

**SIMULATING CHANGES IN RIVERINE PHYTOPLANKTON
ASSEMBLAGES UNDER INTENSIFIED ENVIRONMENTAL
STRESSORS – A DATA MODELLING STUDY IN A GERMAN
LOWLAND CATCHMENT**

By
Szewing To

Master thesis

Submitted in partial fulfillment of the requirement for the degree of
Master of Science in Environmental Management

First reviewer: Prof. Dr. Nicola Fohrer

Second reviewer: Dr. Naicheng Wu

Department of Hydrology and Water Resources Management

Faculty of Agricultural and Nutritional Sciences

Faculty of Mathematics and Natural Sciences

Christian-Albrechts-Universität zu Kiel

December 2018

ACKNOWLEDGEMENT

Without data, technical and emotional supports from different people, this project would not have been completed. I thank Prof. Dr. Nicola Fohrer for being my first supervisor. I express my deepest gratitude to Dr. Naicheng Wu, who suggested this interesting topic as my master thesis. He gave me constructive comments and warm encouragement until the end of my thesis. I am gratefully indebted to Yueming Qu for the permission of using her collected data in the Treene catchment. Her inspirational discussion and valuable comments throughout the project have contributed in various parts of the thesis. My special thank goes to several of my anonymous friends, who provided me huge motivations for completing this thesis. Lastly, I owe my heartfelt appreciation to my family, who always believe in me with their love.

Table of contents

Acknowledgement.....	I
List of Tables.....	IV
List of Figures.....	V
List of Appendices.....	VI
Abstract.....	VII
 1. Introduction.....	 1
1.1 Phytoplankton as bioindicators.....	1
1.2 Modelling riverine phytoplankton assemblages.....	2
1.3 Variable significance and impact strength.....	5
2. Study area and methods.....	7
2.1 Data collection and management.....	8
2.2 Data exploration.....	9
2.2.1 Predictor variables.....	10
2.2.2 Response variables.....	13
2.3 Ecological modelling.....	14
2.3.1 Generalized Linear Model and Linear mixed effect model.....	15
2.3.2 Model selection.....	16
2.4 Scenario studies.....	19
3. Results.....	22
3.1 Variable importance (VIMP).....	22
3.2 Generalized Linear Model (GLM) and Linear Mixed Effect Model (LMM).....	23
3.3 Final models.....	26
3.4 Scenario projections.....	30
4. Discussion.....	35
4.1 Importance and impact strength of abiotic variables on phytoplankton species richness.....	35
4.2 Importance and impact strength of abiotic variables on phytoplankton abundance and biomass.....	37
4.3 Non-linear subbasin responses towards stressors.....	39

4.4 Potentials and limitations of mixed model applications in explaining phytoplankton community patterns and predicting changes in the Treene catchment.....	41
5. Conclusion.....	43
6. References.....	46
7. Appendices.....	55
List of Publications.....	91
Declaration.....	92

List of Tables

Table 1	Data summary of predictor variables.....	11
Table 2	Data summary of response variables.....	14
Table 3	Model selection parameters or GLMs and GLMMs.....	17
Table 4	Scenario settings.....	21
Table 5	Above-mean variable importance value for response variables.....	22
Table 6	GLM performance of three response variables.....	24
Table 7	LMM performance of three response variables.....	26
Table 8	Wilcoxon test of changes significance of all and specific subbasins.....	30
Table 9	The median of changes in species richness and abundance.....	31

List of Figures

Figure 1	Study location and sampling points in the six subbasins Ki (Kielstau), Tr (for the mainstream Treene), Sa (Sankermark See), Je (Jerrisbek), Bo (Bollingstedter Au), Ju (Juebek) at Schleswig-Holstein, Germany.....	8
Figure 2	A brief working steps of this study (VIMP: Variable importance; GLM: Generalized linear model; LMM: Linear mixed effect model).....	9
Figure 3	Response variables data distribution and correlations.....	14
Figure 4	Variable Importance curve for response variables (Only variables with above-mean importance are shown. The variable codes have been defined in Table 1).....	22
Figure 5	Model coefficients from three final models (The variable codes have been defined in Table 1).....	27
Figure 6	Residual and normality plots of final models (Y-axis represents residuals).....	29
Figure 7	Distribution of predicted species richness and abundance in different scenarios.....	31
Figure 8	Projection of median change in species richness in each subbasin.....	33
Figure 9	Projection of median change in abundance (log cell/L) in each subbasin...	34

List of Appendices

Appendix I	Predictor variables data exploration summary	55
Appendix II	Supplementary information for model building and selections	57
Appendix III	R packages and functions (R version 3.5.1)	70
Appendix IV	Complete summary of abiotic variables	72
Appendix V	Complete list of simulated results	75
Appendix VI	Sample R script	80

Abstract

The pronounced response of riverine phytoplankton to environmental stressors offers a useful indicator for detecting anthropogenic disturbances to freshwater ecosystem. Phytoplankton species richness, abundance and biomass compose an elementary picture for the algal assemblage in an aquatic biota. Lowland river catchment Treene is an agriculture-dominated basin with a special river-lake land form. Their algal community response towards stressors intensification exemplify one of the possible subsequent changes to the riverine food web dynamics. Explaining stream phytoplankton response to abiotic environment with data modelling approach remains rare in the study field. This study was used as an endeavour to understand the effect of stressors to phytoplankton community patterns and model their changes under predicted pressure. The study also explored the possibility and capacity of data modelling on phytoplankton response. The temporal-seasonal and spatial effects on phytoplankton community change were emphasized and modelled. The study concluded that, species richness in the catchment was more sensitive to water level and pH while the abundance and biomass were largely influenced by nutrient levels; the phytoplankton community patterns in different subbasin presented different resilient capacity in more vigorous disturbances; simulated change in species richness ranged from -2.6 to 3.1 (sign. > 0.95) while the change in abundance level ranged from -0.7 to 0.4 (sign. > 0.95) in different subbasins; and there are capacities to explain phytoplankton assemblage with linear mixed effect models. This study highlighted the importance of integrating temporal and spatial management strategies in a river basin under intensified precipitation variations and human pressures.

(Word count: 249)

Key words: Riverine phytoplankton; lowland catchment; stressors; linear mixed effect model; data modelling

1. Introduction

With the intensifying pace of various anthropogenic activities and subsequent human-induced disturbances, freshwater ecosystems are increasingly under threat. To manage these threatened ecosystems, biomonitoring is emerging as preventive measures for conservation of water resources, and phytoplankton is one of the most popular bioindicator for monitoring (Wu et al., 2014). Phytoplankton are free-floating, photosynthetic microorganisms, and are easily found in marine and freshwater ecosystems (Levinton, 2001; Reynolds, 2006). The group comprises of algae, cyanobacteria, diatom and protist (Reynolds, 2006). The different aquatic habitats and food web mechanisms of marine and freshwater ecosystem have constituted different species composition, living and reproduction strategies (Levinton, 2001; Reynolds, 2006). Facing the impacts from environmental changes, the induced changes of phytoplankton population dynamics in freshwater ecosystems have been of great interest and information to ecologists and watershed management units (Hering et al., 2015). Modelling of riverine phytoplankton population dynamics has also been a major subsequent focus in depicting their community changes under spatial and temporal variations (Sellers and Bukaveckas, 2003).

1.1 Phytoplankton as bioindicators

The role of phytoplankton being a major primary producer for the aquatic food webs has made it an exceptionally important indicator for ecosystem health (Paształeniec, 2016). Following the bottom-up trophic cascade effect, any changes in phytoplankton assemblages induce far-reaching impacts to the dynamics and the structure of the entire aquatic food web (Barbosa et al., 2010; Paształeniec, 2016). The rapid and sensitive response of phytoplankton to environmental changes has attracted scientists to use it as an early signal of anthropogenic disturbances and climate change (Hay et al., 2005;

Pasztaleniec, 2016). They are also renowned indicators of the trophic state of water bodies and levels of human impacts on the aquatic ecosystems (Wu et al., 2014; Wu et al., 2017). By assessing and predicting the changes in the pattern of phytoplankton communities such as richness, abundance and biomass, ecologists can detect changes in water quality (Jeong et al., 2008). Concurrently, assessing the relationship of these changes and shifts in phytoplankton community can enhance the understandings on the resilience capacity of the freshwater ecosystem (Wu et al., 2014), which is prone to human-induced disturbances.

The relationship of phytoplankton distribution (Wu et al., 2011), biomass (Desortová and Punčochář, 2011), primary production activities (Sellers and Bukaveckas, 2003; Bussi et al., 2016), species diversity and composition (Anneville et al., 2005) and trait composition (Litchman and Klausmeier, 2008; Lange et al., 2016) to environmental variables have been widely explored in different parts of the world. When environmental variables exceed their acceptable range of variation, they elevate the stress level of the aquatic biota and become stressors (Feld et al., 2016). Intensified stressors will cause changes in phytoplankton community pattern and subsequently threaten their biodiversity and ecosystem functions (Matthaei et al., 2010). Accompanying the increasing rate of anthropogenic activities, the intensity of environmental stressors is progressively strengthened (Feld et al., 2016). Nutrient loading has been one of the most prevalent stressors in lowland rivers, where it worsens the water quality of freshwater ecosystems with eutrophication (Lázár et al., 2016). There are increasing needs to investigate major drivers of riverine phytoplankton assemblage changes.

1.2 Modelling riverine phytoplankton assemblages

Riverine phytoplankton biogeochemical and hydrological models were gradually introduced to assess variations in riverine phytoplankton assemblages related to nutrient loadings and flow condition (Whitehead and Hornberger, 1984; Sellers and Bukaveckas, 2003). Since the first decade of the 21st century, the increasing uncertainties and instabilities of climate variations from the previous decades have urged the development of multi-model analysis of phytoplankton assemblages. Ecologists increasingly considered climate variations when accounting for or modelling riverine algae responses (Trolle et al., 2014; Bussi et al., 2016; Pyo et al., 2017). It was until recent years, that a prominent trend of incorporating novel statistical data analysis in ecological studies was observed (Harrison et al., 2018). Such trend was largely caused by a rapid advancement of biological data modelling techniques such as generalized linear model and linear mixed effects model (Breuer et al., 2017; Arisanti et al., 2017). This certainly provided new perspectives in modelling the changes in phytoplankton community distribution under intensified stressors. As a result, developing innovative predictive models and testing their predictability on algal assemblages under multiple stressors remain one of the major agendas in freshwater resources management (Wu et al., 2017).

1.3 Variables and stressors to phytoplankton assemblages

Due to the particular sensitivity of phytoplankton to environmental alterations, a wide range of variables are held accounted for their population dynamics. There is no clear consensus of the major environmental factors that caused variation in riverine phytoplankton communities (Wu et al., 2011). Nevertheless, physicochemical parameters and nutrient sources were common variables examined in riverine phytoplankton studies (Ebigwai, 2014; Sharma et al., 2016). Additionally, Ferrareze (2012) concluded the importance of land cover in affecting variations in riverine

phytoplankton assemblages. Increased human intervention to ecosystem such as land use modification to intensive cropland in the catchment area caused high amount of nutrients and sediments runoff to the river (Ferrareze, 2012). Such case is particularly remarkable for lowland watershed, whose river course passes through gentle terrains consists of agricultural and residential land use (Lázár et al., 2016). Nutrient influx such as nitrogen and phosphate can easily lead to eutrophication or algae bloom, declining dissolved oxygen level and destroying balances in the aquatic ecosystem (Smith et al., 1999). The significant effects of hydrological variables on lowland riverine algae assemblages have also been highlighted (Wu et al., 2018). Unlike alpine and mountainous catchments with high elevations, lowland rivers are specifically characterized with low hydraulic gradients, muddy and sandy substrate, high water table and high chance of water retention in peatland and lakes (Schmalz and Fohrer, 2010; Wu et al., 2011;). Therefore, phytoplankton, instead of benthic algae, served as an important primary producer in a lowland catchment (Wu et al., 2014). Lowland riverine phytoplankton are known to be sensitive to flow variations (Wu et al., 2018; Desortová and Punčochář, 2011). Water level determines the habitat for phytoplankton and fluctuating river hydrology will threaten their survival and reproduction (Sabater et al., 2008).

Current extreme climate events caused variations in the precipitation patterns with intensified episodes of flood and drought occurrences (Bates et al., 2008). The unstable flow conditions caused watershed management to become more challenging. For example, in northern Germany, a rising annual flow rate and higher chance of flooding during peak flow (Asadieh and Krakauer, 2017) with a reduction in flow during low flow seasons (Kakouei et al., 2018) were anticipated. The combined effects of climate change and land use modification would induce steeper environmental gradients on

riverine phytoplankton community. It would also change their assemblage patterns and subsequently destabilizing the ecosystem structure. Hence, the urgency of identifying major environmental factors that govern variations in phytoplankton community pattern is increasingly important, for a successful catchment-based management (Wu et al., 2018).

1.4 Variable significance and impact strength

Major environmental variables that regulate phytoplankton community can be evaluated in terms of significance and impact strength. The former represents qualitative analysis while the latter denotes quantitative analysis. Impact strength is particularly important in a quantitative effect study (Sullivan and Feinn, 2012). Variable significance reflects how likely the impact will take place on the responding object (Sullivan and Feinn, 2012). The impact strength estimates the magnitude of an impact imposed by the variable on the phytoplankton community (Sullivan and Feinn, 2012; Nakagawa and Cuthill, 2007). It enables researchers to identify significant variables and quantify the positive and negative impact strength of those variables on the responding object. Investigating both the significance level and the impact strength of an impact are crucial in ecological studies (Sullivan and Feinn, 2012; Nakagawa and Cuthill, 2007) because they report different aspects of an impact towards the change in phytoplankton community patterns.

1.5 Environmental factors for the Treene catchment

Environmental and hydrological factors have influential effects in structuring pelagic phytoplankton community patterns in lowland streams of the Treene catchment (Wu et al., 2018; Qu et al., 2018a; Qu et al., 2018b). While the effect of environmental variables highly depends on spatial factors such as land use (Qu et al., 2018b), the importance of

hydrological variables tends to dominate during high flow period (Sun et al., 2017). Nonetheless, the diatom-based phytoplankton community in the Treene catchment is primarily sensitive to temporal-seasonal variations (Wu et al., 2011; Wu et al., 2016). In this study, the importance and strength of environmental variables influencing phytoplankton community assemblages in the lowland Treene catchment were investigated and evaluated. Variables with high importance were used to predict changes in phytoplankton richness, abundances and biomass under climate and land use change scenarios.

This study aims to use a modelling approach to explain how phytoplankton community patterns are affected by abiotic factors. After which, we provided an insight into potential changes in phytoplankton assemblages under intensified environmental stressors in the studied subbasins. Three research questions were investigated: (i) What are the significant variables for phytoplankton assemblage in terms of richness, abundance and biomass in the Treene catchment? (ii) Which significant variables have strong impacts on phytoplankton assemblages? (iii) How do phytoplankton assemblages change in different subbasins under intensified stressors? We tested three hypotheses: (i) Abiotic variables have different significance on phytoplankton assemblages; (ii) Significant abiotic variables impose different impact strength on phytoplankton assemblages; (iii) Phytoplankton assemblages responded differently in different subbasins under intensified stressors. Additionally, the importance of spatial and temporal effects in determining the pattern of phytoplankton assemblages in the Treene catchment, as emphasized in previous studies (Wu et al., 2016; Qu et al., 2018a; Qu et al., 2018b), are incorporated in this study. This study was based on the assumption that the phytoplankton community patterns in the studied catchment were influenced by temporal and spatial effects.

2. Study area and methods

The study was conducted in the catchment of the river Treene located between Flensburg and Kiel in the North-west of Schleswig-Holstein, northern Germany (Figure 1). The Treene catchment is a subbasin of the Eider River, which flows into the North Sea, and under temperate marine climate. The Treene catchment is 481km² with six subbasins: Tr (Treia), Bo (Bollingstedter Au), Je (Jerrisbek), Ju (Juebek), Ki (Kielstau), Sa (Sankermark See). There are several lakes in the catchment in the upstream. The three largest lakes are: Sueden See (0.64 km²), Sankermarker See (0.56 km²) and Winderatter See (0.24 km²). Treene catchment is a lowland basin featured with groundwater-dominated river with low hydrological gradients (Guse et al., 2015; Schmalz and Fohrer, 2010; Pfannerstill et al., 2014). It has mild relief with maximum altitude of about 80m above sea level. It is a predominantly agricultural catchment, comprised of 48% agricultural areas and 31% pastoral land while forest and urban-residential areas have relatively small share of around 20% (Haas et al., 2017). The comparatively mild slope in the northeast of the catchment (i.e. the headstream) allows favourable conditions for arable farming, consisted mainly of silage corn, winter wheat, barley, rye and rape (Haas et al., 2017). The southwest of the catchment (i.e. the lower course) is featured with grassland and pastoral activities (Guse et al., 2015).

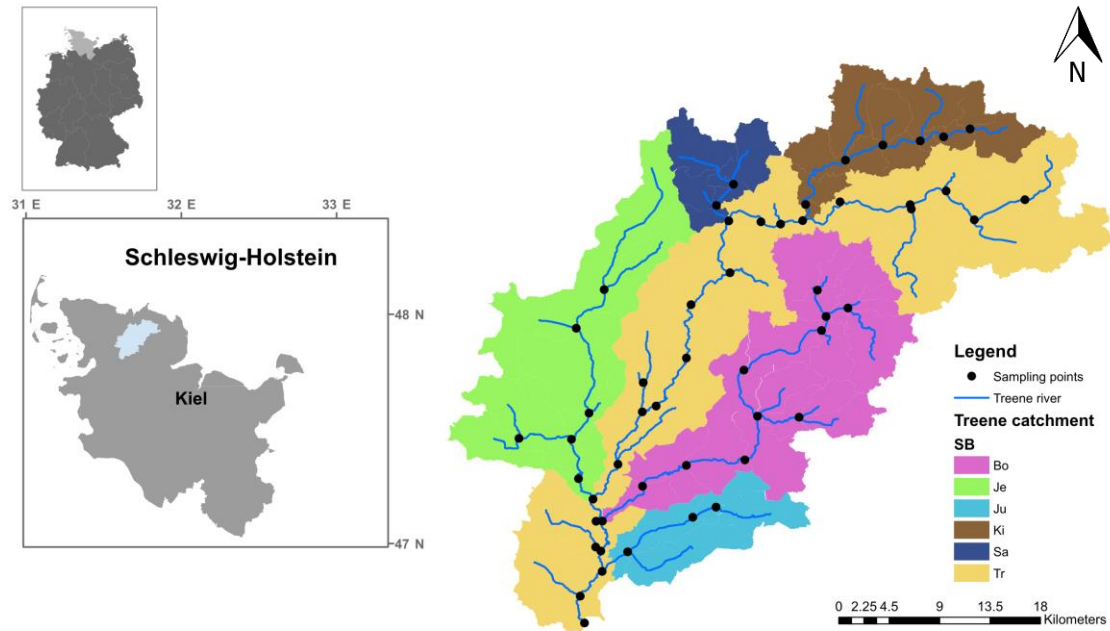


Figure 1 Study location and sampling points in the six subbasins Ki (Kielstau), Tr (for the mainstream Treene), Sa (Sankermark See), Je (Jerrisbek), Bo (Bollingstedter Au), Ju (Juebek) at Schleswig-Holstein, Germany

2.1 Data collection and management

We set up 59 sampling sites and one sample was taken per site in December 2014 and September 2015 respectively. A total of 118 samples were collected and adopted in this study. The location, sampling points, Treene river and subbasin distributions are shown in Figure 1. The number of samples from distinct subbasins were different. Physicochemical parameters data was collected on the same day with water samples. All collection practice and sample treatments were reported in Qu et al. (2018a, 2018b). The sample treatments and identification were conducted by the same person to minimize variation and maintain data consistency. In the following, a brief work flow of this study were illustrated in Figure 2. Each analysis stage was further described and elaborated in following sub-sections. Statistical analysis and data modelling were conducted in R Studio Environment (R Core Team, 2018). All packages and functions adopted in this study are listed in Appendix III.

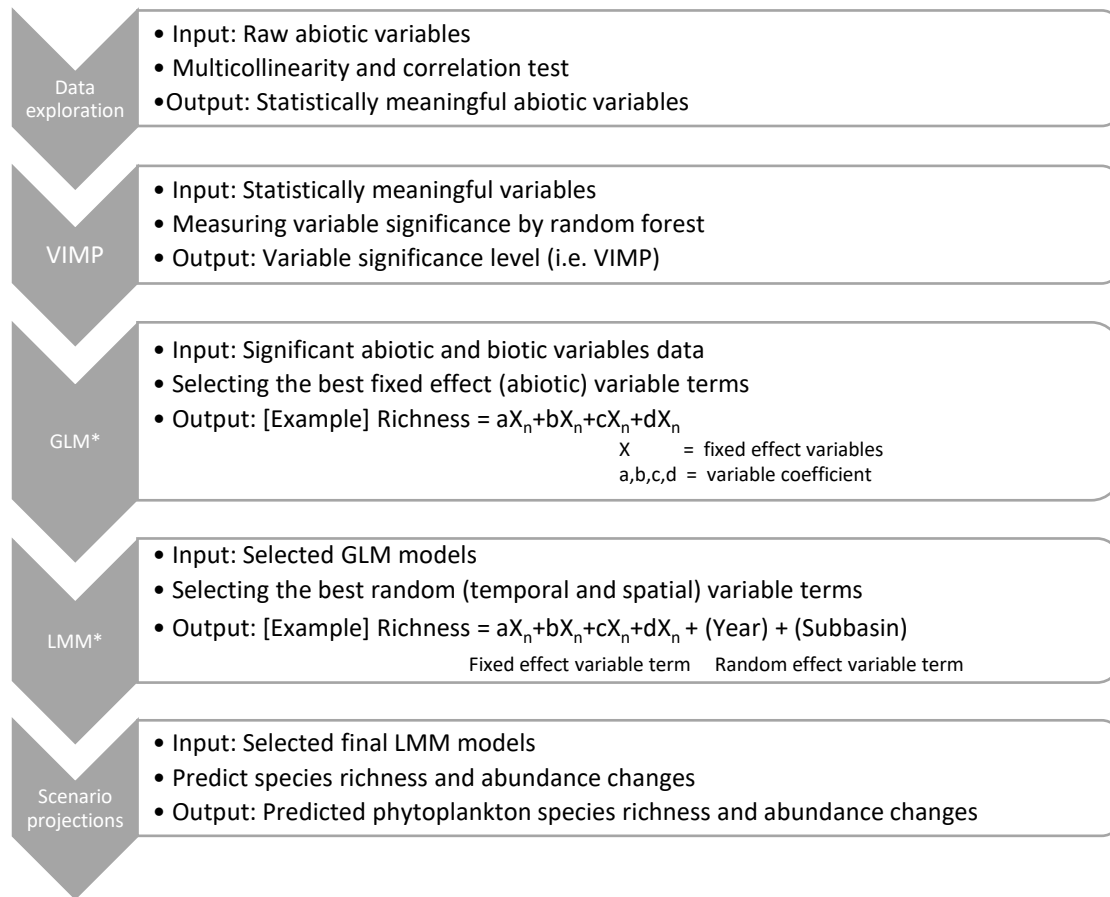


Figure 2 A brief working steps of this study (VIMP: Variable importance; GLM: Generalized linear model; LMM: Linear mixed effect model)

2.2 Data exploration

The phytoplankton cell density (cell/L), biomass ($\mu\text{g/L}$) and physicochemical data were the raw data adopted in this study. Input variables for the modeling analysis included abiotic (predictor variables) and biotic factors (response variables). Abiotic variables were: 1) hydrological variables (Hv) including flow velocity, river depth and width ($n=60$); 2) physicochemical variables (Ev) ($n=13$) and 3) land use variables (LU) ($n=11$). The full list of abiotic variables is attached in Appendix IV. Hv were calculated according to Olden and Poff (2003) in Wu et al., (2018) from a long-term flow discharge data (2010-2016) simulated for each sampling site (for details, see Wu et al., 2018). Ev were obtained from *in-situ* and laboratory measurements. LU was calculated via the Geographical Information System data as in Qu et al. (2018a). Afterwards, the land use

was assigned to each sampling point according to the subbasin that they were located in. All abiotic variables were used to select useful variables as predictor variables through statistical tests.

2.2.1 *Predictor variables*

A total of 86 abiotic factors were used as input to the correlation and multicollinearity test to select statistically consistent variables as predictor variables for the analysis (Figure 2, step 1). Covariates and correlated variables often cause uninterpretable results (Zuur et al., 2010). Multicollinearity has been the biggest obstacle for generalized linear modelling (Feld et al., 2016). Neglecting it would cause a lack of observable significance on any abiotic variables (Zuur et al., 2010). A high correlation between abiotic variables could deteriorate model interpretation because the variables effects were not independently explained, ending up with a similar problem as collinearity (Zuur et al., 2009). Variance Inflation Factor (VIF) and Pearson's correlation coefficient were used to test collinearity and correlation (Harrison et al., 2018). It considers non-linear relationships between variables (Feld et al., 2016) and is useful in detecting multicollinearity (Zuur et al., 2010). A sequential removal of variables was applied by excluding the highest VIF in the collinearity test each time after recalculation. This process was repeated as a loop until all variables attained the threshold VIF (Naimi et al., 2014). The correlation and collinearity threshold were set at $R^2 \leq 0.7$ (Harrison et al., 2018) and $VIF \leq 5$ (Zuur et al., 2010) respectively. After this process, 30 remaining predictor variables were screened and used for this study. The data summary of the selected predictors including their abbreviation and names are summarized in Table 1. The maximum correlation rate is 0.69 between NPR and NO₃ (Appendix I) while the maximum VIF is 4.81 from NO₃ (Table 1). The data distribution and correlation plot are shown in Appendix I.

