.1016/j.jhydrol.2009.06.009

GDAL Development Team, 2017. GDAL - Geospatial Data Abstraction Library, Version 2.2.0.

Gleeson, T., Wada, Y., Bierkens, M.F.P.P., Van Beek, L.P.H.H., 2012. Water balance of global aquifers revealed by groundwater footprint. Nature 488, 197–200. doi:10.1038/nature11295

Gosling, S.N., Arnell, N.W., 2011. Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. Hydrol. Process. 25, 1129–1145. doi:10.1002/hyp.7727

GRDC, 2017. Long-Term Mean Monthly Discharges and Annual Characteristics of GRDC Stations / Online provided by the Global Runoff Data Centre of WMO.

GRDC, 2011. Watershed Boundaries of GRDC Stations / Global Runoff Data Centre.

Greve, P., Kahil, T., Mochizuki, J., Schinko, T., Satoh, Y., Burek, P., Fischer, G., Tramberend, S., Burtscher, R., Langan, S., Wada, Y., 2018. Global assessment of water challenges under uncertainty in water scarcity projections. Nat. Sustain. 1, 486–494. doi:10.1038/s41893-018-0134-9

Grill, G., Khan, U., Lehner, B., Nicell, J., Ariwi, J., 2016. Risk assessment of down-the-drain chemicals at large spatial scales: Model development and application to contaminants originating from urban areas in the Saint Lawrence River Basin. Sci. Total Environ. 541, 825–838. doi:10.1016/j.scitotenv.2015.09.100

Grill, G., Lehner, B., Lumsdon, A.E., MacDonald, G.K., Zarfl, C., Reidy Liermann, C., 2015. An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales. Environ. Res. Lett. 10, 015001. doi:10.1088/1748-9326/10/1/015001

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M.E., Meng, J., Mulligan, M., Nilsson, C., Olden, J.D., Opperman, J.J., Petry, P., Reidy Liermann, C., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R.J.P., Snider, J., Tan, F., Tockner, K., Valdujo, P.H., van Soesbergen, A., Zarfl, C., 2019. Mapping the world's free-flowing rivers. Nature 569, 215-221. doi:10.1038/s41586-019-1111-9

Gudmundsson, L., Do, H.X., Leonard, M., Westra, S., 2018. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. Earth Syst. Sci. Data 10, 787–804. doi:10.5194/essd-10-787-2018

Hack, J.T., Seaton, F.A., Nolan, T.B., 1957. Studies of Longitudinal Stream Profiles in Virginia and Maryland UNITED STATES DEPARTMENT OF THE INTERIOR.

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voß, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S.N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G.P., Yeh, P., 2011. Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results. J. Hydrometeorol. 12, 869–884. doi:10.1175/2011JHM1324.1

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z.D., Wada, Y., Wisser, D., 2014. Global water resources affected by human interventions and climate change. Proc. Natl. Acad. Sci. 111, 3251–3256. doi:10.1073/pnas.1222475110

Hall, D.G., Cherry, S.J., Reeves, K.S., Lee, R.D., Carroll, G.R., Sommers, G.L., Verdin, K.L., 2004. Water Energy Resources of the United States with Emphasis on Low Head/ Low Power Resources. U.S. Department of Energy Efficiency and Renewable Energy, Wind and Hydropower Technologies, Idaho Falls, Idaho.

Hamon, W.R., 1963. Computation of direct runoff amounts from storm rainfall. Int. Assoc. Sci. Hydrol. Publ. 63, 52–62.

Hanafiah, M.M., Xenopoulos, M.A., Pfister, S., Leuven, R.S.E.W., Huijbregts, M.A.J., 2011. Characterization Factors for Water Consumption and Greenhouse Gas Emissions Based on Freshwater Fish Species Extinction. Environ. Sci. Technol. 45, 5272–5278. doi:10.1021/es1039634

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., Tanaka, K., 2008. An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. Hydrol. Earth Syst. Sci. 12, 1007–1025. doi:10.5194/hess-12-1007-2008

Hancock, G.R., 2005. The use of digital elevation models in the identification and characterization of catchments over different grid scales. Hydrol. Process. 19, 1727–1749. doi:10.1002/hyp.5632

Hannah, D.M., Demuth, S., van Lanen, H.A.J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L.M., 2011. Large-scale river flow archives: importance, current status and future needs. Hydrol. Process. 25, 1191–1200. doi:10.1002/hyp.7794

Hansen, M.C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science (80-.). 342, 850–853. doi:10.1126/science.1244693

Harfoot, M., Tittensor, D.P., Newbold, T., Mcinerny, G., Smith, M.J., Scharlemann, J.P.W., 2014. Integrated assessment models for ecologists: The present and the future. Glob. Ecol. Biogeogr. 23, 124–143. doi:10.1111/geb.12100

Harfoot, M.B.J., Newbold, T., Tittensor, D.P., Emmott, S., Hutton, J., Lyutsarev, V., Smith, M.J., Scharlemann, J.P.W., Purves, D.W., 2014. Emergent Global Patterns of Ecosystem Structure and Function from a Mechanistic General Ecosystem Model. PLoS Biol. 12. doi:10.1371/journal.pbio.1001841

Hargreaves, G.L., Hargreaves, G.H., Riley, J.P., 1985. Irrigation Water Requirements for Senegal River Basin. J. Irrig. Drain. Eng. 111, 265–275. doi:10.1061/(ASCE)0733-9437(1985)111:3(265)

Harris, I., Jones, P.D.D., Osborn, T.J.J., Lister, D.H.H., 2014. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int. J. Climatol. 34, 623–642. doi:10.1002/joc.3711

Haykin, S.S., 1994. Neural networks : a comprehensive foundation. Macmillan.

Heiberger, R.M., 2015. HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 3.1-15. URL http://CRAN.R-project.org/package=HH.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236. doi:10.5194/esd-4-219-2013

Hendriks, A.J., Schipper, A.M., Caduff, M., Huijbregts, M.A.J., 2012. Size relationships of water inflow into lakes: Empirical regressions suggest geometric scaling. J. Hydrol. 414, 482–490. doi:10.1016/j.jhydrol.2011.11.025

Herbert, M.E., Gelwick, F.P., 2006. Spatial Variation of Headwater Fish Assemblages Explained by Hydrologic Variability and Upstream Effects of Impoundment. Copeia 2003, 273–284. doi:10.1643/0045-8511(2003)003[0273:SVOHFA]2.0.CO;2

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. doi:10.1002/joc.1276

Hoekstra, A.Y., Chapagain, A.K., Aldaya, M.M., Mekonnen, M.M., 2011. The Water Footprint Assessment Manual, Febrero 2011. doi:978-1-84971-279-8

Hof, C., Araújo, M.B., Jetz, W., Rahbek, C., 2011. Additive threats from pathogens, climate and land-use change for global amphibian diversity. Nature 480, 516–519. doi:10.1038/nature10650

Holmlund, C.M., Hammer, M., 1999. Ecosystem services generated by fish populations. Ecol. Econ. 29, 253–268. doi:10.1016/S0921-8009(99)00015-4

Hortness, J.E., Berenbrock, C., 2001. Estimating monthly and annual streamflow statistics at ungaged sites in Idaho. US Geological Survey.

Intralawan, A., Wood, D., Frankel, R., Costanza, R., Kubiszewski, I., 2018. Tradeoff analysis between electricity generation and ecosystem services in the Lower Mekong Basin. Ecosyst. Serv. 30, 27–35. doi:10.1016/j.ecoser.2018.01.007

Isaak, D.J., Muhlfeld, C.C., Todd, A.S., Al-chokhachy, R., Roberts, J., Kershner, J.L., Fausch, K.D., Hostetler, S.W., 2012. The Past as Prelude to the Future for Understanding 21st-

Century Climate Effects on Rocky Mountain Trout. Fisheries 37, 542–556. doi:10.1080/03632415.2012.742808

Ishiyama, N., Ryo, M., Kataoka, T., Nagayama, S., Sueyoshi, M., Terui, A., Mori, T., Akasaka, T., Nakamura, F., 2018. Predicting the ecological impacts of large-dam removals on a river network based on habitat-network structure and flow regimes. Conserv. Biol. 32, 1403–1413. doi:10.1111/cobi.13137

Istanbulluoglu, E., Tarboton, D.G., Pack, R.T., Luce, C., 2002. A probabilistic approach for channel initiation. Water Resour. Res. 38, 61-1-61–14. doi:10.1029/2001WR000782

IUCN, 2018. The IUCN Red List of Threatened Species. Version 2018-2 [WWW Document]. URL http://www.iucnredlist.org (accessed 11.13.18).

