



Estimating water residence time distribution in river networks by boosted regression trees (BRT) model

Meili Feng^{1,2,3}, Martin Pusch², Markus Venohr²

5 ¹School of Geographical Sciences, University of Nottingham Ningbo China, Ningbo, 31500, China

²Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, 12587, Germany

³Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, 38123, Italy

Correspondence to: Meili Feng (meili.feng@nottingham.edu.cn)

Abstract. In-stream water residence time (WRT) in river networks is a crucial driver for biogeochemical processes in riverine ecosystems. Dynamics of the WRT are critical for understanding and modelling nutrient retention in lakes and rivers, in particular during flood events when riparian areas are inundated. This study illustrates the potential utility of integrating spatial landscape analysis with machine learning statistics to understand the effects of hydrology and geomorphology on WRT in river networks, especially at large scales. We applied the Boosted Regression Trees (BRT) approach to estimate water residence, a promising multi-regression spatial distribution model with consistent cross-validation procedure, and identified the crucial factors of influence. Reach-average WRTs were estimated for the annual mean hydrologic conditions as well as the flood and drought month, respectively. Results showed that the three most contributing factors in shaping the WRT distribution are river discharge (57%), longitudinal slope (21%), and the drainage area (15%). This study enables the identification of key controlling factors of the reach-average WRT and estimation of WRT under varying hydrological conditions. The resulting distribution model of WRT is an easy to apply and sound approach helping to improve water quality modelling at larger scales and water management approaches aiming to estimate nutrient fluxes in river systems.

Keywords: Water residence time; river networks; spatial distribution model; Boosted Regression Trees (BRT).

1 Introduction

Water residence time (WRT) (also known as in-stream water residence time, Worrall et al., 2014) refers to the average time that a certain amount of water travels through a defined river reach. Reach-average WRT represents one of the most important determinants for in-stream biogeochemistry recycling processes (Catalán et al., 2016; Drummond et al., 2016; Ensign and Doyle, 2006; Gibson, 2000; Hrachowitz et al., 2016; Stanley and Doyle, 2002). The response of river flow to precipitation is highly nonlinear, and so are the in-stream processes of water retention (Heidbüchel et al., 2012). Water residence time in river networks differs due to the variability of inflow rates, river topologic and geomorphologic parameters such as slope (Doyle et al., 2005; Wang et al., 2015). Residence time studies especially for extreme hydrologic conditions (i.e. flood and drought events) are of particular importance for water management practice as here significant share of annual fluxes can be transported or retained during a short period of time.



Studies on WRT are often based on process-based, deterministic models for hydrological processes including groundwater, precipitation and surface runoff in a river basin (i.e. SWAT (Grizzetti et al., 2003), SPARROW (Preston Seitzinger et al., 2011)). However, these deterministic models are data demanding and time consuming when applying to networks of large river systems. Besides modelling, another widely applied way of measuring water residence time is by introducing solutes and measured residence time and flow velocity within specific river reach. WRT can be estimated based on the travel time of dissolved solute tracers that are experimentally added to the river, which have been used to analyze retention efficiency and up-take length of dissolved nutrients may be retained by bio-chemical processes (Drummond et al., 2016; Nieuwenhuysse, 2005; Soulsby et al., 2006). Further improvements of the process-based models will require addressing spatial heterogeneities within basins (Mayorga et al., 2010) and a better understanding of river network retention and the driving factors controlled by runoff within watershed (Arora et al., 2016; Dumont et al., 2005).

Computational and empirical methods (e.g. MONERIS, Venohr et al., 2011) offer diversified options in combining them to statistical and process-based models at different scales (Gottschalk et al., 2006; Nieuwenhuysse, 2005; Soulsby et al., 2006). The one-dimensional hydraulic Manning-Strickler formula, which calculates flow velocity in dependence of channel slope and cross-section variations, has been widely used to estimate flow velocity and thus water residence time (Verzano et al., 2012; Worrall et al., 2014). However, distributed data to model the WRT in river networks, especially at larger scales, is often lacking or only available in insufficient spatial resolution. Consequently, any modelling approach addressing WRT or related in-stream processes at larger scales is limited by the quality of observed or alternatively on estimated data on geomorphological (roughness, slope, sediment) parameters. The estimation of hydrological regimes in complex river systems still remain controversial between detailed process-based models on one side, and over simplified empirical methods on the other side. This gap appears to be even larger when it comes to the application of large-scale river basins.

To improve the understanding of WRT as carrier and as driving force for instream processes, while considering impacts hydro-morphological of river channels characteristics (Poole, 2010), Behrendt and Opitz (1999) discussed the dependence of nutrient retention on specific runoff and hydraulic load in river systems. Gücker and Boëchat (2004) investigated the ammonium retention in tropical headwater streams with differing channel morphology and hydraulic characteristics such as riffles and pools. Doyle and Stanley (2003) highlighted the controlling position of hydro-geomorphological factors in nutrient cycling processes. Helton et al. (2018) reemphasized the importance of structures of the whole stream network in nitrogen transformation and removal with varied but integrated spatial distribution from headwaters to downstream. To determine how differences in geomorphologic settings influence spatial heterogeneity in transport and retention of nutrients, a hierarchical network perspective is needed, comprising connectivity, residence times, and reactivity interactions (Lin et al., 2016; Stewart et al., 2011).

Beyond the traditional insights of nonlinear processes using 1-D, 2-D or 3-D hydrodynamic equations, other nonlinear statistical approach such as the Boosted Regression Trees (BRT) is becoming to play a part in hydrodynamic studies (Ouedraogo and Vanclooster, 2016; Toprak et al., 2014; Toprak and Cigizoglu, 2008). The BRT model, which combines advantages of regression trees and boosted adaptive method, has recently been applied in studies on ecological traits and



species distributions (Elith and Leathwick, 2017; Wyse and Dickie, 2017; Zimmermann et al., 2010), as well as in other environmental research fields, such as natural flow regimes, groundwater and hydraulic conductivity (Jorda et al., 2015; Mousavi et al., 2017; Naghibi et al., 2016; Snelder et al., 2009), soil science (Martin et al., 2009; Jalabert et al., 2009), air pollution (Carslaw et al., 2009), energy (Kusiak et al., 2010), or climate change (Shabani et al., 2016) and so on. With consistent cross-validation procedure and the feature of easy application, the BRT model provides a strong potential for applying large-scale WRT analysis while considering multiple hydro-geomorphological parameters.

In this study, we investigate the spatial distribution of water residence time across a wide range of hydro-geomorphologic settings by applying the machine-learning approach known as boosted regression trees (BRT). The selection of leading indicators for predicting the reach-average WRT in 82 river networks in Germany are analyzed.

75 **2 Methods and Materials**

2.1 Study area and dataset

We collected discharge data for the years 2008-2014 from 132 gauging stations in Germany that are recorded with a temporal resolution of 15 minutes (Bavarian State Office for the Environment, and the ITZ Bund www.pegelonline.wsv.de). Using these stations, 82 river reaches were identified (Figure 1), which are delimited at both (upstream and downstream) ends by gauging stations. These reaches are geographically widely distributed across Germany, although underrepresenting northern low lands, and represent the hydro-morphological conditions of 13 stream types that differ in their biogeochemical conditions, too. To be noted, discontinuities in the river systems, as lakes and impoundments (produced by weirs or dams) are not explicitly considered in this paper, as water residence time is much longer and is controlled by other additional mechanisms such as the features of stratification, water body internal currents and volume changes by input-output controls (Heidbüchel et al., 2012; Ji, 2008; Rueda et al., 2006).

2.2 Sources for factors assumed to affect water residence time

In this paper we evaluate the average discharge, drainage area, mean river width, length, and slope, along with sediment composition as hydro-geomorphologic factors for predicting WRT in the selected river reaches. Geomorphologic parameters are averaged over the reach between the upstream and downstream stations to represent the mean situation of the selected river reach (Table 1). Substrate class of the sediment type for each river reach is represented in percentage (100% all classes in sum) according to their length that falls into each class. According to the German soil classification system (Working Group on Soil Classification of the German Soil Science Society, 1997), six substrate classes are derived as Sand (S), Clay (C), Silt (U), Loam (L), Peat bog (HM), Fen (NM), respectively. Furthermore, classifications of stream types with integrated geomorphic features (Table 2) are applied as an indicator for specific landscape characteristics. The stream types are aligned with the official German stream and river type classification systems based on physio-chemical parameters and geological classifications of ecological zones that contain similar environmental characteristics, such as stream size, stream order, altitude,



flow velocity, streambed substrates, water temperature, or stream width etc. (Pottgiesser and Sommerhäuser, 2004; Arle et al., 2014). All geographic analyses and calculations were performed in ArcGIS Desktop (Version 10.0, Environmental Systems Research Institute (ESRI). Redlands, CA, USA).

100 2.3 Spatial distributions of water residence time

We introduced the nonmetric multidimensional scaling (NMDS) plots (Agarwal et al., 2007) to obtain insight into the patterns of hydro-morphological conditions as well as WRT distributions for the studied river reaches. NMDS allows integrating different data formats, such as continuous monitoring data, discrete parameters, binary data or binomial category datasets. We used the Gower's generalized coefficient of dissimilarity approach (Gower and Legendre, 1986) to standardize the continuous variables against the discrete ones and got the standardized Euclidean distance for the NMDS plots.