Table 1 Data summary of predictor variables

Abbreviation	Names	Unit	Max.	Min.	Mean	SD	VIF
pH	pH	-	9.73	6.73	7.79	0.53	2.05
DO	Dissolved oxygen	mg/L	12.98	2.93	9.13	1.84	2.26
PO4	Phosphate	mg/L	0.63	0.01	0.07	0.09	2.41
NH4	Ammonium	mg/L	1.51	0.02	0.23	0.28	2.65
NO3	Nitrate	mg/L	37.49	0.1	6.4	5.68	4.81
NO2	Nitrite	mg/L	0.73	0	0.04	0.09	2.74
NPR	Nitrogen-phosphate ratio	-	234.16	0.23	50.37	56.8	4.69
Cl	Chlorine	mg/L	98.54	14.23	28.48	10.02	1.66
SO4	Sulfate	mg/L	84.92	10.32	37.95	14.16	3.20
TSS	Total suspended solids	mg/L	87.88	0	11.31	11.68	1.62
H11	Variability flows of 3 day before (not including the sampling day)	-	1.09	0	0.04	0.15	2.44
H12	Skewness in 3 day flows (including the sampling day)	-	1.73	-1.73	0.34	1.02	1.40
H13	Skewness of 3 day before (not including the sampling day)	-	1.73	-1.72	0.62	1.08	1.56
H20	Skewness in 7 day flows (including the sampling day)	-	2.44	-0.78	0.88	0.69	3.28
H21	Skewness of 7 day before (not including the sampling day)	-	2.64	-1.58	0.95	0.86	3.13
H29	Skewness of 14 day before (not including the sampling day)	-	3.37	-0.53	0.69	0.66	2.80
H36	Skewness in 30 day flows (including the sampling day)	-	3.28	-0.31	1.29	0.56	2.72
H41	Low flood pulse count 14 days	Days	30	0	17.87	8.88	2.23

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors:
A data modelling study in a German lowland catchment – Szewing To – December 2018
Chapter 2: Study area and methods

H55	Rate of change in flow events	-	0.09	-0.31	-0.03	0.07	3.51
WIDTH	River width	m	17.1	1.05	4.94	3.63	3.77
DEPTH	River depth	m	2	0.1	0.47	0.32	2.36
VELOCITY	Flow velocity	m/s	10.24	0	0.62	1.61	1.95
FRSE	Evergreen forest	%	9.02	0.02	1.02	1.37	2.05
FR	Total forest	%	15.13	0.86	5.71	3.4	1.76
RNGE	Rangeland	%	4.33	0	0.7	0.8	1.69
UIDU	Industrial land	%	8.41	2.98	4.21	0.9	3.07
UR	Urban-residential land	%	12.26	1.75	5.65	2.01	2.34
WATR	Water	%	5.42	0.62	1.71	1.1	2.80
WETL	Wetland	%	7.19	0	1.01	1.11	1.63
WPAS	Winter pasture	%	70.97	7.22	29.18	16.66	2.24

2.2.2 *Response variables*

The study focused on three response variables: species richness, abundance and biomass. They are ideal indicators for aquatic ecosystem health (Wu et al., 2014; Wu et al., 2017) and have been widely used to test for phytoplankton response to environmental variables (Macedo et al., 2001; Gharib et al., 2011; El Gammal et al., 2017). Species richness is a basic and straight-forward measurement for community diversity and integrity of a habitat (Hassany, 2012). Identifications and assessments of species richness can be used to observe primary changes in a community (Levinton, 2001). In our sampling, a total of 386 taxa found. The abundance was interpreted by cell density of each site while the biomass was calculated out of the biovolume of identified species at each site. Due to the high variance and non-normality of cell density and biomass data, logarithmic transformation was applied. Data transformation has been a heated debate over years and no best approach has been universally concluded for ecological researches (Harrison et al., 2018). Log-transformation compresses large values of phytoplankton cell density data for better analysis quality (McCune and Grace, 2002). It also expands the small values of biomass into analyzable range. Ives (2015) evidently demonstrated a better performance of log-transformed count model over a non-transformed model. Log-transformation is a popular method because of its ability of backward transformation after the analysis so the scale of original data can be preserved (Gelman and Hill, 2007). Hence, log-transformation was applied to abundance and biomass data. The covariance of species richness, abundance and biomass were tested to validate their statistical independence (Zuur et al., 2007). Their low variation inflation factor (VIF) implied that they are unlikely covariates (Zuur et al., 2010). Shapiro test reflected the non-normality of these response variables (Table 2). Their data distribution and correlation are illustrated in Figure 3.

Table 2 Data summary of response variables

Response variables	Units	General				Shapiro test	VIF test
		Min.	Max.	Mean	SD	p.value	vif value
Richness	Count	7.0	62.0	33.7	14.7	0.00	1.24
Abundance [#]	Cell/L	7.4	18.7	12.0	2.3	0.04	2.92
Biomass [#]	µg/L	-5.5	2.6	-2.4	1.4	0.00	2.76

[#] Log-transformed data

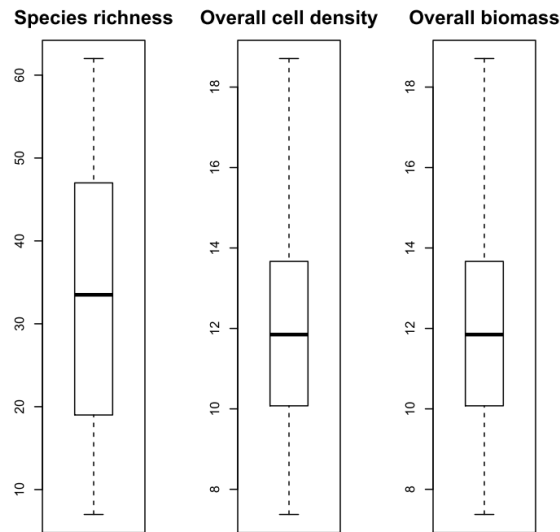


Figure 3 Response variables data distribution and correlations

2.3 Ecological modelling

Ecological data is complex and has been acknowledged openly among scholars and scientists that, species occurrence and abundance are not in linear but in unimodal relationship with environmental gradient (Austin, 1987; Whittaker, 1967). Such feature has hindered the use of simple statistical methods for ecology and evolution because they often involve unrealistic assumptions (Zuur et al., 2009). For example, linear model ignores nested data, temporal and spatial correlation, landscape differences and other possible noise in the analysis (Zuur et al., 2007), of which are usually observed in ecological data. Ecological data is also often characterized with hierarchical structure that involves various non-independent observations such as repeated measurements

(Harrison et al., 2018). Such structure gives rise to fixed and random effects. Fixed effect refers to variables that are expected to affect the response variables that, their variations would be experimentally explained and proved (Bolker et al., 2009). Random effects, which are categorical in nature, are attributed from unexplained or uncontrolled variables that existed in the reality. Linear models exclude random effects or treat them as fixed effects (Bolker et al., 2009; Zuur et al., 2010), resulting in unrealistic conclusions (Feld et al., 2016). Therefore, a generalized linear model and linear mixed effect model would be more suitable to meaningfully analyze ecological data.

2.3.1 *Generalized Linear Model and Linear mixed effect model*

The classical linear regression is inadequate in explaining complex ecological data. Generalized linear model (GLM), a non-linear regression tool, was adopted in this study to derive the best combination of fixed effect variables (i.e. the predictor variables). It is an extension of the linear model that handles nonlinear and non-normal data (Bolker et al., 2009). It calculates variables coefficients by maximum likelihood (Olsson, 2002), which is powerful in testing fixed effect variables terms (Zuur et al., 2009). Data distribution (e.g. Poisson, Gaussian, Binomial) of the response variable must be specified to build a GLM (Olsson, 2002). The species richness was a count data so it was fitted into a poisson model while the log-transformed abundance and biomass were continuous data so a gaussian model was applied on them.

Linear mixed effect model (LMM), considered both fixed and random effect variables, was used to retain the independence of categorical data (i.e. Subbasins) (Harrison et al., 2018). There are no precise rules to classify fixed and random effects because it depends largely on the testing objects (Gelman and Hill, 2007; Bolker et al., 2009). In this study, all predictor variables were regarded as fixed effects variables while temporal (i.e.

seasons) and spatial (i.e. subbasin) factors were treated as random effects variables as they were not our testing objects. Different from fixed effect variables, the coefficients of random effect variables were determined categorically (Harrison et al., 2018). They represented the unexplained effect from different subbasin such as elevation and morphological features towards the response variables. Combining with the selected fixed effect variable terms, LMM tested the best random effect variables terms. It evaluates simultaneously the fixed and random effect variables so all coefficients can be assessed in one model (Bates et al., 2015). Further details related to the structure and application of GLM and LMM are provided in Appendix II. All the created candidate models, were assessed and screened afterwards.

2.3.2 *Model selection*

Model comparison and selection were conducted for GLM and LMM candidate models to obtain the best model. It is a crucial step to enhance model explanatory power and inferences (Burnham and Anderson, 2002). Numerous metrics have been developed and their functions were debated for model comparison and selections (Olsson, 2002; Burnham and Anderson, 2002; Johnson and Omland, 2004; Brewer et al., 2016; Harrison et al., 2018). There is no unique metric that is applicable to all regression models. It is important to employ a range of metrics that assess both model inference and goodness-of-fit to avoid the tendency of biased selection (Burnham and Anderson, 2002; Bolker et al., 2009). The metrics used in this study for evaluation is listed in Table 3. A detailed description and calculation formula of each metric are provided in Appendix II. Since the selection objectives of GLM and LMM were different (i.e. selecting fixed and random effect variable terms respectively), they adopted different model assessment methods to eliminate poor performance models.

Table 3 Model selection parameters or GLMs and GLMMs

	Parameters	Description	Application
GLM	$\Delta AICc$	<p>Small-sample-sized corrected Akaike Information Criterion</p> <ul style="list-style-type: none"> - It should be used if the ratio of sampling size (n) and fixed effect variables (K) (n/K ratio) is <40 (Burnham and Anderson, 2002, pp.66). - It offers bias adjustments to nonlinear regression model (Hurvich and Tsai, 1989) - It penalizes over-fitting models (Burnham and Anderson, 2002) - For model comparison purposes, the difference between AICc of models are considered instead of the absolute AICc value (Burnham and Anderson, 2002) 	$\Delta AICc$ was used as the first screening of all candidate models. $\Delta AICc < 2$ means models having same level of performance while $\Delta AICc > 10$ implies models can be excluded (Burnham and Anderson, 2002). Models with same performance were further assessed with RSE and Adjusted r^2 described below.
	RSE	<p>Residual Standard Error</p> <ul style="list-style-type: none"> - It is one of the parameters for testing model goodness-of-fit - It is based on error values of model residuals 	RSE and Adjusted r^2 were used as second screening for models with same performance (i.e. $\Delta AICc < 2$). These two indices focus on assessing model goodness-of-fit.
	Adjusted r^2 (R^2_a)	Normal regression coefficient would biasedly favour global model when the number of observations is low and explanatory variables are high e.g. n/K ratio < 40 (Burnham and Anderson, 2002). Adjusted regression coefficient considers and balances off small sample effects to reduce selection bias.	

LMM	$\Delta AICc / \Delta(q)AICc$	<p>Small-sample-sized corrected Akaike Information Criterion</p> <ul style="list-style-type: none"> - The same as abovementioned <p>Quasi-corrected Akaike Information Criterion (qAICc)</p> <ul style="list-style-type: none"> - Used by over-dispersion models (Burnham and Anderson, 2002) - Included an over-dispersion parameter in the calculation (\hat{c}) - Same as AICc, the difference between qAICc was considered for model comparison purposes 	<p>$\Delta qAICc$ was used for assessing over-dispersed models (Burnham and Anderson, 2002).</p> <p>$\Delta AICc / \Delta qAICc$ was used as the first screening. $\Delta AICc / \Delta qAICc < 2$ were further assessed with marginal r^2 and conditional r^2 described below.</p>
	Marginal r^2 (R^2_m)	Developed by Nakagawa et al. (2013), R^2_m is a useful fitness metric for mixed model evaluation, because it reported how much random effect variance and error are explained by the fixed effect variables. The greater the random effect variance and error, the smaller the R^2_m . It is more useful to use R^2_m for assessing random effect variables (Nakagawa et al., 2013).	Models with same performance were compared with R^2_m and R^2_c .
	Condition r^2 (R^2_c)	Comparable to R^2_m , R^2_c measured how much error variance are explained by both fixed and random effect variables (Nakagawa et al., 2013). The greater the error variance, the lower the R^2_c .	A higher R^2_m value indicates lower random effect variance and errors while a greater R^2_c value suggests the model better explaining the error variance. Both of them imply higher modeling power and performance.
Final model			In final model evaluation between richness, abundance and biomass, AIC-derived products were no longer available for comparison because of different model dataset (Burnham and Anderson, 2002). Conditional r^2 was employed to compare the total fitness of the three models.

2.4 Scenario studies

The final models were used to conduct eight projected scenarios to explore potential changes of phytoplankton community under various intensification of environmental stressors. Extreme climate variations and land use modifications are the two most likely challenges to the Treene catchment (Kakouei et al., 2018, Haas et al., 2017). Therefore, the scenarios are established accordingly. The scenarios set ups are briefly described below and tabulated in Table 4.

Increased fertilization or leaching rate of nutrients: Phosphate-phosphorus (PO₄-P)

Phosphorus was considered one of the limiting factors for phytoplankton biodiversity in freshwater ecosystem (Hutchinson, 1967) and ecologists have highlighted the disruption of excess phosphate to phytoplankton biodiversity (Torrecilla et al., 2005; Veronica et al., 2014). A high phosphate concentration in water bodies can result from diffuse point-source of sewage effluent discharge and agricultural runoff (Torrecilla et al., 2005). Phosphate-phosphorus (PO₄-P) acts as a direct available nutrient form for phytoplankton uptake, is especially prone to an outbreak of eutrophication (Matthaei et al., 2010) under a mass influx. Non-point source cause of phosphate influx can be resulted from their excessive accumulation in soil leaching to groundwater or seepage over a long time periods (Haygarth and Jarvis, 1999). Schleswig-holstein, where Treene catchment situates, has a history of intensive agricultural activities although the stringent regulation enforcement has gradually relieving her land intensity in recent decades, dairy and arable farming remains prominent (Farmer, 2004). Application of phosphate has highly reduced nowadays due to regulations and an increase in the cost of phosphate (Farmer, 2004). Nonetheless, the accumulated phosphate surpluses in soil caused by the intensive farming over the last half century remain a threatening source

of PO₄-P runoff in European regions (Haygarth and Jarvis, 1999). Lowland river basin represents a high proximity of farmland to surface water. The drain tiles and surface runoff constituted pathways for nutrients entering the water bodies. The mild slope also implies delayed mobilization of phosphorus, which could lead to a drastic increase of available nutrient for phytoplankton bloom (Haygarth and Jarvis, 1999). Scenario 1 and 2 simulated an increase scene of PO₄-P of 15 % and 30% respectively. They were referenced from reduced fertilization scenarios from Haas et al (2017).

Land use modification

Caused by increasing human activities, rural region gradually changed their natural or forest land to plantation areas (Veronica et al., 2014). Winter wheat cropped area in Germany has increased by about one-third during the last 25 years (Weiser et al., 2018). Different best management practices have been studied including increasing pasture land (Haas et al., 2017) to reduce croplands, which are fertilizer-intensive. Two different replacing rate (i.e. 10% and 20%) of agricultural land (AGRL) with winter pastoral land (WPAS) was simulated in scenario 3 and 4. The assumption was that lesser agricultural land directly reduced the same amount of fertilizers, nitrate (NO₃) and phosphorus (P) runoff to the water. The nutrient mobilization and runoff rate were not considered in this model. Scenario 4 additionally modelled a reduction of lake area in the catchment.

Precipitation variation

Climate change will change precipitation patterns and variability and flow regime would be growingly unstable (Kakouei et al., 2018). Flood or drought are the results of extreme weathers (Bates et al., 2008), and flood events imposed strong modifications on river habitat and biota (Vörösmarty et al., 2010). A general decreased of up to 20%

precipitation is anticipated in summer months (June, July and August) in the distant future (i.e. 2100) by the Germany Federal Environmental Agency (Buth et al., 2015). Wagner et al. (2013) has predicted a general 5% increase of precipitation for Germany in the year of 2050 and Jacob et al. (2012) expected a greater increasing rate of rainfall for northern Germany to 15%. Scenario 5 to 8 simulated the range of decline and increase of precipitation of 5% and 15% (Table 4).

Table 4 Scenario settings

Scenarios	Descriptions	Reference
Scenario 1	Increase fertilization and leaching of phosphate-phosphorus of 15%*	Haas et al. (2017)
Scenario 2	Increase fertilization and leaching of phosphate-phosphorus of 30%*	Haas et al. (2017)
Scenario 3	10% increase in WPAS land use** 10% decrease in PO4 10% decrease in NO3 10% decrease in AGRL land use	Haas et al. (2017)
Scenario 4	20% increase of WPAS land use 10% decrease in PO4 10% decrease in NO3 10% decrease in AGRL land use 10% decrease in WATR land use	Haas et al. (2017)
Scenario 5	5% reduction in precipitation	Buth et al., 2015 (FEA Germany)
Scenario 6	15% reduction in precipitation	
Scenario 7	5% increase in precipitation (Germany, year 2050)	Wagner et al., 2013
Scenario 8	15% increase in precipitation (Northern Germany, year 2050)	Jacob et al., 2012

*15% and 30% of reduction in fertilization was simulated in Haas et al. (2017) but increments were adopted in the scenario setting because of stressors intensification projection

** 20% increment in pastoral land was similarly modelled in Haas et al. (2017)

3. Results

3.1 Variable importance (VIMP)

The variable importance (VIMP) measured the significance of the impact from predictor variables taking place on the response variables. From our results, different VIMP rankings implied that the response variables had different sensitivities to the predictor variables (Figure 4). The VIMP curve showed both how and how much do the VIMP decline. The gaps among the curve indicate a significant drop of VIMP. The three response variables had varying curve structure but generally they showed a concave shape (Figure 4). They had rapid depressions of importance after the first or second most importance variables. Variables with above-mean importance were extracted and listed in Table 5.

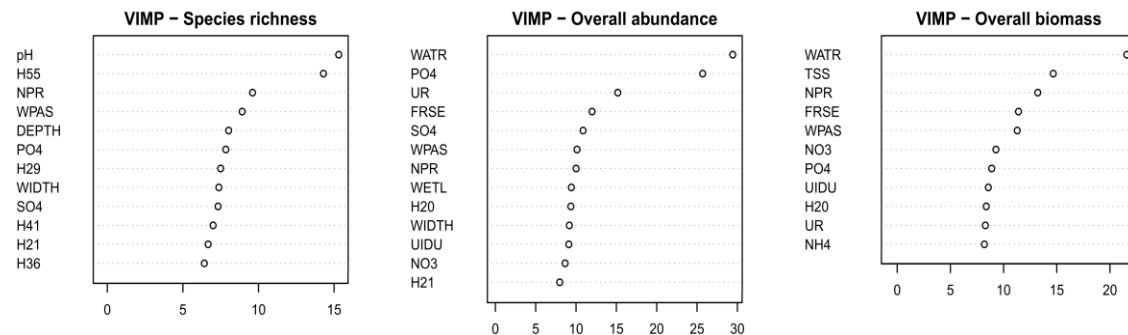


Figure 4 Variable Importance curve for response variables (Only variables with above-mean importance are shown. The variable codes have been defined in Table 1)

Table 5 Above-mean variable importance value for response variables (The variable codes have been defined in Table 1)

Species richness		Overall abundance		Overall biomass	
Variables	Importance	Variables	Importance	Variables	Importance
pH	15.30	WATR	29.42	WATR	21.58
H55	14.29	PO4-P	25.69	TSS	14.66
NPR	9.60	UR	15.17	NPR	13.21
WPAS	8.91	FRSE	11.97	FRSE	11.4

DEPTH	8.02	SO4	10.87	WPAS	11.28
PO4-P	7.84	WPAS	10.11	NO3	9.27
H29	7.49	NPR	10.00	PO4-P	8.88
WIDTH	7.37	WETL	9.42	UIDU	8.56
SO4	7.32	H20	9.35	H20	8.36
H41	6.99	WIDTH	9.16	UR	8.28
H21	6.66	UIDU	9.10	NH4	8.19
H36	6.41	NO3	8.64		
		H21	7.97		

Species richness was sensitive to a mixture of physicochemical, hydrological and land use variables but hydrological variables played slightly greater roles (Table 5). pH and rate of change of flow (H55) were the two most important factors that far outweighed other components (Figure 4). Hydrological factors had comparatively lower importance to abundance, which was more land use- and physicochemical- oriented. In Table 5, 10 out of 13 above-average important variables of abundance, were related to land use and physicochemical parameters. Water area (WATR) and phosphate-phosphorus (PO4-P) had distinctively greater importance over the other variables (Figure 4). Biomass had a slightly different VIMP structure compared to that of richness and abundance where, one single variable, water area (WATR), had considerably higher importance than other variables (Figure 4). This suggested that WATR was highly accountable for biomass pattern. Similar to abundance, biomass was also significantly affected by land use and nutrients-related physicochemical variables (e.g. NPR, TSS, NO3, PO-P, NH4) (Table 5). Similar to abundance, the impact of hydrological variables was less significant to biomass.

3.2 Generalized Linear Model (GLM) and Linear Mixed Effect Model (LMM)

The group of selected important variables (Table 5) was inputted into GLM as local

models (i.e. Smaller models contained only selected variables) to determine the best structure of fixed effect variables (i.e. predictor variables). Each response variable was tested with six GLMs in different structures (Appendix II). The GLM performance of the response variables are summarized in Table 6. $\Delta AICc < 2$ implies models with equal performance (For details, see Appendix II). In our results, model 5 performed distinctively well for species richness, and model 5 and 6 of both abundance and biomass performed equally well (Table 6). The $\Delta AICc$ between their model 5 and 6 were very small, being 1.49 for abundance and 1.88 for biomass (Table 6). Therefore, they were further evaluated by RSE and adjust R^2 . However, they remain competitive in RSE and adjust R^2 with comparable or same values (Table 6). Randomly selecting either one of them would risk losing some information from the models. Thus, a model 7 was developed for both abundance and biomass. In their model 7, the variables from their corresponding model 5 and 6 were combined correspondingly. Both of their model 7 had achieved better results in $\Delta AICc$, RSE and adjust R^2 (Table 6) and have been selected as the best GLM for abundance and biomass. Details on model selection procedures and information are provided in Appendix II. The residual plots of all GLMs are illustrated in Appendix II.

Table 6 GLM performance of three response variables

sp = species richness ab = abundance bio = biomass	$\Delta AICc$	RSE	adjR2
GLMsp1	31.53	1.20	0.55
GLMsp2	10.85	1.09	0.50
GLMsp3	38.18	1.05	0.35
GLMsp4	43.86	1.04	0.32
GLMsp5*	0.00	1.07	0.48
GLMsp6	9.66	1.05	0.45
GLMab1	23.32	1.43	0.61
GLMab2	16.61	1.60	0.52
GLMab3	18.61	1.68	0.47
GLMab4	12.40	1.66	0.48

GLMab5	5.29	1.60	0.52
GLMab6	3.80	1.59	0.52
GLMab7*	0.00	1.54	0.55
GLMbio1	34.24	1.07	0.45
GLMbio2	12.37	1.13	0.37
GLMbio3	44.59	1.33	0.13
GLMbio4	39.74	1.33	0.14
GLMbio5	4.09	1.13	0.38
GLMbio6	5.97	1.15	0.35
GLMbio7*	0.00	1.10	0.40

Remarks: '*' implies selected model

The selected fixed effect structure for the three response variables is interpreted below.

They were inputted to LMM to test random effect variables.

$$\text{Richness} = \text{pH} + \text{DO} + \text{PO4} + \text{NPR} + \text{SO4} + \text{H2O} + \text{DEPTH} + \text{FR} + \text{WPAS}$$

$$\text{Abundance} = \text{DO} + \text{PO4} + \text{H36} + \text{H55} + \text{WATR} + \text{UR} + \text{WIDTH}$$

$$\text{Biomass} = \text{NO3} + \text{TSS} + \text{H55} + \text{FR} + \text{WATR} + \text{PO4}$$

Over-dispersion of species richness in the LMM was compensated by using qAICc (Table 3) as evaluation parameters. Competitive LMM models of different response variables were derived by $\Delta\text{qAICc}/\Delta\text{AICc} < 2$. They are: model 1 and 3 for species richness (ΔqAICc of 1.86), model 3 and 5 for abundance (ΔAICc of 1.83) and model 2, 3 and 5 for biomass (Highest ΔAICc of 0.89) (Table 7). They were further evaluated with marginal r^2 (R^2_m) and conditional r^2 (R^2_c). Model 3, 5 and 2 were subsequently selected as the final model for species richness, abundance and biomass, respectively. Details of the model settings and graphics are provided in Appendix II. The final model terms including their corresponding variable coefficients are described in next section.

Table 7 LMM Performance of three response variables

sp = species richness ab = abundance bio = biomass	$\Delta(q)AICc$	R^2_m	R^2_c
LMMsp1	0.00	0.34	0.57
LMMsp2	2.25	-	-
LMMsp3*	1.86	0.56	0.59
LMMsp4	59.52	-	-
LMMsp5	2.19	-	-
LMMab1	4.26	-	-
LMMab2	8.30	-	-
LMMab3	1.83	0.44	0.54
LMMab4	47.60	-	-
LMMab5*	0.00	0.48	0.61
LMMbio1	2.39	-	-
LMMbio2*	0.00	0.41	0.44
LMMbio3	0.89	0.42	0.42
LMMbio4	52.31	-	-
LMMbio5	0.86	0.42	0.43

Remarks: ‘*’ implies selected model; Only $\Delta(q)AICc < 2$ were entitled to R^2_m and R^2_c comparison.

3.3 Final models

The final model terms for examining phytoplankton community patterns including predictor coefficients had the following terms:

Richness = $-7.68pH - 0.08NPR - 24.53PO_4 - 0.32WPAS + 15.41DEPTH + 3.25H_2O + (1 | \text{Year:Subbasin})$

Abundance = $-0.22DO + 9.06PO_4 - 0.10H_36 - 2.66H_55 + 0.75WATR + 0.11UR + 0.08WIDTH + (1 | \text{Year:Subbasin})$

Biomass = $0.04NO_3 + 0.05TSS - 3.37H_55 + 0.05FR + 0.53WATR + 3.28PO_4 + (1 | \text{Year})$

Spatial and temporal-seasonal factors have been included in the LMMs. The coefficient values of model predictors indicated the impact strength of the predictor variables to the response variables. They were extracted from the model and plotted in Figure 5. The width of the bar in the plot represented the standard deviation range of the fitted

values. In Figure 5, pH had a fairly strong negative effect (sign. > 0.99) on species richness but water depth (15.41, sign. > 0.99) and dissolve oxygen (2.24, sign. >0.99) both facilitate their richness development. Their positive impacts were even able to compensate the distress from pH (Figure 5). PO₄ had strong positive impacts on both abundance (9.06, sign. > 0.99) and biomass (3.28, sign. >0.95). Apart from DO, there were no strong discouraging impacts on phytoplankton abundance (Figure 5). Nevertheless, the favourable impact of PO₄ on biomass was likely to be balanced by the suppressing effects of rate of change (-3.37, sign. >0.95).

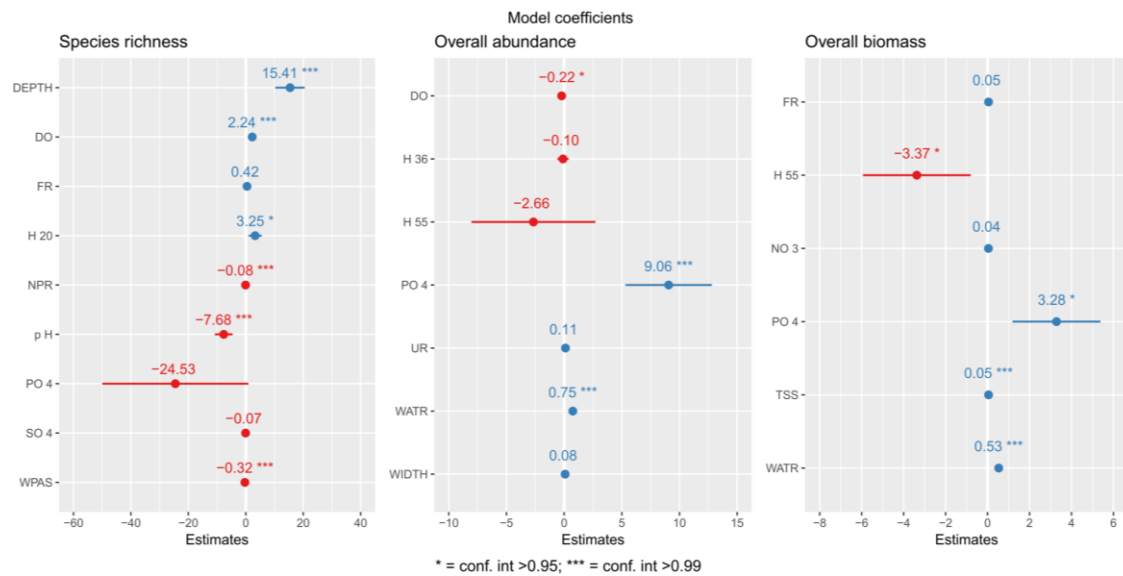


Figure 5 Model coefficients from three final models (The variable codes have been defined in Table 1)

The overall fit of the final models was evaluated numerically and graphically by conditional r^2 (R^2_c), residual plots and QQ plot (Figure 6). The residual plots demonstrated a heterogeneity of variance (i.e. nonlinear blue curve), implying the assumption of homoscedasticity was rejected. The R^2_c indicated abundance (0.61) and richness (0.59) model had better explanatory power on the variables (Figure 6), with a lower error variance from their model. Biomass had the poorest performance among

the three models ($R^2_c = 0.44$), as shown by its fairly unbalanced residual plots (Figure 6). The QQ plot of richness and abundance reflected that their deviates were closer to normal distribution (blue line) while the deviates of biomass had a lower normality (Figure 6). Richness and abundance final model shown sufficient model performance ($R^2_c > 0.5$), they were used for the scenario studies. Biomass was not projected due to its relatively low model capacity on prediction.