Iwasaki, Y., Ryo, M., Sui, P., Yoshimura, C., 2012. Evaluating the relationship between basin-scale fish species richness and ecologically relevant flow characteristics in rivers worldwide. Freshw. Biol. 57, 2173–2180. doi:10.1111/j.1365-2427.2012.02861.x

Janse, J.H., Kuiper, J.J., Weijters, M.J., Westerbeek, E.P., Jeuken, M.H.J.L., Bakkenes, M., Alkemade, R., Mooij, W.M., Verhoeven, J.T.A., 2015. GLOBIO-Aquatic, a global model of human impact on the biodiversity of inland aquatic ecosystems. Environ. Sci. Policy 48, 99–114. doi:10.1016/j.envsci.2014.12.007

Janssen, A.B., Janse, J.H., Beusen, A.H., Chang, M., Harrison, J.A., Huttunen, I., Kong, X., Rost, J., Teurlincx, S., Troost, T.A., van Wijk, D., Mooij, W.M., 2019. How to model algal blooms in any lake on earth. Curr. Opin. Environ. Sustain. doi:10.1016/j.cosust.2018.09.001

Jefferies, D., Muñoz, I., Hodges, J., King, V.J., Aldaya, M., Ercin, A.E., Milà i Canals, L., Hoekstra, A.Y., 2012. Water Footprint and Life Cycle Assessment as approaches to assess potential impacts of products on water consumption. Key learning points from pilot studies on tea and margarine. J. Clean. Prod. 33, 155–166. doi:10.1016/j.jclepro.2012.04.015

Jia, B., Tang, Y., Tian, L., Franz, L., Alewell, C., Huang, J.-H., 2015. Impact of Fish Farming on Phosphorus in Reservoir Sediments. Sci. Rep. 5, 16617. doi:10.1038/srep16617

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N., 2009. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Res. Appl. n/a-n/a. doi:10.1002/rra.1249

King, S., O'Hanley, J.R., Newbold, L.R., Kemp, P.S., Diebel, M.W., 2017. A toolkit for optimizing fish passage barrier mitigation actions. J. Appl. Ecol. 54, 599–611. doi:10.1111/1365-2664.12706

Knouft, J.H., Ficklin, D.L., 2017. The Potential Impacts of Climate Change on Biodiversity in Flowing Freshwater Systems. Annu. Rev. Ecol. Evol. Syst. 48, annurev-ecolsys-110316-022803. doi:10.1146/annurev-ecolsys-110316-022803

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259-263. doi:10.1127/0941-2948/2006/0130

Krider, L.A., Magner, J.A., Perry, J., Vondracek, B., Ferrington, L.C., 2013. Air-Water temperature relationships in the trout streams of southeastern Minnesota's carbonate-sandstone landscape. J. Am. Water Resour. Assoc. 49, 896–907. doi:10.1111/jawr.12046

Labay, B.J., Hendrickson, D.A., Cohen, A.E., Bonner, T.H., King, R.S., Kleinsasser, L.J., Linam, G.W., Winemiller, K.O., 2015. Can Species Distribution Models Aid Bioassessment when Reference Sites are Lacking? Tests Based on Freshwater Fishes. Environ. Manage. 56, 835–846. doi:10.1007/s00267-015-0567-0

Latrubesse, E.M., Arima, E.Y., Dunne, T., Park, E., Baker, V.R., D'Horta, F.M., Wight, C., Wittmann, F., Zuanon, J., Baker, P.A., Ribas, C.C., Norgaard, R.B., Filizola, N., Ansar, A., Flyvbjerg, B., Stevaux, J.C., 2017. Damming the rivers of the Amazon basin. Nature 546, 363–369. doi:10.1038/nature22333

Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R., Bartlein, P.J., 2009. Projected climate-induced faunal change in the Western Hemisphere. Ecology 90, 588–97.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241. doi:10.1029/1998WR900018

Lehner, B., Grill, G., 2013. Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27. doi:10.1002/hyp.9740

Lehner, B., Liermann, C.R., Revenga, C., Vörömsmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Front. Ecol. Environ. 9, 494-502. doi:10.1890/100125

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. Eos (Washington. DC). 89, 93–94. doi:10.1029/2008EO100001

Leuven, R.S., Posthuma, L., Huijbregts, M.A., Struijs, J., De Zwart, D., 2010. Field sensitivity distribution of macroinvertebrates for phosphorus in inland waters. Integr. Environ. Assess. Manag. 7, 280–286. doi:10.1002/ieam.141

Liermann, C.R., Nilsson, C., Robertson, J., Ng, R.Y., 2012. Implications of Dam Obstruction for Global Freshwater Fish Diversity. Bioscience 62, 539–548. doi:10.1525/bio.2012.62.6.5

Lima, C.H.R., Lall, U., 2010. Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow. J. Hydrol. 383, 307–318. doi:10.1016/j.jhydrol.2009.12.045

Lin, P., Pan, M., Beck, H.E., Yang, Y., Yamazaki, D., Frasson, R., David, C.H., Durand, M., Pavelsky, T.M., Allen, G.H., Gleason, C.J., Wood, E.F., 2019. Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resour. Res. 2019WR025287. doi:10.1029/2019WR025287

Lu, J., Sun, G., McNulty, S.G., Amatya, D.M., 2005. A COMPARISON OF SIX POTENTIAL EVAPOTRANSPIRATION METHODS FOR REGIONAL USE IN THE SOUTHEASTERN UNITED STATES. J. Am. Water Resour. Assoc. 41, 621–633. doi:10.1111/j.1752-1688.2005.tb03759.x

MacDonald, L.H., Coe, D., 2007. Influence of headwater streams on downstream reaches. For. Sci. 53, 148–168.

Maheu, A., Poff, N.L., St-Hilaire, A., 2016. A Classification of Stream Water Temperature Regimes in the Conterminous USA. River Res. Appl. 32, 896–906. doi:10.1002/rra.2906

McClain, M.E., Naiman, R.J., 2008. Andean Influences on the Biogeochemistry and Ecology of the Amazon River. Bioscience 58, 325–338. doi:10.1641/B580408

McCully, P., 2001. Silenced rivers: the ecology and politics of large dams, G - Reference, Information and Interdisciplinary Subjects Series. Zed Books.

McIntyre, P.B., Reidy Liermann, C.A., Revenga, C., 2016. Linking freshwater fishery management to global food security and biodiversity conservation. Proc. Natl. Acad. Sci. 113, 12880–12885. doi:10.1073/pnas.1521540113

McManamay, R.A., Griffiths, N.A., DeRolph, C.R., Pracheil, B.M., 2018. A Synopsis of Global Mapping of Freshwater Habitats and Biodiversity: Implications for Conservation. Pure Appl. Biogeogr. 2, 64. doi:10.5772/intechopen.70296

Merow, C., Wolf, F., Leclère, D., Ware, C., Di Marco, M., Arneth, A., Titeux, N., Espinoza, F., Purvis, A., Baisero, D., Havlík, P., Gonzalez, R.E., Sharp, R., Takahashi, K., Obersteiner, M., Harwood, T.D., Ohashi, H., Hellweg, S., Kim, H., Popp, A., Gueguen, M., Jetz, W.,

Pereira, H.M., Rosa, I.M.D., De Palma, A., Visconti, P., Martins, I.S., Hill, S.L.L., van Vuuren, D.P., Janse, J.H., Di Fulvio, F., Rondinini, C., Hasegawa, T., Johnson, J.A., Haverd, V., Chini, L., Chaplin-Kramer, R., Schipper, A.M., Hoskins, A.J., Fujimori, S., Ferrier, S., Caton, E., Harfoot, M., Quesada, B., Poulter, B., Leadley, P., Guerra, C., Alkemade, R., Matsui, T., Krause, A., Hurtt, G., Thuiller, W., Anthoni, P., Hirata, A., 2018. A protocol for an intercomparison of biodiversity and ecosystem services models using harmonized land-use and climate scenarios. Geosci. Model Dev. 11, 4537–4562. doi:10.5194/gmd-11-4537-2018

Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-Being: Synthesis, Ecosystems.

Morrissey-McCaffrey, E., Shephard, S., Kelly, F.L., Kelly-Quinn, M., 2018. Non-native species and lake warming negatively affect Arctic char Salvelinus alpinus abundance; deep thermal refugia facilitate co-existence. J. Fish Biol. doi:10.1111/jfb.13837

Mulligan, M., Saenz-Cruz, L., van Soesbergen, A., Smith, V.T., Zurita, L., 2009. Global dams database and geowiki. Version 1 [WWW Document]. URL http://geodata.policysupport.org/dams

Nelson, J.S., 2006. Fishes of the world, Fourth. ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Nilsson, C., Reidy, C.A., Dynesius, M., Revenga, C., 2005. Fragmentation and Flow Regulation of the World ' s Large River Systems. Science (80-.). 308, 405–408. doi:10.1126/science.1107887

O'Connor, J.E., Duda, J.J., Grant, G.E., 2015. 1000 dams down and counting. Science (80-.). 348, 496–497. doi:10.1126/science.aaa9204

Oberdorff, T., Guegan, J.-F., Hugueny, B., 1995. Global scale patterns of fish species richness in rivers. Ecography (Cop.). 18, 345–352. doi:10.1111/j.1600-0587.1995.tb00137.x

Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., Dürr, H.H., 2011. Global and regional patterns in riverine fish species richness: A review. Int. J. Ecol. 2011, 1-12. doi:10.1155/2011/967631

Okkan, U., Serbes, Z.A., 2012. Rainfall-runoff modeling using least squares support vector machines. Environmetrics 23, 549–564. doi:10.1002/env.2154

Olden, J.D., Hogan, Z.S., Zanden, M.J. Vander, 2007. Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. Glob. Ecol. Biogeogr. 16, 694–701. doi:10.1111/j.1466-8238.2007.00337.x

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

Oldenkamp, R., Hoeks, S., Čengić, M., Barbarossa, V., Burns, E.E., Boxall, A.B.A., Ragas, A.M.J., 2018. A High-Resolution Spatial Model to Predict Exposure to Pharmaceuticals in European Surface Waters: ePiE. Environ. Sci. Technol. 52, 12494–12503. doi:10.1021/acs.est.8b03862

Opperman, J.J., Royte, J., Banks, J., Rose Day, L., Apse, C., 2011. The Penobscot River, Maine, USA: a Basin-Scale Approach to Balancing Power Generation and Ecosystem Restoration. Ecol. Soc. 16, art7. doi:10.5751/ES-04117-160307

Ord, K., 1975. Estimation Methods for Models of Spatial Interaction. J. Am. Stat. Assoc. 70, 120–126. doi:10.1080/01621459.1975.10480272

Ove Hoegh-Guldberg, Jacob, D., Taylor, M., et al, 2018. Impacts of 1.5ºC global warming on natural and human systems. Glob. Warm. 1.5 °C - IPCC's Spec. Assess. Rep. doi:10.1093/aje/kwp410

Parson, E.A., Fisher-Vanden, and K., 2002. Integrated Assessment Models of Global Climate Change. Annu. Rev. Energy Environ. 22, 589–628. doi:10.1146/annurev.energy.22.1.589

Parson, E.A., Fisher-vanden, K., Keith, D., 2007. Global-Change Scenarios : Their Development and Use US Department of Energy US Department of Energy Publications Global-Change Scenarios : Their Development and Use.