105 Water residence time for studied river reaches are estimated for the average discharge conditions during 2008-2013. The spatial distribution of water residence time has been estimated by applying the nonlinear model of Boosted Regression Trees (BRT). The feature of nonlinearity in controlling WRT, and interactions among multiple predictive factors are analyzed via the BRT model (Elith and Leathwick, 2016). In the first place, the fitted model is generated by the known average values of predictive datasets for the complete river networks. Furthermore, the generated model is applied to estimate the distribution of WRT under different environmental scenarios such as flood or drought hydrologic conditions. A measurement of relative importance (in percentage) is calculated by the model to facilitate the comparisons of term-wise contributions. In addition, partial dependence plots and fitted link functions for each variable were produced. Fitted BRT models were obtained by the sum of all trees multiplied by the learning rate (Elith et al., 2008):

$$115 \quad f(x)=g[\sum_i T_i(x)] \quad (1)$$

where f is the fitted model, x is the independent variable, T_i are the individual learners, and g is the link function that grows optimum trees. All calculations and modelling were performed in R (R Core Team, 2016) by using the package 'dismo' (Hijmans et al., 2016) and 'gbm' (Greg Ridgeway with contributions from others, 2015).

2.4 Model validation

120 We here applied the empirical equation of previous studies $t = aQ - bxc$ proposed by Graf (1986), in which t represents the water residence time, Q is the discharge, x is the traveled distance in downstream direction, and a , b , c are the coefficients.

To compare the results of WRT estimated by the BRT model (WRT_{pred}) against those of the empirical equations and observed values, the error of the prediction is calculated using the Root Mean Squared Errors (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (WRT_{pred} - WRT_{obs})^2}{n}} \quad (2)$$



125 where WRT_{pred} is the predicted water residence time (h/km) and WRT_{obs} is the original calculated value according to observation
at the river reach of i , and n is the number of studied river reaches.

3. Results

The elaboration of the results is structured in a) the spatial dissimilarity of geomorphology and hydrological factors for the
studied river reaches, followed by b) the results of relative importance of variables calculated by the BRT model. Furthermore,
130 we discussed the spatial distribution of estimated WRT under long-term annual average discharge conditions as well as during
the extreme hydrological month of flood and drought.

3.1 Governing factors for water residence time

The multidimensional Euclidean distance between the studied river reaches derived from varying hydro-geomorphology
channel characteristics is shown in the Nonmetric Multi-Dimensional Scaling (NMDS) plot (Figure 2). River reaches are
135 grouped and colored according to their stream type classification and revealed clustering patterns in agreement with the river
size. The river type classification of ecoregion independent streams include lake outlets (type 21) and small organic substrate-
dominated rivers (type 11).

Fitted BRT models were obtained by the sum of all trees multiplied by the learning rate of each predictive variables. The fitted
model accounted for 54.5% of the mean total deviance of the monitored dataset (1-mean residual deviance / mean total
140 deviance). The optimal fit was achieved with the following variable setting: interaction depth = 10, tree complexity = 10,
learning rate = 0.001, bag fraction = 0.5 and cross-validation = 10-folds, optimal number of trees = 1680. For this fit, the
training data correlation coefficient was 0.668, and cross-validation correlation coefficient was 0.614.

The predictive variable of mean discharge represented the most influential variable (57.4%) in the BRT model, followed by
slope (21.5%) and the sum of drainage area (15.6%). Mean river width and river types together only explained less than 4%
145 to the model variance. Similarly, substrate classes did not significantly influence water residence time (< 2%). In particular,
the substrates of clay, peat bog and fen showed zero statistical contribution (Table 3). Although the latter predictive variables
have little or no importance in our study, we did not exclude them from the set of the predictive variables while remaining the
complete dataset for further analyses of any other scenarios.

After accounting for the average effects of the predictive variables for all river reaches using the BRT model, we used partial
150 dependence plots to identify the relative influence of the dominant eight variables on WRT_{pred} in the individual reaches (Figure
3). The y-axis shows the fitted WRT_{pred} , optimized on basis of link function Eq. (1). At low values, mean discharge, drainage
area and width (statistically not significant) has a strong negative impact on WRT. The influence of these three predictive
variable decrease with increasing respective values and only cause smaller changes in WRT. Longitudinal slope of riverbed is
found to have a positive relationship with WRT. One possible reason for this is, that with slope also turbulence increases, but



155 not the mean flow velocity. This is revealed by the less dominant position of slope in comparison with discharge, area and
river width. The effects of river topography and sediment composition appear to be largely mediated by their implemented
correlation with hydrological characteristics and geographic distribution.

3.2 WRT distribution

3.2.1 WRT distribution under annual average hydrologic conditions

160 River reaches categorize as large rivers (type 9.2, 10, 15, 20) usually have high levels of the average discharge and area as
well as low slope values, which consequently show a synchronized distributions in the NMDS plot (Figure 4). Water residence
time at smaller rivers is in turn stronger impacted by distinct topologic features, resulting in a less uniform pattern than the
group of large river.

Scattered from the Euclidian distance to the spatial dimension, water residence time distribution under the mean discharge
165 conditions during 2008-2013 are displayed Figure 5. River discharge above 300 m³/s results in WRT of less than 4 h/km,
which equals to flow velocity of more than 0.07 m/s. Comparing the results of BRT model (WRT_{pred}) with that of the observed
values (WRT_{obs}) showed 46% of the mean squared errors are less than 0.1 h/km (Figure 6). Poor model performance under
low flows demonstrates the increased impact of individual predictive variables and the need for further testing and data
collection to support the inclusion of additional biogeochemistry processes. Site-specific uncertainties might arise from
170 unknown flow paths and mixing dynamics that significantly affect management strategies.

Through comparing the results of predicted WRT (WRT_{pred}) with that of Graf's empirical equation (WRT_{emp}), the calculated
flow velocity by BRT model showed a less correlated linear relationship with discharge (Figure 7). WRT_{pred} and WRT_{emp}
delivered water residence times in the same magnitude of order, but with different gradients in their variability and dependence
with changing Q_{mean}. A lower tendency of a linear relationship for smaller discharge levels below 500 m³/s indicates a
175 potential geomorphological influence manifested at small rivers and non-bankfull conditions. Possible explanations for the
described differences between WRT_{pred} and WRT_{emp} could be that in this study, the BRT model is built to explain variables
through multiple boosted regressions by including the nonlinear interactional effects among predictive variables. This
interpretation is in accordance with the partial dependency analysis of each variable that an overall consideration of all
predictive variables at varied levels are needed by applying the systematic or network approach (Dumont et al., 2005).

180 3.2.2 WRT distribution under hydrologic extremes

The response of water residence time on changing and in particular extreme discharge is complex, especially for distinct
geomorphic sites. In order to facilitate more intuitive understanding, we did parallel studies for the extreme flood event in June
2013 and the driest month of November 2011 in Germany. The May/June 2013 flood was the most severe large-scale flood
events in Germany during the last 6 decades (Merz et al., 2014). Compared with the flood events in June 2013, the median
185 discharge in November 2011 is 80.2% lower with the estimated water residence time is 20.7% (0.17 hours) longer per kilometer



(Figure 8). Spatial variation is shown through the bivariate map of mean discharge and water residence times. The contrasting effect is more clearly observed in the Elbe river basin where the most severe floods occurred (Figure 9).

4. Discussions

4.1 Interactional effects of predictive variables

190 Getting to know the interactional effects among predictive variables would facilitate the empirical estimation of WRT in a river reach on basis of generally available information. Among all the predictive variables, river hydrology ranks the first place of relative importance together with slope in shaping the variation of water residence time. Therefore, hydrological variations in the river reach have to be the paramount element of discussion.

The 2-dimensional partial dependence plot in Figure 10 shows the interactional effects between river discharge and drainage
195 area. Another important geomorphological factor is the river width that has great contribution to the distribution of WRT. Furthermore, the river type classification, which implements generalized geomorphologic and geographic attributes, could act as a substitute, helping to simplify the process of WRT estimation especially under limited data availability. The interactional effect between river type and mean discharge is expressed in Figure 11. For a river reach with known substrate class and river topology, water residence time under different discharge levels can be estimated.

200 4.2 Impacts of groyne fields on water residence time

As mentioned earlier in methodology part, the discontinuities in the river systems, potentially impacting water residence time, such as lakes, impoundments or river groyne fields were not included in the original dataset for model generation. River groyne fields (also called wing dams) usually made from rocks or woods are constructed from the riverbank, transversely to the main flow direction to prevent from lateral bank erosion by reducing flow velocity (Yossef, 2002). Due to simple construction,
205 long-term durability and major functions, groyne fields (GF) are very widely applied in the lowland rivers of Germany. At present there are approximately 6900 groynes, covering 92% of the banks along the Middle Elbe River section (Schwartz, 2006). Because of the considerable reduction of water depth and flow velocity relative to the main stream, the prolonged retention time of water in the GF has important impact on the nutrient uptake dynamics and phytoplankton growth (Engelhardt et al., 2004; Guhr et al., 2000; Ockenfeld and Guhr, 2003). Hydraulic waves attenuation and increasing water residence time
210 are two main effects of groyne fields and potentially have strong impacts on nutrients retention and phytoplankton growth. Describing the specific hydraulic characteristics of flow velocity and residence time patterns in GF is the key for understanding the ecological significance of these retention zones.

At 14 out of 82 studied river reaches groyne fields are installed. Distribution of water residence time at groyne fields are linked to the variables and factors as discussed above, however, the shapes of the hydrograph reveal the different attributes of
215 attenuation, which indicate the potential for nutrient retention. In order to exclude the influence of distinct scales, river reaches



from the alpine stream and central highlands in Bavaria (in total 39 reaches) are not considered for the comparison. Among the rest 43 river reaches of comparable discharge level, two groups of 14 river reaches with groyne fields and 29 free-flowing rivers are compared.

We plot the cumulative distribution functions of the mean hydro-width (in hours) for the two groups (Figure 12). The cdf plot shows that the probability level of hydro-width less than 90% are up to 44.4 hours at groyne fields, compared with that of only 21.4 hours at free-flowing rivers. There is very little chance ($< 2.5\%$) that the probability of hydro-width in free-flowing rivers will be less than one hour and there is also small chance ($< 5\%$) that it could be as high as 33.7 hours. The groyne fields show pronounced wider hydro-width than the free-flowing rivers: with 59.4% (87.6 hours) larger maximum value and more than 2 folds' (4.7 hours) at the median level. Not surprisingly, the estimated water residence time for groyne fields show higher probabilities for WRT less than 1.5 h/km in comparison with the free flow rivers (Figure 13).