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors:
A data modelling study in a German lowland catchment – Szewing To – December 2018
Chapter 3: Results

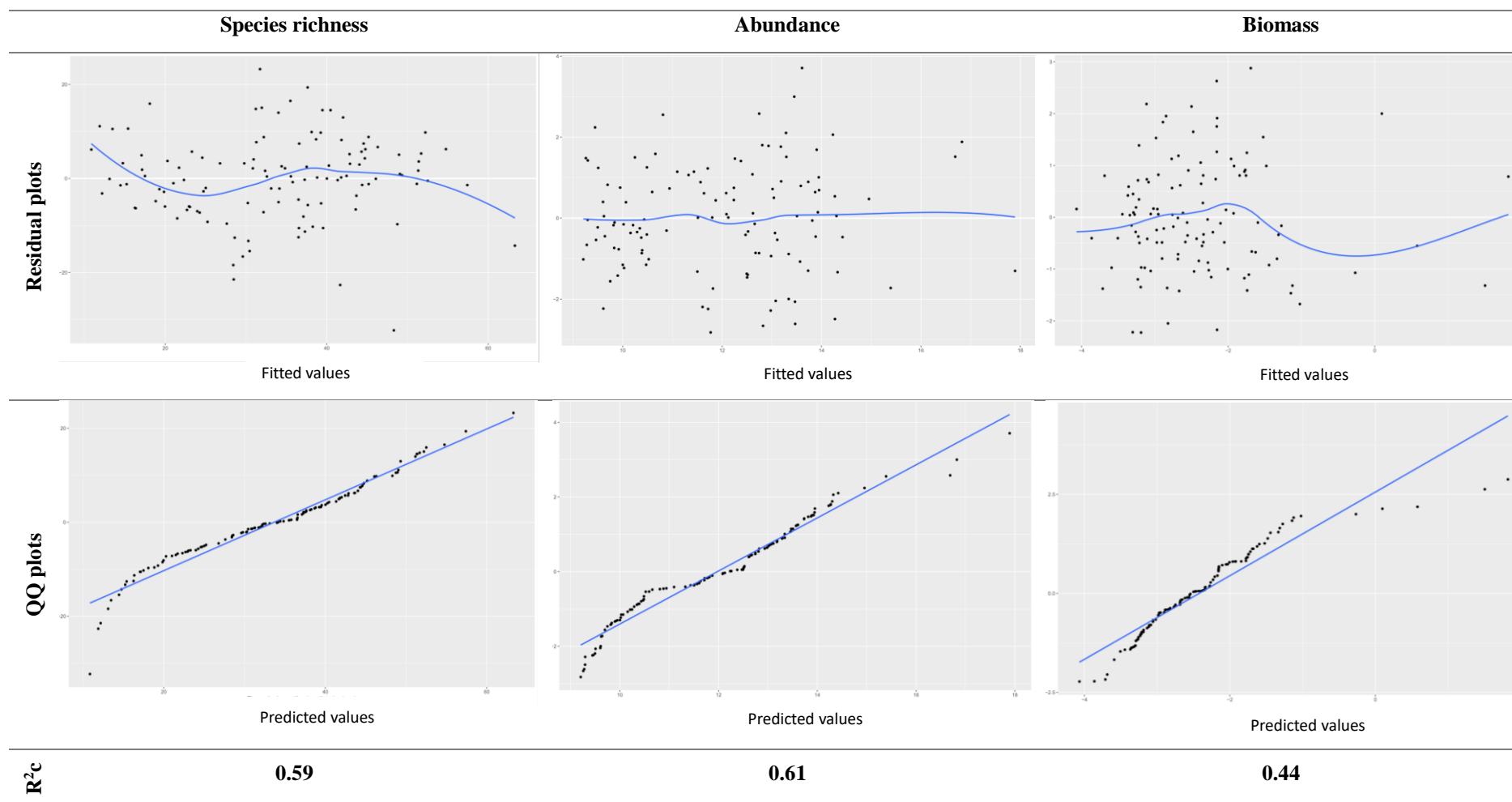


Figure 6 Residual and normality plots of final models (Y-axis represents residuals)

3.4 Scenario projections

The final models of species richness and abundance were used to project potential changes of phytoplankton community patterns under eight scenarios. The scenarios described potential intensifications of significant and influential stressors under nutrient enrichment, land use modifications and varying precipitation patterns (Table 4). The complete list of simulated results is attached in Appendix V. The simulated results were compared with the observed data by statistical tests and graphical illustration for assessing change significance and change magnitude. The Wilcoxon test showed that the simulated changes in species richness and abundance of the catchment were not significant ($p \geq 0.05$, Table 8: All subbasins). It can also be reflected by the similar mean of observed and simulated data in all scenarios (Figure 7). Nevertheless, the boxplot in Figure 7 showed generally high standard deviation among the data, denoting high spatial variation of phytoplankton community dynamics within the Treene catchment. When comparing simulated results with observed data within each subbasins, we found significant changes in both species richness and abundance for numerous subbasins. The changes in species richness were significant for subbasin Tr, Ju and Je and all but subbasin Ju had significant changes in abundance ($p \leq 0.05$, Table 8). Only significant changes were evaluated and discussed in the following.

Table 8 Wilcoxon test of changes significance of all and specific subbasins

P values		Scen1	Scen2	Scen3	Scen4	Scen5	Scen6	Scen7	Scen8
Species richness	All subbasins	0.89	0.77	0.69	0.33	0.93	0.84	0.99	0.85
	wtest_Tr	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01
	wtest_Sa	0.35	0.35	0.31	0.20	0.39	0.36	0.39	0.43
	wtest_Ki	0.94	0.76	0.84	0.99	0.96	0.97	0.89	0.78
	wtest_Ju	0.05	0.04	0.02	0.01	0.05	0.05	0.06	0.06
	wtest_Je	0.02	0.02	0.01	0.00	0.02	0.02	0.02	0.03
	wtest_Bo	0.70	0.69	0.52	0.38	0.70	0.64	0.73	0.77
Abundance	All subbasins	0.50	0.34	0.90	0.94	0.72	0.69	0.74	0.76
	wtest_Tr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	wtest_Sa	0.01	0.01	0.02	0.03	0.02	0.02	0.02	0.02

	wtest_Ki	0.01	0.00	0.06	0.04	0.02	0.02	0.02	0.03
	wtest_Ju	0.17	0.15	0.21	0.27	0.19	0.18	0.20	0.20
	wtest_Je	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	wtest_Bo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

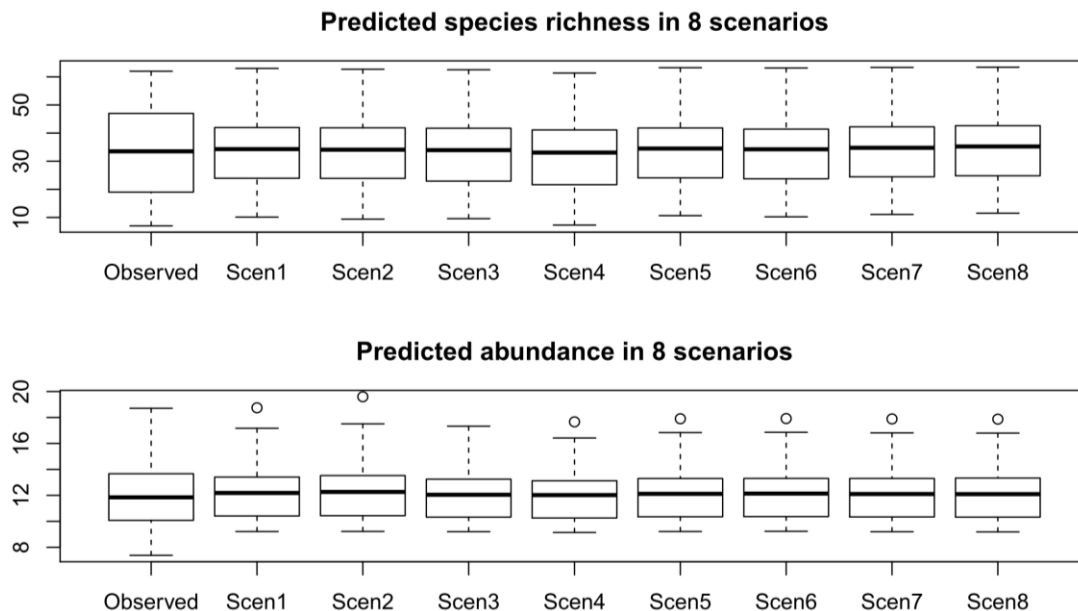


Figure 7 Distribution of predicted species richness and abundance in different scenarios

Table 9 The median of changes in species richness and abundance

		Tr	Sa	Ki	Ju	Je	Bo
Species richness	Scen1	-1.64			2.53	0.87	
	Scen2	-1.98			2.46	0.68	
	Scen3	-1.60			-2.63	-0.04	
	Scen4	-2.45			-1.43	-1.55	
	Scen 5	-1.32			2.45	0.97	
	Scen 6	-1.34			2.15	0.67	
	Scen 7	-1.29			2.75	1.28	
	Scen 8	-1.27			3.05	1.69	
Abundance	Scen1	-0.01	-0.06	-0.03			0.24
	Scen2	0.08	0.09	0.34			0.26
	Scen3	-0.12	-0.30	-0.65			0.20
	Scen4	-0.18	-0.45	-0.53			0.13
	Scen 5	-0.07	-0.20	-0.39			0.22
	Scen 6	-0.07	-0.17	-0.37			0.23
	Scen 7	-0.07	-0.22	-0.41			0.21
	Scen 8	-0.05	-0.24	-0.43			0.19

*Blank grids denote result not significant and hence not evaluated in the study

The median changes of species richness and abundance were extracted from subbasins with significant changes and tabulated in Table 9. Graphical presentations of the magnitude of changes in species richness and abundances of all subbasins were illustrated in Figure 8 and Figure 9 respectively. The range of significant changes in species richness from all scenarios was from -2.6 to +3.1 (Table 9) while the range of significant changes in abundance was between -0.7 and 0.4 (Table 9).

Simulated changes in species richness

There were three subbasins with significant changes on species richness under different stressor intensifications. They were subbasin Tr, Ju and Je (Table 9). Each of these three subbasins reflected high, medium and low level of tolerance level. Subbasin Ju was the most resilient subbasin compared to the other two in nutrient enrichment and precipitation variation scenarios (Scenario 1, 2, 5 to 8) but it was more sensitive to land use modification (Scenario 3 and 4, Figure 8). Subbasin Je had generally medium level of resilience capacity among different stressors while subbasin Tr demonstrated high sensitivity to stressors in most scenarios (Figure 8). The buffer capacity of subbasin Tr was slightly higher for lower level of increase in winter pastoral land (Scenario 3, Figure 8). The changing trend of more intensified stressors was not noticeable for nutrient enrichment (e.g. +15%, scenario 1 and +30%, scenario 2; Figure 8) and precipitation reduction (e.g. -5%, scenario 5 and -15%, scenario 6; Figure 8) in these three subbasins. Trends were not noticeable in land use modifications and increased precipitation scenarios. Compared to subbasin Ju, subbasin Je was more sensitive to higher level of land use modification although both of their species richness responded to land use change. Interestingly, more intensive increase in precipitation tended to enhance the species richness of subbasin Je, implying its higher tolerance to stronger precipitation, or potentially high flow condition.

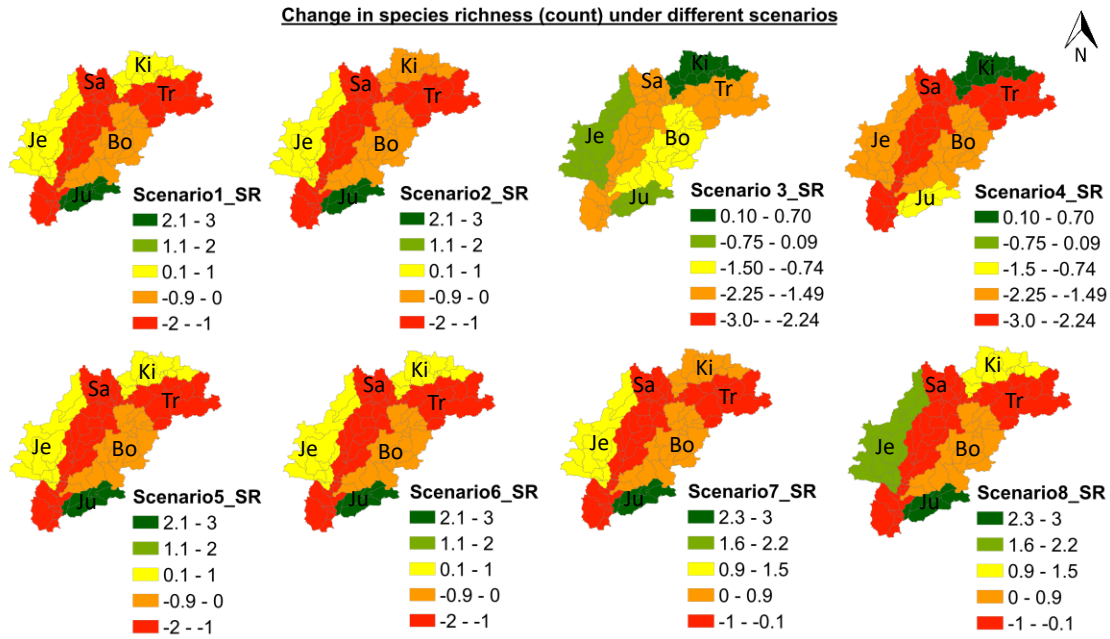


Figure 8 Projection of median change in species richness in each subbasin

Simulated changes in abundance

Except for subbasin Ju, all other subbasins had significant changes of their phytoplankton abundance (Table 9) so it would be excluded from discussion. Compared with the change in species richness, the variations in abundance were more stable trend under intensified stressors. Due to log-transformation of the actual abundance data, the projected data are proxies for the abundance. Subbasin Je and Bo remained highly stable and resilient in all scenarios (Figure 9). Subbasin Ki, Sa and Tr reflected different changing trend under different stressors. Firstly, for nutrient enrichment (Scenario 1 and 2), these three subbasins reflected positive changes in different magnitude (Figure 9), implying the favourable effect of nutrients to phytoplankton abundance. Compared to subbasin Sa and Tr, subbasin Ki showed a higher positive change (Figure 9), therefore more sensitive to nutrient enrichment. Secondly, when considering land use modification (Scenario 3 and 4), these three subbasins had very different response. Both subbasin Sa and Tr showed a declining trend under the introduction of more winter

pastoral land but subbasin Ki reflected a reduction in abundance decline (Figure 9). Lastly, in scenarios with varying precipitation patterns (Scenario 5 to 8), all subbasins showed very little changes (Table 8) that, were not visible in the projection (Figure 9). Nonetheless, some subbasins were more sensitive under the increasing precipitation pattern than the decreasing. In subbasin Bo, an increasing precipitation will reduce their phytoplankton abundance while for subbasin Sa and Ki, such climatic change will further lower their abundance (Figure 9), suggesting their sensitivity towards variation in rainfall patterns.

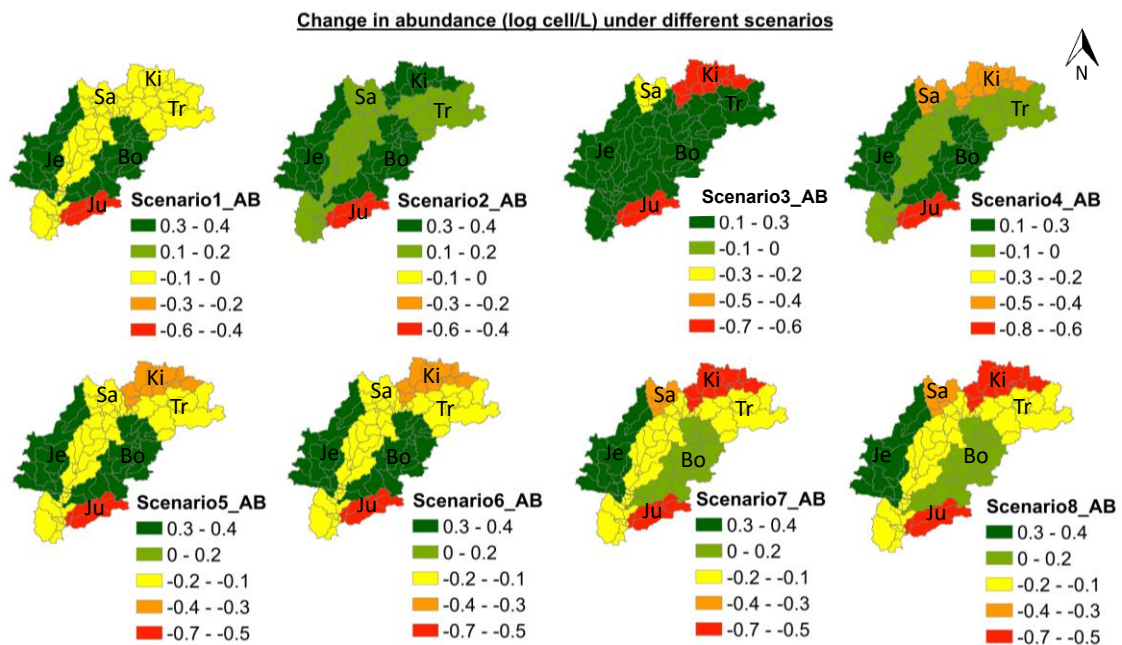


Figure 9 Projection of median change in abundance (log cell/L) in each subbasin

4. Discussion

4.1 *Importance and impact strength of abiotic variables on phytoplankton species richness*

Environmental variables have different importance to phytoplankton assemblages. The effect of a variable becomes more significant when it is important and exert high impact strength at the same time. In the following, we focused our discussion towards the significant effects of pH and flow alteration related variables, rate of change of flow and water depth, because they had both high importance and impact strength in shaping the species richness in the Treene catchment. pH disturbance of inland waters has been largely attributed to aerosol deposition (Siegried et al., 1989; Geelen and Leuven, 1986), which often lower the pH level. Meanwhile, the weak alkali condition of freshwater was also reported and has been another challenge for riverine biota (Medupin, 2011; Sharma et al., 2016). Phytoplankton groups have different optimum, adaptable and resilience range to pH (Richmond, 2004). Instability of pH level would shift phytoplankton composition structure to a more monopolistic one, thereby increasing the dominance of a certain species (Geelen and Leuven, 1986). It explains the importance of an optimal pH range in phytoplankton richness development. Concurrently, our model derived an immense negative impact of pH on species richness in the catchment (Figure 5). Unlike the pH level of marine water, which is mainly mediated by high amount of calcium carbonate, pH level in freshwater is formed from the water solution dissolved from river substrate or geomorphological factors such as rock types and salt saturation in the rocks (Sui et al., 2016). The streams in the Treene catchment recorded a minimum and maximum pH of 6.7 and 9.7 respectively. An average pH of 7.8 suggested a weak alkalinity of the water. A further increase of pH level will be highly unfavourable for phytoplankton groups reproduction. In the Treene

catchment, diatom (Bacillariophyceae) is the dominant phytoplankton group (Qu et al., 2018a). Their maximum growth occurs at pH 8 (El Gammal et al., 2017). Therefore, an excess or inadequate acidity would limit their succession rate or even threaten their survival. Some diatom species had declining cell division activities at pH 9.4 (e.g. Taraldsvik and Mykkestad, 2000). Following the unsuccessful regeneration under pH stress, the phytoplankton community would shift towards homogeneity, limiting richness expansion (Geelen and Leuven, 1986; Taraldsvik and Mykkestad, 2000).

Hydrological-related variables were weekly rate of change of flow (H55) with high importance and water depth with remarkable impact strength. H55 is a combination of change in water depth and flow velocity, in which water depth indicated for water volume or water level. Hence, H55 is a proxy indicator for the change in mass flow rate in a river. The importance of rate of change of flow has been agreed by Qu et al. (2018) with seasonal dataset of the entire year. Nevertheless, disregard of the high importance of H55, only water depth was considered as a significant effect on species richness in the final model (Figure 5). It is possibly because velocity has been observed with very low importance in our data (Figure 1, Appendix II). Instead of a change in water depth, water depth *per se* generated huge impacts on species richness. In the Treene catchment, water level has positive impact on increasing species richness.

Water level plays a more important role in controlling sedimentation losses of algae in small lowland river (Bahnwart et al., 1999). When water level is low, planktonic algae contacts more frequently with the riverbed (Bahnwart et al., 1999) and it accelerates the rate of sedimentation loss of phytoplankton (Reynolds et al., 1990; Reynolds et al., 1994). The Treene river is small with a mean width of 5 to 6 meters and a mean depth 0.5 to 0.6 meters respectively. The water depth became a more crucial factor than water

velocity because of a lower species sedimentation loss in deeper water (Bahnwart et al., 1999). Some unicellular diatom species required a minimum water depth to suspend in water (Reynolds et al., 1990; Reynolds et al., 1994). In addition to river substrate and its spines, the filtering effect of submerged macrophytes also speeds up the sedimentation losses in shallow area (Bahnwart et al., 1999), acting as an indirect effect to hinder species expansion in low water level condition.

4.2 Importance and impact strength of abiotic variables on phytoplankton abundance and biomass

The abundance and biomass of phytoplankton in the Treene catchment had similar important variables. They were largely depended on agricultural or non-natural land use and nutrient-related pollutants. The results agreed with Qu et al. (2018) that land use type has profound implications to phytoplankton density. Interestingly, both of them had high sensitivity to water area (WATR), which is comprised of some glacial riverine lakes in the watershed. The Treene catchment is a river-lake system (Bahnwart et al., 1999). Such system is characterized with lentic-lotic aquatic environment continuum, consists of permanent flow streams and glacial fluvial lakes, which constitutes very different flow features in different course of the river (Bahnwart et al., 1999; Qu et al., 2018a). Under this system, the fluvial lake provides a favourable condition for the accumulation of phytoplankton abundance and biomass (Bahnwart et al., 1999) because of the lower flow and higher retention time (Bahnwart et al., 1999; Wu et al., 2011). This is particularly noticeable for certain types of phytoplankton with a short reproduction cycle that, their abundance flourishing is in multiplicative scale within a limited period of time (Søballe and Kimmel, 1987). A longer residency time in fluvial lakes also contribute to the promising development of cyanobacteria community (e.g. Köhler, 1993; Gosselain et al., 1994). Therefore, the water area (i.e. fluvial lakes) have

high importance and impact strengths in explaining the phytoplankton abundance and biomass in the Treene catchment. Nonetheless, benefitted from this lentic-lotic landscape, physical-chemical variables provided an expellant effect to the algal abundance and biomass distribution.

Phosphate-phosphorus (PO₄-P) and total suspended solids (TSS) had significant effect on phytoplankton abundance and biomass respectively, containing both high importance and impact strength. Phosphate (PO₄) is the direct available form of phosphorus to phytoplankton and it is the limiting nutrient for phytoplankton growth. Different from marine phytoplankton, freshwater phytoplankton depended highly on phosphate instead of nitrogen (Hutchinson, 1967; Veronica et al., 2014), implying their high demand on PO₄. The nutrient loads were largely derived from the runoff from nearby agricultural land drainage tiles or human settlements (Qu et al., 2018b). Manure, fertilizer and pesticide application in arable lands (Guse et al. 2015) and medium-scaled sewage plant (Hayes et al. 2015) offered sufficient available resources for phytoplankton bloom. Combining with the effect of high residence time in the fluvial lakes, both factors provided a synergistic force on the mass development of phytoplankton abundance in the catchment. TSS denotes water turbidity and has induced negative effects to phytoplankton biomass development in the Treene catchment. Its controlling effect has been recognized and predominantly linked to the light limitation for photosynthesis (Cloern, 1987; Søballe and Kimmel, 1987). TSS was reported to have less importance to biomass in some coastal system because the vertical mixing of water provides light absorption opportunities for pelagic algae (Søballe and Kimmel, 1987). In lowland rivers, there is low occurrence of vertical mixing of water (Veronica et al., 2014), restricting the light acquisition of phytoplankton. The amount of TSS can be influenced by both natural events and anthropogenic pressures. First,

during high flow season, precipitation can enhance the amount of suspended solid runoffs to the stream (Blakar et al., 1990). Second, agricultural land use reinforced the sediment runoff effects due to soil plowing activities (Collins and Jenkins, 1996). There were two possibilities for the effect of turbidity, which became more significant in the Treene catchment. First, the highest flow in the catchment occurred in winter (Guse et al., 2015). The sampling period of this study was in September and December respectively, both of them were close to the winter season. Second, the agricultural-dominant landscape of this catchment has hinted a higher vulnerability of phytoplankton biomass formation to TSS in the Treene catchment.

4.3 Non-linear subbasin responses towards stressors

The species richness and algal abundance in different sub-catchments were susceptible to various intensified stressor. The abiotic variables in the Treene catchment were subjected to high spatial variation (Qu et al., 2018b). The distribution of nutrients showed a clear spatial heterogeneity across the water basin, which was in accordance to the land use pattern and their derived activities (Qu et al., 2018b). The lentic-lotic aquatic landscape of the catchment also contributed to a significant variation in hydrological regimes (Qu et al., 2018a). These implied that abiotic variables contained different significance to different subbasins. Meanwhile, the seasonal effect reinforced the variation on the high-low flow regime (Guse et al., 2015; Qu et al., 2018a). In a watershed management perspective, the resilience level of subbasins to different stressors provides implications on different temporal and spatial management strategies (Wu et al., 2014). We did not explain distinct subbasin reactions to different stressor variables in detail because it is upon the scope of this study. The model quality is also subjected to further optimization and verifications through calibration and validation (Section 4.4) before a precise prediction could be made. Despite that, there were two

important implications from the projections.

The first implication is the different resilience level of the subbasins on the species richness and abundance under multiple-stressor intensifications in the Treene catchment. In the previous section, we identified variables with high significance and effect towards species richness and abundance in the Treene catchment (Section 4.1 and 4.2). From the projection, we predicted the changes in various subbasins for evaluations. The changes had reflected non-linear patterns, which collided with the conclusion of some previous aquatic ecosystem stress modelling (e.g. Gordon et al., 2008; Davis et al., 2010). They have concluded the nonlinear reactions of biota towards stressors. Davis et al. (2010) emphasized such nonlinear response of under stressors may trigger ‘surprising’ detrimental effects on the aquatic ecosystem. Through predicting changes in distinct subbasins, their vulnerability can be identified and evaluated. Such information with higher spatial resolutions (i.e. compared with catchment level) on the effect of intensified stressors are often essential to support the local or regional decision-making process with regard to catchment conservation and management (Hering et al., 2015).