Peel, M.C., Blöschl, G., 2011. Hydrological modelling in a changing world. Prog. Phys. Geogr. Earth Environ. 35, 249–261. doi:10.1177/0309133311402550

Pfister, S., Koehler, A., Hellweg, S., 2009. Assessing the Environmental Impacts of Freshwater Consumption in LCA. Environ. Sci. Technol. 43, 4098–4104. doi:10.1021/es802423e

Piccolroaz, S., Calamita, E., Majone, B., Gallice, A., Siviglia, A., Toffolon, M., 2016. Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. Hydrol. Process. 30, 3901–3917. doi:10.1002/hyp.10913

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

Poff, N.L., Schmidt, J.C., 2016. How dams can go with the flow. Science (80-.). 353, 1099–1100. doi:10.1126/science.aah4926

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

Poff, N.L.R., 2018. Beyond the natural flow regime? Broadening the hydro-ecological foundation to meet environmental flows challenges in a non-stationary world. Freshw. Biol. 63, 1011–1021. doi:10.1111/fwb.13038

Pokhrel, Y., Burbano, M., Roush, J., Kang, H., Sridhar, V., Hyndman, D., 2018. A Review of the Integrated Effects of Changing Climate, Land Use, and Dams on Mekong River Hydrology. Water 10, 266. doi:10.3390/w10030266

Pokhrel, Y., Yeh, P.J.-F., Kim, H., Oki, T., Hanasaki, N., Koirala, S., Kanae, S., Cho, J., 2012. Incorporating anthropogenic water regulation modules into a land surface model. J. Hydrometeorol. 13, 255–269. doi:10.1175/JHM-D-11-013.1

Powers, R.P., Jetz, W., 2019. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. Nat. Clim. Chang. doi:10.1038/s41558-019-0406-z

Rabi, A., Hadzima-Nyarko, M., Šperac, M., 2015. Modelling river temperature from air temperature: case of the River Drava (Croatia). Hydrol. Sci. J. 60, 1490–1507. doi:10.1080/02626667.2014.914215

Radinger, J., Essl, F., Hölker, F., Horký, P., Slavík, O., Wolter, C., 2017. The future distribution of river fish: The complex interplay of climate and land use changes, species dispersal and movement barriers. Glob. Chang. Biol. 23, 4970–4986. doi:10.1111/gcb.13760

Radinger, J., Wolter, C., 2015. Disentangling the effects of habitat suitability, dispersal, and fragmentation on the distribution of river fishes. Ecol. Appl. 25, 914–927. doi:10.1890/14-0422.1

Radinger, J., Wolter, C., 2014. Patterns and predictors of fish dispersal in rivers. Fish Fish. 15, 456–473. doi:10.1111/faf.12028

Razavi, T., Coulibaly, P., 2013. Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. J. Hydrol. Eng. 18, 958–975. doi:10.1061/(ASCE)HE.1943-5584.0000690

Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., Kidd, K.A., MacCormack, T.J., Olden, J.D., Ormerod, S.J., Smol, J.P., Taylor, W.W., Tockner, K., Vermaire, J.C., Dudgeon, D., Cooke, S.J., 2019. Emerging threats and persistent

conservation challenges for freshwater biodiversity. Biol. Rev. 94, 849–873. doi:10.1111/brv.12480

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Glob. Environ. Chang. 42, 153–168. doi:10.1016/j.gloenvcha.2016.05.009

Richter, B.D., Baumgartner, J. V, Powell, J., Braun, D.P., 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. Conserv. Biol. 10, 1163–1174. doi:10.2307/2387152

Ripple, W.J., Wolf, C., Newsome, T.M., Hoffmann, M., Wirsing, A.J., McCauley, D.J., 2017. Extinction risk is most acute for the world's largest and smallest vertebrates. Proc. Natl. Acad. Sci. 114, 10678–10683. doi:10.1073/pnas.1702078114

Rodriguez-Iturbe, I., Rinaldo, A., 2001. Fractal river basins: chance and self-organization. Cambridge University Press.

Rogelj, J., Schleussner, C.F., Hare, W., 2017. Getting It Right Matters: Temperature Goal Interpretations in Geoscience Research. Geophys. Res. Lett. 44, 10,662-10,665. doi:10.1002/2017GL075612

Russell, P.P., Gale, S.M., Muñoz, B., Dorney, J.R., Rubino, M.J., 2015. A Spatially Explicit Model for Mapping Headwater Streams. JAWRA J. Am. Water Resour. Assoc. 51, 226–239. doi:10.1111/jawr.12250

Sangireddy, H., Stark, C.P., Kladzyk, A., Passalacqua, P., 2016. GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. Environ. Model. Softw. 83, 58–73. doi:10.1016/j.envsoft.2016.04.026

Santini, M., di Paola, A., 2015. Changes in the world rivers' discharge projected from an updated high resolution dataset of current and future climate zones. J. Hydrol. 531, 768–780. doi:10.1016/j.jhydrol.2015.10.050

Sassolas-Serrayet, T., Cattin, R., Ferry, M., 2018. The shape of watersheds. Nat. Commun. 9, 1–8. doi:10.1038/s41467-018-06210-4

Schaphoff, S., Forkel, M., Müller, C., Knauer, J., von Bloh, W., Gerten, D., Jägermeyr, J., Lucht, W., Rammig, A., Thonicke, K., Waha, K., 2018. LPJmL4 - a dynamic global vegetation model with managed land - Part 1: Model description. Geosci. Model Dev. 11, 1343–1375. doi:10.5194/gmd-2017-146

Schellekens, J., Dutra, E., Martínez-De La Torre, A., Balsamo, G., Van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., Van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., Weedon, G.P., 2017. A global water resources ensemble of hydrological models: The eartH2Observe Tier-1 dataset. Earth Syst. Sci. Data 9, 389–413. doi:10.5194/essd-9-389-2017

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J., Gosling, S.N., Kim, H., Liu, X., Masaki, Y., Portmann, F.T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., Kabat, P., 2014. Multimodel assessment of water scarcity under climate change. Proc. Natl. Acad. Sci. 111, 3245–3250. doi:10.1073/pnas.1222460110

Schmitt, R.J.P., Bizzi, S., Castelletti, A., Kondolf, G.M., 2018. Improved trade-offs of hydropower and sand connectivity by strategic dam planning in the Mekong. Nat. Sustain. 1, 96–104. doi:10.1038/s41893-018-0022-3

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2015. GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. doi:10.5676/DWD_GPCC/FD_M_V7_050

Seneviratne, S.I., Rogelj, J., Séférian, R., Wartenburger, R., Allen, M.R., Cain, M., Millar, R.J., Ebi, K.L., Ellis, N., Hoegh-Guldberg, O., Payne, A.J., Schleussner, C.F., Tschakert, P., Warren, R.F., 2018. The many possible climates from the Paris Agreement's aim of 1.5 °c warming. Nature 558, 41–49. doi:10.1038/s41586-018-0181-4

Shiklomanov, a. I., Lammers, R.B., Vörösmarty, C.J., 2002. Widespread decline in hydrological monitoring threatens Pan-Arctic Research. Eos, Trans. Am. Geophys. Union 83, 13–17. doi:10.1029/2002EO000007

Silva, A.T., Lucas, M.C., Castro-Santos, T., Katopodis, C., Baumgartner, L.J., Thiem, J.D., Aarestrup, K., Pompeu, P.S., O'Brien, G.C., Braun, D.C., Burnett, N.J., Zhu, D.Z., Fjeldstad, H.-P., Forseth, T., Rajaratnam, N., Williams, J.G., Cooke, S.J., 2018. The future of fish passage science, engineering, and practice. Fish Fish. 19, 340–362. doi:10.1111/faf.12258

Sivapalan, M., 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrol. Process. 17, 3163–3170. doi:10.1002/hyp.5155

Steel, E.A., Beechie, T.J., Torgersen, C.E., Fullerton, A.H., 2017. Envisioning, Quantifying, and Managing Thermal Regimes on River Networks. Bioscience 67, 506–522. doi:10.1093/biosci/bix047

Stuckey, M.H., 2006. Low-flow, base-flow, and mean-flow regression equations for Pennsylvania streams. US Geological Survey.