4.3 Hydraulic residence time as a lens for nutrient retention time

It should be noted of the existing differences between flow velocities in the system (that set the velocity of conservative solutes) and the celerity (or speed with which hydraulic perturbations are conveyed, which control the hydrograph), are to be the velocity of conservative solutes, expected since they are controlled by different mechanisms. The nutrient transportation velocity in streams is always slower than the kinematic flow celerity of gravity-driven hydraulic waves. Studies on the differentiation and translation of these two velocities under varying flow conditions have been thoroughly discussed by McDonnell and Beven (2014). The water residence time discussed in this paper is coherently referred to flow velocity. However, the biogeochemical functioning of a river ecosystem is largely dependent on the transportation processes of water and dissolved substances within the geomorphic context of river networks (Benettin et al., 2015; Withers and Jarvie, 2008). Nutrient dynamics are controlled by the interaction of several key parameters, i.e. river discharge, channel geometry and vertical exchanges of water (Maazouzi et al., 2013). The amount of biologically labile dissolved nutrient effectively taken up by primary producers depends, next to other limiting factors, such as light availability or water temperature, on the WRT in a river section during which processes may take place (Drummond et al., 2016; Ensign and Doyle, 2006; Nieuwenhuyse, 2005; Soulsby et al., 2006).

Water flow velocity in a river reach controls the time during which bio-chemical processes can reduce or transform contained nutrients. Nutrient transport and transformation in streams involves both physical dynamics and biological uptake processes along the longitudinal course of rivers (Kronvang et al., 1999; Runkel, 2007). The transport mechanisms are mainly shaped by hydro-morphological parameters such as river discharge, water depth and velocity, and by other related physical ones such as sediment composition. River hydro-morphological shapes those processes and plays a major role in structuring the hydrological, ecological and biogeochemical dynamics in streams and rivers that are essential to ecosystem functioning (Doyle et al., 2003).



Traditional insights into the processes of nutrient spiraling process by experiments are sometimes biased and conditional dependent due to the fact that nutrient addition often brings much higher concentration than the background level, which results with overestimated nutrient uptake length (Mulholland et al., 2002). Modelling studies on nutrient export are mostly based on steady state hydrologic conditions assuming variations in pressure from pollution sources (Ingestad and Ågren, 1988; Powers et al., 2009; Runkel, 2007; Runkel and Bencala, 1995). This assumes that the hydraulic gradients that drive the transports are maintained the whole time the stream water remains in the water body, which is unrealistic (McCallum and Shanafield, 2016). Therefore, the current study adds to improve our understanding on the impact of water residence time on nutrient retention time and transportation processes on medium to large scales. This is in particular required to tackle the challenge in quantifying changes of hydro-morphological parameters in space and time (Ambrosetti et al., 2003; Bouwman et al., 2013; Tong and Chen, 2002).

5. Conclusions

Understanding the dynamics of instream water residence time does not only assist water quality and instream nutrient retention modelling, but also water management practices especially under extreme hydrologic conditions. Through application of the BRT model for estimating WRT in river networks, we identified that river discharge weights the most compared with river topologic and geomorphic attributes. At the scale of river networks, water residence time is primarily affected by river discharge, followed by river width and river channel slope. Geomorphological attributes are more influential on small rivers in the Alpine mountainous areas. By taking the example of river discharge during flood and drought events, the BRT modelling is useful for water residence time estimation under extreme hydrological scenarios.

We conclude that the BRT approach has the potential to be used for addressing how timescales of the hydrological cycle change at different scales. The spatial distribution model contributes to an advanced methodology in WRT estimation in between of complex deterministic process models and empirical statistical models, and can be applied to study areas of diversified scales. The research element of water residence time is elaborated on the temporal scale of multi decades, though it is based on reach-scale information. It has the foremost implications on river ecology by linking nutrient retention time in streams at larger spatial and temporal scales. While applied to few representative catchments. It can set the basis for analyses of large spatial scales and can be relevant at longer time scales associated with climate variability.

The results underline the relative importance of hydro-geomorphological features, which has clear implications to optimize the efficiency of river restoration effects on runoff processes. In combination with developed nonlinear spatial statistics it could be a novel approach in solving hydro-geophysical or even social economic distribution related questions. A further investigation of retention time maxima could help identifying thresholds at which potential restoration measures, land use changes, drought or floods, and climatic stressors affect physical water body conditions and bio-chemical responses.



Acknowledgements

This work has been carried out within the SMART Joint Doctorate programme “Science for the Management of Rivers and their Tidal systems” funded by the Erasmus Mundus program of the European Union. The authors thank the Bavarian State Office for the Environment to provide hydrological dataset for Bavaria, and the ITZBund for providing hydrological monitoring dataset for Germany (www.pegelonline.wsv.de). A special thanks to Judith Mahnkopf and Annett Wetzig in the working group of "nutrient balances in river systems" at the Leibniz Institute of Freshwater Ecology and Inland Fisheries Berlin for their support with environmental data collection and preparation. The authors thank the colleagues and friends who offered their valuable comments and proof readings for the manuscript.

References

- Agarwal, S., Wills, J., Cayton, L., Lanckriet, G. R., Kriegman, D. J., & Belongie, S., 2007. Generalized Non-metric Multidimensional Scaling. In AISTATS, pp. 11-18.
- Aguilera, R., Marcé, R., & Sabater, S., 2013 Modeling nutrient retention at the watershed scale: Does small stream research apply to the whole river network? *J. Geophys. Res.: Biogeosci.* 118(2), 728-740.
- 290 Arle J, Blondzik K, Claussen U, Duffek A, Grimm S, Hilliges F, Hoffmann A, Leujak W, Mohaupt V, Naumann S, et al., 2014. Environmental policy, water resources management in Germany. Part 2. Water quality, federal ministry of environment, nature conservation, building and nuclear safety, 53048 Bonn. Bonn: Federal Ministry for the Environment, Nature Conservation, Construction and Nuclear Safety; p. 37–58.
- Arora, R., Tockner, K., & Venohr, M., 2016. Changing river temperatures in northern Germany: trends and drivers of change. *Hydrological Processes*, 30(17), 3084-3096.
- 295 Bergstrom, A., McGlynn, B., Mallard, J., & Covino, T., 2016. Watershed structural influences on the distributions of stream network water and solute travel times under baseflow conditions. *Hydrol. Process.* 30: 2671–2685. <http://dx.doi.org/10.1002/hyp.10792>
- Behrendt, H., & Opitz, D., 1999. Retention of nutrients in river systems: dependence on specific runoff and hydraulic load. In *Man and River Systems* (pp. 111-122). Springer, Dordrecht.
- 300 Carslaw, D. C., & Taylor, P. J., 2009. Analysis of air pollution data at a mixed source location using boosted regression trees. *Atmos. Environ.* 43(22), 3563-3570.
- Catalán, N., Marcé, R., Kothawala, D. N., & Tranvik, L. J., 2016. Organic carbon decomposition rates controlled by water retention time across inland waters. *Nature Geoscience*.
- 305 Doyle, M. W., 2005. Incorporating hydrologic variability into nutrient spiraling. *J. Geophys. Res.: Biogeosci.* 110(G1).
- Doyle, M. W., Stanley, E. H., & Harbor, J. M., 2003. Hydrogeomorphic controls on phosphorus retention in streams. *Water Resources Research*, 39(6).



- Drummond, J. D., Bernal, S., Schiller, D. V., & Martí, E., 2016. Linking in-stream nutrient uptake to hydrologic retention in two headwater streams. *Freshw Sci.* 35(4), 000-000.
- 310 Dumont, E., Harrison, J. A., Kroeze, C., Bakker, E. J., & Seitzinger, S. P., 2005. Global distribution and sources of dissolved inorganic nitrogen export to the coastal zone: Results from a spatially explicit, global model. *Global Biogeochemical Cycles*, 19(4).
- Elith, J., Leathwick, J. R., & Hastie, T., 2008. A working guide to boosted regression trees. *J Anim Ecol.* 77(4), 802-813.
- Elith, J., & Leathwick, J., 2016. Boosted Regression Trees for ecological modeling.
- 315 Engelhardt, C., Krüger, A., Sukhodolov, A., & Nicklisch, A., 2004. A study of phytoplankton spatial distributions, flow structure and characteristics of mixing in a river reach with groynes. *J. Plankton Res.* 26(11), 1351-1366.
- Ensign, S. H., & Doyle, M. W., 2006. Nutrient spiraling in streams and river networks. *J. Geophys. Res.: Biogeosci.* 111(G4).
- Gibson, G., 2000. Nutrient criteria technical guidance manual.
- Gottschalk, L., Krasovskaia, I., Leblois, E., & Sauquet, E., 2006. Mapping mean and variance of runoff in a river basin. *Hydrol.*
- 320 *Earth Syst. Sci. Discuss.* 3(2), 299-333.
- Gower, J. C., & Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. *J CLASSIF.* 3(1), 5-48.
- Graf, J. B., 1986. Traveltime and longitudinal dispersion in Illinois streams (Vol. 2269). Dept. of the Interior, US Geological Survey.
- Greenacre, M. Measures of distance between samples: non-Euclidean.
- 325 Greg Ridgeway with contributions from others, 2015. gbm: Generalized Boosted Regression Models. R package version 2.1.1. <https://CRAN.R-project.org/package=gbm>
- Grizzetti, B., Bouraoui, F., Granlund, K., Rekolainen, S., & Bidoglio, G., 2003. Modelling diffuse emission and retention of nutrients in the Vantaanjoki watershed (Finland) using the SWAT model. *Ecol. Model.* 169(1), 25-38.
- Guhr, H., Karrasch, B., & Spott, D., 2000. Shifts in the processes of oxygen and nutrient balances in the river Elbe since the
- 330 transformation of the economic structure. *Acta Hydroch. Hydrob.* 28(3), 155-161.
- Gücker, B., & Boëchat, I. G., 2004. Stream morphology controls ammonium retention in tropical headwaters. *Ecology*, 85(10), 2818-2827.
- Heidbüchel, I., Troch, P. A., Lyon, S. W., & Weiler, M., 2012. The master transit time distribution of variable flow systems. *Water Resour Res.* 48(6).
- 335 Helton, A. M., Hall, R. O., & Bertuzzo, E. (2018). How network structure can affect nitrogen removal by streams. *Freshwater Biology*, 63(1), 128-140.
- Hrachowitz, M., Benettin, P., Breukelen, B. M., Fovet, O., Howden, N. J., Ruiz, L., ... & Wade, A. J., 2016. Transit times—the link between hydrology and water quality at the catchment scale. *WIREs Water*. <http://dx.doi.org/10.1002/wat2.1155>.
- Huiliang, W., Zening, W., Caihong, H., & Xinzhong, D., 2015. Water and nonpoint source pollution estimation in the
- 340 watershed with limited data availability based on hydrological simulation and regression model. *Environ Sci Pollut R.* 22(18), 14095-14103.