The second implication is the potential of adopting other proxies to evaluate different perspective of the catchment. Species richness and abundance were used to symbolize phytoplankton community dynamics. Introducing other response variables can explain the phytoplankton community assemblages in a more comprehensive manner. In terms of watershed management, phytoplankton abundance or blue-green algae abundance could act as an indicator for the volatility of a subbasin to algae bloom. With available time-series data, phytoplankton population dynamics can be extrapolated to predict an outbreak of eutrophication or algae bloom (In et al., 2015). Numerous works had used

this proxy for algal bloom management and forecast in Eurasia and Asia lake systems (Karul et al., 2000; Wilson and Rechnagel, 2001; In et al., 2015).

4.4 Potentials and limitations of mixed model applications in explaining phytoplankton community patterns and predicting changes in the Treene catchment

This study focused on providing an insight for the feasibility and the relevance of spatial management of subbasins in the Treene catchment. The application of mixed model in explaining phytoplankton community assemblage has demonstrated a possible method in taking spatial and seasonal variations into account. Admittedly, some model performance was not ideal and can be improved in future studies. Firstly, generalized linear model and linear mixed effect model are likely to perform better with a larger database. An extensive dataset enhances model inferences (Burnham and Anderson, 2002), thereby elevating model prediction capacity. Secondly, an unbalanced data size within random effect groups (i.e. subbasin) is susceptible to unstable mixed effects models (Harrison et al., 2018). Our model contained different number of samples from different subbasins (i.e. Tr = 22; Bo = 12; Je = 10; Ki = 7; Sa = 4; Ju = 4). Nonetheless, unifying the number of samples across subbasins could possibly improve the consistency of random effects calculation in the model. Thirdly, a well-developed linear mixed effect model requires model training and validation process (Harrison et al., 2018). This again, demands extensive dataset (Dochtermann and Jenkins, 2011). Zuur and Ieno (2016) suggested to subset 10% of the data for validating model prediction performance. They were not conducted in this study due to the limited size of our dataset. Johnson and Omland (2004), Grueber et al. (2011) and Harrison et al. (2018) has provided a comprehensive guidance on model selection, model inference and mixed effect modelling respectively, which have a high reference value.

Lastly, another potential for future improvement relates to the inclusion of more environmental variables for more comprehensive studies. Some abiotic variables such as silicon, were not encountered in this study due to resources limitation. Silicon is an important nutrient for diatom unique cell wall formation (Shrestha and Hildebrand, 2015). This regulatory features of dissolved reactive silica on phytoplankton (Tavernini, 2011) suggested silicon might induce some insightful explanations on the phytoplankton assemblages, given the majority group of identified phytoplankton in the Treene catchment belongs to diatom (Qu et al., 2018a). Moreover, grazing and competition effects between phytoplankton and zooplankton were not considered in this study due to their negligible effects in lowland stream ecosystem (Basu and Pick, 1997). The variable interactions, such as the suppression effect of temperature on dissolved oxygen (Rajwa et al., 2015), were also excluded in the study.

5. Conclusion

The environmental, hydrological and land use variables in the Treene catchment were important drivers for phytoplankton species, abundance and biomass. The significance of these variables and the corresponding impact strengths of those significant variables on phytoplankton community patterns were investigated in this study. Such attempt targeted to explain the major driving forces of phytoplankton richness, abundance and biomass pattern in the catchment by a modelling approach. Species richness was shaped by more diverse set of significant variables than those of abundance and biomass, which were predominantly explained by nutrient-related physicochemical parameters and land use. Additional to physicochemical variables and land use, species richness was driven by a series of hydrological variables. The previously concluded close relationships between environmental variables and spatial-seasonal variations in the Treene catchment indicated the necessity of considering these uncontrolled variables into the analysis. This study presented an attempt in explaining phytoplankton species richness, abundance and biomass with consideration of spatial and seasonal effects via mixed effect modelling in a lowland catchment. The mixed effect models showed that species richness in the Treene catchment was shaped predominantly by pH level and water depth. Nevertheless, phytoplankton abundance and biomass were both largely attributed to the lake area, which acted as favourable habitats for algal cell formation and development. The abundance and biomass were also under remarkable effects from phosphate-phosphorus and total suspended solid respectively.

Simultaneously, this study had offered an outlook on the feasibility of a modelling approach in forecasting changes in phytoplankton species richness and abundance under more intensive stressors. The simulation results demonstrated different reactions from subbasins, which implied various resilience level of subbasins towards stress

conditions. The simulated changes in species richness ranged from -2.6 to 3.1 (sign. > 0.95) while the change in abundance level ranged from -0.7 to 0.4 (sign. > 0.95) among different subbasins. Prediction for biomass was not conducted because of poor model performance. The major significant changing trends for species richness in different subbasins were (i) subbasin Tr had generally high sensitivities towards nutrient enrichments, land use modifications and varying precipitation and flow pattern; (ii) subbasin Ju was highly tolerant to most stressors except land use modifications; (iii) subbasin Je had mostly medium level of tolerance toward stressors but more sensitive to high degree of winter pastoral land expansion. Its species richness expanded in larger extent under greater amount of precipitation. Meanwhile, some significant findings for phytoplankton abundance changes under stress were (i) subbasin Tr and Sa had medium to high level of resilience in abundance towards nutrient enrichments, land use modifications and varying precipitation and flow pattern; (ii) subbasin Ki was generally sensitive to precipitation variation; (iii) phytoplankton abundance in subbasin Bo and Je were blooming under stressors of nutrient enrichments, land use modifications and varying precipitation and flow pattern.

These projection scenarios not only quantified possible changes, but also demonstrated nonlinear trends upon subbasin responses on species richness and abundance under a steeper environmental gradient. By this, it provided a preview of changes in phytoplankton population dynamics in various sub-catchments. Vulnerable subbasins could be identified and consolidated with appropriate watershed management strategies. Integrating the understandings of subbasin physicochemical and hydrological characteristics with implementations of appropriate temporal-spatial modelling methods would likely improve eutrophication or algae bloom management during peak seasons.

The assemblage patterns of richness, abundance and biomass were interpreted in this study and other important patterns such as composition and diversity will be another interesting research subject in the future. The species richness data in the study was measured taxonomically, which was insufficient in explaining ecosystem functioning in the Treene catchment. Nonetheless, there are often unexplained hidden attributes within phytoplankton biota such as physiological, morphological and ecological features (Reynolds et al. 2002), and the exploration of ecosystem functioning of phytoplankton in the Treene catchment is still limited. Explaining functional groups composition of phytoplankton by data modeling approach will be potential upcoming focuses to disentangle these attributes. A more comprehensive understanding on the changes in phytoplankton community or functional groups patterns under intensified stressors is needed in this ever-changing age.

6. References

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In Second International Symposium on Information Theory (Petrov, B.N. and Csaki, F., eds), pp. 267–281, Akademiai Kiado
- Anderson, D.R., Burnham, K.P., Thompson, W.L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *J. Wildl. Manage.* 64, 912–923
- Anneville, O., Gammeter, S., & Straile, D. (2005). Phosphorus decrease and climate variability: Mediators of synchrony in phytoplankton changes among European peri-alpine lakes. *Freshwater Biology*. <https://doi.org/10.1111/j.1365-2427.2005.01429.x>
- Arisanti, R., Notodiputro, K.A., Sadik, K., Lim, A. (2017). Bias Reduction in Estimating Variance Components of Phytoplankton Existence at Na Thap River Based on Logistics Linear Mixed Models. *IOP Conf. Ser.: Earth Environ. Sci.* 58:012014. <http://doi:10.1088/1755-1315/58/1/012014>
- Asadieh, B. and Krakauer, N. Y. (2017). Global change in streamflow extremes under climate change over the 21st century, *Hydrol. Earth Syst. Sci.*, 21, 5863-5874, <https://doi.org/10.5194/hess-21-5863-2017>.
- Austin, M. P. (1987). Models for the analysis of species' response to environmental gradients. *Vegetatio*. <https://doi.org/10.1007/BF00038685>
- Bahnwart, M., Hübener, T., & Schubert, H. (1998). Downstream changes in phytoplankton composition and biomass in a lowland river-lake system (Warnow river, Germany). *Hydrobiologia*. <https://doi.org/10.1023/A:1003558209411>
- Barbosa, A.B., Domingues, R.B., Galvão, H.M., (2010). Environmental forcing of phytoplankton in a Mediterranean estuary (Guadiana estuary, south-western Iberia): a decadal study of climatic and anthropogenic influences. *Estuaries and Coasts* 33, 324-341.
- Basu, B. K. and Pick, F. R. (1996) Factors regulating phytoplankton and zooplankton biomass in temperate rivers. *Limnol. Oceanogr.*, 41, 1572–1577.
- Basu, B. K. and Pick, F. R. (1997). Phytoplankton and zooplankton development in a lowland, temperate river. *J. Plankton Res.*, 19, 237 –253
- Bates, B.C., Kundzewicz, Z.W., Wu, S. and Palutikof, J.P. (2008) Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, 210 pp.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1 - 48. doi:<http://dx.doi.org/10.18637/jss.v067.i01>

- Blakar, I. A., Digernes, I., & Seip, H. (1990). Precipitation and streamwater chemistry at an alpine catchment in central Norway, 69–73.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*.
<https://doi.org/10.1016/j.tree.2008.10.008>
- Breuer, F., Janz, P., Farrelly, E., & Ebke, K. P. (2017). Environmental and structural factors influencing algal communities in small streams and ditches in central Germany. *Journal of Freshwater Ecology*.
<https://doi.org/10.1080/02705060.2016.1241954>
- Brewer, M. J., Butler, A., & Cooksley, S. L. (2016). The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12541>
- Burnham, K.P. and Anderson, D.R. (2002). Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach, Springer
- Bussi, G., Whitehead, P. G., Bowes, M. J., Read, D. S., Prudhomme, C., & Dadson, S. J. (2016). Impacts of climate change, land-use change and phosphorus reduction on phytoplankton in the River Thames (UK). *Science of the Total Environment*.
<https://doi.org/10.1016/j.scitotenv.2016.02.109>
- Buth, M., Kahlenborn, W., Savelsberg, J., Becker, N., Bubeck, P., Kabisch, S., Kind, C., Tempel, A., Tucci, F. (2015) Germany's vulnerability to Climate Change. Climate Change: 24/2015. Dessau-Roßlau, Germany ISSN 1862-4359 [Extracted from: <https://www.umweltbundesamt.de/publikationen/vulnerabilitaet-deutschlands-gegenueber-dem> 26-09-2018 15:44]
- Cloern, J. E. (1987). Turbidity as a control on phytoplankton biomass and productivity in estuaries. *Continental Shelf Research*.
[https://doi.org/10.1016/0278-4343\(87\)90042-2](https://doi.org/10.1016/0278-4343(87)90042-2)
- Collins, R., & Jenkins, A. (1996). The impact of agricultural land use on stream chemistry in the Middle Hills of the Himalayas, Nepal. *Journal of Hydrology*.
[https://doi.org/10.1016/0022-1694\(95\)03008-5](https://doi.org/10.1016/0022-1694(95)03008-5)
- Davis, J., Sim, L., & Chambers, J. (2010). Multiple stressors and regime shifts in shallow aquatic ecosystems in antipodean landscapes. *Freshwater Biology*.
<https://doi.org/10.1111/j.1365-2427.2009.02376.x>
- Desortová, B., & Punčochář, P. (2011). Variability of phytoplankton biomass in a lowland river: Response to climate conditions. *Limnologica - Ecology and Management of Inland Waters*. <https://doi.org/10.1016/j.limno.2010.08.002>
- Ebigwai, J. K., Imedimfon, I. E., Bright, H. A., Olowu, C., & Ekanem, F. A. (2014). Physico chemical parameters and phytoplankton assemblages along spatial and temporal gradients in great Kwa River, Calabar, Nigeria. *International Journal of*

- Biological Chemistry*. <https://doi.org/10.3923/ijbc.2014.1.20>
- El Gammal, M. A. M., Nageeb, M., & Al-Sabeb, S. (2017). Phytoplankton abundance in relation to the quality of the coastal water – Arabian Gulf, Saudi Arabia. *The Egyptian Journal of Aquatic Research*. <https://doi.org/10.1016/j.ejar.2017.10.004>
- Farmer, A.M. (2004) Phosphate pollution: a global overview of the problem In Valsami-Jones, E. (Ed.), *Phosphorus in Environmental Technology* pp. 176-180. IWA Publishing. London. UK. ISBN: 1843390019
- Feld, C. K., Segurado, P., & Gutiérrez-Cánovas, C. (2016). Analysing the impact of multiple stressors in aquatic biomonitoring data: A ‘cookbook’ with applications in R. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2016.06.243>
- Ferreze, M. (2012). The effect of the land use on phytoplankton assemblages of a Cerrado stream (Brazil). *Acta Limnologica Brasiliensia*. <https://doi.org/10.1590/S2179-975X2012005000025>
- Geelen, J. F. M., & Leuven, R. S. E. W. (1986). Impact of acidification on phytoplankton and zooplankton communities. *Experientia*, 42(5), 486–494. <https://doi.org/10.1007/BF01946686>
- Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press.
- Gharib, S. M., El-Sherif, Z. M., Abdel-Halim, A. M., & Radwan, A. A. (2011). Phytoplankton and environmental variables as a water quality indicator for the beaches at Matrouh, south-eastern Mediterranean Sea, Egypt: An assessment. *Oceanologia*. <https://doi.org/10.5697/oc.53-3.819>
- Gosselain, V., J. -P. Descy & E. Everbecq, (1994). The phytoplankton community of the River Meuse, Belgium: Seasonal dynamics (year 1992) and the possible incidence of zooplankton grazing. *Hydrobiologia* 289: 179–191.
- Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*. <https://doi.org/10.1111/j.1420-9101.2010.02210.x>
- Guse, B., Pfannerstill, M., & Fohrer, N. (2015). Dynamic Modelling of Land Use Change Impacts on Nitrate Loads in Rivers. *Environmental Processes*. <https://doi.org/10.1007/s40710-015-0099-x>
- Haas, M. B., Guse, B., & Fohrer, N. (2017). Assessing the impacts of Best Management Practices on nitrate pollution in an agricultural dominated lowland catchment considering environmental protection versus economic development. *Journal of Environmental Management*. <https://doi.org/10.1016/j.jenvman.2017.02.060>
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N.,

- Goodwin, C. E. D., ... Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*.
<https://doi.org/10.7717/peerj.4794>
- Hassany, J. S. A., Zahraw, Z., Murtadeh, A., Ali, H., & Sulaaïman, N. (2012). Study of the Effect of Himreen Dam on the Phytoplankton Diversity in Dyala River, Iraq. *Journal of Environmental Protection*.
<https://doi.org/10.4236/jep.2012.328109>
- Hayes, N. M., Vanni, M. J., Horgan, M. J., & Renwick, W. H. (2015). Climate and land use interactively affect lake phytoplankton nutrient limitation status. *Ecology*. <https://doi.org/10.1890/13-1840.1>
- Haygarth, P.M. and Jarvis, S.C., 1999. Transfer of phosphorus from agricultural soils. *Advances in Agronomy* 66, 196– 249.
- Hays, G. C., Richardson, A. J., & Robinson, C. (2005). Climate change and marine plankton. *Trends in Ecology & Evolution*.
<https://doi.org/http://dx.doi.org/10.1016/j.tree.2005.03.004>
- Hering, D., Carvalho, L., Argillier, C., Beklioglu, M., Borja, A., Cardoso, A. C., ... Birk, S. (2015). Managing aquatic ecosystems and water resources under multiple stress - An introduction to the MARS project. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2014.06.106>
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*. <https://doi.org/10.1093/biomet/76.2.297>
- Hutchinson, G. E., (1967). *A Treatise on Limnology*. Vol. 2. Introduction to lake biology and the limnoplankton. Wiley Intersci. Publ., New York.
- In Jeong, I., Lou, I., Ung, W. K., & Mok, K. M. (2015). Using Principle Component Regression, Artificial Neural Network, and Hybrid Models for Predicting Phytoplankton Abundance in Macau Storage Reservoir. *Environmental Modeling and Assessment*. <https://doi.org/10.1007/s10666-014-9433-3>
- IPCC (2014) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 688.
- Ives, A.R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution* 6(7):828–835 DOI 10.1111/2041-210x.12386.
- Jacob, D., Bulow, K., Kotova, L., Moseley, C., Petersen, J., Rechid, D. (2012). Regionale Klimaprojektionen für Europa und Deutschland: Ensemble-

- Simulationen für die Klimafolgenforschung, CSC Report 6: http://www.climate-service-center.de/imperia/md/content/csc/csc_report6.pdf
- Jeong, K. S., Joo, G. J., Kim, H. W., Ha, K., & Recknagel, F. (2001). Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Ecological Modelling*.
[https://doi.org/10.1016/S0304-3800\(01\)00300-3](https://doi.org/10.1016/S0304-3800(01)00300-3)
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*. <https://doi.org/10.1016/j.tree.2003.10.013>
- Kakouei, K., Kiesel, J., Domisch, S., Irving, K. S., Jähnig, S. C., & Kail, J. (2018). Projected effects of Climate-change-induced flow alterations on stream macroinvertebrate abundances. *Ecology and Evolution*.
<https://doi.org/10.1002/ece3.3907>
- Karul, C., Soyupak, S., Çilesiz, A. F., Akbay, N., & Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling*.
[https://doi.org/10.1016/S0304-3800\(00\)00360-4](https://doi.org/10.1016/S0304-3800(00)00360-4)
- Köhler, J. (1993). Growth, production and losses of phytoplankton in the lowland River Spree. I. Population dynamics. *Journal of Plankton Research*.
<https://doi.org/10.1093/plankt/15.3.335>
- Lázár, A. N., Wade, A. J., & Moss, B. (2016). Modelling Primary Producer Interaction and Composition: An Example of a UK Lowland River. *Environmental Modeling and Assessment*. <https://doi.org/10.1007/s10666-015-9473-3>
- Leland, H. V. (2003). The influence of water depth and flow regime on phytoplankton biomass and community structure in a shallow, lowland river. In *Hydrobiologia*.
<https://doi.org/10.1023/B:HYDR.0000008596.00382.56>
- Levinton, J. S. (2001). “Marine Biology. Function, Biodiversity, Ecology,” 2nd Edition, Oxford University Press, Inc., Oxford.
- Litchman, E., & Klausmeier, C. A. (2008). Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics*.
<https://doi.org/10.1146/annurev.ecolsys.39.110707.173549>
- Macedo, M. F., Duarte, P., Mendes, P., & Ferreira, J. G. (2001). Annual variation of environmental variables phytoplankton species composition and photosynthetic parameters in a coastal lagoon. *Journal of Plankton Research*.
<https://doi.org/10.1093/plankt/23.7.719>
- Matthaei, C. D., Piggott, J. J., & Townsend, C. R. (2010). Multiple stressors in agricultural streams: Interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology*.
<https://doi.org/10.1111/j.1365-2664.2010.01809.x>

- McCune, B. and Grace, J.B. (2002). Analysis of Ecological Communities. MJM Software Design, Gleneden Beach, Oregon, USA.
- Medupin, C. (2011). Phytoplankton community and their impact on water quality: An analysis of Hollingsworth Lake , UK. *Appl. Sci. Environ. Manage.*
- Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., and Toxopeus, A.G. (2014). Where is positional uncertainty a problem for species distribution modelling?. *Ecography* 37 (2): 191-203.
- Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*.
<https://doi.org/10.1111/j.1469-185X.2007.00027.x>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*.
<https://doi.org/10.1002/rra.700>
- Olsson U. (2002). Generalized linear models: an applied approach. Studentlitteratur. Lund, Sweden. Palm T. 1951. Die Holz- und Rinden-Käfer der nordschwedischen Laubbäume. Meddelanden från Statens Skogsforskningsinstitut 40: 1–242.
- Pasztaleniec, A. (2016). Phytoplankton in the ecological status assessment of European lakes – advantages and constraints. *Ochrona Środowiska i Zasobów Naturalnych*, 27(1), pp. 26-36. Retrieved 17 Mar. 2018, from doi:10.1515/oszn-2016-0004
- Pfannerstill, M., Guse, B., & Fohrer, N. (2014). A multi-storage groundwater concept for the SWAT model to emphasize nonlinear groundwater dynamics in lowland catchments. *Hydrological Processes*. <https://doi.org/10.1002/hyp.10062>
- Pyo, J., Pachepsky, Y. A., Kim, M., Baek, S.-S., Lee, H., Cha, Y., ... Cho, K. H. (2017). Simulating seasonal variability of phytoplankton in stream water using the modified SWAT model. *Environmental Modelling & Software*.
<https://doi.org/https://doi.org/10.1016/j.envsoft.2017.11.005>
- Qu, Y., Wu, N., Guse, B., & Fohrer, N. (2018a). Riverine phytoplankton shifting along a lentic-lotic continuum under hydrological, physiochemical conditions and species dispersal. *Science of the Total Environment*.
<https://doi.org/10.1016/j.scitotenv.2017.10.139>
- Qu, Y., Wu, N., Makareviciute, K., Guse, B., & Fohrer, N. (2018b). Effects of land-use pattern and physiochemical conditions on phytoplankton communities in a German lowland catchment. *Fundamental and Applied Limnology / Archiv Für Hydrobiologie*, 191, 175–187. <https://doi.org/10.1127/fal/2018/1125>

- Reynolds, C. S., Descy, J. P., & Padisák, J. (1994). Are phytoplankton dynamics in rivers so different from those in shallow lakes? *Hydrobiologia*.
<https://doi.org/10.1007/BF00007404>
- Reynolds, C. S., White, M. L., Clarke, R. T., & Marker, A. F. (1990). Suspension and settlement of particles in flowing water: comparison of the effects of varying water depth and velocity in circulating channels. *Freshwater Biology*.
<https://doi.org/10.1111/j.1365-2427.1990.tb00304.x>
- Reynolds, C.S. (2006). *The Ecology of Phytoplankton*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511542145>
- Richmond, A. (2004). *Handbook of microalgal culture: biotechnology and applied phycology*. Oxford, OX, UK ; Ames, Iowa, USA :Blackwell Science
- RStudio Team (2018). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- Sabater, S., Artigas, J., Durán, C., Pardos, M., Romaní, A. M., Tornés, E., & Ylla, I. (2008). Longitudinal development of chlorophyll and phytoplankton assemblages in a regulated large river (the Ebro River). *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2008.06.013>
- Schmalz, B., & Fohrer, N. (2010). Ecohydrological research in the German lowland catchment Kielstau. In *IAHS-AISH Publication*.
- Sellers, T., & Bukaveckas, P. A. (2003). Phytoplankton production in a large, regulated river: A modeling and mass balance assessment. *Limnology and Oceanography*. <https://doi.org/10.4319/lo.2003.48.4.1476>
- Sharma, R. C., Singh, N., & Chauhan, A. (2016). The influence of physico-chemical parameters on phytoplankton distribution in a head water stream of Garhwal Himalayas: A case study. *Egyptian Journal of Aquatic Research*.
<https://doi.org/10.1016/j.ejar.2015.11.004>
- Shrestha, R. P., & Hildebrand, M. (2015). Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryotic Cell*.
<https://doi.org/10.1128/EC.00209-14>
- Siegfried, C. A., Bloomfield, J. A., & Sutherland, J. W. (1989). Acidity status and phytoplankton species richness, standing crop, and community composition in Adirondack, New York, U.S.A. lakes. *Hydrobiologia*.
<https://doi.org/10.1007/BF00008472>
- Smith, V. H., Tilman, G. D., & Nekola, J. C. (1998). Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. In *Environmental Pollution*. [https://doi.org/10.1016/S0269-7491\(99\)00091-3](https://doi.org/10.1016/S0269-7491(99)00091-3)
- Søballe, D. M., & Kimmel, B. L. (1987). A large-scale comparison of factors influencing phytoplankton abundance in rivers, lakes, and impoundments.