Sutanudjaja, E.H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J.H.C., Drost, N., van der Ent, R.J., de Graaf, I.E.M., Hoch, J.M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M.W., Vannametee, E., Wisser, D., Bierkens, M.F.P., 2018. PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. Geosci. Model Dev. 11, 2429–2453. doi:10.5194/gmd-11-2429-2018

Syvitski, J.P.., Peckham, S.D., Hilberman, R., Mulder, T., 2003. Predicting the terrestrial flux of sediment to the global ocean: a planetary perspective. Sediment. Geol. 162, 5–24. doi:10.1016/S0037-0738(03)00232-X

Syvitski, J.P.M., 2005. Impact of Humans on the Flux of Terrestrial Sediment to the Global Coastal Ocean. Science (80-.). 308, 376–380. doi:10.1126/science.1109454

Tarboton, D.G., 2008. Terrain Analysis Using Digital Elevation Models (Taudem).

Tarboton, D.G., Bras, R.L., Rodriguez-iturbe, I., 1992. A physical basis for drainage density. pdf. Geomorphology 5, 59–76.

Tarboton, D.G., Bras, R.L., Rodriguez-Iturbe, I., 1991. On the extraction of channel networks from digital elevation data. Hydrol. Process. 5, 81–100. doi:10.1002/hyp.3360050107

Tarboton, D.G., Schreuders, K.A.T., Watson, D.W., Baker, M.E., 2009. Generalized terrain-based flow analysis of digital elevation models, in: The 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Cairns, Australia from 13–17 July 2009. Cairns, Australia, pp. 2377–2383.

Tedesco, P.A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., Cornu, J.F., Dias, M.S., Grenouillet, G., Hugueny, B., Jézéquel, C., Leprieur, F., Brosse, S., Oberdorff, T., 2017. Data Descriptor: A global database on freshwater fish species occurrence in drainage basins. Sci. Data 4, 1–6. doi:10.1038/sdata.2017.141

Tedesco, P.A., Oberdorff, T., Cornu, J.-F., Beauchard, O., Brosse, S., Dürr, H.H., Grenouillet, G., Leprieur, F., Tisseuil, C., Zaiss, R., Hugueny, B., 2013. A scenario for impacts of

water availability loss due to climate change on riverine fish extinction rates. J. Appl. Ecol. 50, n/a-n/a. doi:10.1111/1365-2664.12125

Tendall, D.M., Hellweg, S., Pfister, S., Huijbregts, M.A.J., Gaillard, G., 2014. Impacts of river water consumption on aquatic biodiversity in life cycle assessment-a proposed method, and a case study for Europe. Environ. Sci. Technol. 48, 3236–3244. doi:10.1021/es4048686

Tesfa, T.K., Tarboton, D.G., Watson, D.W., Schreuders, K.A.T.T., Baker, M.E., Wallace, R.M., 2011. Extraction of hydrological proximity measures from DEMs using parallel processing. Environ. Model. Softw. 26, 1696–1709. doi:10.1016/j.envsoft.2011.07.018

Tewksbury, J., Huey, R., Deutsch, C., 2008. Putting heat on tropical animals. Science (80-.). 320, 1296–1297.

Thomas, D.M., Benson, M.A., 1970. Generalization of streamflow characteristics from drainage-basin characteristics.

Thompson, J.A., Bell, J.C., Butler, C.A., 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma 100, 67–89. doi:10.1016/S0016-7061(00)00081-1

Thornthwaite, C.W., 1948. An Approach toward a Rational Classification of Climate. Geogr. Rev. 38, 55. doi:10.2307/210739

Thorp, J.H., Thoms, M.C., Delong, M.D., 2006. The riverine ecosystem synthesis: biocomplexity in river networks across space and time. River Res. Appl. 22, 123–147. doi:10.1002/rra.901

Tisseuil, C., Cornu, J.-F., Beauchard, O., Brosse, S., Darwall, W., Holland, R., Hugueny, B., Tedesco, P.A., Oberdorff, T., 2013. Global diversity patterns and cross-taxa convergence in freshwater systems. J. Anim. Ecol. 82, 365–376. doi:10.1111/1365-2656.12018

Toffolon, M., Piccolroaz, S., 2015. A hybrid model for river water temperature as a function of air temperature and discharge. Environ. Res. Lett. 10. doi:10.1088/1748-9326/10/11/114011

Tran, L.T., O'Neill, R. V., Bruins, R.J.F., Smith, E.R., Harden, C., 2015. Linking land use/land cover with climatic and geomorphologic factors in regional mean annual streamflow models with geospatial regression approach. Prog. Phys. Geogr. Earth Environ. 39, 258–274. doi:10.1177/0309133314562441

Urban, M.C., 2015. Accelerating extinction risk from climate change. Science (80-.). 348, 571–573. doi:10.1126/science.aaa4984

Van Beek, L.P.H., Bierkens, M.F.P., 2009. The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification. Utr. Univ. Utrecht, Netherlands 1, 25–26.

van Beek, L.P.H., Wada, Y., Bierkens, M.F.P., 2011. Global monthly water stress: 1. Water balance and water availability. Water Resour. Res. 47. doi:10.1029/2010WR009791

Van Der Knijff, J.M., Younis, J., De Roo, A.P.J., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. Int. J. Geogr. Inf. Sci. 24, 189–212. doi:10.1080/13658810802549154

van Puijenbroek, P.J.T.M., Buijse, A.D., Kraak, M.H.S., Verdonschot, P.F.M., 2019. Species and river specific effects of river fragmentation on European anadromous fish species. River Res. Appl. 35, 68–77. doi:10.1002/rra.3386

van Vliet, M.T.H., Ludwig, F., Kabat, P., 2013. Global streamflow and thermal habitats of freshwater fishes under climate change. Clim. Change 121, 739–754. doi:10.1007/s10584-013-0976-0

van Vuuren, D.P., Stehfest, E., Gernaat, D.E.H.J., Doelman, J.C., van den Berg, M., Harmsen, M., de Boer, H.S., Bouwman, L.F., Daioglou, V., Edelenbosch, O.Y., Girod, B., Kram, T., Lassaletta, L., Lucas, P.L., van Meijl, H., Müller, C., van Ruijven, B.J., van der Sluis, S., Tabeau, A., 2017. Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. Glob. Environ. Chang. 42, 237–250. doi:10.1016/j.gloenvcha.2016.05.008

Verdin, K.L., Godt, J.W., Funk, C., Pedreros, D., Worstell, B., Verdin, J., 2007. Development of a Global Slope Dataset for Estimation of Landslide Occurrence Resulting from Earthquakes, Colorado: U.S. Geological Survey, Open-File Report.

Verdin, K.L., Worstell, B., 2008. A fully distributed implementation of mean annual streamflow regional regression equations. J. Am. Water Resour. Assoc. 44, 1537–1547. doi:10.1111/j.1752-1688.2008.00258.x

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., Alcamo, J., 2012. Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe. J. Hydrol. 424–425, 238–251. doi:10.1016/j.jhydrol.2012.01.005

Visconti, P., Bakkenes, M., Baisero, D., Brooks, T., Butchart, S.H.M., Joppa, L., Alkemade, R., Di Marco, M., Santini, L., Hoffmann, M., Maiorano, L., Pressey, R.L., Arponen, A., Boitani, L., Reside, A.E., van Vuuren, D.P., Rondinini, C., 2016. Projecting Global Biodiversity Indicators under Future Development Scenarios. Conserv. Lett. 9, 5–13. doi:10.1111/conl.12159

Vogel, R.M., Sankarasubramanian, A., 2010. Spatial scaling properties of annual streamflow in the United States. Hydrol. Sci. J. 45, 465–476. doi:10.1080/02626660009492342

Vogel, R.M., Wilson, I., Daly, C., 1999. Regional Regression Models of Annual Streamflow for the United States. J. Irrig. Drain. Eng. 125, 148–157. doi:10.1061/(ASCE)0733-9437(1999)125:3(148)

Vorosmarty, C.J., Fekete, B.M., Tucker, B.A., 1998. Global river discharge, 1807–1991, V. 1.1 (RivDIS). Data set. Available on-line [http//www. daac. ornl. gov] from Oak Ridge Natl. Lab. Distrib. Act. Arch. Center, Oak Ridge, TN, USA.