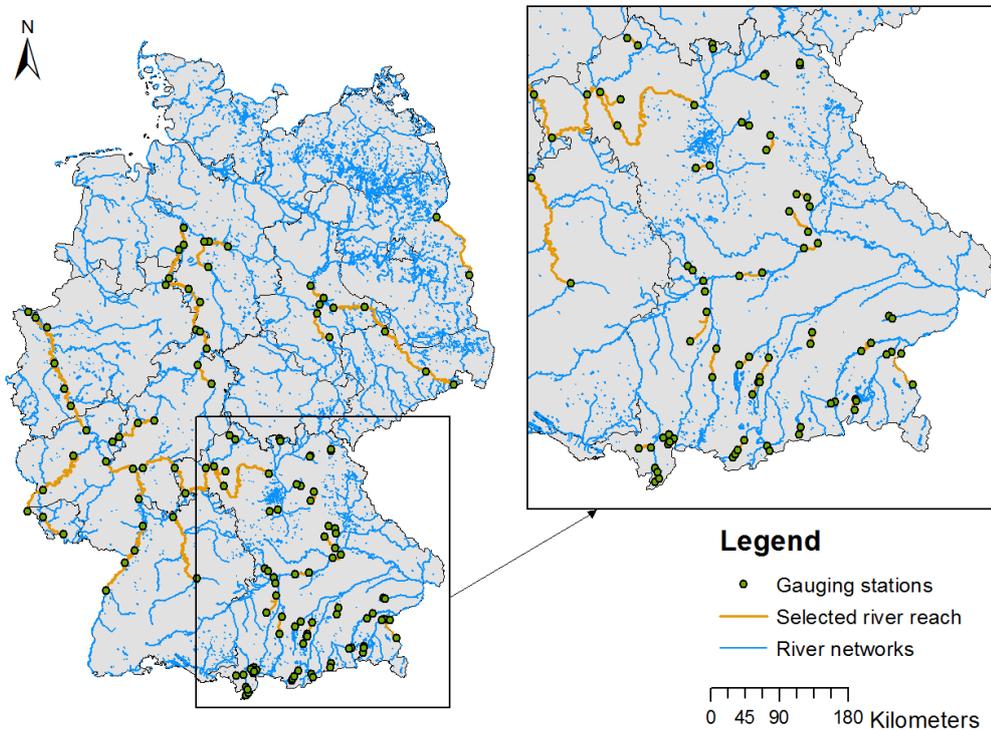
- Jalabert, S. S. M., Martin, M. P., Renaud, J. P., Boulonne, L., Jolivet, C., Montanarella, L., & Arrouays, D., 2010. Estimating forest soil bulk density using boosted regression modelling. *Soil Use Manage.* 26(4), 516-528.
- Ji, Z. G., 2008. *Hydrodynamics and water quality: modeling rivers, lakes, and estuaries.* John Wiley & Sons.
- 345 Jorda, H., Bechtold, M., Jarvis, N., & Koestel, J., 2015. Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. *Eur. J. Soil Sci.* 66(4), 744-756.
- Kronvang, B., Hezlar, J., Boers, P., Jensen, J. P., Behrendt, H., Anderson, T., ... & Nielsen, C. B., 2004. Nutrient retention handbook. Software Manual for EUROHARP-NUTRET and Scientific review on nutrient retention, EUROHARP report, 9-2004.
- 350 Kusiak, A., Li, M., & Tang, F., 2010. Modeling and optimization of HVAC energy consumption. *Appl. Energy.* 87(10), 3092-3102.
- Leopold, L.B., Wolman, M.G., and Miller, J.P., 1964. *Fluvial processes in geomorphology.* W.H. Freeman and Co., San Francisco, Calif.
- Lin, L., Davis, L., Cohen, S., Chapman, E., & Edmonds, J. W., 2016. The influence of geomorphic unit spatial distribution on nitrogen retention and removal in a large river. *Ecological modelling*, 336, 26-35.
- 355 Maavara, T., Parsons, C. T., Ridenour, C., Stojanovic, S., Dürr, H. H., Powley, H. R., & Van Cappellen, P., 2015. Global phosphorus retention by river damming. *Proc. Natl. Acad. Sci. U.S.A.* 112(51), 15603-15608.
- Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H., Bouwman, A. F., ... & Van Drecht, G. (2010). Global nutrient export from WaterSheds 2 (NEWS 2): model development and implementation. *Environmental Modelling & Software*, 25(7), 837-853.
- 360 Martin, M. P., Lo Seen, D., Boulonne, L., Jolivet, C., Nair, K. M., Bourgeon, G., & Arrouays, D., 2009. Optimizing pedotransfer functions for estimating soil bulk density using boosted regression trees. *Soil Sci Soc Am J.* 73(2), 485-493.
- MATLAB and Signal Processing Toolbox Release R2016a, The MathWorks, Inc., Natick, Massachusetts, United States.
- McDonnell, J. J., & Beven, K., 2014. Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resour. Res.* 50(6), 5342-5350.
- 365 Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., & Uhlemann-Elmer, S. (2014). The extreme flood in June 2013 in Germany. *La Houille Blanche*, (1), 5-10.
- Mousavi, S. M., Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B., 2017. GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. *AIMS Geosciences*, 3.
- 370 Naghibi, S. A., Pourghasemi, H. R., & Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 188(1), 1-27.
- Newcomer Johnson, T. A., Kaushal, S. S., Mayer, P. M., Smith, R. M., & Sviridchi, G. M., 2016. Nutrient Retention in Restored Streams and Rivers: A Global Review and Synthesis. *Water.* 8(4), 116.
- 375



- Ouedraogo, I. and Vanclooster, M., 2016. A meta-analysis and statistical modelling of nitrates in groundwater at the African scale. *Hydrol. Earth Syst. Sci.*, 20, 2353-2381, <http://dx.doi.org/10.5194/hess-20-2353-2016>
- Pottgiesser, T., & Sommerhäuser, M., 2004. Profiles of German stream types. Länder-Arbeitsgemeinschaft Wasser (LAWA). Umweltbundesamt, Dessau, Germany.[Available at: <http://www.wasserblick.net/servlet/is/18727>.
- 380 Preston, S. D., Alexander, R. B., Schwarz, G. E., & Crawford, C. G., 2011. Factors affecting stream nutrient loads: a synthesis of regional SPARROW model results for the continental united states1. *J Am Water Resour Assoc.* 47(5), 891-915.
- Ockenfeld, K., & Guhr, H., 2003. Groyne fields–sink and source functions of “flow-reduced zones” for water content in the River Elbe (Germany). *Wat Sci Tech.* 48(7), 17-24.
- Quiel, K., Becker, A., Kirchesch, V., Schöl, A., & Fischer, H., 2011. Influence of global change on phytoplankton and nutrient
385 cycling in the Elbe River. *Reg Environ Change.* 11(2), 405-421.
- Poole, G. C., 2010. Stream hydrogeomorphology as a physical science basis for advances in stream ecology. *Journal of the North American Benthological Society*, 29(1), 12-25.
- R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 390 Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith, 2016. dismo: Species Distribution Modeling. R package version 1.1-1. URL "<https://CRAN.R-project.org/package=dismo>."
- Rueda, F., Moreno-Ostos, E., & Armengol, J., 2006. The residence time of river water in reservoirs. *Ecol. Model.* 191(2), 260-274.
- Schwartz, R., 2006. Geochemical characterisation and erosion stability of fine-grained groyne field sediments of the Middle
395 Elbe River. *Acta Hydroch. Hydrob.* 34(3), 223-233.
- Shabani, F., Kumar, L., & Solhjoui-fard, S., 2016. Variances in the projections, resulting from CLIMEX, Boosted Regression Trees and Random Forests techniques. *Theor. Appl. Climatol.* 1-14.
- Siebert, W. M., 1986. Circuits, signals, and systems (Vol. 2). MIT press.
- Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E., & Shankar, U., 2009. Predictive mapping of the natural
400 flow regimes of France. *J. Hydrol.* 373(1), 57-67.
- Stewart, R. J., Wollheim, W. M., Gooseff, M. N., Briggs, M. A., Jacobs, J. M., Peterson, B. J., & Hopkinson, C. S. (2011). Separation of river network–scale nitrogen removal among the main channel and two transient storage compartments. *Water Resources Research*, 47(10).
- Working Group on Soil Classification of the German Soil Science Society, 1997. Soil classification of the federal republica of
405 Germany. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft.* Bd. 84, p 253-275.
- Soulsby, C., Tetzlaff, D., Rodgers, P., Dunn, S., & Waldron, S., 2006. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: an initial evaluation. *J. Hydrol.* 325(1), 197-221.
- Stanley, E. H., & Doyle, M. W., 2002. A Geomorphic Perspective on Nutrient Retention Following Dam Removal Geomorphic models provides a means of predicting ecosystem responses to dam removal. *BioScience.* 52(8), 693-701.