- Ecology*. <https://doi.org/10.2307/1939885>
- Sui, F., Zang, S., Fan, Y., & Ye, H. (2016). Effects of different saline-alkaline conditions on the characteristics of phytoplankton communities in the Lakes of Songnen Plain, China. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0164734>
- Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Taraldsvik, M., & Mykkestad, S. (2000). The effect of pH on growth rate, biochemical composition and extracellular carbohydrate production of the marine diatom *Skeletonema costatum*. *European Journal of Phycology*. <https://doi.org/10.1080/09670260010001735781>
- Tavernini, S., Pierobon, E., & Viaroli, P. (2011). Physical factors and dissolved reactive silica affect phytoplankton community structure and dynamics in a lowland eutrophic river (Po river, Italy). *Hydrobiologia*. <https://doi.org/10.1007/s10750-011-0688-2>
- Torrecilla, N. J., Galve, J. P., Zaera, L. G., Retamar, J. F., & Álvarez, A. N. A. (2005). Nutrient sources and dynamics in a mediterranean fluvial regime (Ebro river, NE Spain) and their implications for water management. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2004.07.029>
- Trolle, D., Elliott, J. A., Mooij, W. M., Janse, J. H., Bolding, K., Hamilton, D. P., & Jeppesen, E. (2014). Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling and Software*. <https://doi.org/10.1016/j.envsoft.2014.01.032>
- Veronica, E., Setyo Leksono, A., & Arfiati, D. (2014). Effect Of Water Quality On Phytoplankton Abundance In Hampalam River And Fish Pond Of Batanjung Village. *IOSR Journal of Environmental Science Toxicology And Food Technology Ver. I*.
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*. <https://doi.org/10.1038/nature09440>
- Wagner, S., Berg, P., Schädler, G., & Kunstmann, H. (2013). High resolution regional climate model simulations for Germany: Part II-projected climate changes. *Climate Dynamics*. <https://doi.org/10.1007/s00382-012-1510-1>
- Weiser, C., Fuß, R., Kage, H., & Flessa, H. (2018). Do farmers in Germany exploit the potential yield and nitrogen benefits from preceding oilseed rape in winter wheat cultivation? *Archives of Agronomy and Soil Science*. <https://doi.org/10.1080/03650340.2017.1326031>
- Whitehead, P. G., & Hornberger, G. M. (1984). Modelling algal behaviour in the river thames. *Water Research*. [https://doi.org/10.1016/0043-1354\(84\)90244-6](https://doi.org/10.1016/0043-1354(84)90244-6)

- Whittaker, R. H. (1967). Gradient analysis of vegetation. Biological Reviews of the Cambridge Philosophical Society. <https://doi.org/10.1111/j.1469-185X.1967.tb01419.x>
- Wilson, H., & Reznagel, F. (2001). Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. *Ecological Modelling*, 146, 69–84.
- Wu, N., Dong, X., Liu, Y., Wang, C., Baattrup-Pedersen, A., & Riis, T. (2017). Using river microalgae as indicators for freshwater biomonitoring: Review of published research and future directions. *Ecological Indicators*. <https://doi.org/10.1016/j.ecolind.2017.05.066>
- Wu, N., Faber, C., Sun, X., Qu, Y., Wang, C., Ivetic, S., ... Fohrer, N. (2016). Importance of sampling frequency when collecting diatoms. *Scientific Reports*, 6, 36950. Retrieved from <https://doi.org/10.1038/srep36950>
- Wu, N., Qu, Y., Guse, B., Makarevičiūtė, K., To, S., Riis, T., & Fohrer, N. (2018). Hydrological and environmental variables outperform spatial factors in structuring species, trait composition, and beta diversity of pelagic algae. *Ecology and Evolution*. <https://doi.org/10.1002/ece3.3903>
- Wu, N., Schmalz, B., & Fohrer, N. (2011). Distribution of phytoplankton in a German lowland river in relation to environmental factors. *Journal of Plankton Research*. <https://doi.org/10.1093/plankt/fbq139>
- Wu, N., Schmalz, B., & Fohrer, N. (2014). Study progress in riverine phytoplankton and its use as bio-indicator—a review. *Austin J. Hydrol.* <https://doi.org/10.1101/209569>
- Yang, S. L., Zhao, Q. Y., & Belkin, I. M. (2002). Temporal variation in the sediment load of the Yangtze river and the influences of human activities. *Journal of Hydrology*. [https://doi.org/10.1016/S0022-1694\(02\)00028-8](https://doi.org/10.1016/S0022-1694(02)00028-8)
- Zuur, A. F. (2009). Mixed effects models and extensions in ecology with R. New York: Springer.
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12577>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>
- Zuur, A. F., Ieno, E. N., & Smith, G. M. (2007). Analyzing Ecological Data. *Methods*. <https://doi.org/10.1016/B978-0-12-387667-6.00013-0>

7. Appendices

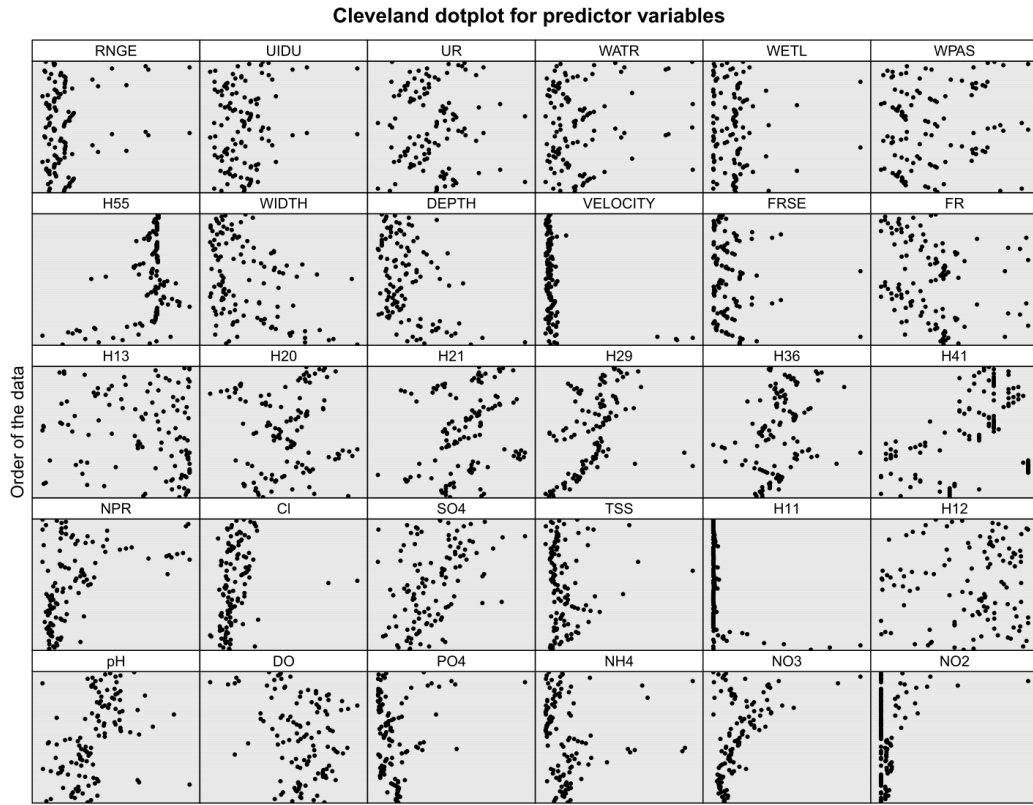
7.1 Appendix I

1. Data summary for predictor data (n=30)

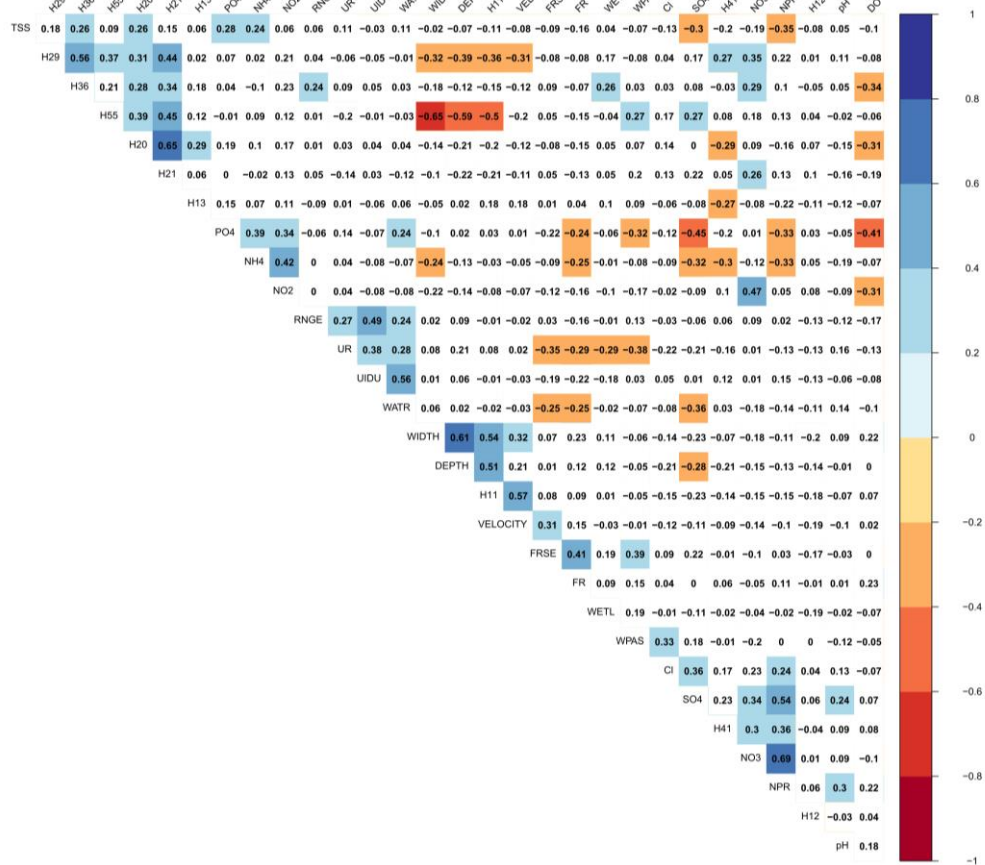
Predictor var.	General				Shapiro test	VIF test
	Min.	Max.	Mean	SD	p.value	vif value
pH	6.73	9.73	7.79	0.53	0.025	2.05
DO	2.93	12.98	9.13	1.84	0.003	2.26
PO4	0.01	0.63	0.07	0.09	0.000	2.41
NH4	0.02	1.51	0.23	0.28	0.000	2.65
NO3	0.10	37.49	6.40	5.70	0.000	4.81
NO2	0.00	0.73	0.04	0.09	0.000	2.74
NPR	0.23	234.16	50.37	57.04	0.000	4.69
CI	14.23	98.54	28.48	10.07	0.000	1.66
SO4	10.32	84.92	37.95	14.23	0.016	3.20
TSS	0.00	87.88	11.31	11.73	0.000	1.62
H11	0.00	1.09	0.04	0.16	0.000	2.44
H12	-1.73	1.73	0.34	1.03	0.000	1.40
H13	-1.72	1.73	0.62	1.08	0.000	1.56
H20	-0.78	2.44	0.88	0.69	0.541	3.28
H21	-1.58	2.64	0.95	0.87	0.072	3.13
H29	-0.53	3.37	0.69	0.66	0.000	2.80
H36	-0.31	3.28	1.29	0.56	0.007	2.72
H41	0.00	30.00	17.87	8.92	0.000	2.23
H55	-0.31	0.09	-0.03	0.07	0.000	3.51
WIDTH	1.05	17.10	4.94	3.64	0.000	3.77
DEPTH	0.10	2.00	0.47	0.32	0.000	2.36
VELOCITY	0.00	10.24	0.62	1.61	0.000	1.95
FRSE	0.02	9.02	1.02	1.37	0.000	2.05
FR	0.86	15.13	5.71	3.42	0.000	1.76
RNGE	0.00	4.33	0.70	0.80	0.000	1.69
UIDU	2.98	8.41	4.21	0.90	0.000	3.07
UR	1.75	12.26	5.65	2.02	0.005	2.34
WATR	0.62	5.42	1.71	1.10	0.000	2.80
WETL	0.00	7.19	1.01	1.12	0.000	1.63
WPAS	7.22	70.97	29.18	16.73	0.000	2.24

Appendix I Predictor variables data exploration summary

2. Predictor variable distribution plot (a) and correlation plot (b)



(a)



(b)

7.2 Appendix II

1. Random forest

Random forest (RF) is used for estimating variables importance (VIMP) in this study. It is highly flexible in dealing with different types of data such as binary, continuous and count and allows large amounts of independent variables with small samples (Feld et. al., 2016). VIMP was measured in terms of percentage increment in mean standard error (%IncMSE). A low %IncMSE will lead to low VIMP, because the removal or inclusion of this variable would have insignificant changes to the model quality (Strobl, 2007). The RF models were refined by tuning the number of regression trees (*ntree*) to minimize over-fitting and under-fitting. Each response variable had four RF models with different *ntree* (i.e. 500, 1000, 5000, 8000). Too little trees tend to be instable with high variance (Louppe et al., 2013), while too many trees could lead to over-fitting of model. Comparing different *ntree* in the model helps to optimize the model efficiency (Feld et al., 2016). The mean residual square was used to evaluate model performance. The performance of RF models is shown in the Table 1 below. VIMP was extracted from the most stable RF model of each response variable. Variables with importance above-mean were screened and tested in generalized linear model (GLM) to select the best fixed variables structure. The complete VIMP curve is also illustrated in Figure 1. The complete VIMP value is listed in Table 2.

Table 1 Random forest model performance

	ntree	Mean residual square	% of Variable explained	Best ntree
Species richness	8000	104.6131	51.32	500
	5000	104.4229	51.41	
	1000	103.1878	51.99	
	500	106.3382	52.09	
Overall abundance	8000	1.865665	64.4	1000
	5000	1.861052	64.49	
	1000	1.858746	64.53	
	500	1.877572	64.17	
Overall biomass	8000	1.207764	40.49	5000
	5000	1.205002	40.63	
	1000	1.192034	41.27	
	500	1.208602	40.45	

Appendix II Supplementary information to model building and model selection

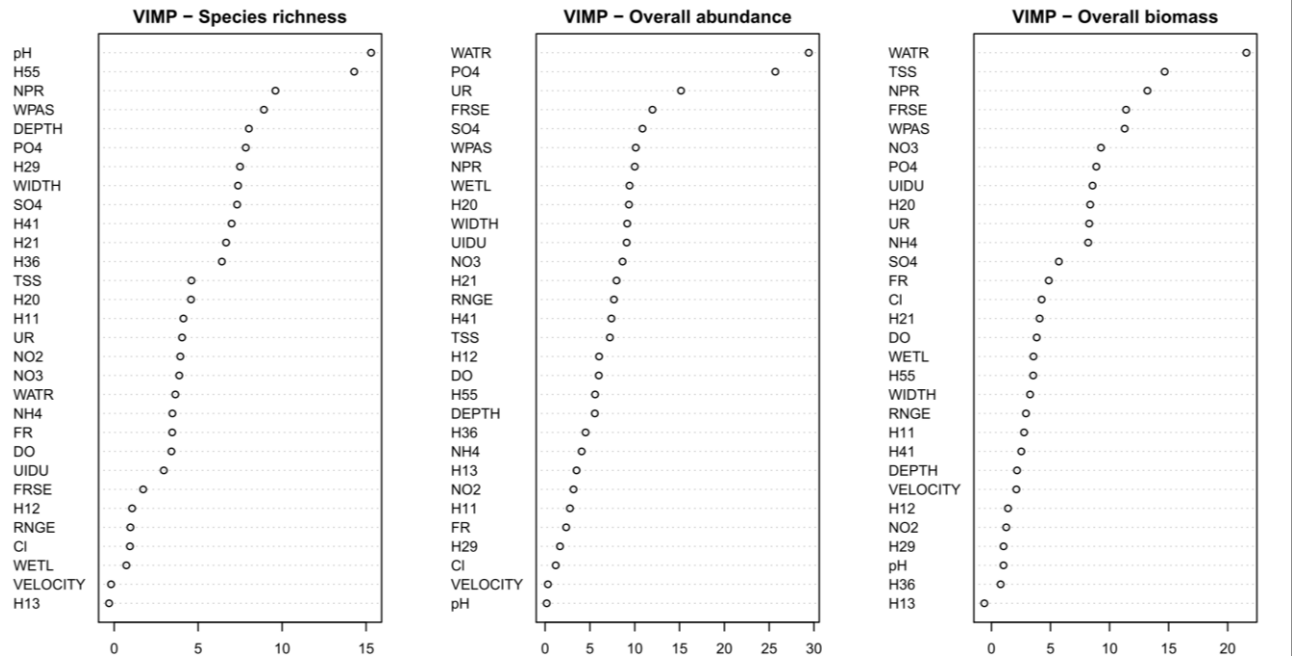


Figure 1 Variable importance curve

Table 2 RF models – VIMP values (Complete)

Variables	Species richness	Overall abundance	Overall biomass
	%IncMSE	%IncMSE	%IncMSE
pH	15.30	0.18	1.02
DO	3.41	5.99	3.83
WATR	3.64	29.42	8.28
NH4	3.47	4.08	14.66
H55	14.29	5.58	3.54
NO2	3.94	3.17	1.25
UR	4.04	15.17	8.36
CI	0.94	1.18	4.25
FRSE	1.73	11.97	8.88
TSS	4.60	7.23	11.28
H11	4.12	2.77	2.76
H12	1.07	6.02	1.4
H13	-0.30	3.51	-0.6
NPR	9.60	10.00	11.4
WPAS	8.91	10.11	8.19
DEPTH	8.02	5.55	2.17
PO4	7.84	25.69	21.58
H29	7.49	1.66	1.02
WIDTH	7.37	9.16	3.27
SO4	7.32	10.87	5.71

Appendix II Supplementary information to model building and model selection

H41	6.99	7.40	2.53
VELOCITY	-0.19	0.31	2.11
WETL	0.72	9.42	3.56
FR	3.45	2.35	4.86
RNGE	0.97	7.68	2.92
H20	4.57	9.35	9.27
H21	6.66	7.97	4.08
UIDU	2.95	9.10	8.56
NO3	3.87	8.64	13.21
H36	6.41	4.52	0.78

2. Generalized Linear Models (GLM)

In total 6 GLMs in different sizes were conducted for each response variables. The modeling structure of GLMs was listed in Table 3. The structure was the same for all three response variables. Variable interactions should be considered in the model (Zuur et al., 2009; Burham and Anderson, 2002) so Model 3 was designed with interaction terms. It is common to generate a global model that considers all possible variables in the set of candidate models (Crawley, 2013). Model 1 was deigned as the global model.

The structure of GLM is shown in below formula. It consists of a response variable, a predictor intercept, a series of predictor variables, a link function for corresponding familiar structure and residuals or errors. The link function defines the dependence of the response variables on the linear predictors in order to attain linearity and homoscedasticity between them (Bolker et al., 2009; Johnson et al., 2015). We identified link function by the distribution of response variables. Species richness is a count data so it used poisson family and log link function while Shannon-wiener's diversity and overall abundance are continuous data (after log-transformed), they used gaussian family which adopted identity link function.

$$\gamma = \eta_0 + \eta_1 + \eta_2 + \dots \eta_i + \varepsilon_i$$

where

γ = response variable

η_0 = mean predictor intercept

η = linear predictors

ε_i = residuals / errors

Table 3 GLM candidate model structure

Model	Description
Model 1	Global model with all potential predictor variables
Model 2	Local model with only predictor variables with VIMP above average
Model 3	Local model with first 3 important predictors and their interactions
Model 4	Local model with first 3 important predictors
Model 5	Selected significant parameters from global model – Model 1 (p<0.05)
Model 6	Selected significant parameters from the best model in Model 2 (p<0.05)

3. Linear Mixed Models (LMM)

The structure of LMM is shown in below formula. It contains similar structure to GLM except an addition of a random effect component and an exclusion of the link function to the response variable (γ) (Johnson et al., 2015). Table 4 tabulated the model structure of LMMs. All response variables were tested against same set of LMM models.

$$\gamma = \eta_0 + \eta_1 + \eta_2 + \dots \eta_i + \sigma_i + \varepsilon_i$$

where

γ = response variable

η_0 = mean predictor intercept

η = linear predictors / fixed effect variables

σ_i = random effect variables

ε_i = residuals / errors

Table 4 LMM candidate model structure

Model	Description
Model 1	Crossed random effects – Year and Subbasin
Model 2	Single random effect - Year
Model 3	Single random effect - Subbasin
Model 4	Random slope – (Subbasin Year)
Model 5	Nested random variables (1 Subbasin:Year)

4. Model selection

Model selection is an advancement to classical null hypothesis model testing methods that, allows multiple competing hypothesis or models to be assessed at the same time (Johnson and Omland, 2004). Model selection is based on a robust likelihood theory framework that is applicable to majority of statistical

Appendix II Supplementary information to model building and model selection

models. To perform model selection, one needs to generating a set of reasonable candidate models, as denoted above with GLM and LMM. There are two model selection considerations, model fit and model complexity (Johnson and Omland, 2004). Multi-model inference parameters such as Akaike Information Criterion (AIC), quasi-AIC and other AIC-derivatives (For details, please see below) enable both edges to be considered. Another major merits selecting models with AIC-derived parameters over null-hypothesis methods is that they consider model selection uncertainty (Harrison et al., 2018), which means several competing models can match similarly well to the data (Burnham and Anderson, 2002; Burnham et al., 2011). The major purpose of model selection is to identify the best model among the candidate models by settling the trade-off between model fit and model complexity for onward analysis (Harrison et al., 2018). The best parsimonious model that has desired level of predictability and explanatory power with minimum predictor variables is the ultimate goal for model selection approach. (Johnson and Omland, 2004; Harrison et al., 2018) In this study, the best model was used for prediction. The model selection parameters used in this study is explained below.

Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) is an estimate of the relative amount of information expected to loss in the process of using a model to generate the true observed data (Burnham and Anderson, 2002). Therefore, it is more meaningful to assess the relative differences among AIC values although the lowest absolute AIC value represents the best model (Burnham and Anderson, 2002; Harrison et al., 2018). AIC is a concrete model assessment method that has laid foundations for other improved unbiased parameters such as quasi AIC (qAIC) and corrected AIC (AICc) (Burnham and Anderson, 2002; Zuur et al., 2007). AICc is a second order variant of AIC and is corrected for dataset with small sampling size (Hurvich and Tsai, 1989). AICc can be calculated from AIC through adding a bias-penalty term at the end as following:

$$AICc = -2LogLik + 2K + \frac{2K(K + 1)}{n - K - 1}$$

(Hurvich and Tsai, 1989)

The difference between AICc and qAICc is the inclusion of an overdispersion factor (\hat{c}) in the calculation. When $\hat{c} > 1$, there is overdispersion in the real data and quasi-AIC methods should be adopted (Burnham and Anderson, 2002). When $\hat{c} \leq 1$, there is no overdispersion and qAICc will be equal to AICc. qAICc can be calculated as follows:

$$qAICc = -\left(\frac{2\text{LogLik}}{\hat{c}}\right) + 2K + \frac{2K(K+1)}{n-K-1}$$

(Burnham and Anderson, 2002)

Application of AIC-derived parameters

Similar to AIC, AICc and qAICc indicate information loss in the fitted model. It selects the best model upon a set of models in terms of the lowest value (as it implies the least information loss) and it can only be applied to the set of models that employed the same data (Burnham and Anderson, 2002). Nevertheless, AIC, AICc and qAICc cannot be compared against each other so a set of hypothetical models can only be compared on the same parameter. All competing models should also be based on the same dataset for meaningful selections (Burnham and Anderson, 2002). A particular remark is that all AIC-derived parameter should not be evaluated by its absolute size but the differences among the set of models ($\Delta AICc/\Delta qAICc$) because it is merely relative values over the set of competing models (Burnham and Anderson, 2002). However, a relatively best model can still be poor in absolute sense. Therefore, other evaluation criteria should be involved to enhance model selection consistency (Burnham and Anderson, 2002; Bolker et al., 2009).

Overdispersion

Overdispersion is common in non-normal data or binary data and its presence indicates a bad model fit (Harrison, 2014) and ignoring overdispersion can cause over-estimation on the model fit (Crawley, 2013). To investigate overdispersion for GLM, the theta parameter was calculated by dividing the residual

Appendix II Supplementary information to model building and model selection

deviance by residual degrees of freedom (df.residual) (Feld et al., 2016; Bolker et al., 2017) as below:

$$\theta = \frac{\text{model residual deviance}}{\text{model df.residual}}$$

When GLM θ ranges between 2-15, it suggests the response variable variance is much greater than its mean (Crawley et al., 2013; Feld et al., 2016). Several methods could be adopted to fix overdispersion (Harrison, 2014) and negative binomial models were used to fix over-dispersion of count data in GLMs of species richness. Overdispersion on gaussian data model is not required for adjustment (Harrison et al., 2018). The overdispersion results of the count data model of species richness is presented in Table 5. After fitting with negative binomial GLMs, the overdispersion was stabilized and model selections could be conducted.

Table 5 Overdispersion factor of Poisson GLMs

Poisson GLMS	θ^*		Negative binomial GLMs	θ^*
GLMsp1	2.88	→	nbGLMsp1	1.43
GLMsp2	3.18		nbGLMsp2	1.18
GLMsp3	4.25		nbGLMsp3	1.11
GLMsp4	4.59		nbGLMsp4	1.08
			nbGLMsp5	1.15
			nbGLMsp6	1.11

* θ represents an overdispersion factor that is calculated by the sum of residual squares dividing degree of freedom of residuals (Bolker et al., 2017)

The 7th GLM Model for abundance

As described above, every response variable had respectively 6 GLMs. However, due to the very close performance of two GLMs for abundance and biomass, the 7th model was derived for both of them (Table 6). Subsequently, their 7th models were proved with the best performance and was adopted for abundance (Table 6).

Table 6 GLM performance of overall abundance with and without model 7

	logLik	AICc	delta	RSE	deviance	adjusted r2
<i>Model performance without model 7</i>						
GLMab6	-219.91	452.57	0.00	1.59	287.16	0.52
GLMab5	-219.52	454.06	1.49	1.60	285.28	0.52
<i>Model performance with model 7</i>						
GLMab7	-214.55	448.80	0.00	1.54	262.25	0.55
GLMab6	-219.91	452.57	3.80	1.59	287.16	0.52
GLMab5	-219.52	454.06	5.29	1.60	285.28	0.52
<i>Model performance without model 7</i>						
GLMbio5	-178.76	372.54	0.00	1.13	142.98	0.38
GLMbio6	-181.94	374.42	1.88	1.15	150.89	0.35
<i>Model performance with model 7</i>						
GLMbio7	-214.55	448.80	0.00	1.10	135.43	0.40
GLMbio5	-219.91	452.57	4.09	1.13	142.98	0.38
GLMbio6	-219.52	454.06	5.97	1.15	150.89	0.35

Adjusted r^2 , marginal r^2 and conditional r^2

Coefficient of determination (r^2) is a common measure for assessing linear model fit (Harrison et al., 2018) to report the percentage of variance explained. Nevertheless, since GLM no longer assume a normal error structure as linear model, r^2 can no longer meaningfully evaluate GLMs, instead, the adjusted r^2 reported in Zhang (2017) was used to report the model fit of GLMs. The adjusted r^2 includes a function that accounts for non-normal mean and variance that has proved useful for GLMs (Zhang, 2017). Concurrently, it includes a sample size correction, hence reduces bias on models with different predictor variables, which is common in GLMs (Zhang, 2017). In LMM, the inclusion of random effect results in random effect variance that the original r^2 does not account for (Nakagawa et al., 2013). Marginal r^2 and conditional r^2 were used to calculate model fit for LMMs. The calculation methods of these three modified r^2 are interpreted below:

Appendix II Supplementary information to model building and model selection

$$Adjusted\ r^2 = 1 - \frac{\frac{\sum_{i=0}^n dv(yi, fitted)}{n - p}}{\frac{\sum_{i=0}^n dv(yi, observed\ mean)}{n - 1}}$$

Zhang (2017)

where

dv = specified mean and variance function (For details, please see Zhang (2017))

yi = observed data

n = total number of observations

p = number of predictor variables accounted in the model

$$Marginal\ r^2 = \frac{\sigma_{f^2}}{(\sigma_{f^2} + \sigma_{\alpha^2} + \sigma_{\varepsilon^2})}$$

$$Conditional\ r^2 = \frac{(\sigma_{f^2} + \sigma_{\alpha^2})}{(\sigma_{f^2} + \sigma_{\alpha^2} + \sigma_{\varepsilon^2})}$$

Nakagawa et al. (2013)

where

σ_{f^2} = variance of the fixed effect variables

σ_{α^2} = variance of the random effect variables

σ_{ε^2} = error variance (For details, please see Nakagawa et al. (2013))

5. Model performance

The residual plots of all GLMs and LMMs are illustrated in Figure 2 and Figure 3.

Appendix II Supplementary information to model building and model selection

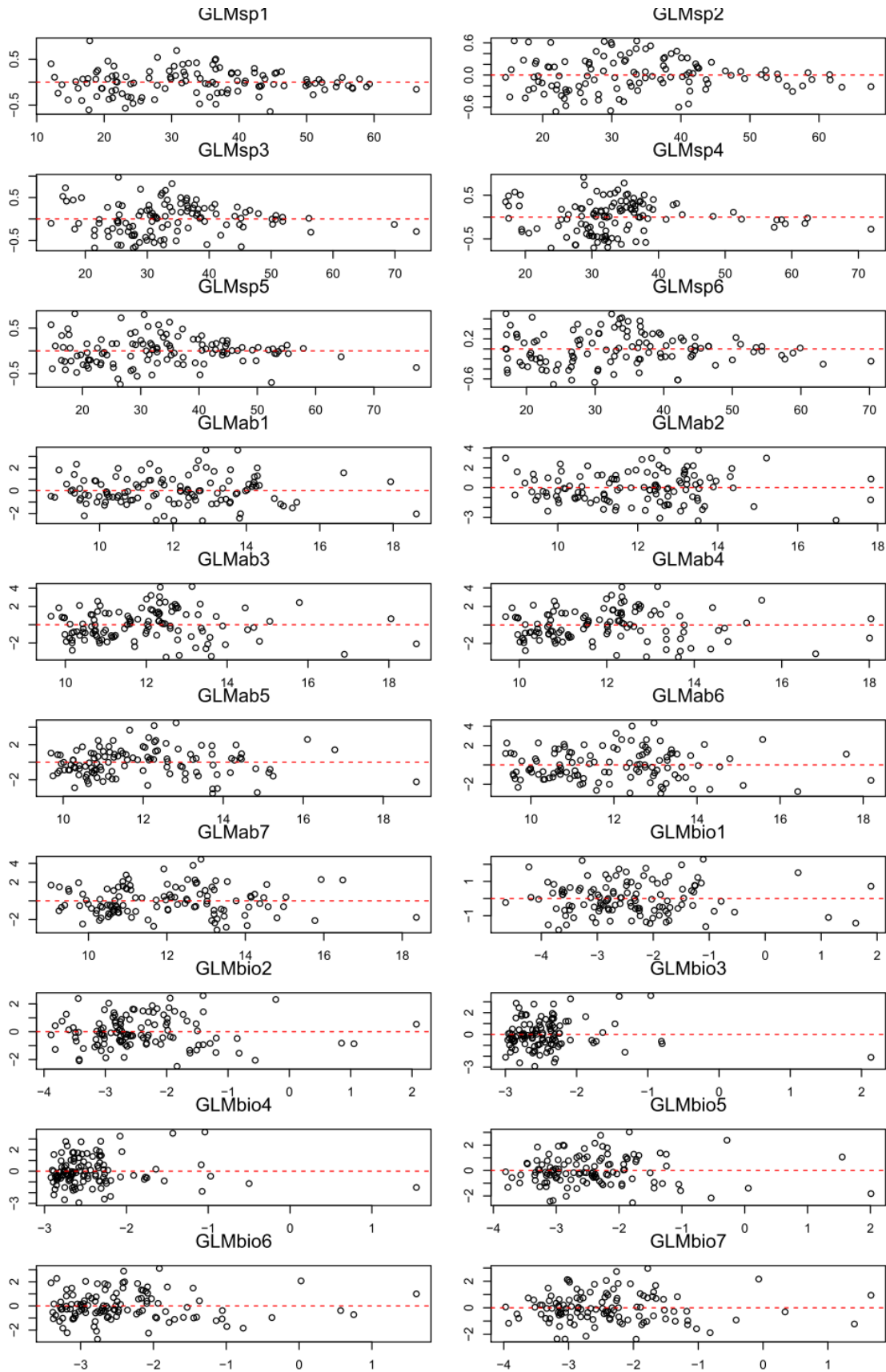


Figure 2 Residual plots of generalized linear models (sp: Species richness; ab: abundance; bio: biomass)

Appendix II Supplementary information to model building and model selection

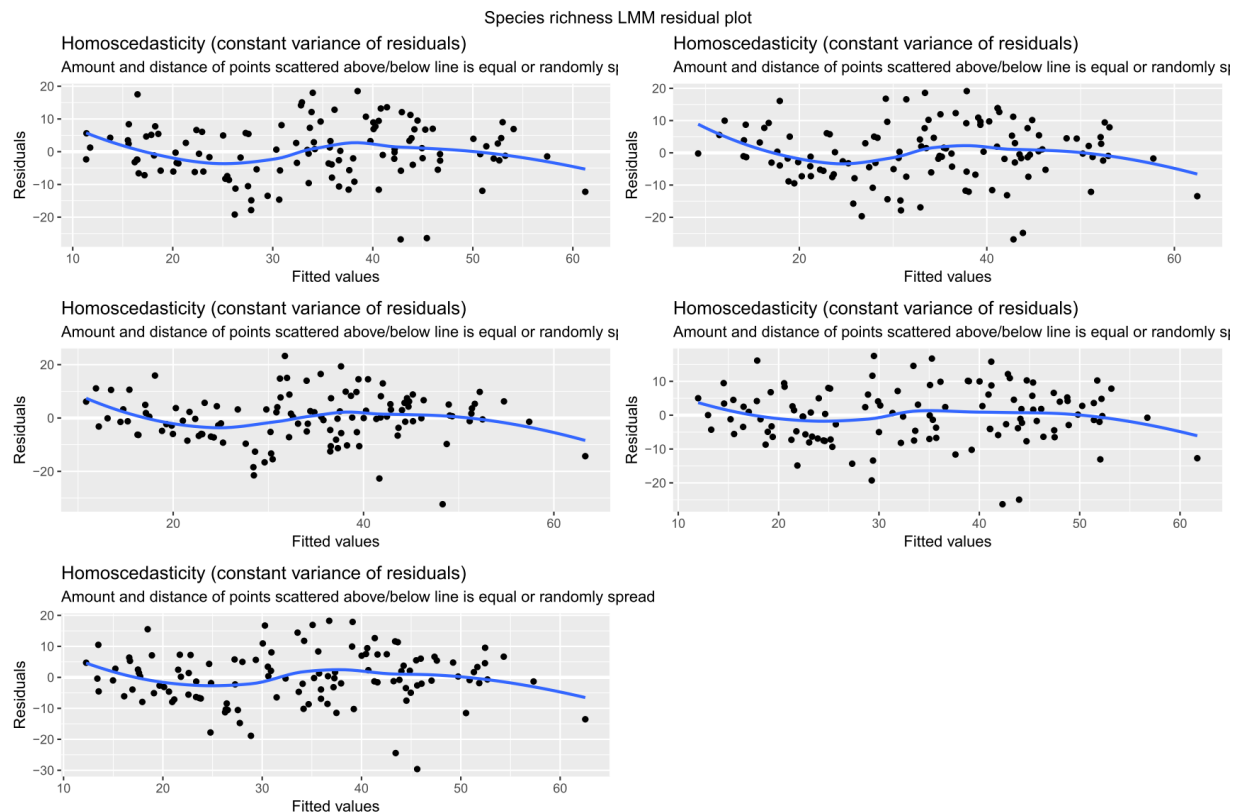


Figure 3a Linear mixed effect models of species richness residual plots

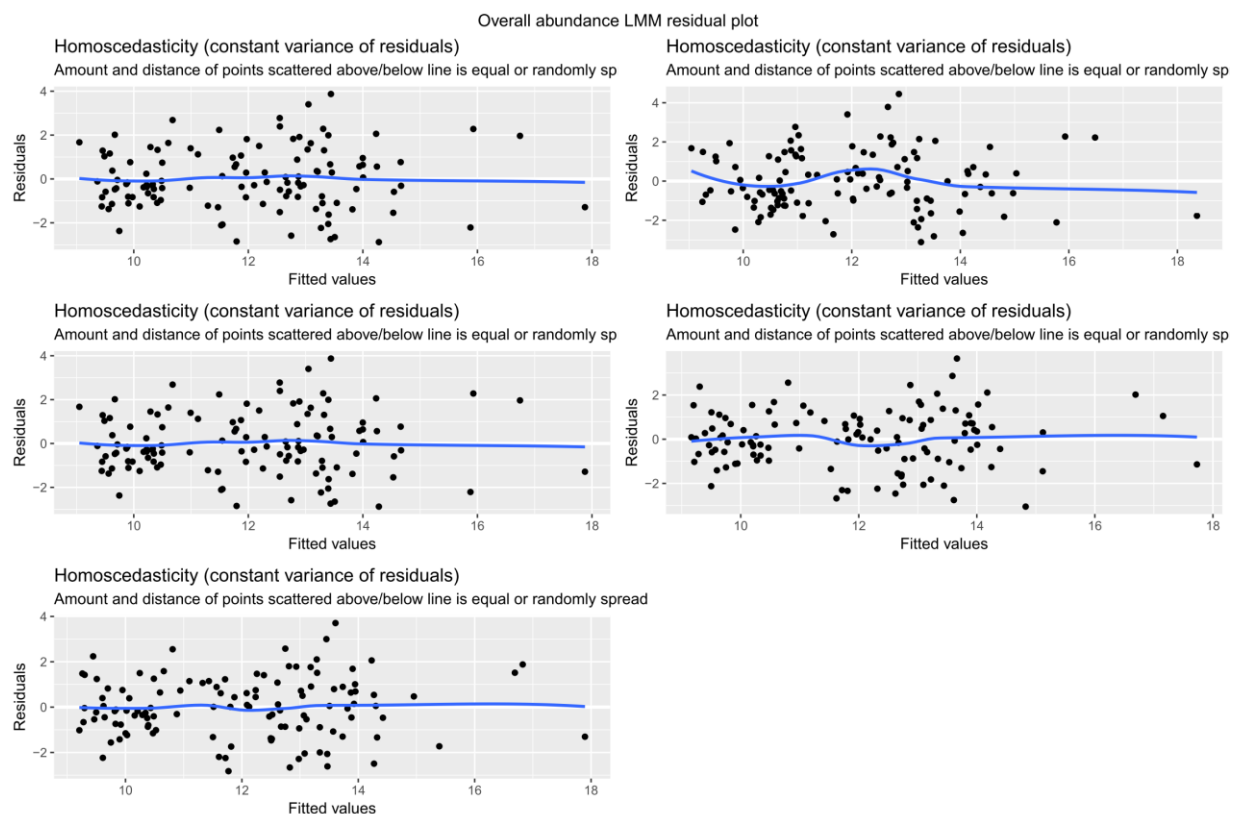


Figure 3b Linear mixed effect models of abundance residual plots

Appendix II Supplementary information to model building and model selection

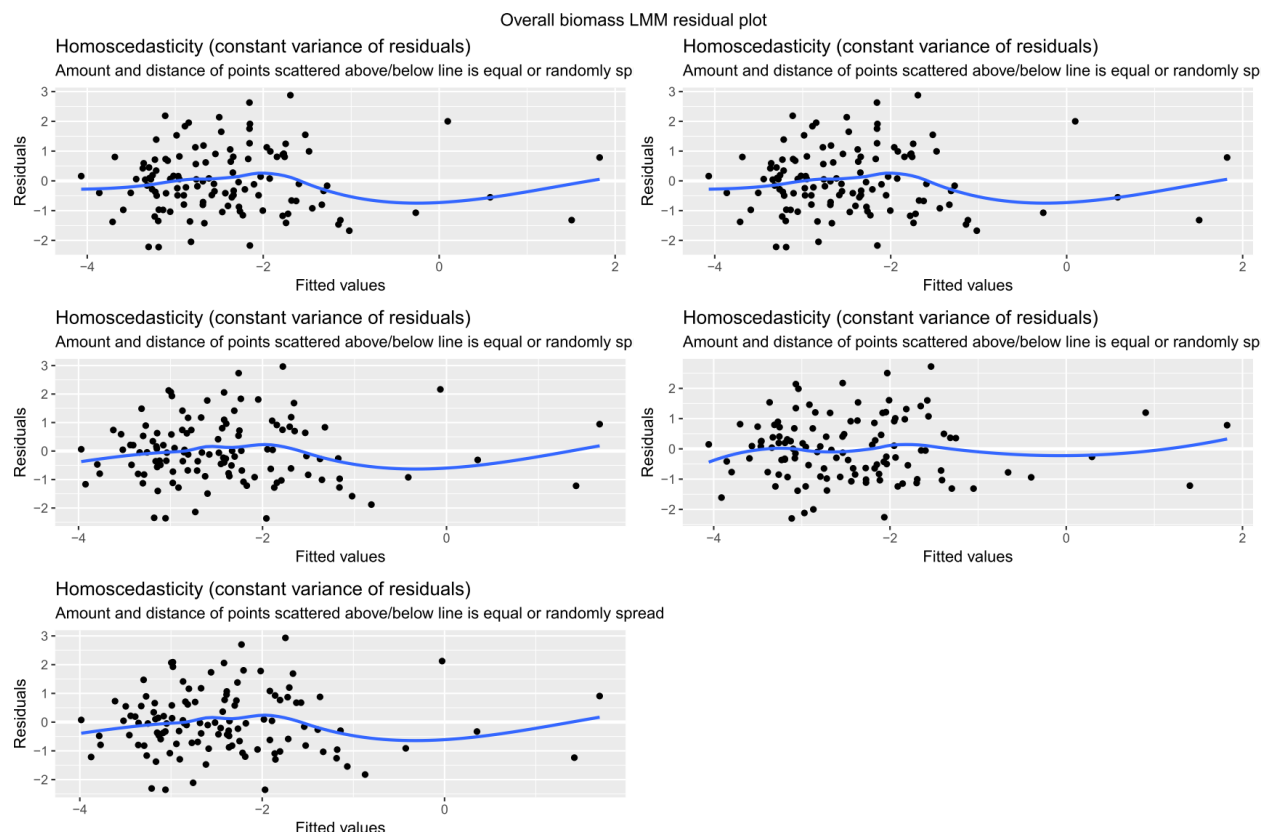


Figure 3c Linear mixed effect models of biomass residual plots

Reference

- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*. <https://doi.org/10.1016/j.tree.2008.10.008>
- Burnham KP, Anderson DR, Huyvaert KP. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65(1):23–35 DOI 10.1007/s00265-010-1029-6.
- Burnham, K. P. editor, & Anderson, D. R. (2002). Model Selection and Multimodel Inference A Practical Information-Theoretic Approach *MODEL SELECTION & MULTIMODEL INFERENCE*.
- Crawley, M. J. (2013). *The R book*. (2nd Ed.) Chichester, England: Wiley.
- Feld, C. K., Segurado, P., & Gutiérrez-Cánovas, C. (2016). Analysing the impact of multiple stressors in aquatic biomonitoring data: A ‘cookbook’ with applications in R. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2016.06.243>
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N.,

Appendix II Supplementary information to model building and model selection

- Goodwin, C. E. D., ... Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*.
<https://doi.org/10.7717/peerj.4794>
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*. <https://doi.org/10.1093/biomet/76.2.297>
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*. <https://doi.org/10.1016/j.tree.2003.10.013>
- Johnson, P. C. D., Barry, S. J. E., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12306>
- Louppe, G., Wehenkel, L., Suter, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Neural Information Processing Systems*. https://doi.org/NIPS2013_4928
- Strobl, C. (2007). Why and how to use random forest variable importance measures (and how you shouldn't). Statistik.Tu-Dortmund.De.
- Zhang, D. (2017). A Coefficient of Determination for Generalized Linear Models, *The American Statistician*, 71:4, 310-316, DOI: 10.1080/00031305.2016.1256839
- Zuur, A. F. (2009). Mixed effects models and extensions in ecology with R. New York: Springer.

7.3 Appendix III

Appendix III R packages and functions adopted in the study (R version 3.5.1)

Packages	Version	Functions	Description	Citation
usdm	1.1-18	vifstep	To perform multicollinearity testing by calculating Variable Inflation Factor (VIF) of the variables. Variables with VIF higher than the set threshold will be eliminated. The calculation will initiate every time when there is a variable being excluded, until all variables have a VIF lower than the threshold.	Naimi B, Hamm Na, Groen TA, Skidmore AK, Toxopeus AG (2014). "Where is positional uncertainty a problem for species distribution modelling." <i>Ecography</i> , *37*, 191-203. doi: 10.1111/j.1600-0587.2013.00205.x URL: http://doi.org/10.1111/j.1600-0587.2013.00205.x
randomForest	4.6-14	randomForest	To build random forest models	A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. <i>R News</i> 2(3),18--22.
		importance	To calculate the variable permutative importance (VIMP) value.	
Base	0.1-3	glm	To build generalized linear models (GLM)	R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
		predict	To predict response variable values from linear mixed models (LMM)	
MuMIn	1.42.1	model.sel	To tabulate the performance criteria of model for selections.	Kamil Barton (2018). MuMIn: Multi-Model Inference. R package version 1.42.1. https://CRAN.R-project.org/package=MuMIn
		QAICc	To calculate AICc value for overdispersed models	

		r.squaredGLMM	To calculate conditional r^2 for linear mixed effect models	<p>Nakagawa, S., Schielzeth, H. (2013) A general and simple method for obtaining R^2 from Generalized Linear Mixed-effects Models. Methods in Ecology and Evolution 4: 133–142</p> <p>Johnson, P.C.D. (2014) Extension of Nakagawa & Schielzeth's R_GLMM^2 to random slopes models. Methods in Ecology and Evolution 5: 44-946</p>
MASS	7.3-51	glm.nb	Building negative binomial distribution GLM	Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
rsq	1.1	rsq	Calculating adjusted r^2	<p>Dabao Zhang (2018). rsq: R-Squared and Related Measures. R package version 1.1. https://CRAN.R-project.org/package=rsq</p>
lme4	1.1-18-1	lmer	To build linear mixed effect models	<p>Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.</p>
		BIC	To calculate the Bayesian Information Criterion value	

7.4 Appendix IV

Appendix IV Complete summary of physicochemical, hydrological and land use variables

Abiotic variables (Predictors)	Description	Unit	Max.	Min.	Mean	SD
<i>Physico-chemical variables</i>						
WT	Water temperature	°C	19.20	0.20	9.96	4.56
pH	pH	-	9.73	6.73	7.79	0.53
DO	Dissolved oxygen	mg/L	12.98	2.93	9.13	1.84
TP	Total phosphorus	mg/L	1.10	0.04	0.22	0.17
PO4	Orthophosphate-phosphorus	mg/L	0.63	0.01	0.07	0.09
NH4	Ammonium-nitrogen	mg/L	1.51	0.02	0.23	0.28
NO3	Nitrate-nitrogen	mg/L	37.49	0.10	6.40	5.68
NO2	Nitrite-nitrogen	mg/L	0.73	0.00	0.04	0.09
DIN	Dissolve inorganic nitrogen	mg/L	37.90	0.15	6.67	5.70
NPR	Nitrogen-phosphate ratio	-	234.16	0.23	50.37	56.80
Cl	Chloride	mg/L	98.54	14.23	28.48	10.02
SO4	Sulfate	mg/L	84.92	10.32	37.95	14.16
TSS	Total suspended solid	mg/L	87.88	0.00	11.31	11.68
<i>Hydrological variables</i>						
H01	Intraday flows	m ³ /s	18.30	0.00	1.29	3.25
H02	the first day before	m ³ /s	16.80	0.01	1.19	2.98
H03	the second day before	m ³ /s	15.50	0.00	1.11	2.72
H04	The Third day before	m ³ /s	15.20	0.01	1.15	2.74
H05	The four day before	m ³ /s	16.10	0.01	1.24	2.93
H06	mean in 3 day flows (including the sampling day)	m ³ /s	16.87	0.01	1.20	2.98
H07	mean flows of 3 day before (not including the sampling day)	m ³ /s	15.83	0.01	1.15	2.81
H08	Median in 3 day flows (including the sampling day)	m ³ /s	16.80	0.00	1.19	2.98
H09	Median flows of 3 day before (not including the sampling day)	m ³ /s	15.50	0.01	1.13	2.74
H10	Variability in 3 days flows (including the sampling day)	-	2.25	0.00	0.09	0.34
H11	Variability flows of 3 day before (not including the sampling day)	-	1.09	0.00	0.04	0.15
H12	Skewness in 3 day flows (including the sampling day)	-	1.73	-1.73	0.34	1.02
H13	Skewness of 3 day before (not including the sampling day)	-	1.73	-1.72	0.62	1.08
H14	mean in 7 day flows (including the sampling day)	m ³ /s	17.09	0.01	1.25	3.04
H15	mean flows of 7 day before (not including the sampling day)	m ³ /s	17.03	0.01	1.24	3.03
H16	Median in 7 day flows (including the sampling day)	m ³ /s	16.80	0.00	1.22	3.04
H17	Median flows of 7 day before (not including the sampling day)	m ³ /s	16.80	0.01	1.21	3.00
H18	Variability in 7 days flows (including the sampling day)	-	2.66	0.00	0.13	0.41

Appendix IV Complete summary of abiotic variables

H19	Variability flows of 7 day before (not including the sampling day)	-	2.52	0.00	0.15	0.44
H20	Skewness in 7 day flows (including the sampling day)	-	2.44	-0.78	0.88	0.69
H21	Skewness of 7 day before (not including the sampling day)	-	2.64	-1.58	0.95	0.86
H22	mean in 14 day flows (including the sampling day)	m ³ /s	12.49	0.01	0.95	2.20
H23	mean flows of 14 day before (not including the sampling day)	m ³ /s	11.32	0.01	0.88	2.00
H24	Median in 14 day flows (including the sampling day)	m ³ /s	15.35	0.01	1.09	2.69
H25	Median flows of 14 day before (not including the sampling day)	m ³ /s	14.15	0.01	0.99	2.49
H26	Variability in 14 days flows (including the sampling day)	-	38.98	0.00	1.53	5.85
H27	Variability flows of 14 day before (not including the sampling day)	-	43.59	0.00	1.69	6.55
H28	Skewness in 14 day flows (including the sampling day)	-	3.45	-0.70	0.68	0.76
H29	Skewness of 14 day before (not including the sampling day)	-	3.37	-0.53	0.69	0.66
H30	mean in 30 day flows (including the sampling day)	m ³ /s	7.12	0.01	0.63	1.26
H31	mean flows of 30 day before (not including the sampling day)	m ³ /s	6.58	0.01	0.60	1.17
H32	Median in 30 day flows (including the sampling day)	m ³ /s	3.29	0.01	0.37	0.63
H33	Median flows of 30 day before (not including the sampling day)	m ³ /s	3.10	0.01	0.36	0.60
H34	Variability in 30 days flows (including the sampling day)	-	43.78	0.00	1.64	6.62
H35	Variability flows of 30 day before (not including the sampling day)	-	39.99	0.00	1.51	6.04
H36	Skewness in 30 day flows (including the sampling day)	-	3.28	-0.31	1.29	0.56
H37	Skewness of 30 day before (not including the sampling day)	-	4.40	-0.40	1.34	0.59
H38	Frequency of flow events	-	3.00	0.00	1.78	1.38
H39	Low flood pulse count 3 days	Days	7.00	0.00	3.76	2.82
H40	Low flood pulse count 7 days	Days	14.00	0.00	7.53	5.02
H41	Low flood pulse count 14 days	Days	30.00	0.00	17.87	8.88
H42	Low flood pulse count 30 days	Days	3.00	0.00	0.75	1.27
H43	High flood pulse count 3 days	Days	7.00	0.00	1.68	2.93
H44	High flood pulse count 7 days	Days	12.00	0.00	2.20	3.78
H45	High flood pulse count 14 days	Days	12.00	0.00	2.21	3.78
H46	High flood pulse count 30 days	Days	100.00	0.00	59.32	45.95
H47	Low flood pulse count 3 days	%	100.00	0.00	53.75	40.22
H48	Low flood pulse count 7 days	%	100.00	0.00	53.75	35.83
H49	Low flood pulse count 14 days	%	100.00	0.00	59.58	29.59
H50	Low flood pulse count 30 days	%	100.00	0.00	24.86	42.34
H51	High flood pulse count 3 days	%	100.00	0.00	23.97	41.87
H52	High flood pulse count 7 days	%	85.71	0.00	15.80	26.98
H53	High flood pulse count 14 days	%	40.00	0.00	7.37	12.59
H54	High flood pulse count 30 days	%	1.50	-0.20	0.09	0.28
H55	Rate of change in flow events	-	0.09	-0.31	-0.03	0.07

Appendix IV Complete summary of abiotic variables

H56	Rate of change 3 days	-	1.27	-0.11	0.08	0.24
H57	Rate of change 7 days	-	0.62	-0.01	0.04	0.11
<i>Land use variables</i>						
WIDTH	River width	m	17.10	1.05	4.94	3.63
DEPTH	River depth	m	2.00	0.10	0.47	0.32
VELOCITY	Flow velocity	m/s	10.24	0.00	0.62	1.61
AGRL	Agricultural land	%	79.65	15.04	51.83	17.14
FRSD	Diciduous forest	%	9.89	0.01	2.23	1.92
FRSE	Evergreen forest	%	9.02	0.02	1.02	1.37
FRST	Mixed Forest	%	13.47	0.00	2.46	2.56
FR	Total forest	%	15.13	0.86	5.71	3.40
RNGE	Rangeland	%	4.33	0.00	0.70	0.80
UIDU	Industrial land	%	8.41	2.98	4.21	0.90
URLD	Residential - Low density	%	3.98	0.00	0.43	0.59
URMD	Residential - Middle density	%	9.79	1.74	5.22	1.74
UR	Urban-residential land	%	12.26	1.75	5.65	2.01
WATR	Water	%	5.42	0.62	1.71	1.10
WETL	Wetland	%	7.19	0.00	1.01	1.11
WPAS	Winter pasture	%	70.97	7.22	29.18	16.66

7.5 Appendix V Complete list of observed and simulated results

SS*	SB*	Species richness									Abundance								
		Obs	Scen1	Scen2	Scen3	Scen4	Scen5	Scen6	Scen7	Scen8	Obs	Scen1	Scen2	Scen3	Scen4	Scen5	Scen6	Scen7	Scen8
1	Tr	61	54.426	54.091	53.983	52.759	54.699	54.576	54.822	54.946	12.432	13.855	13.979	13.649	13.614	13.702	13.643	13.761	13.819
2	Tr	54	44.863	44.546	43.525	41.449	44.815	44.088	45.543	46.271	12.249	11.753	11.870	11.558	11.553	11.637	11.639	11.635	11.633
3	Tr	49	63.002	62.704	62.529	61.361	63.254	63.163	63.346	63.438	11.034	13.189	13.299	13.005	12.960	13.062	13.029	13.095	13.128
4	Tr	46	30.919	30.621	29.257	26.899	30.974	30.487	31.460	31.947	12.454	11.676	11.786	11.492	11.500	11.569	11.576	11.562	11.556
5	Tr	46	42.560	42.240	42.217	41.127	42.836	42.749	42.923	43.010	11.342	13.458	13.576	13.261	13.224	13.332	13.316	13.348	13.364
6	Tr	44	44.832	44.468	44.604	43.526	45.105	44.922	45.288	45.471	13.426	14.018	14.152	13.793	13.763	13.862	13.819	13.905	13.948
7	Tr	56	57.071	56.733	57.068	56.275	57.375	57.306	57.444	57.514	12.453	13.465	13.590	13.256	13.203	13.329	13.306	13.351	13.373
8	Tr	40	37.199	36.883	36.636	35.335	37.350	37.019	37.681	38.012	12.939	11.828	11.945	11.633	11.632	11.722	11.745	11.700	11.677
9	Tr	52	52.193	51.877	52.178	51.425	52.473	52.401	52.546	52.619	14.072	14.044	14.160	13.849	13.787	13.891	13.820	13.962	14.034
10	Tr	55	40.308	40.142	40.019	39.343	40.248	39.796	40.699	41.150	12.186	12.150	12.211	12.048	12.042	12.086	12.082	12.091	12.095
11	Tr	55	51.062	50.731	51.102	50.370	51.397	51.406	51.389	51.380	14.515	13.995	14.118	13.792	13.725	13.848	13.799	13.898	13.948
12	Tr	50	48.733	48.409	48.752	48.016	49.061	49.068	49.053	49.046	14.953	14.064	14.184	13.865	13.794	13.918	13.865	13.971	14.024
13	Tr	54	48.611	48.269	48.661	47.912	48.961	48.976	48.945	48.930	14.642	14.077	14.204	13.867	13.796	13.929	13.884	13.973	14.018
14	Tr	53	50.986	50.640	51.045	50.297	51.303	51.245	51.360	51.418	14.808	14.397	14.525	14.184	14.105	14.246	14.199	14.292	14.339
15	Tr	50	50.832	50.508	50.917	50.247	51.051	50.841	51.260	51.470	15.390	13.406	13.525	13.206	13.149	13.271	13.240	13.302	13.333
16	Tr	57	51.403	51.113	51.457	50.831	51.586	51.369	51.802	52.019	13.544	13.149	13.257	12.970	12.912	13.026	12.994	13.058	13.090
17	Tr	55	31.482	31.224	31.451	30.820	31.647	31.464	31.831	32.015	14.720	13.030	13.125	12.871	12.793	12.930	12.922	12.938	12.946
18	Tr	50	49.102	48.878	49.088	48.551	49.256	49.115	49.397	49.538	11.815	11.884	11.967	11.746	11.733	11.802	11.804	11.801	11.799
19	Tr	41	41.016	40.685	41.088	40.387	41.218	40.959	41.477	41.736	14.096	13.310	13.432	13.106	12.997	13.187	13.185	13.189	13.191
20	Tr	39	48.411	48.080	48.318	47.453	48.605	48.330	48.879	49.154	11.039	12.626	12.748	12.422	12.418	12.504	12.505	12.503	12.502
21	Tr	19	41.227	40.789	41.728	41.207	41.571	41.384	41.758	41.945	13.666	15.554	15.716	15.285	14.986	15.398	15.409	15.387	15.377

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors: A data modelling study in a German lowland catchment
Szewing To – December 2018

Appendix V Complete list of observed and simulated results

22	Tr	62	51.927	51.640	52.131	51.664	52.159	52.048	52.270	52.381	11.525	11.619	11.725	11.443	11.454	11.517	11.525	11.509	11.502
23	Bo	55	41.970	41.915	41.070	40.041	41.821	41.413	42.230	42.638	11.741	10.508	10.528	10.474	10.386	10.492	10.501	10.483	10.474
24	Bo	53	46.263	46.208	45.386	44.380	46.090	45.634	46.546	47.002	11.239	10.613	10.634	10.579	10.491	10.597	10.605	10.589	10.581
25	Bo	42	41.738	41.686	40.907	39.956	41.564	41.115	42.014	42.463	9.508	10.543	10.562	10.511	10.420	10.528	10.536	10.520	10.512
26	Bo	37	43.471	43.353	42.768	41.791	43.351	42.874	43.827	44.303	13.363	10.854	10.897	10.781	10.714	10.814	10.821	10.807	10.799
27	Bo	43	42.351	42.244	41.858	41.116	42.242	41.810	42.673	43.104	12.245	10.697	10.737	10.632	10.559	10.661	10.667	10.655	10.649
28	Bo	40	39.953	39.887	38.953	37.798	39.841	39.484	40.198	40.556	10.748	9.531	9.556	9.491	9.451	9.515	9.531	9.499	9.483
29	Bo	26	37.230	37.157	35.881	34.360	37.281	37.236	37.327	37.372	8.924	9.490	9.517	9.445	9.412	9.469	9.481	9.457	9.445
30	Bo	51	44.677	44.566	44.276	43.619	44.587	44.186	44.987	45.388	10.398	10.474	10.515	10.406	10.344	10.435	10.438	10.432	10.428
31	Bo	43	40.155	40.045	39.581	38.749	40.259	40.245	40.273	40.286	9.473	11.759	11.800	11.691	11.460	11.732	11.759	11.705	11.678
32	Bo	47	43.990	43.934	43.571	43.024	43.860	43.489	44.230	44.600	9.598	10.413	10.433	10.379	10.322	10.396	10.403	10.389	10.382
33	Bo	54	39.352	39.235	38.823	38.019	39.324	39.031	39.616	39.909	9.518	10.428	10.471	10.355	10.319	10.384	10.384	10.385	10.385
34	Bo	29	39.469	39.388	39.186	38.713	39.586	39.656	39.515	39.444	7.376	9.641	9.671	9.591	9.556	9.613	9.618	9.608	9.604
35	Je	34	32.141	31.909	30.996	29.309	32.055	31.421	32.690	33.325	9.791	10.243	10.329	10.101	10.080	10.178	10.219	10.138	10.097
36	Je	32	36.292	36.082	35.150	33.520	36.172	35.513	36.831	37.490	9.891	10.451	10.529	10.322	10.295	10.390	10.422	10.357	10.325
37	Je	47	31.567	31.174	30.565	28.645	31.667	31.077	32.256	32.845	9.945	10.435	10.580	10.193	10.219	10.304	10.333	10.275	10.247
38	Je	29	36.981	36.848	35.579	33.868	36.735	35.978	37.492	38.248	10.004	9.653	9.702	9.571	9.539	9.617	9.643	9.591	9.565
39	Je	33	32.448	32.305	31.208	29.633	32.261	31.599	32.923	33.585	9.149	9.973	10.026	9.884	9.842	9.936	9.970	9.903	9.869
40	Je	35	30.790	30.636	29.517	27.884	30.619	29.966	31.271	31.924	9.729	9.893	9.950	9.798	9.755	9.851	9.883	9.820	9.789
41	Je	16	28.363	28.124	27.074	25.226	28.379	27.934	28.825	29.271	10.520	9.785	9.873	9.638	9.631	9.714	9.750	9.679	9.644
42	Je	26	36.507	36.455	35.254	33.881	36.191	35.457	36.926	37.660	9.269	9.516	9.535	9.484	9.397	9.512	9.541	9.482	9.453
43	Je	32	34.096	34.048	32.990	31.773	33.800	33.112	34.488	35.176	9.864	10.035	10.053	10.006	9.874	10.030	10.056	10.005	9.980
44	Je	16	25.213	25.187	24.496	23.720	24.843	24.051	25.635	26.426	8.479	9.909	9.919	9.893	9.850	9.912	9.937	9.887	9.862
45	Ki	46	45.688	45.298	45.964	45.331	45.930	45.635	46.225	46.520	13.788	12.522	12.666	12.282	12.265	12.374	12.367	12.382	12.390
46	Ki	52	44.117	43.665	44.480	43.786	44.409	44.088	44.731	45.053	13.693	12.779	12.946	12.500	12.487	12.608	12.600	12.615	12.623

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors: A data modelling study in a German lowland catchment
Szewing To – December 2018

Appendix V Complete list of observed and simulated results

47	Ki	48	42.350	41.875	42.881	42.304	42.665	42.346	42.984	43.304	11.121	12.671	12.846	12.379	12.400	12.499	12.506	12.493	12.486
48	Ki	48	37.416	36.706	38.096	37.119	37.956	37.615	38.296	38.637	15.593	14.166	14.428	13.729	13.742	13.900	13.892	13.907	13.914
49	Ki	47	37.437	36.193	38.936	37.534	38.508	38.163	38.853	39.199	11.777	14.730	15.189	13.965	14.177	14.277	14.291	14.264	14.251
50	Ki	40	43.034	42.394	43.715	42.901	43.507	43.172	43.842	44.177	14.629	13.970	14.207	13.576	13.503	13.734	13.735	13.734	13.733
51	Ki	43	43.770	43.156	44.477	43.750	44.224	43.903	44.545	44.867	13.728	12.485	12.712	12.107	12.189	12.266	12.283	12.250	12.233
52	Sa	36	35.359	35.161	34.769	33.715	35.457	35.257	35.658	35.859	12.988	14.397	14.471	14.275	13.984	14.329	14.340	14.319	14.308
53	Sa	37	34.477	34.120	34.231	33.152	34.686	34.389	34.983	35.280	13.955	14.556	14.687	14.336	14.077	14.429	14.438	14.419	14.410
54	Sa	37	33.741	33.068	33.366	31.420	34.293	34.051	34.535	34.777	15.428	15.204	15.453	14.790	14.716	14.966	14.988	14.945	14.923
55	Sa	57	36.904	36.171	37.540	36.469	37.527	37.310	37.745	37.962	14.612	13.079	13.349	12.629	12.735	12.816	12.830	12.802	12.788
56	Ju	32	37.416	37.176	35.907	33.841	37.557	37.360	37.753	37.949	12.140	12.213	12.302	12.066	11.960	12.136	12.158	12.114	12.092
57	Ju	17	22.752	22.554	20.837	18.458	22.833	22.596	23.069	23.306	11.675	11.015	11.089	10.893	10.859	10.950	10.966	10.934	10.918
58	Ju	41	32.013	31.818	30.645	28.822	32.180	32.123	32.237	32.294	12.683	12.308	12.380	12.188	12.019	12.244	12.260	12.229	12.213
59	Ju	23	24.909	24.799	23.308	21.450	24.982	24.908	25.057	25.131	11.401	13.508	13.549	13.440	13.062	13.472	13.482	13.462	13.452
60	Tr	39	38.693	38.575	37.888	36.809	38.686	38.436	38.936	39.186	12.494	13.612	13.656	13.540	13.451	13.550	13.511	13.588	13.626
61	Tr	17	24.157	24.050	22.469	20.533	24.081	23.715	24.446	24.812	9.412	11.646	11.685	11.580	11.524	11.610	11.616	11.604	11.598
62	Tr	39	31.146	31.010	30.403	29.343	31.164	30.927	31.401	31.638	12.049	13.036	13.087	12.952	12.868	12.972	12.945	13.000	13.027
63	Tr	22	17.013	16.958	14.946	12.750	16.963	16.752	17.173	17.384	10.574	10.903	10.923	10.869	10.817	10.887	10.895	10.878	10.870
64	Tr	29	33.889	33.742	33.258	32.284	33.837	33.440	34.235	34.632	11.879	12.798	12.852	12.708	12.628	12.742	12.739	12.746	12.749
65	Tr	24	36.265	36.026	35.829	34.834	36.429	36.278	36.580	36.731	12.059	12.560	12.649	12.413	12.352	12.475	12.481	12.469	12.463
66	Tr	28	38.054	37.841	37.843	37.133	38.181	38.009	38.354	38.527	12.741	12.700	12.779	12.569	12.484	12.626	12.634	12.617	12.608
67	Tr	17	23.922	23.878	22.905	21.785	23.975	23.993	23.957	23.939	10.181	11.519	11.535	11.492	11.423	11.506	11.512	11.500	11.493
68	Tr	34	39.086	38.868	38.905	38.218	39.222	39.062	39.383	39.543	12.513	12.736	12.816	12.602	12.516	12.661	12.670	12.651	12.641
69	Tr	23	11.813	11.739	11.371	10.757	11.742	11.451	12.032	12.322	8.946	11.797	11.824	11.752	11.723	11.774	11.783	11.766	11.757
70	Tr	50	44.164	44.021	43.892	43.285	44.149	43.830	44.467	44.785	12.697	13.118	13.171	13.029	12.917	13.065	13.065	13.065	13.064
71	Tr	48	33.842	33.665	33.616	32.977	33.842	33.487	34.196	34.550	12.575	13.175	13.240	13.066	12.959	13.110	13.110	13.110	13.110

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors: A data modelling study in a German lowland catchment
Szewing To – December 2018

Appendix V Complete list of observed and simulated results

72	Tr	49	44.580	44.440	44.293	43.679	44.560	44.240	44.880	45.201	13.734	13.067	13.119	12.981	12.860	13.016	13.017	13.015	13.014
73	Tr	35	35.618	35.448	35.382	34.752	35.600	35.227	35.973	36.346	13.558	13.569	13.631	13.465	13.342	13.509	13.513	13.504	13.499
74	Tr	52	35.310	35.123	35.169	34.589	35.285	34.857	35.712	36.139	14.806	13.362	13.432	13.247	13.156	13.297	13.304	13.289	13.282
75	Tr	50	41.589	41.353	41.550	40.962	41.606	41.170	42.043	42.479	15.326	12.833	12.920	12.688	12.616	12.755	12.773	12.737	12.720
76	Tr	44	36.323	36.058	36.305	35.668	36.374	35.945	36.802	37.231	16.452	13.550	13.648	13.387	13.311	13.461	13.477	13.444	13.427
77	Tr	16	47.969	47.626	48.151	47.535	48.116	47.725	48.506	48.897	11.116	12.642	12.768	12.431	12.447	12.525	12.545	12.505	12.486
78	Tr	37	36.876	36.486	37.046	36.306	37.048	36.611	37.484	37.921	17.316	13.752	13.896	13.512	13.418	13.616	13.632	13.600	13.585
79	Tr	10	28.255	28.082	27.899	27.139	28.241	27.869	28.614	28.987	12.310	11.935	11.999	11.829	11.786	11.878	11.892	11.865	11.851
80	Tr	25	31.252	30.328	32.562	31.718	31.900	31.349	32.450	33.001	18.709	17.167	17.508	16.599	16.420	16.836	16.857	16.816	16.795
81	Tr	49	39.068	38.880	39.105	38.705	38.981	38.432	39.530	40.079	12.700	12.151	12.221	12.036	12.023	12.092	12.113	12.072	12.051
82	Bo	17	19.242	19.202	18.318	17.298	19.157	18.906	19.409	19.660	11.746	10.261	10.276	10.236	10.144	10.250	10.259	10.241	10.232
83	Bo	13	29.537	29.467	28.685	27.669	29.520	29.347	29.694	29.867	10.081	10.511	10.537	10.468	10.383	10.490	10.500	10.481	10.471
84	Bo	22	22.287	22.257	21.419	20.483	22.245	22.102	22.388	22.531	10.035	10.219	10.230	10.201	10.104	10.214	10.224	10.203	10.193
85	Bo	29	23.245	23.186	22.444	21.506	23.246	23.129	23.362	23.479	10.457	10.087	10.109	10.051	9.970	10.071	10.082	10.060	10.049
86	Bo	14	15.172	15.120	14.587	13.882	15.152	15.009	15.294	15.437	8.848	10.021	10.040	9.989	9.903	10.009	10.022	9.995	9.981
87	Bo	24	13.436	13.400	12.387	11.251	13.473	13.474	13.473	13.472	11.685	9.458	9.471	9.435	9.388	9.447	9.452	9.441	9.436
88	Bo	29	24.518	24.448	23.162	21.644	24.309	23.751	24.866	25.424	10.741	9.285	9.311	9.242	9.208	9.263	9.272	9.255	9.247
89	Bo	30	26.696	26.604	26.265	25.619	26.807	26.844	26.769	26.732	10.091	10.378	10.412	10.321	10.254	10.351	10.365	10.337	10.322
90	Bo	15	30.062	29.672	29.954	28.935	30.260	29.876	30.644	31.028	10.163	12.968	13.112	12.728	12.566	12.830	12.842	12.818	12.806
91	Bo	20	20.984	20.951	20.529	19.996	21.075	21.192	20.959	20.842	9.640	9.827	9.839	9.807	9.744	9.821	9.834	9.809	9.796
92	Bo	7	28.462	28.444	27.767	27.030	28.427	28.319	28.534	28.642	8.189	9.214	9.221	9.203	9.142	9.215	9.229	9.200	9.186
93	Bo	16	22.639	22.584	22.312	21.857	22.612	22.446	22.777	22.942	8.616	9.298	9.319	9.264	9.223	9.285	9.300	9.271	9.256
94	Je	34	18.033	17.977	16.593	15.024	18.214	18.465	17.962	17.711	10.702	9.969	9.989	9.935	9.871	9.947	9.945	9.950	9.952
95	Je	13	12.647	12.187	11.923	10.125	13.145	13.221	13.069	12.993	12.391	11.495	11.665	11.212	11.247	11.329	11.335	11.323	11.316
96	Je	18	17.350	17.210	15.925	14.174	17.577	17.752	17.402	17.227	8.795	10.079	10.131	9.993	9.956	10.029	10.034	10.025	10.021

Simulating changes in riverine phytoplankton assemblages intensified environmental stressors: A data modelling study in a German lowland catchment
Szewing To – December 2018

Appendix V Complete list of observed and simulated results

97	Je	24	21.691	21.669	20.105	18.468	21.779	21.910	21.647	21.516	9.253	9.304	9.312	9.290	9.231	9.300	9.308	9.291	9.283
98	Je	19	17.113	17.091	15.671	14.177	17.169	17.237	17.101	17.033	9.670	9.631	9.639	9.617	9.545	9.625	9.630	9.620	9.615
99	Je	18	14.725	14.695	13.243	11.693	14.791	14.866	14.716	14.642	9.194	9.652	9.663	9.634	9.561	9.645	9.651	9.639	9.633
100	Je	14	18.757	18.687	17.185	15.450	18.686	18.404	18.968	19.249	10.722	9.324	9.350	9.281	9.233	9.312	9.341	9.284	9.255
101	Je	13	21.464	21.438	20.168	18.812	21.508	21.544	21.472	21.435	9.092	9.839	9.848	9.823	9.729	9.832	9.838	9.826	9.821
102	Je	17	19.796	19.744	18.696	17.476	19.867	19.907	19.827	19.787	9.320	10.491	10.510	10.460	10.328	10.480	10.495	10.465	10.450
103	Je	10	16.318	16.292	15.601	14.825	16.360	16.394	16.327	16.294	8.189	9.758	9.768	9.743	9.699	9.753	9.763	9.744	9.735
104	Ki	31	32.583	32.046	33.105	32.374	32.893	32.438	33.348	33.803	12.581	11.632	11.830	11.301	11.320	11.442	11.460	11.425	11.407
105	Ki	33	29.877	28.905	31.104	30.065	30.635	30.209	31.061	31.487	12.193	12.881	13.240	12.284	12.398	12.531	12.548	12.514	12.497
106	Ki	33	28.996	28.201	30.060	29.270	29.604	29.231	29.977	30.350	12.986	12.531	12.824	12.042	12.142	12.248	12.271	12.226	12.204
107	Ki	14	18.751	17.529	20.283	18.965	19.762	19.342	20.182	20.603	14.355	14.753	15.204	14.001	14.141	14.310	14.326	14.294	14.278
108	Ki	18	26.768	25.914	27.617	26.475	27.428	27.041	27.815	28.202	11.814	12.988	13.304	12.463	12.579	12.681	12.698	12.665	12.649
109	Ki	10	13.946	11.639	17.404	15.480	15.964	15.387	16.542	17.119	16.591	18.747	19.599	17.328	17.664	17.903	17.918	17.888	17.873
110	Ki	22	23.576	22.414	25.197	24.105	24.514	24.065	24.964	25.413	10.694	13.408	13.837	12.692	12.910	12.986	13.002	12.971	12.955
111	Sa	17	30.231	30.168	29.414	28.451	30.087	29.674	30.500	30.913	16.290	14.253	14.276	14.214	13.890	14.236	14.248	14.224	14.212
112	Sa	13	14.439	14.417	13.635	12.779	14.215	13.724	14.707	15.198	10.858	13.479	13.487	13.465	13.124	13.482	13.503	13.460	13.438
113	Sa	25	30.101	29.976	28.812	27.232	30.010	29.579	30.442	30.873	13.755	13.862	13.908	13.785	13.576	13.826	13.848	13.805	13.783
114	Sa	26	15.288	15.178	14.889	14.231	15.079	14.441	15.717	16.356	12.244	11.139	11.180	11.071	11.025	11.109	11.130	11.088	11.067
115	Ju	24	20.182	20.086	18.435	16.464	20.002	19.451	20.553	21.105	14.946	13.217	13.252	13.158	13.016	13.188	13.199	13.176	13.164
116	Ju	9	12.154	12.125	9.957	7.691	11.923	11.400	12.445	12.968	10.076	11.827	11.838	11.809	11.733	11.822	11.833	11.811	11.799
117	Ju	17	22.906	22.758	21.458	19.667	22.772	22.211	23.333	23.894	14.380	13.638	13.692	13.547	13.367	13.590	13.602	13.577	13.565
118	Ju	17	10.121	9.389	9.557	7.284	10.639	10.209	11.069	11.499	18.208	16.962	17.232	16.511	16.286	16.697	16.708	16.686	16.674

*SS: Sampling sites; SB: Subbasins

7.6 Appendix VI

Sample R script for data exploration, random forest, generalized linear model, linear mixed effect model and projection.

Part I – Data exploration (Shapiro test, correlationship and collineation)

```
rm(list=ls())      # cleaning data
gc()
setwd("D:/Thesis/Thesis_wd/New/")
# Installing required packages
install.packages(c("readr", "ggplot2", "ggpubr", "usdm", "randomForest", "MuMIn",
"MASS", "sjstats", "lmtest", "rsq", "lme4", "MASS", "car", "r2glmm"))
library(readr)
library(ggplot2)
library(ggpubr)
# Species richness and abundance data-----
options(scipen = 999)
Resp.var <- read.csv(paste0(getwd(), "/InputData.csv"), sep = ",", header =
T)
Resp.var$Biomass = log(Resp.var$Biomass)
# Log-transformation to cell density for better model-fitting

#####
### Resp.var data exploration ###
#####
# Shapiro test on response variables
Resp.var.Shapiro <- lapply(Resp.var, shapiro.test)
Resp.var.Shapiro <- sapply(Resp.var.Shapiro, `[,`, c("statistic", "p.value"))
(as.data.frame(t(Resp.var.Shapiro)))
rm(Resp.var.Shapiro)

# vif
library(usdm)
(vifstep(Resp.var, th = 5))
# Boxplot
pdf("Boxplot_RespVar.pdf", width = 5, height = 5)
par(mar=c(3,3,3,3))
par(mfrow=c(1,3))
boxplot(Resp.var$Richness, main = "Species richness", cex.main = 1.5, xlab
= NULL)
boxplot(Resp.var$Abundance, main = "Relative abundance", cex.main = 1.5,
xlab = NULL)
boxplot(Resp.var$Biomass, main = "Overall biomass", cex.main = 1.5, xlab =
NULL)
dev.off()
#####
### Predictor var. data exploration ###
#####
# Reading data (n=86) -----
Av <- read.csv(paste0(getwd(), "/Av_input.csv"), sep = ",", header = T)
# Av = Ev + Hv (118 obs. of 86 variables)
Av <- Av[c(2:87)]
# Calculating VIF - - VIF_step (threshold value = 5 [Sheather, Simon
(2009)]) -----
(vifstep(Av, th = 5))

# Random effects (Year and Subbasin)-----
t = matrix(data = NA, nrow = 1, ncol = 118)
t[c(1:59)] = c("2014")
t[c(60:118)] = c("2015")
Year = as.data.frame(t(t))
colnames(Year) = c("Year")
```



```
s = matrix(data = NA, nrow = 1, ncol = 118)
s[c(1:22)] = c("Tr")
s[c(23:34)] = c("Bo")
s[c(35:44)] = c("Je")
s[c(45:51)] = c("Ki")
s[c(52:55)] = c("Sa")
s[c(56:59)] = c("Ju")
s[c(60:118)] = s[c(1:59)]
SB = as.data.frame(t(s))
colnames(SB) = c("Subbasin")
rm(t,s)
```

Part II – Random forest

```
library(randomForest)
# 1. Species richness (500) -----
rfsprch <- Resp.var$Richness
rfsprch = cbind(rfsprch, Av)

set.seed(1234)
(randomForest(rfsprch~.,data=rfsprch, ntree=8000, importance = T))
set.seed(1234)
(randomForest(rfsprch~.,data=rfsprch, ntree=5000, importance = T))
set.seed(1234)
(randomForest(rfsprch~.,data=rfsprch, ntree=1000, importance = T))
set.seed(1234)
(randomForest(rfsprch~.,data=rfsprch, ntree=500, importance = T))
set.seed(1234)
VIMPSr <- as.data.frame(randomForest::importance(
  randomForest(rfsprch~.,data=rfsprch,ntree=500,importance = T), type = 1))
  ## VImportance

# 2. Overall abundance (1000) -----
rfAbun <- Resp.var$Abundance
rfAbun = cbind(rfAbun, Av)

set.seed(1234)
(randomForest(rfAbun~.,data=rfAbun, ntree=8000, importance = T))
set.seed(1234)
(randomForest(rfAbun~.,data=rfAbun, ntree=5000, importance = T))
set.seed(1234)
(randomForest(rfAbun~.,data=rfAbun, ntree=1000, importance = T))
set.seed(1234)
(randomForest(rfAbun~.,data=rfAbun, ntree=500, importance = T))
set.seed(1234)
VIMPab <- as.data.frame(randomForest::importance(
  randomForest(rfAbun~.,data=rfAbun, ntree=1000,importance = T), type =
1))
  ## VImportance

# 3. Overall Biomass (1000) -----
rfBiom <- Resp.var$Biomass
rfBiom = cbind(rfBiom, Av)

set.seed(1234)
(randomForest(rfBiom~.,data=rfBiom, ntree=8000, importance = T))
set.seed(1234)
(randomForest(rfBiom~.,data=rfBiom, ntree=5000, importance = T))
set.seed(1234)
(randomForest(rfBiom~.,data=rfBiom, ntree=1000, importance = T))
set.seed(1234)
(randomForest(rfBiom~.,data=rfBiom, ntree=500, importance = T))
set.seed(1234)
VIMPbio <- as.data.frame(randomForest::importance(
  randomForest(rfBiom~.,data=rfBiom, ntree=1000,importance = T), type =
1))
  ## VImportance

# Plotting VIMP -----
pdf("VIMP_new.pdf", width = 10, height = 5)
par(mar=c(2,2,2,2))
```

```
par(mfrow=c(1,3))
set.seed(1234)
varImpPlot(randomForest(rfsprch~,data=rfsprch, ntree=500, importance =
T), type = 1,main = "VIMP - Species richness")
set.seed(1234)
varImpPlot(randomForest(rfAbun~, data=rfAbun,ntree=1000, importance =
T), type = 1, main = "VIMP - Overall abundance")
set.seed(1234)
varImpPlot(randomForest(rfBiom~, data=rfBiom,ntree=1000, importance =
T), type = 1, main = "VIMP - Overall biomass")
# Shouldn't use MDI because it is based towards some predictor variables
(Strobl et al.,2007)
dev.off()
# Writting VIMP -----
write.table(round(cbind(VIMPSr,VIMPab,VIMPbio),2), "VIMP_new.txt")
save(Resp.var,Av,SB,Year, file = "Input.RData")
```

Part III – Generalized Linear Model

```
load("Input.RData")
#####
### Fixed Effects Model - glm ###
### Testing fixed effect terms via ML is recommended ###
### GLM1 - Global model (Complete fixed var.) ###
### GLM2 - Regional model (Only important parameters) ###
### GLM3 - Local interaction model ###
### (Important parameters with interactions) ###
### GLM4 - Local model (First 3 important parameters) ###
### GLM5 - Selected significant parameters from global model (GLM1) ###
### GLM6 - Selected significant parameters from regional model (GLM2) ###
#####
library(MuMIn)
library(MASS)
library(rsq)

# 1. Species richness ----
SRdata <- as.data.frame(Resp.var$Richness)
colnames(SRdata) <- c("Richness")
SRdata <- cbind(SRdata,Av)
(GLMsp1 <- glm(Richness ~ pH+DO+PO4+NH4+NO3+NO2+NPR+Cl+S04+TSS+H11+H12+H13+
H20+H21+H29+H36+H41+H55+WIDTH+DEPTH+VELOCITY+FRSE+FR+RNGE+
UIDU+UR+WATR+WETL+WPAS, family = poisson(link = "log"), data
= SRdata))
(GLMsp2 <- glm(Richness ~
pH+H55+NPR+WPAS+DEPTH+PO4+H29+WIDTH+S04+H41+H21+H36,
family = poisson(link = "log"), data = SRdata))
(GLMsp3 <- glm(Richness ~ pH+H55+NPR+(pH:H55)+(pH:NPR)+(H55:NPR),
family = poisson(link = "log"), data = SRdata))
(GLMsp4 <- glm(Richness ~ pH+H55+NPR, family = poisson(link = "log"), data
= SRdata))

# Overdispersion check ----
theta <- function(x){x$deviance/x$df.residual}
as.data.frame(matrix(unlist(list(
theta(GLMsp1),theta(GLMsp2),theta(GLMsp3),theta(GLMsp4))))))

# Over-dispersion --> negative binomial GLM (Harrison, 2014)
(GLMsp1 <- glm.nb(Richness ~ pH+DO+PO4+NH4+NO3+NO2+NPR+Cl+S04+TSS+H11+H12+
H13+H20+H21+H29+H36+H41+H55+WIDTH+DEPTH+VELOCITY+FRSE+
FR+RNGE+UIDU+UR+WATR+WETL+WPAS, data = SRdata))
(GLMsp2 <- glm.nb(Richness ~
pH+H55+NPR+WPAS+DEPTH+PO4+H29+WIDTH+S04+H41+H21+H36,
data = SRdata))
(GLMsp3 <- glm.nb(Richness ~ pH+H55+NPR+(pH:H55)+(pH:NPR)+(H55:NPR),
data = SRdata))
(GLMsp4 <- glm.nb(Richness ~ pH+H55+NPR, data = SRdata))
# Mod5 = selecting significant var. (<0.05) from GLMsp1 (i.e. global mod)
as predictors
summary(GLMsp1) # --> pH,DO,PO4,NPR,S04,H20,DEPTH,FR,WPAS
```

```
(GLMsp5 <- glm.nb(Richness ~ pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS,
  data = SRdata))
# Mod6 = selecting significant var. from the GLMsp2 as predictors
summary(GLMsp2) # --> pH,NPR,WPAS,PO4,WIDTH,H41
(GLMsp6 <- glm.nb(Richness ~ pH+NPR+WPAS+PO4+WIDTH+H41, data = SRdata))

# Select the best model
GLMsp <-
as.data.frame(model.sel(GLMsp1,GLMsp2,GLMsp3,GLMsp4,GLMsp5,GLMsp6)) #562134
as.data.frame(matrix(unlist(list(
  theta(GLMsp1),theta(GLMsp2),theta(GLMsp3),theta(GLMsp4),
  theta(GLMsp5),theta(GLMsp6))))))
# no overdispersion model

# 2. Overall abundance ----
Abdata <- as.data.frame(Resp.var$Abundance)
colnames(Abdata) <- c("Abundance")
Abdata <- cbind(Abdata, Av)

# WATR+PO4+UR+FRSE+SO4+WPAS+NPR+WETL+H2O+WIDTH+UIDU+NO3+H21
(GLMab1 <- glm(Abundance ~ pH+DO+PO4+NH4+NO3+NO2+NPR+Cl+SO4+TSS+H11+H12+
  H13+H2O+H21+H29+H36+H41+H55+WIDTH+DEPTH+VELOCITY+
  FRSE+FR+RNGE+UIDU+UR+WATR+WETL+WPAS,data = Abdata,
  family = gaussian(link = "identity"))))
(GLMab2 <- glm(Abundance ~ WATR+PO4+UR+FRSE+SO4+WPAS+NPR+WETL+H2O+WIDTH+
  UIDU+NO3+H21,family = gaussian(link = "identity"),
  data = Abdata))
(GLMab3 <- glm(Abundance ~ WATR+PO4+UR+(WATR:PO4)+(WATR:UR)+(PO4:UR),
  data = Abdata,family = gaussian(link = "identity"))))
(GLMab4 <- glm(Abundance ~ WATR+PO4+UR,data = Abdata,
  family = gaussian(link = "identity"))))

# Gaussian models do not require overdispersion adjustment because they do
# not assume a specific mean-variance relationship like other distributions
# (Harrison et al., 2018)

summary(GLMab1) #DO,PO4,H36,H55,WATR
(GLMab5 <- glm(Abundance ~ DO+PO4+H36+H55+WATR,data = Abdata,
  family = gaussian(link = "identity"))))
summary(GLMab2) # WATR,PO4,UR, WIDTH
(GLMab6 <- glm(Abundance ~ WATR+PO4+UR+WIDTH,data = Abdata,
  family = gaussian(link = "identity"))))
model.sel(GLMab1,GLMab2,GLMab3,GLMab4,GLMab5,GLMab6) #654231
# Mod 6 has been rated as the best but mod 5 has comparable performance
# with dAICc < 2 and 1 more predictor. They have similar logLik so more
# evaluation has to be made.
# Mod 7 is finally generated as a combination of Mod 5 and Mod 6
(GLMab7 <- glm(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH, data = Abdata,
  family = gaussian(link="identity"))))

GLMab <- as.data.frame(model.sel(GLMab1,GLMab2,GLMab3,GLMab4,
  GLMab5,GLMab6, GLMab7)) #7546231

# 3. Overall biomass ----
Biodata <- as.data.frame(Resp.var$Biomass)
colnames(Biodata) <- c("Biomass")
Biodata <- cbind(Biodata, Av)

# PO4+NH4+NO3+NPR+TSS+H2O+FRSE+UIDU+UR+WATR+WPAS
(GLMbio1 <- glm(Biomass ~ pH+DO+PO4+NH4+NO3+NO2+NPR+Cl+SO4+TSS+H11+H12+
  H13+H2O+H21+H29+H36+H41+H55+WIDTH+DEPTH+VELOCITY+FRSE+
  FR+RNGE+UIDU+UR+WATR+WETL+WPAS,data = Biodata,
  family = gaussian(link = "identity"))))
(GLMbio2 <- glm(Biomass ~ PO4+NH4+NO3+NPR+TSS+H2O+FRSE+UIDU+UR+WATR+WPAS,
  family = gaussian(link = "identity"),data = Biodata))
(GLMbio3 <- glm(Biomass ~ PO4+NH4+NO3+(PO4:NH4)+(PO4:NO3)+(NH4:NO3),
  data = Biodata,family = gaussian(link = "identity"))))
(GLMbio4 <- glm(Biomass ~ PO4+NH4+NO3,data = Biodata,
```

```

family = gaussian(link = "identity"))))

# Gaussian models do not require overdispersion adjustment because they do
# not assume a specific mean-variance relationship like other distributions
(Harrison et al., 2018)

summary(GLMbio1) #NO3+TSS+H55+FR+WATR
(GLMbio5 <- glm(Biomass ~ NO3+TSS+H55+FR+WATR,data = Biodata,
               family = gaussian(link = "identity"))))
summary(GLMbio2) # WATR,TSS
(GLMbio6 <- glm(Biomass ~ WATR+TSS+PO4,data = Biodata,
               family = gaussian(link = "identity"))))

model.sel(GLMbio1, GLMbio2, GLMbio3, GLMbio4, GLMbio5, GLMbio6) #562143
rse(GLMbio5) # 1.13
rse(GLMbio6) # 1.15

rsq(GLMbio5, adj = T) # 0.3763317
rsq(GLMbio6, adj = T) # 0.3533497

(GLMbio7 <- glm(Biomass ~ NO3+TSS+H55+FR+WATR+PO4,data = Biodata,
               family = gaussian(link = "identity"))))
model.sel(GLMbio1, GLMbio2, GLMbio3, GLMbio4, GLMbio5, GLMbio6, GLMbio7) #7562143
GLMbio <-
as.data.frame(model.sel(GLMbio1, GLMbio2, GLMbio3, GLMbio4, GLMbio5, GLMbio6, GLMbio7)) #562143

# Performance result output -----
write.table(rbind(
  GLMsp[c("df", "logLik", "AICc", "delta", "weight")],
  GLMab[c("df", "logLik", "AICc", "delta", "weight")],
  GLMbio[c("df", "logLik", "AICc", "delta", "weight")] ), "GLMresult.txt")

# Other evaluation for competitive models (RSE, deviance, adjusted r2) ----
RSE <- as.data.frame(matrix(unlist(list(
  rse(GLMsp1), rse(GLMsp2), rse(GLMsp3), rse(GLMsp4), rse(GLMsp5),
  rse(GLMsp6),
  rse(GLMab1),
  rse(GLMab2), rse(GLMab3), rse(GLMab4), rse(GLMab5), rse(GLMab6), rse(GLMab7),
  rse(GLMbio1),
  rse(GLMbio2), rse(GLMbio3), rse(GLMbio4), rse(GLMbio5), rse(GLMbio6), rse(GLMbio7))),
  nrow = 20, byrow = T))
colnames(RSE) <- c("RSE")
rownames(RSE) <- c("GLMsp1", "GLMsp2", "GLMsp3", "GLMsp4", "GLMsp5",
  "GLMsp6", "GLMab1", "GLMab2", "GLMab3", "GLMab4",
  "GLMab5", "GLMab6", "GLMab7", "GLMbio1", "GLMbio2",
  "GLMbio3", "GLMbio4", "GLMbio5", "GLMbio6", "GLMbio7")
aR2 <- function(x){(rsq(x, adj = T))}
adjR2 <- as.data.frame(matrix(unlist(list(
  aR2(GLMsp1), aR2(GLMsp2), aR2(GLMsp3), aR2(GLMsp4), aR2(GLMsp5), aR2(GLMsp6),
  aR2(GLMab1), aR2(GLMab2), aR2(GLMab3), aR2(GLMab4), aR2(GLMab5), aR2(GLMab6),
  aR2(GLMab7), aR2(GLMbio1), aR2(GLMbio2), aR2(GLMbio3), aR2(GLMbio4), aR2(GLMbio5),
  aR2(GLMbio6), aR2(GLMbio7))),
  nrow = 20, byrow = T))
colnames(adjR2) <- c("adjR2")

save(SRdata, Abdata, Biodata, file = "LMM_input.RData")

```

Part IV – Linear Mixed Model

```

load("Input.RData")
load("LMM_input.RData")
#####
### Mixed Effects Model - lmer ###
#####
library(lme4)

```

```
library(MASS)
library(car)
library(MuMIn)

# 1. Species richness -----
# Richness ~ pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS (GLMsp5)
SRdata <- cbind(SRdata,Year,SB)

(LMMsp1 <- lmer(Richness ~ pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS+(1|Year)+
  (1|Subbasin), data = SRdata))
(LMMsp2 <- lmer(Richness ~ pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS+(1|Year),
  data = SRdata))
(LMMsp3 <- lmer(Richness ~
  pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS+(1|Subbasin),
  data = SRdata))
(LMMsp4 <- lmer(Richness ~
  pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS+(Subbasin|Year),
  data = SRdata))
(LMMsp5 <- lmer(Richness ~
  pH+DO+PO4+NPR+SO4+H2O+DEPTH+FR+WPAS+(1|Year:Subbasin),
  data = SRdata))
# Overdispersion checks -----
dfun <- function(x) {
  with(x,(sum(residuals(x)^2))/df.residual(x))
} # Ben Bolker, 2017
as.data.frame(matrix(unlist(list(
  dfun(LMMsp1),dfun(LMMsp2),dfun(LMMsp3),dfun(LMMsp4),dfun(LMMsp5))))))

# qAICc (because of overdispersion models)
QAICc(LMMsp1, LMMsp2,LMMsp3,LMMsp4,LMMsp5,
  chat = deviance(LMMsp1) / df.residual(LMMsp1))

# Evaluations -----
model.sel(LMMsp1,LMMsp2,LMMsp3,LMMsp4,LMMsp5)
#AICc = 12534 (125 have close AICc i.e. <2)
#qAICc = 25314 (253 have close qAICc i.e. <2)
BIC(LMMsp1,LMMsp2,LMMsp3,LMMsp4,LMMsp5) #BIC
# 25134 (251 have super close BIC i.e. <2)

# Marginal and conditional r2 (Nagakawa, 2013)
# Trigamma-estimate is recommended whenever available
# (Nagakawa et al., 2014; Johnson, 2014).
round(r.squaredGLMM(LMMsp2),3) #0.433 0.573
round(r.squaredGLMM(LMMsp5),3) #0.446 0.569
round(r.squaredGLMM(LMMsp3),3) #0.559 0.587
round(r.squaredGLMM(LMMsp1),3) #0.344 0.574
round(r.squaredGLMM(LMMsp4),3) #0.310 0.579

LMMSprCh <- LMMsp3

# 2. Overall abundance ----
# Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH
Abdata <- cbind(Abdata, Year, SB)

(LMMab1 <- lmer(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH+(1|Year)+
  (1|Subbasin), data = Abdata))
(LMMab2 <- lmer(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH+(1|Year),
  data = Abdata))
(LMMab3 <- lmer(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH+(1|Subbasin),
  data = Abdata))
(LMMab4 <- lmer(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH+(Subbasin|Year),
  data = Abdata))
(LMMab5 <- lmer(Abundance ~ DO+PO4+H36+H55+WATR+UR+WIDTH+(1|Year:Subbasin),
  data = Abdata))

model.sel(LMMab1,LMMab2,LMMab3,LMMab4,LMMab5) #AICc 53124
# 3 and 5 are very close (dAICc~2)
BIC(LMMab1,LMMab2,LMMab3,LMMab4,LMMab5) # BIC 53124
```

```
# 3 and 5 are very close (dBIC <2)

# Evaluation -----
# Marginal and conditional r2 (Nagakawa, 2013)
round(r.squaredGLMM(LMMab5),2) #0.48 0.61
round(r.squaredGLMM(LMMab3),2) #0.44 0.54

LMMAbun <- LMMab5 # Overall abundance final model

# 3. Overall biomass ----
# Biomass ~ NO3+TSS+H55+FR+WATR+PO4
Biodata <- cbind(Biodata, Year, SB)

(LMMbio1 <- lmer(Biomass ~ NO3+TSS+H55+FR+WATR+PO4+(1|Year)+
  (1|Subbasin), data = Biodata))
(LMMbio2 <- lmer(Biomass ~ NO3+TSS+H55+FR+WATR+PO4+(1|Year),
  data = Biodata))
(LMMbio3 <- lmer(Biomass ~ NO3+TSS+H55+FR+WATR+PO4+(1|Subbasin),
  data = Biodata))
(LMMbio4 <- lmer(Biomass ~ NO3+TSS+H55+FR+WATR+PO4+(Subbasin|Year),
  data = Biodata))
(LMMbio5 <- lmer(Biomass ~ NO3+TSS+H55+FR+WATR+PO4+(1|Year:Subbasin),
  data = Biodata))

model.sel(LMMbio1,LMMbio2,LMMbio3,LMMbio4,LMMbio5) #AICc 25314
BIC(LMMbio1,LMMbio2,LMMbio3,LMMbio4,LMMbio5) # BIC 25314

# Marginal and conditional r2 (Nagakawa, 2013)
round(r.squaredGLMM(LMMbio1),3) # 0.411 0.437
round(r.squaredGLMM(LMMbio2),3) # 0.411 0.437
round(r.squaredGLMM(LMMbio3),3) # 0.422 0.422
round(r.squaredGLMM(LMMbio4),3) # 0.43 0.483
round(r.squaredGLMM(LMMbio5),3) # 0.421 0.427

LMMBiom <- LMMbio2 # Overall Biomass final model

save(SRdata,Abdata, Biodata, file = "LMMdata.RData")
save(LMMSpRch,LMMAbun,LMMBiom, file = "LMM.RData")

Part V – Final model projection
load("Input.RData")
load("LMMdata.RData")
load("LMM.RData")

#####
#### Scenarios study (Projection) ####
#####
# Scen 1: 15% increment of PO4----
# creating scen1 data
scen1data <- Av
scen1data$PO4 <- Av$PO4*1.15

# Species richness ----
scen1SRdata <- as.data.frame(cbind(Resp.var$Richness,scen1data,SB,Year))
(scen1SR <- as.data.frame(predict(LMMSpRch, newdata = scen1SRdata,
  type = "response", allow.new.level=T)))
scen1SR <- cbind(scen1SR,Resp.var$Richness)
colnames(scen1SR) <- c("Predicted","Observed")
scen1SR$Site <- 1:118
#####
# Overall abundance ----
scen1ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen1data,SB,Year))
(scen1AB <- as.data.frame(predict(LMMAbun, newdata = scen1ABdata,
  type = "response", allow.new.level=T)))
scen1AB <- cbind(scen1AB,Resp.var$Abundance)
colnames(scen1AB) <- c("Predicted","Observed")
scen1AB$Site <- 1:118
#####
```

```
# Scen 2: 30% increment of P04----
# creating scen2 data
scen2data <- Av
scen2data$P04 <- Av$P04*1.3

# Species richness ----
scen2SRdata <- as.data.frame(cbind(Resp.var$Richness,scen2data,SB,Year))
(scen2SR <- as.data.frame(predict(LMMSpRch, newdata = scen2SRdata,
                                type = "response", allow.new.level=T)))
scen2SR <- cbind(scen2SR,Resp.var$Richness)
colnames(scen2SR) <- c("Predicted","Observed")
scen2SR$Site <- 1:118
#####
# Overall abundance ----
scen2ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen2data,SB,Year))
(scen2AB <- as.data.frame(predict(LMMABun, newdata = scen2ABdata,
                                type = "response", allow.new.level=T)))
scen2AB <- cbind(scen2AB,Resp.var$Abundance)
colnames(scen2AB) <- c("Predicted","Observed")
scen2AB$Site <- 1:118
#####

# Scen 3: 10% increase WPAS; 10% decrease NO3; 10% decrease P04;
# 10% decrease AGRL (Not available in variables)-----
scen3data <- Av
scen3data$WPAS <- Av$WPAS*1.1
scen3data$P04 <- Av$P04*0.9
scen3data$NO3 <- Av$NO3*0.9

# Species richness ----
scen3SRdata <- as.data.frame(cbind(Resp.var$Richness,scen3data,SB,Year))
(scen3SR <- as.data.frame(predict(LMMSpRch, newdata = scen3SRdata,
                                type = "response", allow.new.level=T)))
scen3SR <- cbind(scen3SR,Resp.var$Richness)
colnames(scen3SR) <- c("Predicted","Observed")
scen3SR$Site <- 1:118
#####
# Overall abundance ----
scen3ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen3data,SB,Year))
(scen3AB <- as.data.frame(predict(LMMABun, newdata = scen3ABdata,
                                type = "response", allow.new.level=T)))
scen3AB <- cbind(scen3AB,Resp.var$Abundance)
colnames(scen3AB) <- c("Predicted","Observed")
scen3AB$Site <- 1:118
#####

# Scen 4: 20% increase WPAS; 20% decrease NO3; 20% decrease P04;
# 10% decrease AGRL (Not available in variables);10% decrease WATR-----
scen4data <- Av
scen4data$WPAS <- Av$WPAS*1.2
scen4data$NO3 <- Av$NO3*0.8
scen4data$WATR <- Av$WATR*0.9

# Species richness ----
scen4SRdata <- as.data.frame(cbind(Resp.var$Richness,scen4data,SB,Year))
(scen4SR <- as.data.frame(predict(LMMSpRch, newdata = scen4SRdata,
                                type = "response", allow.new.level=T)))
scen4SR <- cbind(scen4SR,Resp.var$Richness)
colnames(scen4SR) <- c("Predicted","Observed")
scen4SR$Site <- 1:118
#####
# Overall abundance ----
scen4ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen4data,SB,Year))
(scen4AB <- as.data.frame(predict(LMMABun, newdata = scen4ABdata,
                                type = "response", allow.new.level=T)))
scen4AB <- cbind(scen4AB,Resp.var$Abundance)
colnames(scen4AB) <- c("Predicted","Observed")
scen4AB$Site <- 1:118
```

```
#####

# Scen 5: 5% reduction precipitation----
scen5data <- Av
scen5data$H20 <- Av$H20*0.95      # Weekly skewness
scen5data$H36 <- Av$H36*0.95      # Monthly skewness
scen5data$H55 <- Av$H55*0.95      # Rate of change - difference in flow
median

# Species richness ----
scen5SRdata <- as.data.frame(cbind(Resp.var$Richness,scen5data,SB,Year))
(scen5SR <- as.data.frame(predict(LMMSpRch, newdata = scen5SRdata,
                                type = "response", allow.new.level=T)))
scen5SR <- cbind(scen5SR,Resp.var$Richness)
colnames(scen5SR) <- c("Predicted","Observed")
scen5SR$Site <- 1:118
#####

# Overall abundance ----
scen5ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen5data,SB,Year))
(scen5AB <- as.data.frame(predict(LMMAbun, newdata = scen5ABdata,
                                type = "response", allow.new.level=T)))
scen5AB <- cbind(scen5AB,Resp.var$Abundance)
colnames(scen5AB) <- c("Predicted","Observed")
scen5AB$Site <- 1:118
#####

# Scen 6: 15% reduction precipitation----
scen6data <- Av
scen6data$H20 <- Av$H20*0.85      # Weekly skewness
scen6data$H36 <- Av$H36*0.85      # Monthly skewness
scen6data$H55 <- Av$H55*0.85      # Rate of change - difference in flow
median

# Species richness ----
scen6SRdata <- as.data.frame(cbind(Resp.var$Richness,scen6data,SB,Year))
(scen6SR <- as.data.frame(predict(LMMSpRch, newdata = scen6SRdata,
                                type = "response", allow.new.level=T)))
scen6SR <- cbind(scen6SR,Resp.var$Richness)
colnames(scen6SR) <- c("Predicted","Observed")
scen6SR$Site <- 1:118
#####

# Overall abundance ----
scen6ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen6data,SB,Year))
(scen6AB <- as.data.frame(predict(LMMAbun, newdata = scen6ABdata,
                                type = "response", allow.new.level=T)))
scen6AB <- cbind(scen6AB,Resp.var$Abundance)
colnames(scen6AB) <- c("Predicted","Observed")
scen6AB$Site <- 1:118
#####

# Scen 7: 5% increase precipitation----
scen7data <- Av
scen7data$H20 <- Av$H20*1.05      # Weekly skewness
scen7data$H36 <- Av$H36*1.05      # Monthly skewness
scen7data$H55 <- Av$H55*1.05      # Rate of change - difference in flow
median

# Species richness ----
scen7SRdata <- as.data.frame(cbind(Resp.var$Richness,scen7data,SB,Year))
(scen7SR <- as.data.frame(predict(LMMSpRch, newdata = scen7SRdata,
                                type = "response", allow.new.level=T)))
scen7SR <- cbind(scen7SR,Resp.var$Richness)
colnames(scen7SR) <- c("Predicted","Observed")
scen7SR$Site <- 1:118
#####

# Overall abundance ----
scen7ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen7data,SB,Year))
(scen7AB <- as.data.frame(predict(LMMAbun, newdata = scen7ABdata,
```



```

                                type = "response", allow.new.level=T)))
scen7AB <- cbind(scen7AB,Resp.var$Abundance)
colnames(scen7AB) <- c("Predicted","Observed")
scen7AB$Site <- 1:118
#####

# Scen 8: 15% increase precipitation-----
scen8data <- Av
scen8data$H20 <- Av$H20*1.15      # Weekly skewness
scen8data$H36 <- Av$H36*1.15      # Monthly skewness
scen8data$H55 <- Av$H55*1.15      # Rate of change - difference in flow
median

# Species richness ----
scen8SRdata <- as.data.frame(cbind(Resp.var$Richness,scen8data,SB,Year))
(scen8SR <- as.data.frame(predict(LMMSpRch, newdata = scen8SRdata,
                                type = "response", allow.new.level=T)))
scen8SR <- cbind(scen8SR,Resp.var$Richness)
colnames(scen8SR) <- c("Predicted","Observed")
scen8SR$Site <- 1:118
#####

# Overall abundance ----
scen8ABdata <- as.data.frame(cbind(Resp.var$Abundance,scen8data,SB,Year))
(scen8AB <- as.data.frame(predict(LMMAbun, newdata = scen8ABdata,
                                type = "response", allow.new.level=T)))
scen8AB <- cbind(scen8AB,Resp.var$Abundance)
colnames(scen8AB) <- c("Predicted","Observed")
scen8AB$Site <- 1:118
#####

#### 1. Statistical test for significant difference with observed data----
scenSR <- as.data.frame(cbind(
  Resp.var$Richness,scen1SR$Predicted, scen2SR$Predicted,scen3SR$Predicted,
  scen4SR$Predicted, scen5SR$Predicted,scen6SR$Predicted,
  scen7SR$Predicted,
  scen8SR$Predicted))
colnames(scenSR) <- c("Observed","Scen1","Scen2","Scen3","Scen4","Scen5",
  "Scen6","Scen7","Scen8")

scenAB <- as.data.frame(cbind(
  Resp.var$Abundance,scen1AB$Predicted,
  scen2AB$Predicted,scen3AB$Predicted,
  scen4AB$Predicted, scen5AB$Predicted,scen6AB$Predicted,
  scen7AB$Predicted,
  scen8AB$Predicted))
colnames(scenAB) <- c("Observed","Scen1","Scen2","Scen3","Scen4","Scen5",
  "Scen6","Scen7","Scen8")

### Wilcoxon test----
wtestSR <- sapply(
  lapply(scenSR[-1], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))
(wtestSR <- as.data.frame(t(wtestSR)))

wtestAB <- sapply(
  lapply(scenAB[-1], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
(testAB <- as.data.frame(t(wtestAB)))

### 2. Statistical test for significant difference with observed data
### (within subbasins)-----
### Wilcoxon test----
scenSR <- cbind(scenSR, SB)
wtest_Tr <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Tr'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

```

```
wtest_Sa <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Sa'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

wtest_Je <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Je'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

wtest_Bo <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Bo'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

wtest_Ki <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Ki'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

wtest_Ju <- sapply(
  lapply(scenSR[ which(scenSR$Subbasin == 'Ju'),2:9], function(x) {
    wilcox.test(x,scenSR$Observed, data = scenSR, conf.level = 0.95)}),
  `[`, c("p.value"))

write.table(cbind(wtest_Tr, wtest_Sa, wtest_Ki, wtest_Ju, wtest_Je,
  wtest_Bo), "wtest_SB.txt")

scenAB <- cbind(scenAB, SB)
wtest_Tr <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Tr'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
wtest_Sa <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Sa'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
wtest_Je <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Je'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
wtest_Bo <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Bo'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
wtest_Ki <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Ki'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))
wtest_Ju <- sapply(
  lapply(scenAB[ which(scenAB$Subbasin == 'Ju'),2:9], function(x) {
    wilcox.test(x,scenAB$Observed, data = scenAB, conf.level = 0.95)}),
  `[`, c("p.value"))

write.table(cbind(wtest_Tr, wtest_Sa, wtest_Ki, wtest_Ju, wtest_Je,
  wtest_Bo), "wtestAB_SB.txt")
#####

write.table(cbind(scenSR,SB),"scenSR.txt") # output of simulated results
write.table(cbind(scenAB,SB),"scenAB.txt")
```

List of Publications

Paper in peer-reviewed journals:

Naicheng Wu, Yueming Qu, Björn Guse, Kristė Makarevičiūtė, **Szewing To**, Tenna Riis and Nicola Fohrer. 2018. Hydrological and environmental variables outperform spatial factors in structuring species, trait composition and beta diversity of pelagic algae. *Ecology and Evolution* 8: 2947-2961.

Conference abstracts:

Yueming Qu, Naicheng Wu, Björn Guse, Kristė Makarevičiūtė, **Szewing To**, Tenna Riis and Nicola Fohrer. 2018. Einfluss von hydrologischen Bedingungen auf das Auftreten von pelagischen Algen. Tag der Hydrologie. March, 2018, Dresden, Deutschland.

Naicheng Wu, Yueming Qu, Björn Guse, Kristė Makarevičiūtė, **Szewing To**, Tenna Riis and Nicola Fohrer. 2017. HYDROLOGICAL, ENVIRONMENTAL AND SPATIAL VARIABLES INTERACT TO DETERMINE SPECIES, TRAIT COMPOSITION AND BETA DIVERSITY OF PELAGIC ALGAE. Society for Freshwater Science (SFS) Annual Meeting. June 4-8, 2017, Raleigh, USA.

Declarations

I, herewith, declare that this thesis has been completed independently and unaided and that no other sources other than the ones given here have been used. The submitted written version of this work is the same as the one electronically saved and submitted on CD. Furthermore, I declare that this work has never been submitted at any other time and anywhere else as a final thesis.

Date, Place, Signature

Szewing, TO (1113621)