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

Wada, Y., van Beek, L.P.H., Viviroli, D., Dürr, H.H., Weingartner, R., Bierkens, M.F.P., 2011. Global monthly water stress: 2. Water demand and severity of water stress. Water Resour. Res. 47. doi:10.1029/2010WR009792

Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dynam 5, 15–40. doi:10.5194/esd-5-15-2014

Walsh, R.P.D., Lawler, D.M., 1981. Rainfall seasonality: description, spatial patterns and change through time. Weather 36, 201–208. doi:10.1002/j.1477-8696.1981.tb05400.x

Wanders, N., van Vliet, M.T.H., Wada, Y., Bierkens, M.F.P., van Beek, L.P.H., 2019. High-resolution global water temperature modelling. Water Resour. Res. doi:10.1029/2018WR023250

Wanders, N., Wada, Y., 2015. Human and climate impacts on the 21st century hydrological drought. J. Hydrol. 526, 208–220. doi:10.1016/j.jhydrol.2014.10.047

Wang, W., Chau, K., Xu, D., Chen, X.-Y., 2015. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. Water Resour. Manag. 29, 2655–2675. doi:10.1007/s11269-015-0962-6

Warren, R., Price, J., Graham, E., Forstenhaeusler, N., VanDerWal, J., 2018. The projected effect on insects, vertebrates, and plants of limiting global warming to 1.5°C rather than 2°C. Science (80-.). 360, 791–795. doi:10.1126/science.aar3646

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., Schewe, J., 2014. The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. Proc. Natl. Acad. Sci. 111, 3228–3232. doi:10.1073/pnas.1312330110

Widén-Nilsson, E., Halldin, S., Xu, C., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. J. Hydrol. 340, 105–118. doi:10.1016/j.jhydrol.2007.04.002

Wiel, K., Wanders, N., Selten, F.M., Bierkens, M.F.P., 2019. Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. Geophys. Res. Lett. 46, 2093–2102. doi:10.1029/2019GL081967

Winemiller, K.O., McIntyre, P.B., Castello, L., Fluet-Chouinard, E., Giarrizzo, T., Nam, S., Baird, I.G., Darwall, W., Lujan, N.K., Harrison, I., Stiassny, M.L.J., Silvano, R.A.M., Fitzgerald, D.B., Pelicice, F.M., Agostinho, A.A., Gomes, L.C., Albert, J.S., Baran, E., Petrere, M., Zarfl, C., Mulligan, M., Sullivan, J.P., Arantes, C.C., Sousa, L.M., Koning, A.A., Hoeinghaus, D.J., Sabaj, M., Lundberg, J.G., Armbruster, J., Thieme, M.L., Petry, P., Zuanon, J., Vilara, G.T., Snoeks, J., Ou, C., Rainboth, W., Pavanelli, C.S., Akama, A., Soesbergen, A. v., Saenz, L., van Soesbergen, A., Saenz, L., 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. Science (80-.). 351, 128–129. doi:10.1126/science.aac7082

Wisser, D., Fekete, B.M., Vörösmarty, C.J., Schumann, A.H., 2010. Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network-Hydrology (GTN-H). Hydrol. Earth Syst. Sci. 14, 1–24. doi:10.5194/hess-14-1-2010

Wohl, E., 2017. The significance of small streams. Front. Earth Sci. 11, 447–456. doi:10.1007/s11707-017-0647-y

World Comission on Dams, 2000. Damd and development. A new framework for decision-making, Report. Earthscan Publications Ltd.

World Meteorological Organization, 1992. International Meteorological Vocabulary. Geneva.

Wu, C.L., Chau, K.W., Li, Y.S., 2009. Methods to improve neural network performance in daily flows prediction. J. Hydrol. 372, 80–93. doi:10.1016/j.jhydrol.2009.03.038

Wu, H., Kimball, J.S., Li, H., Huang, M., Leung, L.R., Adler, R.F., 2012. A new global river network database for macroscale hydrologic modeling. Water Resour. Res. 48. doi:10.1029/2012WR012313

Wu, H., Kimball, J.S., Mantua, N., Stanford, J., 2011. Automated upscaling of river networks for macroscale hydrological modeling. Water Resour. Res. 47. doi:10.1029/2009WR008871

WWAP (United Nations World Water Assessment Programme), 2015. The United Nations world water development report 2015: water for a sustainable world. UNESCO Publishing, Paris.

WWF, 2018. Living Planet Report 2018: Aiming Higher (Eds M. Grooten and R. E. A. Almond). Gland, Switzerland.

WWF, 2016. Living Planet Report 2016. Risk and resilience in a new era. Gland, Switzerland.

X. Wang, J. R. Williams, P. W. Gassman, C. Baffaut, R. C. Izaurralde, J. Jeong, J. R. Kiniry, 2013. EPIC and APEX: Model Use, Calibration, and Validation. Trans. ASABE 55, 1447–1462. doi:10.13031/2013.42253

Xenopoulos, M.A., Lodge, D.M., 2006. Going with the Flow : Using Species-Discharge Relationships to Forecast Losses in Fish Biodiversity Published by : Ecological Society of America content in a trusted digital archive . We use information technology and tools to increase productivity and fa 87, 1907–1914.

Xenopoulos, M.A., Lodge, D.M., Alcamo, J., Marker, M., Schulze, K., Van Vuuren, D.P., 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. Glob. Chang. Biol. 11, 1557–1564. doi:10.1111/j.1365-2486.2005.001008.x

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. J. Hydrol. doi:10.1016/j.jhydrol.2015.10.038

Yu, K. xia, Gottschalk, L., Xiong, L., Li, Z., Li, P., 2015. Estimation of the annual runoff distribution from moments of climatic variables. J. Hydrol. 531, 1081–1094. doi:10.1016/j.jhydrol.2015.11.012

Zarfl, C., Lumsdon, A.E., Berlekamp, J., Tydecks, L., Tockner, K., 2015. A global boom in hydropower dam construction. Aquat. Sci. 77, 161–170. doi:10.1007/s00027-014-0377-0

Ziv, G., Baran, E., Nam, S., Rodriguez-Iturbe, I., Levin, S.A., 2012. Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. Proc. Natl. Acad. Sci. 109, 5609–5614. doi:10.1073/pnas.1201423109

Zurell, D., Graham, C.H., Gallien, L., Thuiller, W., Zimmermann, N.E., 2018. Long-distance migratory birds threatened by multiple independent risks from global change. Nat. Clim. Chang. 8, 992–996. doi:10.1038/s41558-018-0312-9

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R. Springer Science & Business Media.

Summary
Samenvatting
Riassunto

SUMMARY

Freshwater systems across the globe are subject to significant modification by human activities. Integrated assessments models can help to understand impacts of human activities on natural systems and to inform policy makers about optimal strategies to preserve ecosystems and the essential services they provide. Various key aspects of freshwater systems, however, remain underrepresented in integrated assessments models. The aim of this thesis was to improve the representation of freshwater systems in integrated assessments by establishing global models of 1) hydrological variables at high spatial resolution and 2) human impacts on freshwater fish species, with a focus on dams and climate change. Ultimately, this thesis thus provides tools to advance the understanding of human impacts on freshwater systems.

An adequate spatially explicit representation of hydrological variables at a global extent is key to correctly predict, among others, concentrations of harmful contaminants in fresh water and the integrity of freshwater ecosystems. Chapters 2 and 3 explore data-driven modelling strategies to obtain global spatially continuous layers of streamflow variables. Chapter 2 presents a multiple regression approach to quantify mean annual streamflow based on data from 1,885 monitoring stations with associated catchment basins ranging from 2 to 10^6 km^2 in size. The model relates the streamflow at the mouth of the catchment to a set of catchment-based characteristics expected to influence streamflow, i.e., precipitation, temperature, slope, elevation and area. The regression model explained 89% of the variance among the observations and performed better (RMSE=0.29-0.38) than a process-based global hydrological model (RMSE=0.49-0.57; based on log-transformed values). Motivated by the high predictive accuracy of the regression model, Chapter 3 employs supervised artificial neural networks based on data from 6,600 monitoring stations to map maximum, minimum and mean annual streamflow at ~1x1 km^2 grid-cell resolution across the globe and for every year from 1960 through 2015. FLO1K is the first database reporting year-specific streamflow metrics at such high spatial resolution.

Freshwater biodiversity has so far been underrepresented in global assessments and there is an urgent need to more accurately quantify the potential impacts of human activities on freshwater biodiversity. Chapters 4 and 5 focused on the development of species-based models to quantify impacts of dams and climate change on fish geographical ranges worldwide. Chapter 4 presents an assessment of dam impacts quantified in terms of connectivity loss for ~5,700 riverine freshwater fish species. For each species, a connectivity index between 0 (no connectivity) and 100 (full connectivity) was calculated based on ~40,000 current dams and for ~3,700 dams that are under construction or planned. The results showed lower range connectivity for

fish species that complete their life cycle in freshwater (CI=69 ± 30%; mean ± standard deviation) as opposed to fish that migrate to or from the marine environment (CI=82 ± 22%). Connectivity losses due to the development of new hydropower dams were especially high in the tropics, with mean declines of 20-30 CI points for species in the Amazon, Niger, Congo and Mekong basins. Chapter 5 adopted a similar species-based approach to quantify potential geographical range contractions of ~7,000 freshwater fish species due to future climate change. For each species, a maximum weekly water temperature and a minimum weekly flow threshold were established by overlaying species occurrence ranges with data on flow and water temperature. These thresholds were then used to estimate potential range contractions due to projected changes in water temperature or flow at different levels of global warming, according to four representative concentration pathway scenarios and five global climate models. Water flow and temperature were modelled with the global hydrological model PCR-GLOBWB 2.0 at ~10x10 km^2 spatial resolution. The assessment indicates that 20% of freshwater fish species will lose >50% of their current geographical range at 2°C warming, while intensifying mitigation efforts to limit warming to 1.5°C lowers this percentage to 13% (all warming levels relative to pre-industrial). In comparison, 44% of the fish species will lose >50% of their range in a "current pledges" scenario (3.2°C) and 69% (±7%) in case no mitigation action will take place (4.5°C).