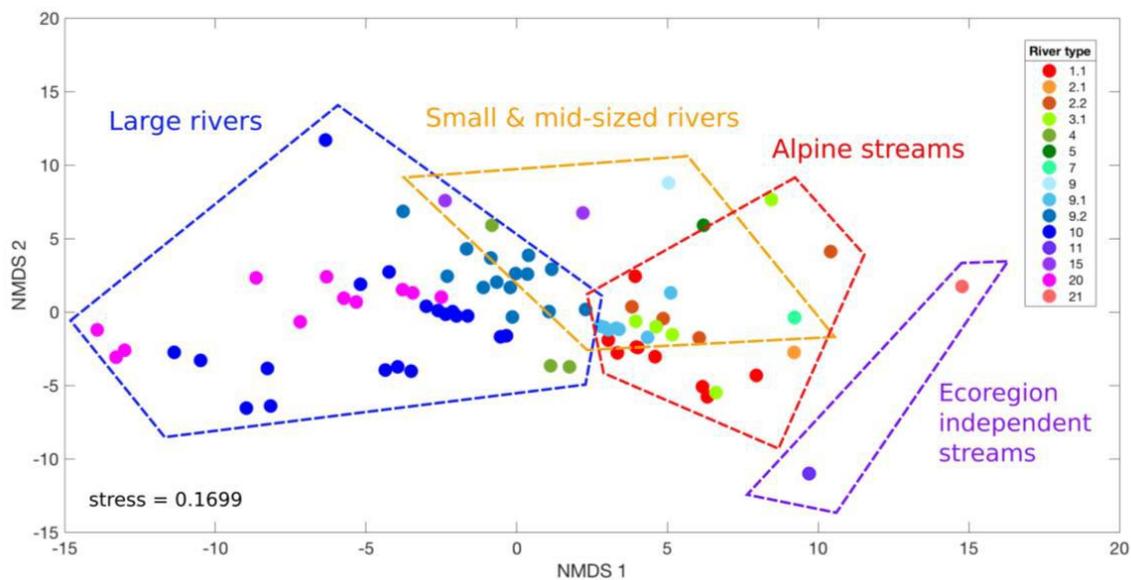


- 410 Toprak, Z. F., & Cigizoglu, H. K., 2008. Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. *Hydrol. Process.* 22(20), 4106-4129.
- Toprak, Z. F., Hamidi, N., Kisi, O., & Gerger, R., 2014. Modeling dimensionless longitudinal dispersion coefficient in natural streams using artificial intelligence methods. *KSCE J Civ Eng.* 18(2), 718-730.
- Van Nieuwenhuysse, E. E., 2005. Empirical model for predicting a catchment-scale metric of surface water transit time in streams. *Can J Fish Aquat Sci.* 62(3), 492-504.
- 415 Venohr, M., Hirt, U., Hofmann, J., Opitz, D., Gericke, A., Wetzig, A., ... & Mahnkopf, J., 2011. Modelling of nutrient emissions in river systems–MONERIS–methods and background. *Int Rev Hydrobiol.* 96(5), 435-483.
- 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, 238-251.
- 420 Worrall, F., Howden, N. J. K., & Burt, T. P., 2014. A method of estimating in-stream residence time of water in rivers. *J. Hydrol.* 512, 274-284.
- Wyse, S. V., & Dickie, J. B., 2017. Taxonomic affinity, habitat and seed mass strongly predict seed desiccation response: a boosted regression trees analysis based on 17 539 species. *Annals of botany.*
- Yossef, M. F., 2002. The effect of groynes on rivers: Literature review. Delft Cluster.
- 425 Zimmermann, N. E., Edwards, T. C., Graham, C. H., Pearman, P. B., & Svenning, J. C., 2010. New trends in species distribution modelling. *Ecography.* 33(6), 985-989.



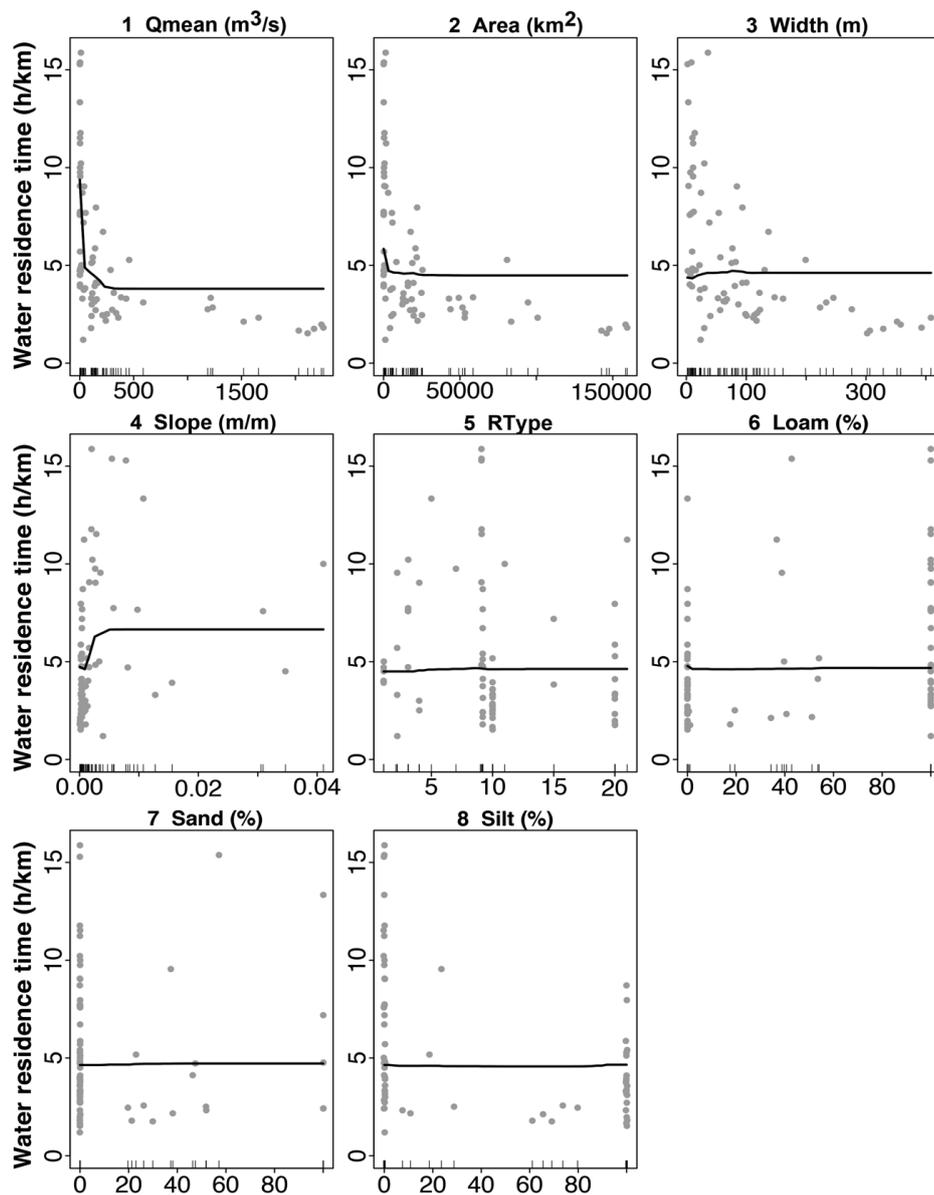
430

Figure 1: Map of river networks in Germany, with selected river reaches (orange) and corresponding upstream-downstream gauging stations (circles).

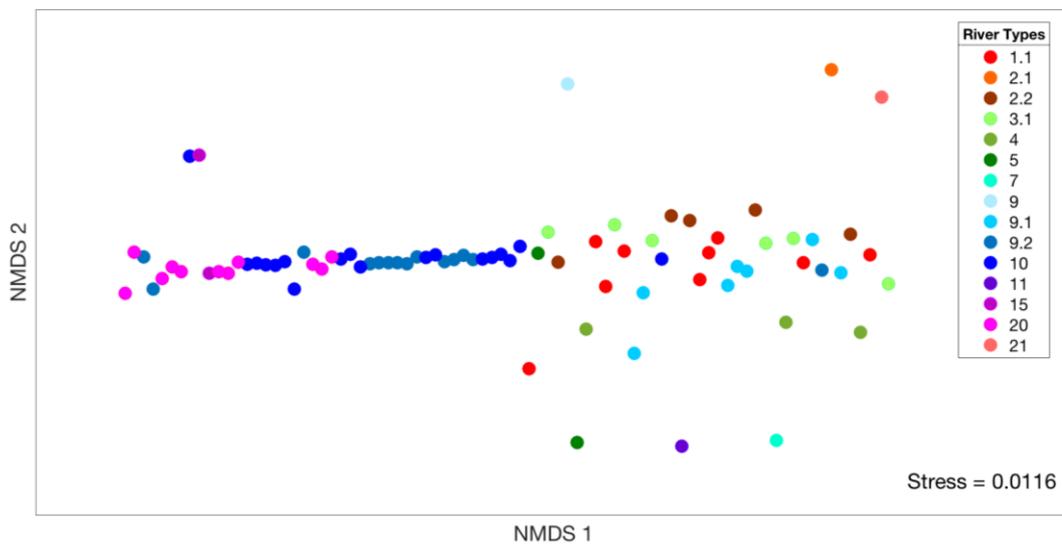


435

Figure 2: Dissimilarities of the studied river reach in the Nonmetric Multi-Dimensional Scaling (NMDS) ordination space according to hydro-geomorphic attributes.



440 Figure 3: Partial dependence plots showing the dependence of water residence time on hydro-geomorphic variables after accounting for the average effects of all the other predictors in boosted regression tree analysis. Each point represents an observed value with rug plots at the bottom of each panel. Y-axes are predicted values of the fitted functions (WRTpred). All panels are plotted on the same scale for comparison.



445 **Figure 4: Dissimilarities of the calculated WRT (h/km) in the Nonmetric Multi-Dimensional Scaling (NMDS) ordination space.**

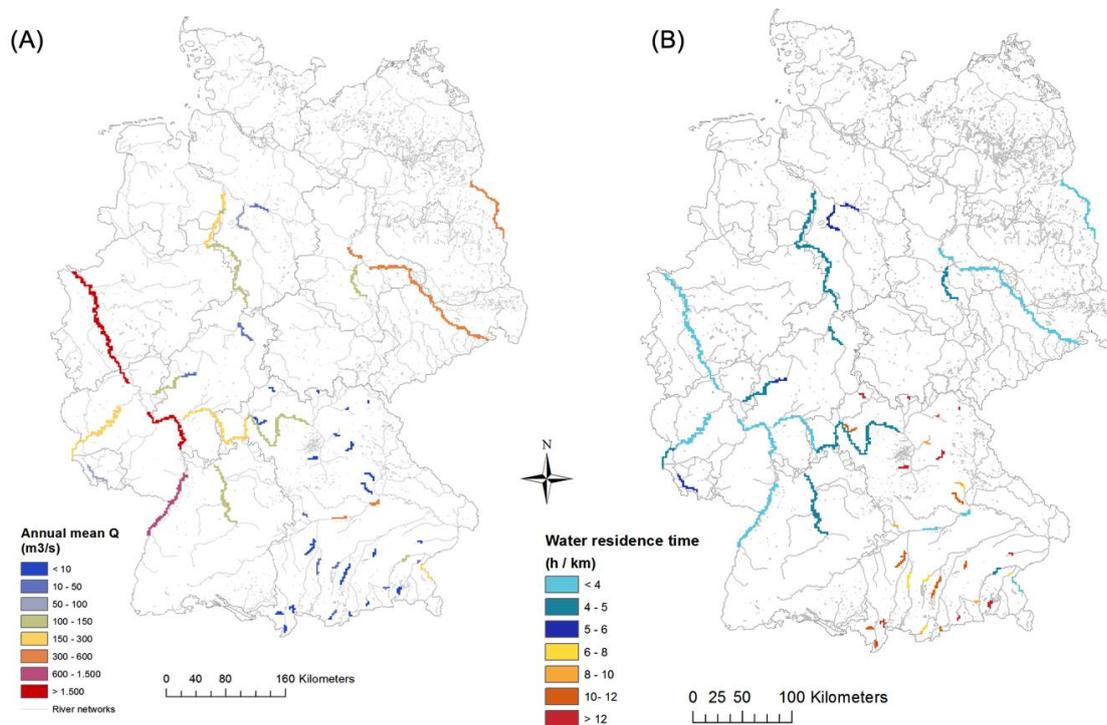


Figure 5: Spatial distribution of annual mean discharge conditions during 2008-2013 (A), and predicted water residence time (h/km) for studied river reaches.



450

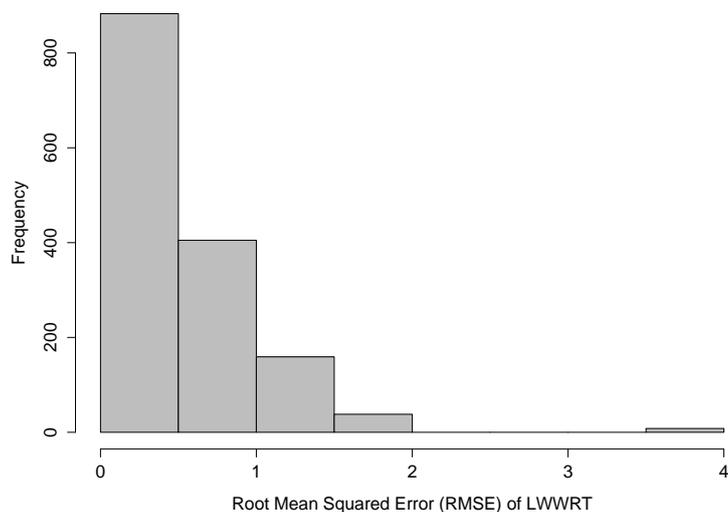


Figure 6: Frequency distribution of RMSE of predicted length weighted water residence time (h/km) against observed values across all sites.

455

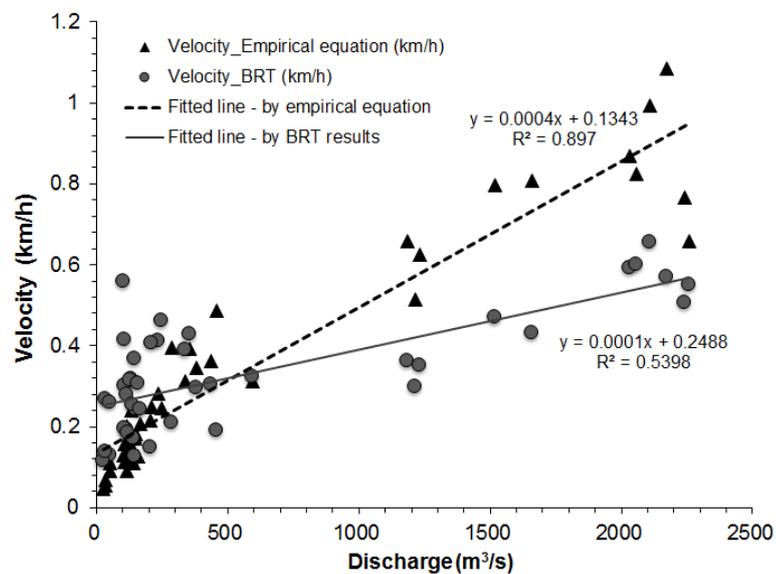
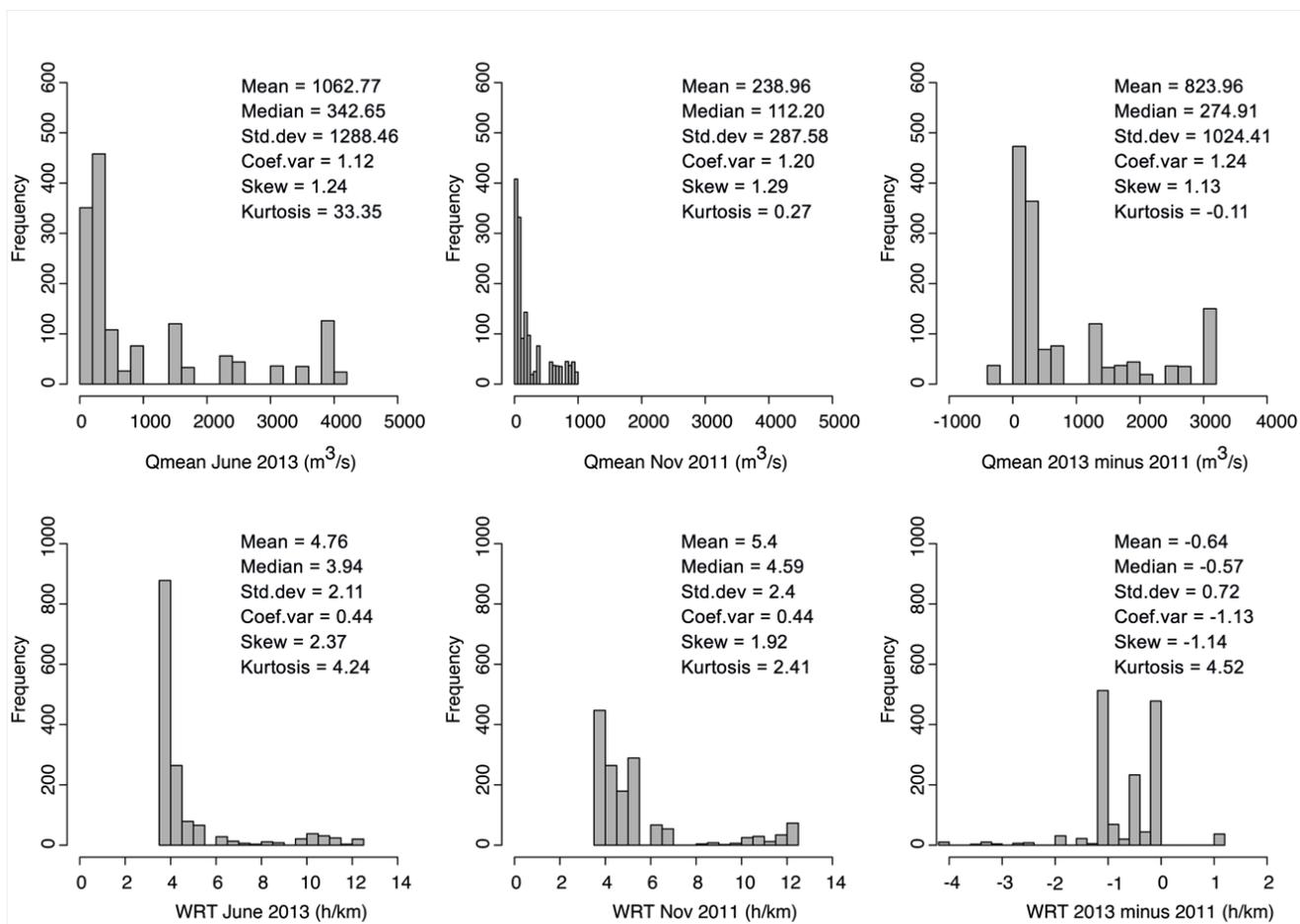


Figure 7: Comparison of the fitted linear relationship between the average discharge (x-axis) and predicted flow velocity (km/h) by equation of Graf (1986), and the BRT model, respectively.