Based on the findings of this thesis it was concluded that data-driven approaches are a powerful tool to estimate hydrological variables at high spatial-resolution and with (potentially) greater accuracy than process-based models. Data-driven models for global applications are only in their infancy and the increasing availability of remote sensing data has the potential to improve the temporal accuracy of data-driven models and allow the quantification of wider range of hydrological variables. The freshwater fish distribution models developed in this thesis constitute a major first step to a global species-level impact assessment-modeling framework for freshwater biodiversity. Future studies may improve the models by including a wider range of stressors, including for example aspects of water quality, and include more realism, such as dispersal ability of fish and barrier passability. The tools and results of this thesis can be linked to broader integrated assessment modeling frameworks to more comprehensively assess human impacts on freshwater systems and, ultimately, better guide water management and policy.

SAMENVATTING

Zoetwatersystemen zijn wereldwijd onderhevig aan veranderingen als gevolg van menselijke activiteiten. Om de omvang van deze veranderingen te kwantificeren en beleidsmakers te helpen bij het ontwikkelen van optimale strategieën voor shet behoud van gezonde zoetwatersystemen, zijn grootschalige integrale beoordelingsmodellen en –methoden nodig. Verschillende belangrijke aspecten van zoetwatersystemen zijn echter ondervertegenwoordigd in bestaande integrale beoordelingsmodellen. Het doel van dit proefschrift was om de representatie van zoetwatersystemen in mondiale integrale beoordelingen te verbeteren door 1) het kwantificeren van hydrologische variabelen op hoge ruimtelijke resolutie en 2) het bepalen van de effecten van dammen en klimaatverandering op de verspreidingsgebieden van zoetwatervissen. Hiermee draagt het proefschrift bij aan een verbetering van grootschalige methoden voor het bepalen van de effecten van menselijk handelen op zoetwatersystemen.

Een gedetailleerde ruimtelijk expliciete weergave van hydrologische variabelen is van cruciaal belang voor het bepalen van, onder andere, concentraties van verontreinigingen in oppervlaktewater en de integriteit van zoetwaterecosystemen. Hoofdstuk 2 en 3 beschrijven datagedreven modelleringsstrategieën om rivierafvoer te berekenen op hoge ruimtelijke resolutie wereldwijd. Hoofdstuk 2 presenteert een meervoudige regressiemodel om de jaargemiddelde rivierafvoer te kwantificeren op basis van gegevens van 1.885 meetstations met bijbehorende stroomgebieden variërend in grootte van 2 tot 10^6 km^2. Het model relateert de gemeten afvoer aan een aantal kenmerken van het stroomgebied die naar verwachting de afvoer beïnvloeden, d.w.z. neerslag, temperatuur, helling, hoogte en oppervlakte. Het regressiemodel verklaarde 89% van de variantie tussen de waarnemingen en bleek nauwkeuriger dan een mondiaal hydrologisch procesmodel (RMSE van 0,29-0,38 t.o.v. 0,49-0,57; gebaseerd op log-getransformeerde waarden). In hoofdstuk 3 is gebruik gemaakt van machine learning (supervised artifical neural network) om rivierafvoer in kaart te brengen op basis van afvoergegevens van 6.600 meetstations. Dit heeft geresulteerd in de dataset FLO1K: maximale, minimale en gemiddelde jaarlijkse rivierafvoer op een resolutie van ~ 1x1 km^2, voor elk jaar vanaf 1960 tot en met 2015. FLO1K is de eerste mondiale database met jaarspecifieke afvoerdata op een dergelijke hoge ruimtelijke resolutie.

De biodiversiteit van zoetwatersystemen is tot dusverre sterk ondervertegenwoordigd in grootschalige integrale milieubeoordelingen en er is een dringende behoefte om de potentiële effecten van menselijke activiteiten op de zoetwaterbiodiversiteit nauwkeuriger te kwantificeren. Hoofdstuk 4 en 5 beschrijven methoden voor het kwantificeren van de effecten van dammen en klimaatverandering op de verspreidingsgebieden van vissen. In hoofdstuk 4 is voor ~ 5.700 vissoorten de mate van

fragmentatie van hun verspreidingsgebied berekend als gevolg van de aanwezigheid van dammen. Voor elk soort werd een connectiviteitsindex berekend tussen 0 (geen connectiviteit; volledig gefragmenteerd) en 100 (volledige connectiviteit; geen fragmentatie), op basis van een mondiale dataset met ~ 40.000 bestaande dammen en ~ 3.700 dammen die in aanbouw of gepland zijn. De verspreidingsgebieden van vissoorten die hun levenscyclus in zoet water voltooien bleken sterker gefragmenteerd (CI = 69 ± 30%; gemiddelde ± standaardafwijking) dan de verspreidingsgebieden van soorten die migreren tussen zoetwater en mariene systemen (CI = 82 ± 22%). Fragmentatie als gevolg van de ontwikkeling van nieuwe waterkrachtdammen bleek vooral op te treden in de tropen, met gemiddelde dalingen van 20-30 CI-punten voor soorten in de stroomgebieden van de Amazone, Niger, Congo en Mekong.

In hoofdstuk 5 is de potentiële inkrimping van het verspreidingsgebied als gevolg van klimaatverandering gekwantificeerd voor ~ 7.000 zoetwatervissoorten. Voor elk soort is eerst de maximale wekelijkse watertemperatuur en de minimale wekelijkse rivierafvoer binnen het verspreidingsgebied vastgesteld op basis van ruimtelijk expliciete gegevens over rivierafvoer en watertemperatuur. Deze drempelwaarden zijn vervolgens gebruikt om potentiële inkrimping van het verspreidingsgebied in te schatten als gevolg van geprojecteerde veranderingen in watertemperatuur of stroming bij verschillende niveaus van opwarming van de aarde, gebaseerd op vier scenario's voor broeikasgasemissies en vijf mondiale klimaatmodellen. De rivierafvoer en temperatuur zijn gemodelleerd met het globale hydrologische model PCR-GLOBWB 2.0, op een ruimtelijke resolutie van ~ 10x10 km^2. Bij een opwarming van 2^{o}C zal voor 20% van de zoetwatervissoorten meer dan 50% van hun huidige verspreidingsgebied mogelijk ongeschikt worden als gevolg van veranderingen in afvoer of temperatuur. Dit percentage daalt van 20 naar 13% indien de opwarming beperkt blijft tot 1,5 °C en stijgt van 20 naar 44% bij een opwarming van 3,2^{o}C (t.o.v. een pre-industriële situatie). Bij een opwarming van 4,5^{o}C zal 69% van de vissoorten mogelijk meer dan 50% van hun verspreidingsgebied verliezen.

Op basis van de bevindingen in dit proefschrift kan worden geconcludeerd dat datagedreven benaderingen in staat zijn om hydrologische variabelen te kwantificeren op een hoge ruimtelijke resolutie en met een (potentieel) grotere nauwkeurigheid dan procesmodellen. Datagedreven hydrologische modellen staan nog in de kinderschoenen en de nauwkeurigheid en temporele resolutie van de schattingen van de hydrologische variabelen zullen naar verwachting kunnen worden verbeterd met behulp van de toenemende beschikbaarheid van data (o.a. vanuit remote sensing). De modellen die in dit proefschrift zijn ontwikkeld voor het bepalen van de effecten van klimaatveranderingen en fragmentatie op de verspreidingsgebieden van vissen vormen

een belangrijke eerste stap naar een mondiaal en soortspecifiek beoordelingsinstrument voor de biodiversiteit in zoet water. Toekomstige studies kunnen de modellen verbeteren door een breder scala aan relevante omgevingsfactoren op te nemen, waaronder bijvoorbeeld aspecten van waterkwaliteit, en meer realisme toe te voegen, zoals het verspreidingsvermogen van vissen en de passeerbaarheid van dammen. De methoden en resultaten in dit proefschrift kunnen worden gekoppeld aan bestaande integrale beoordelingsmodellen om de menselijke effecten op zoetwatersystemen beter te beoordelen en, uiteindelijk, een betere leidraad te vormen voor waterbeheer en waterbeleid.

RIASSUNTO

Gli ecosistemi di acqua dolce presenti sul nostro pianeta sono soggetti a sostanziali modifiche ad opera dell'uomo. I cosiddetti Integrated Assessment Models (IAMs) - modelli progettati per studiare l'interazione tra i settori dell'economia, dell'energia, dell'uso del suolo e il clima - costituiscono uno strumento essenziale per comprendere l'impatto delle attività umane sui sistemi naturali e portare i *policy makers* a conoscenza delle più efficienti strategie per preservare gli ecosistemi e i molteplici benefici che questi forniscono al genere umano. Vari aspetti chiave degli ecosistemi di acqua dolce rimangono tuttavia sottostimati negli IAMs. Con questo lavoro di tesi si intende fornire dei modelli numerici che siano applicabili su scala globale per 1) variabili idrologiche ad alta risoluzione spaziale e 2) l'impatto sulla ittiofauna mondiale d'acqua dolce di attività umane, quali lo sbarramento dei corsi d'acqua ad opera di dighe, e il cambiamento climatico. In definitiva, questa tesi fornisce strumenti essenziali per migliorare la comprensione delle conseguenze delle attività antropogeniche sugli ecosistemi di acqua dolce.