460 **Figure 8:** Statistical comparison of the mean discharge (in cubic meters per second) and corresponding water residence time (in hour per kilometer) in June 2013 (left), November 2011 (middle), and the difference between them (right), respectively.

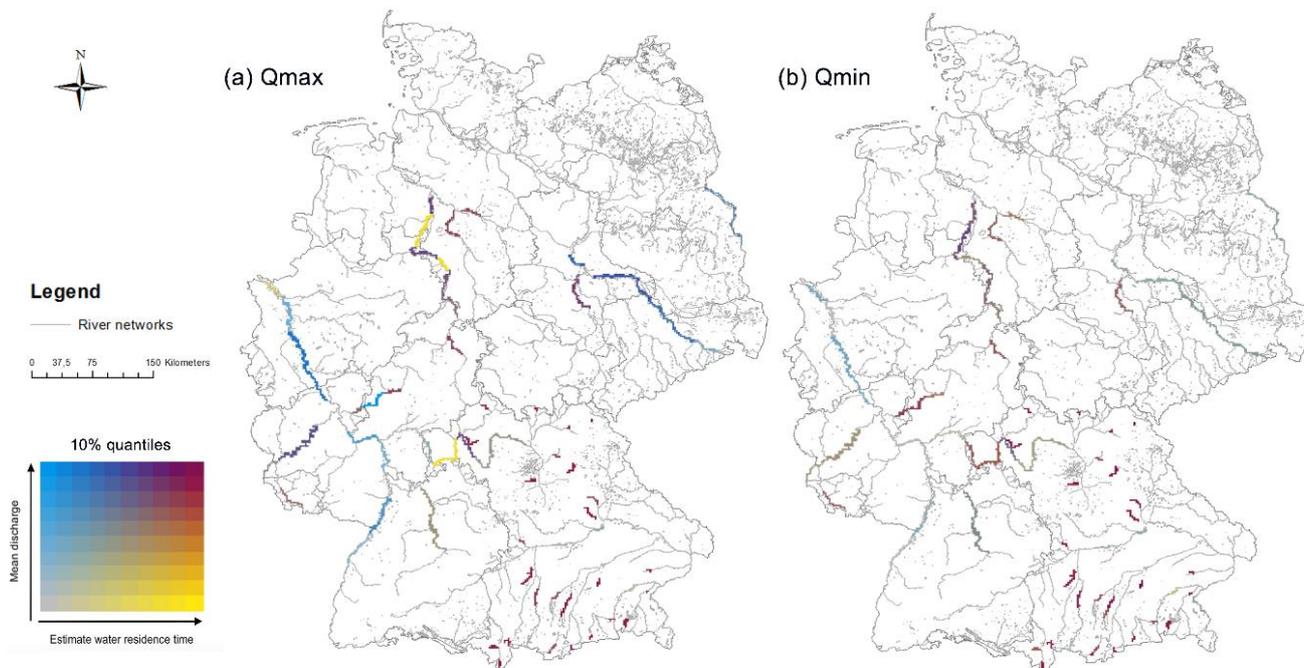


Figure 9: Bivariate map of predicted water residence time (h/km) for (a) floods in June 2013 with Qmax scenarios and (b) drought with Qmin conditions during November 2011.

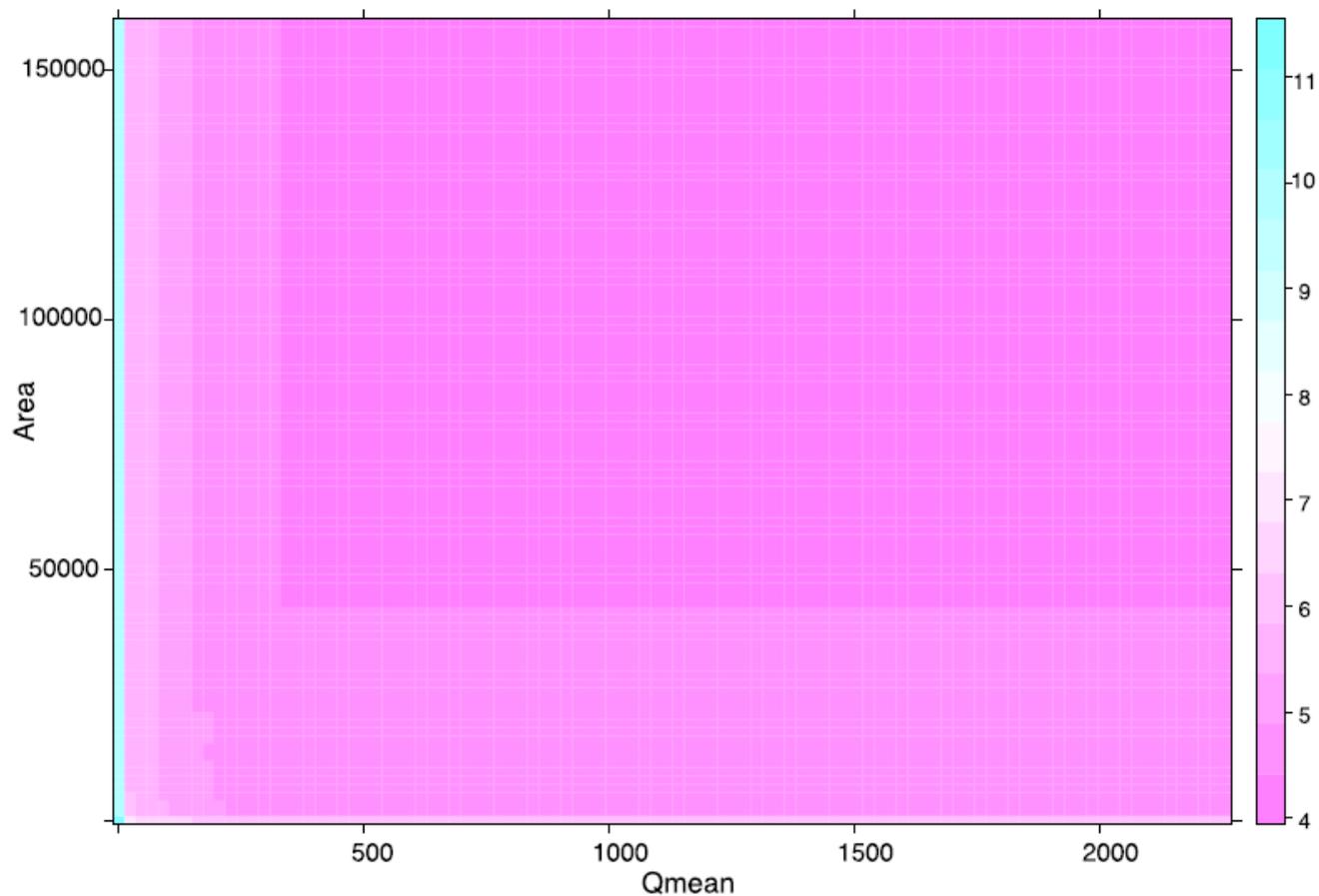
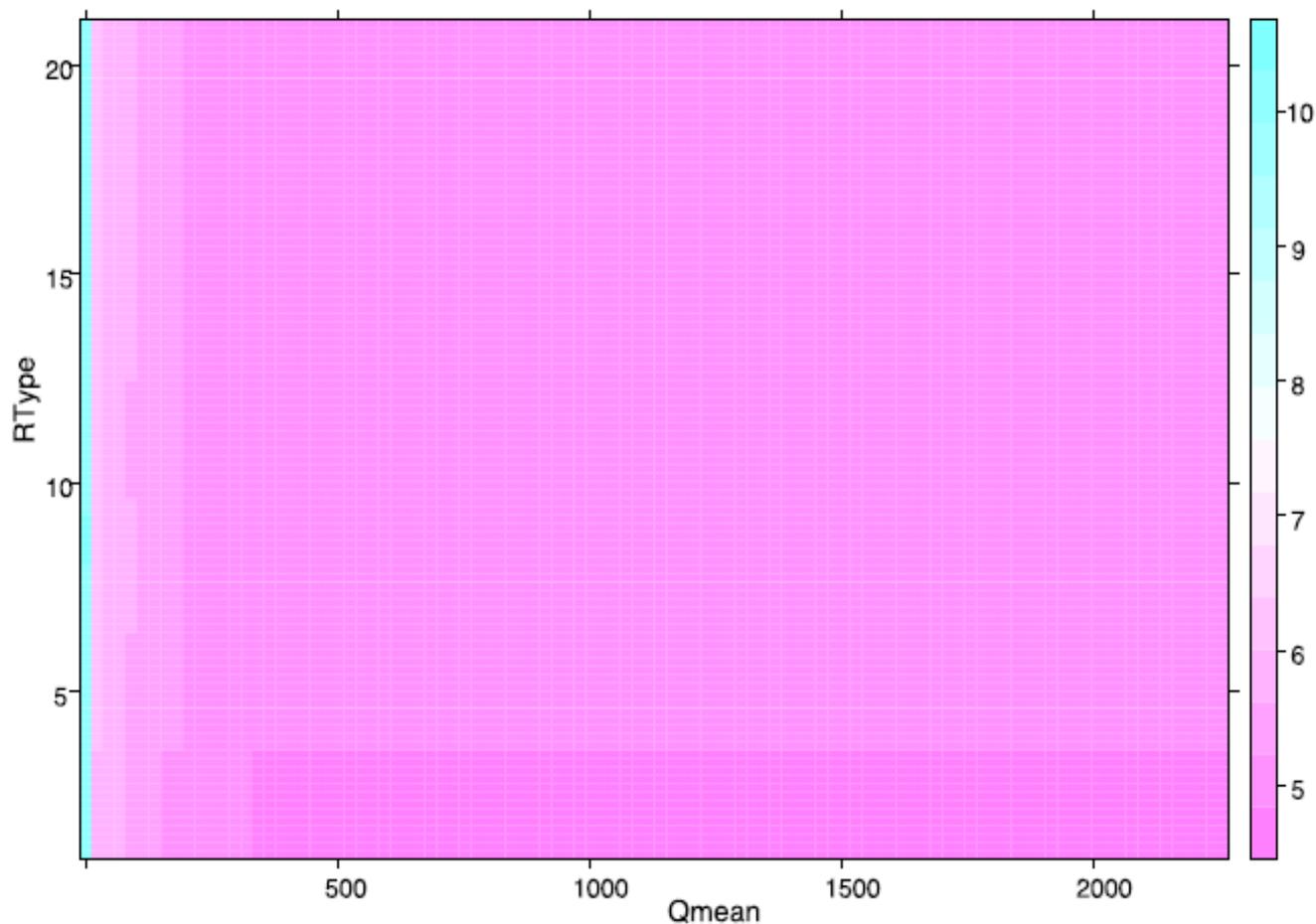
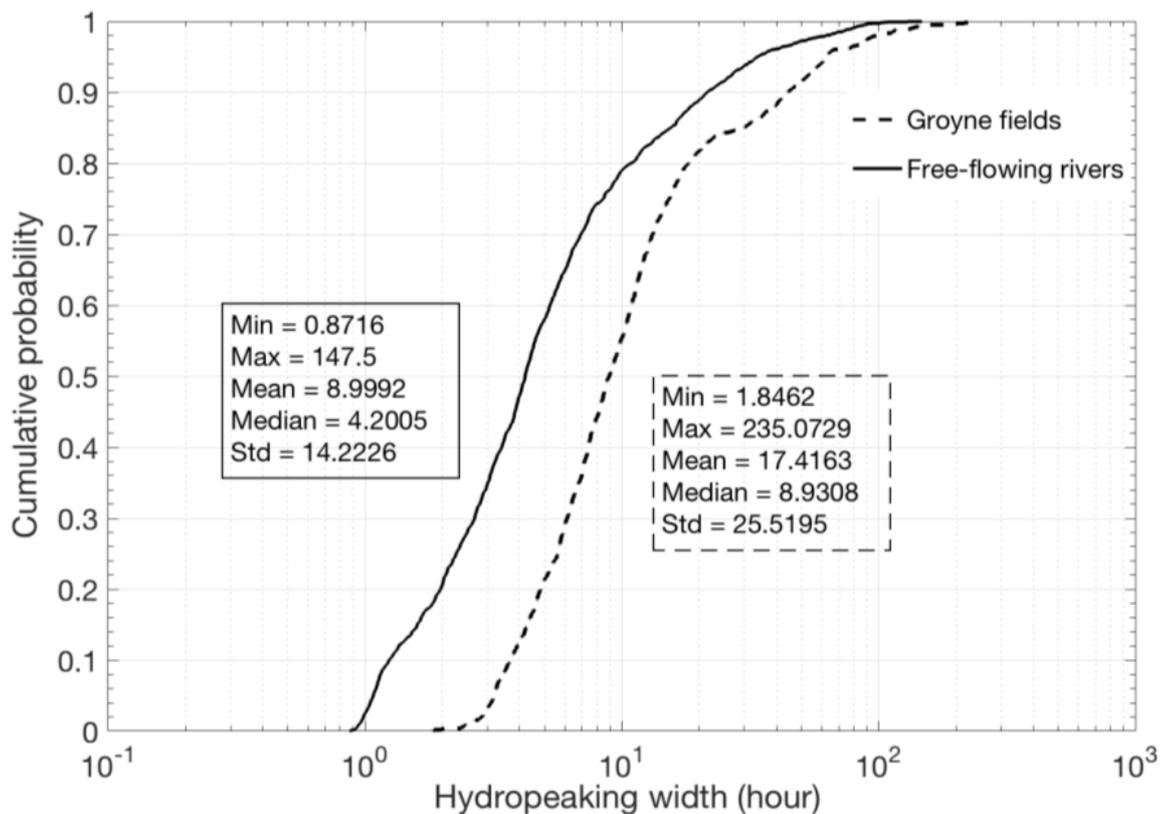


Figure 10: Two-dimensional interaction effects between the mean discharge (x-axis) and drainage area (y-axis). Colored scales are the estimated water residence time (h/km) accordingly.



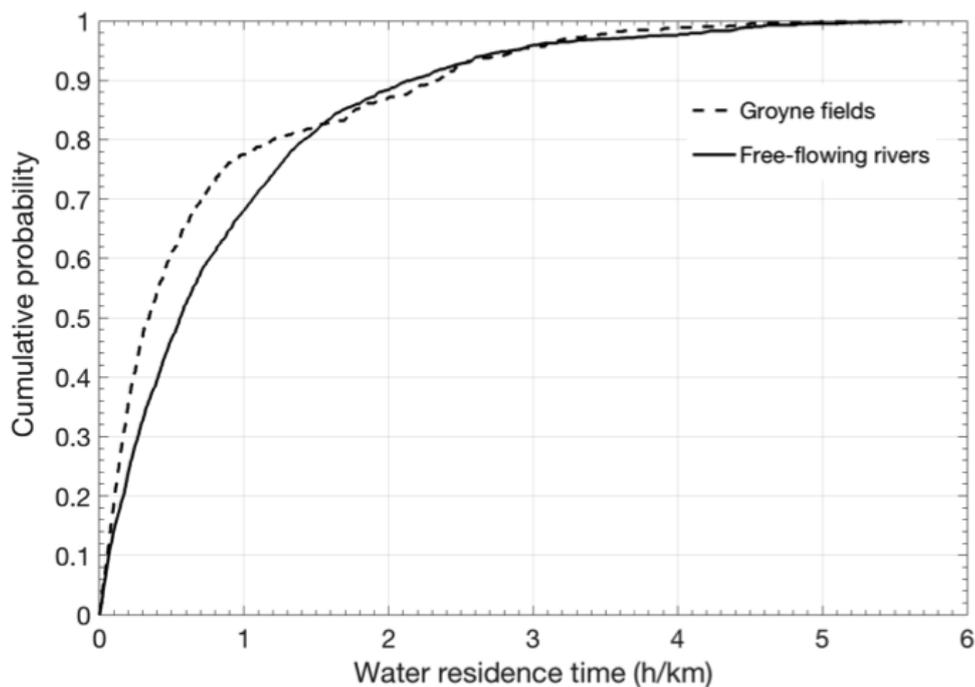
475 **Figure 11: Two-dimensional interaction effects between the mean discharge (x-axis) and river types (y-axis). Colored scales are the estimated water residence time (h/km) accordingly.**

480



485

Figure 12: Cumulative probability plots of the average hydraulic waves half-prominence widths (in hours) at river reaches with groyne fields and the free-flowing ones.



490 **Figure 13: Cumulative probability plots of the estimated water residence time (h/km) for river reaches with groyne fields and the**
495 **free-flowing ones.**

495

500

505



Table 1: Hydrologic and geographic variables of studied river reaches.

Continuous variables	Mean	Range	Std dev
Length (km)	30.8	1.01 - 145.4	30.34
Slope (m/m)	0.00379	0.00005, 0.04104	0.00776
Width (m)	88.40	1.73 - 408.42	105.91
Drainage area (km ²)	25115.28	11.15 - 159427.5	41625.53
Mean discharge (m ³ /s)	327.86	0.253 - 2259.32	610.45

510

Table 2: Stream types covered by our study reaches (Pottgiesser and Sommerhäuser, 2004).

Main category	Sub-category
Alps and Alpine foothills	1.1 = Small and mid-sized rivers
	2.1 = Small rivers in the alpine foothills
	2.2 = Mid-sized rivers in the alpine foothills
	3.1 = Small rivers in the Pleistocene sediments of the alpine foothills
	4 = Large rivers in the alpine foothills
Central highlands	5 = Small coarse substrate dominated siliceous
	7 = Small coarse substrate dominated calcareous highland rivers
	9 = Mid-sized fine to coarse substrate dominated siliceous highland rivers
	9.1 = Mid-sized fine to coarse substrate dominated calcareous highland rivers
	9.2 = Large highland rivers
Central plains	10 = Very large gravel-dominated rivers
	15 = Mid-sized and large sand and loam-dominated lowland rivers
	20 = Very large sand-dominated rivers



Ecoregion independent streams

11 = Small organic substrate-dominated rivers

21 = Lake outflows

Catchment size class:

Small river: 10 - 100 km²

Mid-sized river: 100 - 1,000 km²

Large river: 1000 - 10,000 km²

Very large river: > 10,000 km²

515 **Table 3: The relative influence of predictive variables of river hydro-geomorphology as computed from the fitted BRT model on water retention time.**

Variable	Short name	Relative importance (%)
Mean discharge (m ³ /s)	Qmean	57.42
Slope (m/m)	Slope	21.54
Drainage area (km ²)	Area	15.64
Mean river width (m)	Width	2.41
River type	RType	1.25
Substrate_Sand (%)	Sand	0.70
Substrate_Loam (%)	Loam	0.69
Substrate_Silt (%)	Silt	0.34
Substrate_Clay (%)	Clay	0
Substrate_Peat bog (%)	Peat bog	0
Substrate_Fen (%)	Fen	0