Un'adeguata rappresentazione spaziale delle variabili idrologiche su scala globale è essenziale, ad esempio, per predire concentrazioni nocive di sostanze contaminanti nell'acqua e comprendere l'integrità degli ecosistemi di acqua dolce. Nei capitoli 2 e 3 sono esplorate strategie di modellazione numerica di tipo *data-driven* per ottenere una coerente rappresentazione spaziale della portata di flusso dei fiumi della Terra. Nel capitolo 2 è presentato un approccio di regressione lineare multipla utile a quantificare la portata media annua di flusso basandosi su dati provenienti da 1.885 stazioni di monitoraggio e dei bacini idrografici associati di dimensioni comprese tra 2 e 10^6 km^2. Il modello mette in relazione la portata di flusso allo sbocco del bacino con un insieme di caratteristiche basate sullo stesso, quali la precipitazione atmosferica, la temperatura dell'aria, il pendio medio del versante, l'elevazione e l'area del bacino idrografico. Il modello di regressione ha spiegato l'89% della varianza delle osservazioni ed ha esibito una prestazione più soddisfacente (RMSE = 0,29-0,38) rispetto ad un modello idrologico globale di tipo *process-based* (RMSE = 0,49-0,57; RMSE si riferisce a valori di portata trasformati in scala logaritmica). Considerata la precisione predittiva del modello statistico di regressione sviluppato nel Capitolo 2, nel Capitolo 3 viene impiegata una modellazione numerica basata su *Artificial Neural Networks,* e sui dati provenienti da 6.600 stazioni di monitoraggio, per mappare la portata annua massima, minima e media di flusso ad una risoluzione spaziale di ~1x1 km^2 su scala globale, e per ogni anno dal 1960 al 2015. *FLO1K* è il primo database che riporta stime di vari indici di portata annua con una risoluzione spaziale così elevata.

Allo stato dell'arte attuale non risultano soddisfacenti rappresentazioni della biodiversità di acqua dolce nelle valutazioni su scala globale ed è pertanto urgente quantificare con maggiore precisione a che livello sia stata e sarà impattata dalle attività umane. Nei capitoli 4 e 5 sono sviluppati modelli numerici che sono *species-based* (la modellazione viene fatta per ogni singola specie separatamente), per quantificare gli impatti delle dighe e del cambiamento climatico sull'habitat di specie di pesci d'acqua dolce di tutto il mondo. Nel capitolo 4 viene presentata una valutazione dell'impatto derivante dallo sbarramento dei corsi fluviali ad opera di dighe, impatto quantificato in termini di perdita di connettività longitudinale e modellato per circa 5.700 specie di pesci d'acqua dolce. Per ogni specie è stato calcolato un indice di connettività compreso tra 0 (connettività nulla) e 100 (connettività massima) basato su ~40.000 dighe esistenti e per ~3.700 dighe che al momento sono in costruzione o la cui realizzazione è stata pianificata. I risultati hanno riportato una connettività inferiore per specie che completano il ciclo di vita in acqua dolce (CI = 69 ± 30%; media ± deviazione standard) rispetto alle specie ittiche che migrano da o verso ambienti marini (CI = 82 ± 22%). Le perdite di connettività dovute allo sviluppo di nuove dighe per centrali idroelettriche saranno particolarmente elevate ai tropici, con una riduzione media del 20-30% per le specie che abitano i bacini dell'Amazzonia, del Niger, del Congo e del Mekong. Nel capitolo 5 è stato adottato un simile approccio *species-based* per quantificare potenziali contrazioni delle aree geografiche di circa 7.000 specie ittiche d'acqua dolce in risposta al cambiamento climatico. Per ogni specie, sono state stabilite una massima temperatura dell'acqua e una portata di flusso minima settimanali consentite, sovrapponendo le aree geografiche delle specie con i dati su temperatura dell'acqua e portata. Queste soglie sono state quindi utilizzate per stimare le potenziali contrazioni delle aree geografiche a cui potrebbero condurre ipotetiche variazioni di temperatura dell'acqua o della portata di flusso causate dal riscaldamento globale, a sua volta quantificato relativamente alle ipotesi formulate da quattro *Representative Concentration Pathways* in combinazione con cinque modelli climatici globali. La portata fluviale e la temperatura dell'acqua sono state quantificate con il modello idrologico a scala globale PCR-GLOBWB 2.0, ad una risoluzione spaziale di ~10x10 km^2. La valutazione mostra che il 20% delle specie di pesci d'acqua dolce perderà più della metà della loro area geografica per via di un innalzamento della temperatura di 2°C. In uno scenario di maggiore mitigazione climatica, per cui si può ipotizzare un aumento limitato a 1,5°C, questa percentuale si ridurrebbe al 13% (tutti i livelli di riscaldamento globale sono relativi alla temperatura atmosferica media di circa la seconda meta' del XIX secolo). Di contro, in uno scenario in cui non vi sia alcuna mitigazione climatica, una volta adempiuti i *National Determined Contributions* (3,2°C), il 44% delle specie ittiche rischia di perdere piu' della metà del

proprio spazio vitale, cifra che sale fino al 69% nel caso in cui non vi sia alcun intervento di mitigazione (4,5°C).

A partire dai risultati delle ricerche esposti in questa sede si può concludere che i cosiddetti approcci *data-driven* siano un potente strumento per stimare le variabili idrologiche ad alta risoluzione spaziale e con un'accuratezza potenzialmente maggiore rispetto ai modelli *process-based*. I modelli *data-driven* per le applicazioni globali sono solo agli esordi e grazie alla crescente disponibilità di dati di telerilevamento, si presenta la possibilità di un netto miglioramento dell'accuratezza temporale di modelli di tipo *data-driven*; in aggiunta, ciò consentirebbe la quantificazione di una gamma ancor più ampia di variabili idrologiche. I modelli per l'ittiofauna d'acqua dolce sviluppati in questa tesi rappresentano il primo passo di un processo che porterà ad ottenere un quadro definito degli impatti antropogenici sulla biodiversità di acqua dolce su scala globale. Studi futuri potrebbero migliorare tali modelli includendo una gamma più ampia di fattori di stress, come ad esempio variabili rappresentive della qualità delle acque ed un maggior realismo, come, ad esempio, la capacità di dispersione dei pesci e il grado di insormontabilità delle dighe. Gli strumenti e i risultati di questa tesi possono essere collegati tra di loro ed utilizzati in valutazioni integrate di sostenibilità ambientale con l'intento di migliorare la rappresentazione dell'impatto delle attività umane sui sistemi di acqua dolce e, in definitiva, migliorare la gestione delle risorse idriche.

Acknowledgments
About the author

ACKNOWLEDGMENTS

It has been a long journey, full of satisfactions, critical and inspiring discussions, but also countless stressful periods and unexpected rejections. Looking back, this journey has brought me to work in many offices across different countries, but mostly to meet many new people. This final section is dedicated to all the wonderful people that made the realization of this thesis possible, to all of those who believed in me, cheered with me in good times and remained by my side during tough periods.

Aafke, thank you for the active and thoughtful role you held throughout my PhD, not only as a supervisor but also as a friend. You have always believed in my potential, listened to my voice in the many collaborations and always tried to make room for one of my typical "5 minutes" chats. Since the very start, it has been easy but mostly a lot of fun working with you. You have taught me much of what I know now and I am truly grateful for that. I hope you will forgive me for not putting any fish pictures in this thesis, but please beware that even if my PhD time is over I am still expecting a Christmas fish every year (say until retirement at best!).

Mark, I will never forget the day I came to the department and your immediate trust in me, to me you have been a true mentor and role model. We surely have had our differences, especially when it came to make a final decision for models' setups (not sure I will ever be able to re-pay all those 5.000 euros tickets you issued!), yet, I feel that the often-intense discussions between Aafke, you and myself have really fortified my thought and approach to research. I really hope I will be able to continue collaborating with you in the future and keep making great science together (who knows, we might be finally able to get to talk about actual water footprints!).

Henry, even if topic-wise there was not always a clear overlap of interest, especially when I was pursuing the more theoretical hydro-informatics matters, I always appreciated your insights and constructive feedback. I really enjoyed the one and a half year spent at Unilever and I am truly thankful for having been given that opportunity. Seeing how business operates alongside science has truly enriched my perspective on scientific research. With this occasion, I would also like to extend my thanks to the rest of the Sustainability team of SEAC that made my stay at Unilever a special one. Sarah, Michal, Julie, Hilde, Ed, Fabrizio, Tirma, Niteen, Giles, thanks for your time, sharing your ideas and, of course, casual talks over Sarah's freshly baked sweets (West wing SEAC rules!).

It feels natural and most necessary at this point to acknowledge those who dared to embark on this journey with me. Sadegh, Melanie, Wan Yee and Sandra, thanks for making the RELIEF project a truly collective journey. Sharing with you all the most frustrating moments definitely helped me to get through all the lengthy bureaucracy,

planning and different stages of the PhD. At the time that I am writing this final words, it seems like most of us are going to make it quite on time, and that must certainly be a RELIEF (I know it's a lame joke, but someone had to say it!).

Then, I would like to extend my thanks to all the people I met at the Environmental Science department of Radboud University. Especially, I would like to thank Ana, for being a real friend (and roomie for a while) from the very early days of my PhD. To me, you represent a clear example that it is possible to be crazy fun and a great scholar at the same time. One further thought goes to the great memories I collected during my stay at Radboud. The kanjam sessions during the lunch breaks with Jelle, Zoran, Thomas and Rik (the "old crew"), the 24/7 salmon grilling with Steef (in Norway), and the casual corridor chats and laughs with Rosalie, Lisette, Juan, Mirza, Joyce, Luca (et al.; I wil stop here as the list is becoming very long). I would also like to thank the people I have been working with and met at PBL, from whom I have learned a lot on how to relate scientific research to policy questions. Serious policy stuff aside, I also recollect the many coffee moments and fun "borrels" here at PBL. So thank you for making my stay a better one, Clara, Arlette, Machteld and the rest of the NLG "youngish" group. Additional "hella" thanks go to the Natural Capital group of Stanford University, that enriched yet another chapter of my PhD experience with high-quality academic talks and great Californian vibes; so thank you Rafael and Becky for having made this experience possible and a special one.

It is time now to thank the different Italian crews, and I would like to start this paragraph by dedicating a few words to the one that formed in Bedford. Fabrizio, Carlo, Antonella, Mattia e Serena, grazie per le innumerevoli serate passate insieme (special thanks to: la pasta fatta in casa di Mattia, i taralli di Antonella, il vino di casa Serena, i prosciutti di Fabrizio e le sedie tagliate a metà di Carlo!), le piccole escursioni e gli infiniti addii. Senza di voi, la mia permanenza a Bedford di certo non sarebbe stata altrettanto divertente. Quindi, agli amici di sempre, al gruppo della "Capa" di Perugia, grazie per essere stati un enorme e costante sostegno morale, festeggiando con me di persona o a distanza ogni piccolo traguardo di questo lungo percorso accademico. Grazie Mark, Rob, Nik, Trot, Babi, Fra, Ricco, Sas^2, Nico, Andrea (al quale vorrei dedicare un ringraziamento speciale per il tempo impiegato a correggere la traduzione del riassunto di questa tesi in Italiano!) & the girls.

In maniera forse ovvia ma non di certo scontata, vorrei a questo punto ringraziare di cuore la mia famiglia, per il supporto incondizionato durante il mio intero percorso accademico che eventualmente mi ha portato a raggiungere, con il compimento di questa tesi, il tanto agognato titolo di dottore. Grazie per esere una insostituibile costante, mamma, papà, zio Carlo e zia Francesca, nonna Adele, Andrea e Fabio.

Speaking of family, I would also like to thank the recently acquired Iranian part of my family, a true aid during my struggles and joyful moments of the last four years, kheyli mamnoon maman Maboubeh and agha Philip, ghorboonet beram! To conclude, my dearest Maral, nothing of this would have ever been possible without your full support and love. Merci for being my everywhere adventure companion, closest friend and a true rock throughout these four long years. Grazie di tutto, azizam!

ABOUT THE AUTHOR

Curriculum Vitae

My passion and curiosity for natural and earth sciences has defined my path since the very beginning of my studies. I attended a science-major high school (the so-called "Liceo Scientifico") in my hometown Perugia, Italy where I grew up. Thereafter I decided to pursue a BSc in Geology, which I graduated from in 2012 ("cum laude"). During my bachelor, I was awarded with a scholarship to write my thesis in Stockholm at the Natural History Museum, which was the official start of my international research endeavors. For my MSc, I was awarded another scholarship to attend a double-degree program between the Milan-Bicocca University (Italy) and the State University of New York at Buffalo (NY, USA). Given my growing interest for water resources and sustainability, I specialized in hydrogeology and for my thesis, I conducted a numerical modelling study to aid the remediation of a contaminated aquifer. Immediately after my graduation in 2015, I started my PhD at Radboud University as part of the EU project RELIability of Environmental Footprints of consumer products (RELIEF). Here, I specialized in the development of large-scale hydrological and ecological numerical models for global environmental sustainability assessments. The high mobility that characterized my PhD allowed me to collaborate with many experienced scholars from different universities as well as industry and policy institutes. In fact, I spent a few months as a visiting researcher at Stanford University joining the Natural Capital project group. Further, the one and a half year stay at Unilever R&D (UK) taught me how environmental sustainability is applied in a multinational company and how science can be effectively communicated to business. Throughout my stay in the Netherlands, I collaborated with the Netherlands Environmental Assessment Agency (PBL) in The Hague as part of the global biodiversity assessment group GLOBIO, where I learned how environmental assessments support policymaking. At PBL, I have been appointed as a post-doctoral researcher since August 2019.

Publications

(To be) peer-reviewed journal articles

Barbarossa, V., Schmitt, R. J. P., Huijbregts, M. A. J., Zarfl, C., King, H., & Schipper A. M. Worldwide impacts of existing and future dams on the connectivity of freshwater fish ranges. *Submitted*.

Barbarossa, V., Bosmans, J., Huijbregts, M. A. J., Wanders, N., Bierkens, M. F. P., King, H., & Schipper A. M. Implications of global warming for the world's freshwater fishes. *Submitted*.

Bjørn, A., Sim, S., Boulay, A. M., King, H., Clavreul, J., Lam, W. Y., Barbarossa, V., Bulle, C., & Margni, M. (2020). A planetary boundary-based method for freshwater use in life cycle assessment: Development and application to a tomato production case study. Ecological Indicators, 110, 105865.

Oldenkamp, R., Hoeks, S., Čengić, M., Barbarossa, V., Burns, E. E., Boxall, A. B., & Ragas, A. M. (2018). A high-resolution spatial model to predict exposure to pharmaceuticals in European surface waters: EPiE. *Environmental science & technology*, *52*(21), 12494-12503.

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

Barbarossa, V., Huijbregts, M. A. J., Hendriks, A. J., Beusen, A. H., Clavreul, J., King, H., & Schipper, A. M. (2017). Developing and testing a global-scale regression model to quantify mean annual streamflow. *Journal of hydrology*, *544*, 479-487.

Conference proceedings

Barbarossa, V., Schmitt, R. J. P., Huijbregts, M. A. J., Zarfl, C., & Schipper A. (2019) Assessing existing and future dam impacts on the connectivity of freshwater fish ranges worldwide. *American Geophysical Union Fall Meeting 2019.* San Francisco, USA

Nguyen, H., Barbarossa, V., Turner, S.W.D., Buckley, B.M., & Galelli, S. (2019) A high-resolution paleo streamflow record for Monsoon Asia with surrogate streamflow data. *23rd International Congress on Modelling and Simulation*. Canberra, Australia

Barbarossa, V., Bosmans, J., Huijbregts, M. A. J., Wanders, N., Bierkens, M. F. P., King, H., & Schipper, A. M. (2019) Global habitat loss of freshwater fish species in a 1.5-degree warmer world. *European Geoscience Union General Assembly Annual Meeting 2019, Vol. 21, p. 8920.* Vienna, Austria

Barbarossa, V., Bosmans, J., Huijbregts, M., King, H., & Schipper, A. (2018) Impacts of climate change on the world's freshwater fish species. *International Conference Water Science for Impact*. Wageningen, The Netherlands

Barbarossa, V., Schipper, A., Bosmans, J., King, H., Bierkens, M., Wanders, N., & Huijbregts, M. (2018) Incorporating climate change impacts on local freshwater fish species richness in LCIA. *SETAC Europe 24th LCA Symposium*. Vienna, Austria

Barbarossa, V., Huijbregts, M., Hendriks, J., Beusen, A., Clavreul, J., King, H., & Schipper, A. (2017) Global-scale high-resolution (1 km) modelling of mean, maximum and minimum annual streamflow. *European Geoscience Union General Assembly Annual Meeting 2017, Vol. 19, p. 6721*. Vienna, Austria

Nazzareni, S., Barbarossa, V., Skogby, H., & Zanon, V. (2014) Clinopyroxenes from Pico Volcano (Azores Island, Portugal): crystal chemistry and water content. *Rend. Online Soc. Geol. It., Suppl. n. 1 al Vol. 31, 368*. Milan, Italy

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Valerio Barbarossa

born on 16 June 1989 in Perugia, Italy

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Nijmegen, 6 February 2020

The Chairman of the SENSE board

Prof. dr. Martin Wassen

the SENSE Director of Education

Dr. Ad van Dommelen

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Valerio Barbarossa** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 46.1 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2016)
- Research in context activity: 'Organizing and chairing session "Impacts of global environmental change on freshwater biodiversity" at conference Water Science for Impact, Wageningen (2018)
- Zero Inflated Models in R (2018)
- Geostatistics (2018)
- Aquatic Ecology: Robustness of aquatic ecosystems in the face of global change (2018)

Other PhD and Advanced MSc Courses

- Life Cycle Assessment, Norwegian University of Science and Technology (2016)
- Scientific Writing for PhD Candidates, Radboud University (2016)
- Career guidance course, Radboud University (2019)

External training at a foreign research institute

- Business awareness course, Unilever (2017)
- Business application course, Unilever (2017)

Management and Didactic Skills Training

- Supervising MSc student with thesis entitled 'A review of models and applications for freshwater species-discharge relationships' (2016)
- Teaching in the MSc courses 'Environmental and ecological modelling' (2015) and 'Research skills' (2016)

Oral Presentations

- *Impacts of climate change on the world's freshwater fish species.* Water Science for Impact, 16-18 October 2018, Wageningen, The Netherlands
- *Global habitat loss of freshwater fish species in a 1.5-degree warmer world.* EGU general assembly, 7-12 April 2019, Vienna, Austria
- *Climate-change and fragmentation impacts on the freshwater fish species worldwide.* GLOBIO day, 5 February 2019, Utrecht, The Netherlands
- *Impacts of major dams on the connectivity of freshwater fish habitat worldwide.* Global dams watch/Europe free flowing rivers workshop, 16-18 April 2019 WWF, Zeist, The Netherlands

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen