CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

Department of Environmental Geosciences

Diploma Thesis

Estimating Optically Inactive Water Quality Parameters in the Upper Vltava Catchment, using Machine Learning on Satellite Imagery

Diploma Thesis

Thesis Supervisor: doc. Dr. Ing. Zhongbing Chen

Author: Jonas Niese

Prague

2024

Author's statement

I hereby declare that I have independently elaborated the diploma/final thesis with the topic of: Estimating Optically Inactive Water Quality Parameters in the Upper Vltava Catchment, using Machine Learning on Satellite Imagery and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources. I am aware that my diploma/final thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendment of some acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis. I am aware that by submitting the diploma/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of some acts, as amended, regardless of the result of its defence. With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

I declare that I have used AI tools in accordance with the university's internal regulations and principles of academic integrity and ethics. Appropriate references to the use of those tools have been made in the thesis.

The used AI-tools are: The large language models Chat-GPT 3.5 and Chat-GPT 4 for specific questions on programming (as stated in methodology chapter), and the free version of the translator DeepL for translating the abstract to the Czech language before having it proofread by a Czech native-speaker.

Prague, Czech Republic, 27.03.2024

Abstract (EN)

Management of river water quality is important for mitigating eutrophication. After the water quality improvements since the 1990s, this trend might have turned in the upper Vltava catchment after 2010: Concentrations of the four optically inactive water quality parameters ammonia nitrogen, nitrate nitrogen, five-day biochemical oxygen demand, and total phosphorus might have increased, as results presented in this thesis suggest. All four parameters follow a land-use and ecosystem degradation gradient in the upper Vltava, increasing in concentration from the headwaters to the lower parts of the research area.

In this study, five different types of algorithms (multiple stepwise linear regression, partial least squares regression, support vector regression, random forest regressor, and backpropagation artificial neural network) were performed on harmonized Landsat-Sentinel-2-data (HLS), in order to predict AN, NN, BOD5, and TP. Such machine learning methods can be useful tools for estimating the four named parameters, but they still bear a lot of challenges, especially given the often narrow streambed. Further improvements at all steps – from input data selection, over model optimization, to model selection – are required for ensuring accurate and robust predictions. The presented results indicate that this goal can be achieved. In most cases, machine learning algorithms clearly outperformed simpler linear models. Especially the method of the random forest regressor can often estimate a big fraction of variance, while also producing comparatively low errors.

Of all four water quality parameters, NN was most effectively-predicted, with $R^2 = 0.555$ by a random forest regressor, whereas AN is the most challenging with a maximum $R^2 = 0.156$, also by a random forest regressor. BOD5 and TP prediction also remains challenging, but some models returned relatively strong metrics, indicating great potential for sound predictions of these parameters as well.

Abstract (CZ)

Řízení kvality říční vody je důležité pro zmírnění eutrofizace. Po zlepšení kvality vody od 90. let 20. století se tento trend mohl v povodí horní Vltavy po roce 2010 obrátit: Koncentrace čtyř opticky neaktivních ukazatelů jakosti vody - amoniakálního dusíku, dusičnanového dusíku, pětidenní biochemické spotřeby kyslíku a celkového fosforu - se mohly zvýšit, jak naznačují výsledky uvedené v této práci. Všechny čtyři parametry sledují v horní Vltavě gradient využití půdy a degradace ekosystémů, přičemž jejich koncentrace se zvyšuje od pramenné oblasti k dolním částem výzkumného území.

V této studii bylo na harmonizovaných datech Landsat-Sentinel-2 (HLS) provedeno pět různých typů algoritmů (vícenásobná kroková lineární regrese, regrese částečných nejmenších čtverců, regrese s podpůrnými vektory, regresor náhodného lesa a umělá neuronová síť s zpětným šířením) za účelem předpovědi AN, NN, BOD5 a TP. Tyto metody strojového učení mohou být užitečnými nástroji pro odhad čtyř jmenovaných parametrů, ale stále s sebou nesou mnoho problémů, zejména vzhledem k často úzkému korytu toku. Pro zajištění přesných a spolehlivých předpovědí je nutné další zdokonalování ve všech krocích - od výběru vstupních dat, přes optimalizaci modelu až po výběr modelu. Prezentované výsledky naznačují, že tohoto cíle lze dosáhnout. Ve většině případů algoritmy strojového učení jednoznačně překonaly jednodušší lineární modely. Zejména metoda regresoru náhodného lesa dokáže často odhadnout velkou část rozptylu a zároveň produkuje relativně nízké chyby.

Ze všech čtyř parametrů kvality vody byl nejefektivněji předpovídán NN s R2 = 0,555 pomocí regresoru náhodného lesa, zatímco AN je nejnáročnější s maximálním R2 = 0,156, rovněž pomocí regresoru náhodného lesa. Predikce BSK5 a TP zůstává rovněž náročná, ale některé modely vrátily poměrně silné metriky, což naznačuje velký potenciál pro spolehlivé predikce i těchto parametrů.

Table of Contents

Table of Figures

Index of Tables

Table of Equations

1. Introduction:

1.1 The Challenge of Eutrophication

Eutrophication, especially of aquatic ecosystems, as a consequence of anthropogenic interferences with the phosphorus and nitrogen cycles is widely considered to be one of the great environmental challenges of the present (Carpenter et al., 1998; Carpenter and Bennett, 2011; Conley et al., 2009; Rockström et al., 2009; Smith et al., 2006). This is because it poses threats to ecosystem integrity and biodiversity loss (Maúre et al., 2021; Wang et al., 2021), safe water supplies, contributes to soil degradation (Vitousek et al., 1997), and even contributes to the release of the potent greenhouse gases nitrous oxide (Vitousek et al., 1997) and methane (Beaulieu et al., 2019). Coupled with climate change, eutrophication is also a main driver of oxygen depletion in freshwater and marine ecosystems (Brush et al., 2020; Foley et al., 2012; Pitcher et al., 2021). The global annual economic damages due to eutrophication amount multiple billion US-dollars, as research from China (Le et al., 2010), the United Kingdom (Pretty et al., 2003), and the United States (Dodds et al., 2009) shows.

N and P are mainly introduced to the environment via the application of artificial fertilizers used in agriculture, by wastewater, by atmospheric deposition, and by various urban activities (Carpenter et al., 1998; Selman and Greenhalgh, 2009). Another major nitrogen source is the combustion of fossil fuels (Gruber and Galloway, 2008; Selman and Greenhalgh, 2009). Phosphorus enters the environment also via the use and production of detergents (Mainstone and Parr, 2002). Overall, non-point sources (such as surface runoff from agriculturally used land) tend to be the more dominant pathways of eutrophication, while point sources (such as wastewater treatment plants) often play a lesser role (Kakade et al., 2021), although they are still of great importance, e.g., due to the high bioavailability of P from point sources (Mainstone and Parr, 2002).

This is also true for the Czech Republic, where the annual P surface runoff ranges between 0.1 and 9.98 $kg\,m^{-2}$ in the Vltava basin; for the upper Vltava, maximum values of 4.08 kg $km^{-2}\,year^{-1}$ have been reported (Rosendorf et al., 2016).

This P input is not considered to be a significant source of inland water eutrophication (Rosendorf et al., 2016), yet monitoring of the trophic state of the river remains important.

Nitrogen input into the Vltava is considered critical in 25% of the cases studies by Rosendorf et al. (2016). Evidence suggests that the trophic state is critical for the assemblage of fish communities in the Vltava River, with different nitrogen compounds and oxygen demand having a negative influence on fish diversity, while total phosphorus (TP) might support some isotopic niche area species (Horka et al., 2023).

On a global and regional scale, rivers play a special role in the context of eutrophication. Yet, research of nutrient and eutrophication studies on river has lagged behind, compared to lentic (i.e., lacustrine/standing water) and coastal and marine ecosystems (Smith, 2003).

By draining their catchments, rivers take up large amounts of various substances from the environment, including potential pollutants such as nitrogen and phosphorus compounds (Beusen et al., 2005; Houser and Richardson, 2010).

Subsequently, the rivers transport the accumulated compounds to the seas (Smith, 2003) and the oceans where they can cause further problems, such as promoting acidification (Cai et al., 2011), hypoxia and anoxia of coastal regions (Howarth et al., 2011), mainly due to harmful "blooms" of phytoplankton, usually cyanobacteria (Anderson et al., 2002). But eutrophication of lotic systems, i.e., of streams and rivers, is not only a concern because of nutrient transport, but also because of the detrimental effects it has on these water bodies themselves.

The formerly widespread belief according to which lotic systems are always nutrient-saturated an that accordingly any algal development is suppressed by light limitation and short water residence times has been refuted (Smith, 2003) by manifold observations of extensive river eutrophication (Köhler and Gelbrecht, 1998; Soana et al., 2024; Yan et al., 2019; Zhang et al., 2015). Additionally, the trophic state of a river is not only relevant for the ecology of the river itself, but also for that of all wetlands fed by it (Mainstone and Parr, 2002). Considering the underappreciated relevance of rivers in the context of eutrophication, researchers from the U.S. Geological survey recently demanded the prioritization of river basins for nutrient studies (Tesoriero et al., 2024).

Although it has been argued by scholars that only control of phosphorus was needed for tackling the detrimental effects of eutrophication (Schindler, 1977; Schindler et al., 2008), effective mitigation of nutrient pollution depends on controlling input of both nutrient elements, P and N (Basu et al., 2022; Conley et al., 2009). This is mainly because while P is indeed the main element causing eutrophication of inland waters (Schindler et al., 2008), N is considered to be the main cause of coastal and marine eutrophication, at least in most temperate systems (Howarth and Marino, 2006). Regardless of the implemented control measures, their effectivity and efficiency fundamentally depends on the available knowledge on nutrient pollution and surface water quality.

Such knowledge can be obtained by an extensive network of field sampling and sample analysis infrastructure. Despite the relevance of this approach, relying on it alone comes with three major downsides (Sagan et al., 2020): 1.) It is connected with high costs and efforts; 2.) The results can only represent the quality at the given point from which a sample was taken at the given time. But water quality parameters are not homogeneously distributed in space and time. Hence, knowledge about their distribution in the area is required. 3.) Not all potentially relevant points can be accessed easily and in a non-invasive way.

To address these requirements, remote sensing methods have been successfully applied for monitoring various water quality parameters since the 1970s (Anding and Kauth, 1970; Gholizadeh et al., 2016; Kondratyev et al., 1998; Topp et al., 2020; Zhu et al., 2022). However, since many water-quality parameters – including concentrations of dissolved phosphorus and of the inorganic nitrogen species in water – do not show straightforward interactions with the electromagnetic radiation emitted by the sun, common knowledge-driven approaches for estimating these parameters are often insufficient. Instead, data-driven approaches based on machine-learning algorithms (Reichstein et al., 2019) have proven to be suitable in such contexts (Dey and Vijay, 2021; Dong et al., 2023; Gao et al., 2024; Peterson et al., 2020; Pu et al., 2019; Sagan et al., 2020; Tan et al., 2023; Tian et al., 2024).

1.2 Introduction to Selected Water Quality Parameters in Rivers

Water quality parameters that tend not to directly influence the optical characteristics measured by remote-sensing (RS) instruments are called non-optically active water quality paramters (Guo et al., 2021; Sagan et al., 2020), many of which are directly linked to eutrophication: These parameters include the different forms of inorganic nitrogen, total phosphorus, and oxygen as dissolved oxygen and the (bio)chemical oxygen demand (e.g., Fu et al., 2022; Gao et al., 2024).

1.2.1 Reactive inorganic Nitrogen: Ammonia and Nitrate Nitrogen

Annual human production of reactive nitrogen (i.e., all N species except N_2) largely exceeds natural production (Razon, 2018; Xia et al., 2018). Reactive inorganic N is present in rivers mainly in the forms of ammonia nitrogen (AN) and nitrate nitrogen (NN).

Ammonia nitrogen, also known as ammoniacal nitrogen, includes both, N in unionized ammonia (NH_3), as well as in cationic ammonium (NH_4^+) (Boyd, 2015; Lin et al., 2019). In virtually all natural waters, ammonium is the dominant species in which ammonia N occurs due to a pH- and temperature dependent equilibrium (Lin et al., 2019). With increasing pH and temperature, the ammonia proportion gets higher (Boyd, 2015). In some rivers, NH_4^+ is considered the main pollutant (Xia et al., 2018).

Ammonia N is transformed to nitrate-N by nitrification, mostly under aerobic conditions (Canfield et al., 2010). Like most N transformations in riverine ecosystems, this process is generally mediated by bacteria (Canfield et al., 2010). Nitrification is usually a chemoautotrophic process, in which ammonium is oxidized to nitrate (NO_3^-), with nitrite (NO_2^-) as an intermediate, by nitrifying bacteria, using CO_2 as carbon source for cell generation (Vymazal, 2007). The oxidation of ammonium to NO_2^- is done by strictly aerobic chemolithotrophs (e.g., *Nitrosomonas*), whereas the second step (the oxidation of nitrite to nitrate) is executed by facultative chemolithotrophs (e.g., *Nitrobacter*) that use nitrite and organic compounds as energy sources for growth.

Due to its comparatively high thermodynamic stability under common environmental conditions, nitrate is the most abundant form of nitrogen in rivers, usually accounting for more than 80% of dissolved inorganic nitrogen (DIN) (Meybeck, 1982; Shuiwang et al., 2000; Xia et al., 2018).

Little is known about how and to which extent rivers contribute to nitrogen loss at different spatial and temporal scales (Boyer and Howarth, 2002) and about other nitrogen transformations in rivers and streams (Xia et al., 2018). Known hotspots of the different N transformation processes include sediment-water interfaces, riparian zones and even interfaces between suspended particles and water (Xia et al., 2018). Also dams can strongly influence river nitrogen transformations (Akbarzadeh et al., 2019).

While having well-documented adverse effects on biodiversity and ecosystem integrity (Baxter, 1977; Wu et al., 2019), dams and impoundments act as net nitrogen sinks on a global scale, as denitrification and nitrogen burial usually exceed N fixation (Akbarzadeh et al., 2019).

1.2.2 Total Phosphorus

In most rivers, the limiting major nutrient to plant growth is phosphorus, which can be indicated by the N/P ratio (Mainstone and Parr, 2002).

Phosphorus concentrations and dynamics in all riverine ecosystems are largely determined by P loads of tributaries, catchment geology and (hydro)geochemistry, climate, channel hydrology and morphology, land use and environmental change (Records et al., 2016; Withers and Jarvie, 2008). Artificially enhanced loads of P pose a risk to riverine ecosystems, as P enrichment can shift competitive balances between primary producers, including both, phytoplankton and higher plants (Mainstone and Parr, 2002).

Increased P levels in faster flowing rivers (such as the upper Vltava) promote increased growth rates of epiphytic diatoms and periphyton, like green algae, while leading to a decrease of shallow-rooted submerged plants, whereas in slower-moving riverine systems, higher P concentrations lead to growth of phytoplankton along the water column, increasing light competition (Withers and Jarvie, 2008). The detrimental direct effects of P-induced eutrophication are generally considered to be lower in riverine ecosystems than in lakes and reservoirs, as the production of suspended algae per mass unit of P is significantly lower in rivers (Smith, 2003; Soballe and Kimmel, 1987).

Elevated riverine P concentrations in sediments and the water column can in principle affect riverine plant communities in four different ways (Mainstone and Parr, 2002):

1. Increasing plant growth rates, and thus standing stock.

2. Promoting higher plant species adapted to higher nutrient levels and thus altering ecological composition/balance.

3. Promoting epiphytic, epibenthic, filamentous and planktonic algae, thus decreasing light availability for higher plants.

4. Reducing root depth, making plants more sensitive to being ripped out by currents.

Throughout the growing season, continuous point sources of P (mainly sewage treatment plants) can be highly important for bioavailable P in rivers (Mainstone and Parr, 2002). It is thus important to reduce P flows from these sources during the growing season and monitor their dynamics throughout the year. When point sources of P are under control, diffuse sources gain relevance for the trophic state of any water body. Diffuse P sources (mainly agricultural runoff) contribute substantially to P levels in river sediments, where they can be used by benthic algae, and by rooted plants. The highest loads of P from diffuse sources occur when animal excretions or soluble fertilizers are washed off.

While acknowledging the relevance of P in riverine sediments, the analysis presented in this thesis can only focus on the TP concentration in the water column.

1.2.3 Five-day biochemical oxygen Demand

Higher concentrations of N and P can contribute to higher microbial pollution; in turn, the biochemical oxygen demand (BOD) can be enhanced by microbial pollution. This measure is therefore also critical for water quality assessment and control. It is a measure of the amount of dissolved oxygen in a given water sample that microorganisms consume in a certain amount of time (Boyd, 2015). Most commonly, BOD is determined for a five-day-period (BOD_5). This parameter is estimated in a bioassay procedure, where samples are placed in a constant temperature chamber at $20\pm1°C$ for five days. At the beginning and the end of the period, dissolved oxygen is instrumentally measured (Delzer and McKenzie, 2003). A detailed description of the method is given by Delzer and McKenzie (2003). Pristine rivers usually have BOD5 concentrations below 1 mg L^{-1}. Concentrations between 2 and 8 mg L^{-1} indicate medium pollution; higher values are considered severe pollution (Li and Liu, 2019).

As BOD represents the amount of dissolved oxygen (DO) in water, higher BOD levels reduce DO concentrations (Penn et al., 2009; Whitehead et al., 2009). BOD5 levels in water bodies are projected to increase globally, due to climate change (Whitehead et al., 2009).

1.3 State of the Art: Utilizing Remote Sensing and Machine Learning for Water Quality Prediction

Water management in general, and specifically eutrophication mitigation, usually depends on environmental models (Records et al., 2016). Traditional approaches largely depend on extensive field measurements that are costly, tedious, labor-intensive, and strongly limited in spatial and temporal resolution (Ahmed et al., 2022; Gao et al., 2024; Guo et al., 2021; Li et al., 2022).

Utilizing remote sensing data for water-quality prediction and modeling can be a remedy to these shortcomings that has been applied since the 1970s (Gao et al., 2024; Wrigley and Horne, 1974). However, using conventional, knowledge-based linear and nonlinear modeling approaches on satellite imagery is not suitable for many – especially optically inactive – water-quality parameters: They usually fail to model the complex nonlinear relationships between the satellite data and the concerning water quality parameters (Niu et al., 2021).

Recent advances in data-driven methodologies can help limiting this issue: Machine-learning approaches have proven to outperform classical modeling strategies and can often give robust estimates of different water quality parameters. Remote sensing-based methods are well-established for the estimation and prediction of optically active water quality parameters, such as chlorophyll-a concentration, total suspended solids, colored dissolved organic matter or turbidity (Sagan et al., 2020). This also reflects in the amount of scientific publications using machine learning techniques to estimate these optically active parameters from remote sensing data (Kupssinsku et al., 2019; Leggesse et al., 2023; Magri et al., 2023; Maier and Keller, 2018; Ruescas et al., 2018; Singhal et al., 2019).

Research on machine learning algorithms being used for estimating optically non-active water quality parameters from remote sensing data is still relatively scarce, but the body of highly promising published literature is steadily growing (e.g., Gao et al., 2024; Guo et al., 2021; Li et al., 2022; Niu et al., 2021; Peterson et al., 2020; Sagan et al., 2020).

More literature has been published on lakes than on rivers. Research focusing on rivers, mostly tries to estimate water quality parameters in the lower reaches, as due to their greater width they are generally easier to monitor with the available spatial solutions of remote sensing satellites. In the following, results of relevant literature are presented, including assessment of model performances on the testing dataset, based on the coefficient of determination (R^2) and, where possible, the root-mean-square error (RMSE) in mg L^{-1}.

Sagan et al. (2020) reviewed and analyzed recent advances in ML methodology using remote sensing for water quality prediction. The authors further analyzed 200 water quality datasets, showing great potentials of ML and RS for water quality prediction of optically active parameters, but challenges for inactive parameters. The study presents various different ML models for predicting a multitude of optically active and inactive water quality parameters, the latter including nitrate nitrogen and phosphate phosphorus (PO4-P). Some of the algorithm types are also used in the present work. The models by Sagan et al. use data from 9 sapling locations in a river and 34 points in lakes in Central Illinois, USA. Partial least squares regression (PLSR) was unsuitable for estimating PO4-P (R^2 = 0.02; RMSE = 0.25), while turning out quite useful for estimating NN (R^2 = 0.26; RMSE = 0.32). Support vector regression (SVR) performed slightly better (PO4-P: R^2 = 0.00; RMSE = 0.25. NN: R^2 = 0.29; RMSE = 0.31). A deep neural network approach also performed similarly for PO4-P (R^2 = 0.00; RMSE = 0.25) and NN (R^2 = 0.29; RMSE = 0.31).

Another study by the same research group utilized surface reflectance data from a virtual constellation of Landsat-8 and Sentinel-2 for constructing ML models in order to estimate various optically active (e.g., fluorescent dissolved organic matter fDOM) and inactive (DO, electric conductivity) water quality parameters (Peterson et al., 2020). The study was performed in four lakes within the Great Rivers Ecological Observatory (GREON) in the USA. Model performance for optically active parameters was generally excellent: A deep neural network model estimated the concentration of fDOM with R^2 = 0.926 and RMSE (µg L^{-1}) = 0.863. A DNN algorithm performed also very well for DO estimation with R^2 = 0.894 (RMSE (mg L^{-1}) = 1.806 mg L^{-1}). Relatively strong performances for DO estimation were also shown by SVR (R^2 = 0.805; RMSE = 2.141) and multiple linear regression (R^2 = 0.440; RMSE = 3.309) algorithms.

Despite minor difficulties of the models to predict DO concentrations in the higher ranges, the algorithms presented in the study are of great value for predicting this non-optically active measure of water quality.

Relatively good model performances for estimating non-optically active water quality parameters have been reported in a recent paper by Gao et al. (2024) who used machine learning imagery on Sentinel-2 imagery in various water bodies in China (Zhejiang province, West Lake, Xin'an River, Fuchun River, Lan River, Changtan Reservoir).

SVR was relatively weak on TP prediction ($R^2 = 0.10$; RMSE = 0.057), while showing decent results for total nitrogen (TN. $R^2 = 0.31$; RMSE = 1.164) and DO ($R^2 = 0.36$; RMSE = 1.99). It was outperformed by random forest regressors for TP ($R^2 = 0.39$; RMSE = 0.047) and TN ($R^2 = 0.42$; RMSE = 1.068), but not for DO ($R^2 = 0.34$; RMSE = 1.02). K-nearest neighbour neural networks (KNN) also quite good performance for TP ($R^2 = 0.35$; RMSE = 0.048), TN ($R^2 = 0.33$: RMSE = 1.151), and DO ($R^2 = 0.35$; RMSE = 2.00). While the authors themselves judge the model performances for DO and TP as relatively "poor", they still conclude that the obtained models are helpful for obtaining water quality information and can be a relevant reference for water management in the Zhejiang Province.

Another study from China, estimating non-optically active water quality parameters in the tropical Nandu River, reported very strong performances of some algorithms for estimating TN, AN, and TP from Landsat-8 imagery (Li et al., 2022). 25 Landsat images were used and 67 water samples analyzed. TN was predicted with $R^2 = 0.2$ by SVR (RMSE =), $R^2 = 0.49$ by RFR, and $R^2 = 0.45$ by an artificial neural network (ANN). AN estimates were substantially less accurate with $R^2 = 0.07$ for SVR, $R^2 = 0.24$ for RFR, and $R^2 = 0.44$ for ANN. TP estimates were very strong for SVR ($R^2 = 0.59$) and ANN ($R^2 = 0.67$), and still quite useful for RFR ($R^2 = 0.21$). It is remarkable that RFR was outperformed by simple regression trees for TP ($R^2 = 0.24$). Yet, it performed a lot worse for the other investigated parameters. The clearly study shows the suitability of machine learning algorithms for monitoring water quality at in the tropical river at relatively low cost, without depending on a deep understanding of the underlying environmental processes.

Water bodies in China are arguably the geographical focus of most published research using ML methodologies for estimating optically inactive water quality parameters from

RS imagery. Guo et al. (2021) applied machine learning algorithms to Sentinel-2 imagery in order to obtain information on TP, and TN concentrations and chemical oxygen demand (COD) in an urban lake in an industrial park in Tianjin, China. The authors compared the performances of SVR, RFR, and NN models. For each target parameter, a different model type gave the best estimates. Some models showed excellent performances: TP was estimated with = 0.94, with a root mean square percentage error (RMPSE) of 16.80% by a back-propagation ANN.

An RFR algorithm gave the best predictions for TN (R^2 = 0.88; RMPSE = 18.39%). COD was best predicted by SVR (R^2 = 0.86; RMPSE = 18.75%). Compared to the ML algorithms, multiple linear regression performed quite poorly. While having substantially lower – but still considerably high – R^2 values, the errors of the linear models are a lot higher for the three parameters, TP (R^2 = 0.65; RMPSE = 30.65), TN (R^2 = 0.76; RMPSE = 36.24), and COD (R^2 = 0.65; RMPSE = 71.65).

The outstanding performances of the models by Guo et al. (2021) in comparison to the previously discussed studies can likely be attributed to the following factors: Water quality in lentic systems is a lot more stable than in lotic systems (Schwoerbel and Brendelberger, 2022). Samplings were performed within four hours of a satellite overpass, when there was no cloud cover above the lake and no considerable plant cover on the lake surface. Furthermore, the model predictions might have profited from the advantages of Sentinel-2 data over other similar products, such as its relatively high spatial, spectral and temporal resolution. The quasi-ideal setup of the study by Guo et al. (2021) is not realistic for water-quality monitoring in most water bodies, let alone in near-natural systems.

The described studies demonstrate the suitability of combining remote sensing and machine learning methods for estimating different optically inactive water quality parameters in inland water bodies. They also reflect on current challenges and limitations of such methodologies. The still relatively small amount of relevant publications points to a vast, under-explored area for scientific inquiry, with the potential to substantially improve water quality management.

At the time of writing the present thesis, no RS-based ML models with optically inactive target parameters in European inland waters have been found. Moreover, the upper reaches of river networks have received only little scientific attention in this context.

The reasons for that can only be speculated about, but it might be due to the challenges of faster flow and thus higher transport rates of dissolved and suspended matter, due to the challenges of narrow streambeds, often covering only fraction of satellite pixels, and due to lower interest because of smaller amounts of water and a lower potential for accumulation of pollutants (Schwoerbel and Brendelberger, 2022) than downstream.

Anyhow, it is still highly desirable to have robust models for estimating water quality – including optically inactive parameters – in headwaters and upper reaches of rivers, due to their pivotal importance for biogeochemical connectivity between terrestrial and aquatic ecosystems and their control function for transporting nutrients and pollutants downstream (Withers and Jarvie, 2008). From this virtually unilateral relationship between upper and lower reaches of rivers, it follows that upstream pollution mitigation and conservation measures can automatically result in protection of downstream environments.

1.4 Identifying Appropriate Remote Sensing Data

The performance of statistical models largely depends on the quality of the used data. Besides using suitable and reliable water quality data, it is critical to select the most suitable remote sensing data product.

Two fundamentally different categories of RS data are available for the purpose of water quality estimations: Airborne and satellite-based data (Richards, 2013). Airborne remote sensing methodologies can be useful for water quality analysis and have been successfully used as input data for ML-based estimation of non-optically active water quality parameters (e.g., Niu et al., 2021). They are highly adaptable to the purposes of individual analyses and come with the great advantage that detrimental atmospheric effects on the data quality are reduced. Yet, airborne data come with multiple major disadvantages, making them unsuitable for the scopes of the present thesis. Unlike satellite-derived data, they are very costly, resource-demanding, and often labor-intensive to obtain. In contrast, satellite-based RS data are readily available in good temporal resolution, covering the entire earth surface over a long period of time (Myers and Miller, 2005).

Both described ways of data acquisition can rely on active or passive methods, where passive refers to the measurement of radiation emitted or reflected by the observed objects, and active to the measurement of time delay between actively emitted radiation back-scattered or reflected by the observed objects (e.g., LiDAR). Passive remote sensing systems have some distinctive advantages over active ones, making them more suitable for most river water quality monitoring tasks: Passively obtained RS data are available in comparatively high spatial and temporal resolutions.

Their reliance on reflection of sunlight further allows to obtain spectral fingerprints that can be related to different aspects of water quality.

As expounded, passively obtained data from earth observation satellites are most suitable for the aims of the thesis. Suitable satellites include Landsat-5, Landsat-8, and the Sentinel-2 satellite pair, for which different data products are available.

1.4.1 Landsat (LS)

Landsat-5 (LS-5) was a remote sensing satellite, launched by NASA, orbiting earth from March 1984 until June 2013 (Riebeek, Holly, 2013). A surface-reflectance dataset (USGS Landsat 5 Level 2, Collection 2, Tier 1) is offered by the USGS (Crawford et al., 2023) and readily available for Google Earth Engine, covering the period from 16. March 1984 to 05. May 2012 (Google, 2024a). Landsat-5 was succeeded by Landsat-8 (LS-8), which entered service on 30. March 2013. Like its predecessor, it has a spatial resolution of 30 m. LS-8 was constructed in a way that allows continuity from LS-5 by ensuring sufficient consistency of geometry, calibration, and spectral band properties (USGS, 2012).

1.4.2 Sentinel (S)

The Sentinel-2 (S-2) earth observation satellite pair was launched by the Copernicus Programme of the European Space Agency 23. June 2015 (Sentinel-2 A) and 7. March 2017 (Sentinel-2 B). S-2 has a spatial resolution of 10 m on three bands – nine times higher than that of LS-5 and LS-8. Most remaining bands resolve at 20 m. The spectral resolution and the amount of spectral bands are also higher than those of the corresponding LS satellites.

Due to the constellation as a satellite pair, and the orbit of the satellites being adapted to Europe, S-2 has a revisit time of less than five days in Northern mid latitudes (Claverie et al., 2018).

1.4.3 Harmonized Landsat Sentinel-2 Surface Reflectance Data

The harmonized Landsat and Sentinel-2 (HLS) surface reflectance dataset is a product offered by the NASA, providing virtual constellation data LS-8 and S-2 (Claverie et al., 2018). The data product HLS L30 offers data with 1. atmospheric correction, 2. geometric resampling and geographic registration, and 3. Bidirectional Reflectance Distribution Function normalization (BRDF) (Claverie et al., 2018).

This adjustment follows the methodology described by Roy et al. (2017, 2016) and aims to normalize the view angle effects, with the sun in zenith angle largely intact (Claverie et al., 2018; Ju et al., 2020). The spatial resolution and pixel geometry of the Sentinel-2 data is adjusted to that of the Landsat-8 data, necessarily coming with a downward correction of the spatial resolution of many bands of the Sentinel-2 data to 30 m (Claverie et al., 2018). The virtual constellation of the satellites allows a mean revisit time of as little as ~3.5 days in mid latitudes (Claverie et al., 2018).

While S-2 has great technological advantages due to its comparatively great spatial and spectral resolution, it cannot compete with the temporal availability and resolution of the HLS L30 product, which is thus used for the analysis presented and discussed in further chapters.

1.5 Relevant Machine Learning Problems and Algorithms

As elaborated, conventional statistical methods usually fail to capture the highly nonlinear relationship between non-optically active water quality parameters and surface reflectance of water bodies, as measured by satellites. Machine learning algorithms, in contrast, are often able to account for those nuances, leading us to the questions: What is ML and what specifically differs such algorithms from conventional ones?

A useful working definition, helping to answer both questions, is given by El Naqa and Murphy (2015, p. 4):

"A machine learning algorithm is a computational process that uses input data to achieve a desired task without being literally programmed (i.e., "hard coded") to produce a particular outcome. These algorithms are in a sense "soft coded" in that they automatically alter or adapt their architecture through repetition (i.e., experience) so that they become better and better at achieving the desired task."

Being a branch of Artificial Intelligence (AI), a central problem of ML algorithms is one, known as the fundamental credit assignment problem. It can be paraphrased as: Which modifiable components of a learning system are responsible for its success or failure? (Minsky, 1961; Schmidhuber, 2015).

ML algorithms should aim for providing the best possible answer to such questions. This, however, is difficult, due to the great amount of learning parts and their often in themselves complex interrelations. Closely related to the fundamental assignment problem is the so-called curse of dimensionality.

1.5.1 Overfitting and Dimensionality Reduction

The course of dimensionality was famously introduced by Bellman (1984, p. ix):

"In the first place, the effective analytic solution of a large number of even simple equations as, for example, linear equations, is a difficult affair. Lowering our sights, even a computational solution usually has a number of difficulties of both gross and subtle nature. Consequently, the determination of this maximum is quite definitely not routine when the number of variables is large. All this may be subsumed under the heading 'the curse of dimensionality.'"

In simpler terms, this means that a high number of input features can cause gross and subtle problems for solving even simple mathematical equations in a modeling problem (Bellman, 1984). While additional dimensions can add valuable information to an investigated space of variables, they also add additional complexity to the space, thus making accurate representation in a model more difficult. Hence, it is important to adapt the number of model inputs to the number and quality of observations, and to engineer and select suitable feature combinations.

The challenges posed by the curse of dimensionality make it relevant to reflect on some aspects of dimensionality reduction, feature engineering, and other preparation steps for constructing the ML models.

1.5.2 Feature Selection and Regularization

One of the most important considerations in ML modeling is: Which variables are relevant for achieving the goal of good prediction (De Mol et al., 2009)? Various algorithms are generally used to achieve this goal of feature selection, including LASSO (Least Absolute Shrinkage and Selection Operator) regression (*L1* regularization), ridge regression (*L2* regularization), and Elastic Nets which are mathematical combinations of the previously mentioned algorithms (De Mol et al., 2009; Zou and Hastie, 2005).

Another problem addressed by many of these algorithms is that of regularization. Like feature selection, regularization serves the purpose of dimensionality reduction. Regularization reduces variance by shrinking the estimated coefficients towards zero, relative to the least square estimates (James et al., 2023). A model with all p parameters is fitted in these shrinkage approaches. Some of the input coefficients may be estimated to be zero, depending on the type of regularization, and therefore shrinkage methods can also function as methods for variable selection (James et al., 2023). *L1* regularization penalizes regression coefficients proportional to their absolute values . Doing so, the LASSO algorithm often shrinks coefficients to exactly zero and can thus be effective for feature selection, especially in high-dimensional feature spaces (Tibshirani, 1996). The L_2 ridge penalty encourages the sum of the squares of the parameters to be small (towards but not exactly 0) by penalizing regression coefficients proportional to their squared values (Ng, 2004). It can further be useful to balance penalties from L1 and L2 in an elastic net, resulting in a tradeoff between their benefits (De Mol et al., 2009). L1, L2, and combined regularization can directly be implemented into neural networks and partial least square regression (Pedregosa et al., 2011; TensorFlow Developers, 2024).

Before regularization, other preparatory steps should be considered. Techniques of normalization (transformation of data to specific range, e.g., 0-1 or –1-1) and standardization (e.g., Z-score standardization: transformation to dataset with mean = 0 and standard deviation = 1) are powerful tools for equalizing feature importance and

increasing model performance (Peshawa and Rezhna, 2014; Santos and Papa, 2022). Combined with other techniques, they can also reduce the risk of overfitting (Santos and Papa, 2022).

1.5.3 Further Model Optimizations

In order to avoid overfitting, the initial dataset is split into a training dataset including 80% of the data, and a test dataset containing the remaining 20%. This is done for all models. However, adapting modifiable components of the model until the best fit on the test dataset is reached still poses a risk of overfitting (Pedregosa et al., 2011). A third partition of the initial dataset, known as validation set, can help to mitigate this so-called leak of data, which comes with the downside of reducing the number of observations in training or test set (Pedregosa et al., 2011).

Therefore, the approach of *k*-fold cross-validation is used for all models constructed for the present thesis. In this procedure, the training set is split into *k* smaller sets of equal size ("folds"). $k - 1$ of the k folds are then used on the training dataset. The remaining subsample is used for calculating an error (Rodríguez et al., 2010). Other types of cross-validation are also commonly used (Yadav and Shukla, 2016) but not further considered here.

For some types of algorithms, including partial least squares regression and the random forest regressor, a different approach called bootstrapping can be beneficial (Egbert and Plonsky, 2020; Ljumović and Klar, 2015). This procedure is based on repeated random sampling with replacement from the original dataset (Egbert and Plonsky, 2020). Appropriate models can then be trained using these bootstrap-samples (Mooney and Duval, 2006). The approach is considered especially useful for small and non-normally distributed samples . For random forest algorithms, oftentimes similar generalization errors can be expected from cross-validation and bootstrapping approaches (Ljumović and Klar, 2015).

Combining cross-validation and bootstrapping algorithms can be a powerful approach (Kohavi, 1995; Tibshirani, 1996; Tsamardinos et al., 2018). This could for instance be achieved by performing cross-validation on further divided bootstrap samples drawn from the original dataset.

The combination allows evaluation of model stability and uncertainty while also contributing to hyperparameter optimization (Tsamardinos et al., 2018).

Hyperparameters are model parameters that are not derived from the input data, but additional components of the model itself (Bakhteev and Strijov, 2020). such as the kernel function of SVR or weights in artificial neural networks or other algorithms. The predictive power of an ML model can be largely enhanced by identifying optimal values of these parameters. Their adaptation with that goal is called hyperparameter optimization or hyperparameter tuning (Bardenet et al., 2013).

1.5.4 Multiple stepwise linear regression (MSLR)

Multiple stepwise linear regression is a linear modeling technique for estimating the relationship between multiple input features and one dependent variable.

There are two different approaches in MSLR: a) forward selection and b) backward elimination (Breaux, 1967). In the more commonly used forward selection method, predictors are iteratively added to an initially empty model, beginning with the one most correlated with the target parameter (Breaux, 1967; Zhan et al., 2013). This procedure continues until new feature addition does not significantly improve the model performance. Backward elimination in contrast, starts with all input features included in the model, and is followed by iterative variable removal, until the model does not further improve its performance (Breaux, 1967). Different criteria (e.g., p-value, Akaike Information Criterion AIC) can be used for predictor selection and performance comparison between the models (Keith, 2019). Analogous to other model types, the performance of the final model should be evaluated based on criteria such as the goodness of fit. It is recommendable to use cross-validation to mitigate the risk of overfitting (Keith, 2019).

Multiple linear regression algorithms can come with some serious constraints for various applications in the environmental sciences and have been criticized for being unsuitable in some cases due to poor performances (Grossman et al., 1996). Caution is imperative when using such models . Despite this, many cases, they have been applied successfully for environmental modeling problems (Liu et al., 2021). useful and widely applied for comparing performances of ML algorithms to linear models (e.g., Guo et al., 2021; Latt and Wittenberg, 2014).

1.5.5 Partial least squares regression (PLSR)

In partial least squares regression algorithms are based on constructing so-called latent variables, also known as components, capturing the maximum variance in predictors and response variables (Geladi and Kowalski, 1986; Tobias, 1995). The components serve the purpose of obtaining a low-dimensional predictor space by maximizing the covariance between predictors and a response variable (Abdi, 2010; Pedregosa et al., 2011). This approach is similar to that of PCA, which aims to maximize the variance only within the space of the predictors (Pedregosa et al., 2011). Iteratively the components explaining the greatest covariance are extracted, while adding a new orthogonal component at each iteration (Abdi, 2010). Predictors are weighted in a way that maximizes the covariance between input and output variables. Cross-validation is commonly performed to reduce overfitting.

Due to its approach of component of constructing components and thus increasing covariance, PLSR can handle great degrees of multicollinearity among the input features (Abdi, 2010; Pedregosa et al., 2011). PLSR can be performed on multiple dependent variables simultaneously. It is a powerful, flexible and versatile transformer and regressor for many learning problems involving. Unlike MSLR, PLSR is often considered a machine learning algorithm, despite its roots in linear regression. PLSR is a supervised algorithm, meaning that the model is trained on a training dataset, meaning each input data point is associated with a corresponding target output (Berry et al., 2020).

1.5.6 Support Vector Regression (SVR)

The support vector regressor is a type of supervised ML algorithms, based on the support vector machine (SVM) classifier, applying its logic to regression problems (Pedregosa et al., 2011). SVR is based on computing linear regression in a higher dimensional feature space and mapping the input data via a nonlinear function (Basak et al., 2007). By this linear combination of weighted input features, the so-called hyperplane is constructed. The data points closest to this plane are called support-vectors (SVs). Using only this SV-subset of the training data for fitting the hyperplane, makes SVR algorithms comparably storage efficient.

The margin of the hyperplane to the support vectors is maintained as large as possible, while maintaining the lowest possible generalization error (Pedregosa et al., 2011). A regularization parameter C controls the trade-off between error minimization and margin maximization (Smola and Schölkopf, 2004). The performance of SVR algorithms depends largely on the selected kernel function (Pedregosa et al., 2011). The kernel function can allow capturing non-linear relationships between predictors and target variables. Common kernel functions include linear, polynomial, sigmoid tanh, and radial basis function (rbf).

SVMs for classification and regression problems are effective in high dimensional spaces (Burges, 1998; Pedregosa et al., 2011). This holds true in cases where the number of dimensions is greater than the sample number. Probability estimates can be calculated using five-fold cross validation (Pedregosa et al., 2011).

1.5.7 Random Forest Regressor

The random forest regressor is based on a type of decision trees, called regression tree. Like SVMs, decision trees can be constructed for classification as well as for regression problems. The logic behind regression trees is hierarchical splitting of the input space into ever narrower regions. Tree growth is limited by a maximum depth, defined by a stopping criterion. Nodes that are not followed by further split are called leaves, each of which represents a prediction value of the output parameter,given the previously followed path of splits.

Growing a random forest (RF) means generating a large number of decision trees from random subsamples (Breiman, 2001). RFs can be applied for classification and regression problems. RFRs are based on trees depending on a random vector Θ so that the tree predictor $h(x, \Theta)$ takes on numerical values instead of categorical ones/classes. Predictions of random forest classifiers are based on "votes" for the most popular class or in the case of RFR for the average metric output. Feature selection in RF algorithms is performed randomly, mitigation intercorrelation between the individual trees (Biau and Scornet, 2016). Random forests can also be useful for feature selection in other models, as they can produce information on feature relevance. Due to the averaging of a great number of trees, random forests are generally not very susceptible to overfitting (Breiman, 2001).

However, it is still advantageous to perform cross-validation or bootstrapping resampling techniques for further reducing the risk of overfitting (Ljumović and Klar, 2015).

RFR is a powerful approach for obtaining robust estimates of metric target variables with the ability to capture nonlinear relationships between input and output features. The predictive power of RFRs can come at the cost of high computational complexity (Hassine et al., 2019). Moreover, like most ML algorithms, random forests are characterized by their limited interpretability (Carvalho et al., 2019; Sagi and Rokach, 2020).

1.5.8 Backpropagation Artificial Neural Network (BP-ANN)

Artificial neural networks (ANNs) are computational models inspired by the human brain (Zou et al., 2009). ANNs are composed of nodes, called neurons that ar arranged in layers:

> "A standard neural network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations.
>
> Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons" (Schmidhuber, 2015, p. 86).

Each individual neuron communicates with neurons in the adjacent layers, but never within its own layer. Between the input and output layer, there are hidden layers of mutually independent size, meaning that specific layers can have any arbitrary number of neurons (Abiodun et al., 2018). The neurons of the hidden layers process their respective input information by assigning weights and applying an activation function to produce their output that is then forwarded to the following layer (Schmidhuber, 2015). Some widespread activation unctions are sigmoid, tanh, Rectified Linear Unit (ReLu), and softmax (Ramachandran et al., 2017; Sharma et al., 2017). The subsequent processing of information propagated from the input layer over the hidden layers to the output layer is known as feedforward operation (Schmidhuber, 2015). ANNs are trained by adjustnig weights and biases of neurons to optimize the output. For this purpose, different gradient descent algorithms are often used (Andrychowicz et al., 2016). Backpropagation (BP) is a common gradient descent method in supervised learning that has been used in discrete, differentiable neural networks since the early 1980s (Schmidhuber, 2015).

BP calculates gradients of the loss function considering the weights and biases, that are then updated in a way that minimizes the loss function (Amari, 1993).

ANNs have shown excellent performances in a great variety of applications. Due to their complexity, they are often very resource demanding. When designing ANNs it is further important to address the challenges of overfitting and of hyperparameter optimization (Bakhteev and Strijov, 2020; Piotrowski and Napiorkowski, 2013).

1.6 Objectives

In this thesis, the presented machine learning methodologies are applied on remote sensing data from the upper Vltava river in order to estimate AN, BOD5, NN, and TP concentrations. For each output variable, two types of linear models (MSLR and PLSR) and three types of more complex machine learning algorithms (SVR, RFR, BP-ANN) are constructed and evaluated.

Overarching aims are to identify fundamental spatiotemporal patterns and dynamics of the investigated water quality parameters in the upper Vltava basin, and to assess the suitability of machine learning methods for modeling the named parameters in this part of a major central-European river. More specifically, the thesis follows three three main objectives of:

1. Analyzing spatial and patterns and temporal developments of concentrations of BOD5, AN, NN, and TP in the upper Vltava river and selected tributaries.

2. Estimating concentrations of the four water quality parameter based on remote sensing imagery, using machine learning methods.

3. Evaluating model performances and select the best algorithms for predicting each of the four water quality parameters.

Based on the aims, and following the relevant literature, three main hypotheses are investigated:

1. There is a decrease of BOD5, AN, NN, and TP concentrations from the decade 2001-2010 and 2011-2020.

2. There is a downstream increase of BOD5, AN, NN, and TP in the upper Vltava River.

3. The used algorithms can give accurate estimates of the investigated water quality parameters.

4. Machine learning algorithms outperform the linear models for estimating the target variables (in terms of R^2, RMSE, MAPE).

2. Study Area, Materials and Methods

2.1 Study Area

The investigated area (fig. 1) is the upper Vltava catchment, including the Lužnice River. The Vltava river sources in the Šumava mountains (Bohemian Forest) in Southwest-Bohemia, Czech Republic. Since 1991, the area is protected as a national park (Mentlík, 2016).
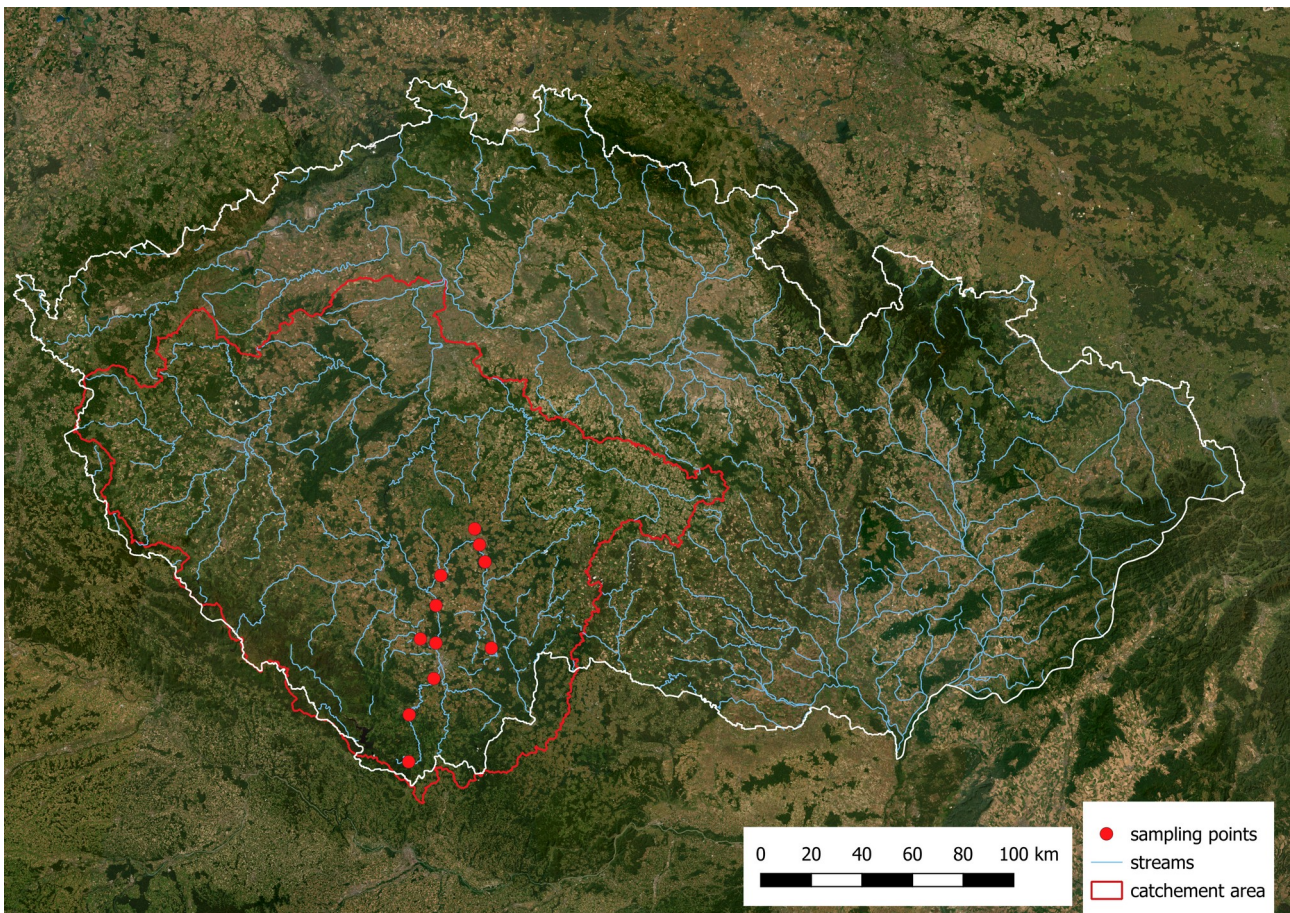
Figure 1: Sampling localities in the upper Vltava catchment

The Bohemian Forest is characterized by felsic plutonic and metamorphic bedrock (mostly granites) and shallow acidic soils (Baburek et al., 2013). The mountains are part of the bohemian massif which formed during the Variscian orogeny. From a hydro-geological perspective, the Šumava mountains are a hard-rock environment. Magmatic (mostly plutonic) and metamorphic rocks dominate. Most notably, there are large bodies of granite, migmatite, paragneiss, granulite, eclogite, amphibolites, marbles, and quartzites.

The soils are mostly acidic and poor in nutrients. The Bohemian forest is among the few European regions naturally dominated by *Picea abies*- (European spruce-)forests.

Downstream of the Šumava region the Lipno reservoir – Czech Republic's largest standing water body – follows (de Moraes et al., 2023). It was created in the 1950s for by damming the Vltava river, serving the purposes of hydropower production, flood mitigation, and recreation. In contrast to the Vltava headwaters, the large (46.5 km^2) but shallow (max. 22 m) Lipno reservoirs are relatively eutrophic (Tesfaye et al., 2023). The area downstream of Lipno is charcterized by a transition to sedimentary rock. Important land cover forms are forests, quite vast floodplains and extensive agriculture. In this section, the Vltava crosses a part of the Třeboň Basin, known for its fishpond system (Bohnet et al., 2022). As the Vltava River flows northwards into the Budějovice basin, agricultural intensity gradually increases to high levels in the Northern part of the study area. Other human activities also intensify, following the river from the source to the lower parts of the investigated region. Major municipalities include Český Krumlov (ca. 13,907 inhabitants; 492 m above NN) and most notably České Budějovice with almost 100,000 inhabitants (Czech Statistical Office, 2023). At 381 m above NN, České Budějovice is only slightly more elevated than the lowest point used for modeling, located in the dam reservoir Vodní nádrž Hněvkovice at 371 m above NN. The Vltava was dammed there in the 1980s for supplying the Temelín nuclear power plant with cooling water (Růžička et al., 2005). In brief, it can be concluded that along the upper Vltava, there is an increasing gradient of land use and agricultural intensity and a corresponding nutrient gradient.

Apart from this, general developments of the trophic state of the upper Vltava can be described. After a long, steady increase of nitrogen in synthetic fertilizers after the second world war until 1989, followed by a short, but very steep decrease from approximately 110 kg ha^{-1} year^{-1} to approximately 50 ha^{-1} year^{-1} within only four years; afterwards, the use slowly increased again to 70 kg ha^{-1} year^{-1} (Kopáček et al., 2013). More recent data were not found. Despite the re-elevated N use in fertilizers, overall observed total nitrogen (TN) concentrations in the Vltava have continued to decrease until 2010 (Kopáček et al., 2013). While more recent information has not been available, the present work presents evidence indicating a re-introduction of increasing N-concentrations in the upper Vltava.

Evidence from the Slapy reservoir north of the study area shows relatively stable concentrations with considerable seasonal fluctuations (maxima in winter, minima in summer), with annual averages fluctuating between ~40 and ~80 µg L^{-1} between 1963 and 2015 (Vystavna et al., 2017).

Until 1992, a P increase has been documented, afterwards, a decrease; only for the summer months, there was an overall increase (Vystavna et al., 2017). However, it must be noted that these values are from a standing water body which is downstream of the investigated catchment area, so the reported trends are only of minor indicative value.

Annual P runoff from agricultural land in the upper Vltava basin has been found to be in the unit range of kg km^{-2}, being close to anthropogenically unaffected areas (Rosendorf et al., 2016). Nitrogen pollution, in contrast, is found to be a major obstacle for achieving and sustaining healthy water quality in the area: In parts of the upper Vltava catchment, surface N runoff of more than 20 kg ha^{-1} has been reported,largely due to the application of animal manure (Rosendorf et al., 2016).

2.2 Identification of relevant literature

Relevant literature was identified by entering various keyword combinations to the Web of Science, Scopus, ScienceDirect, Google Scholar and Pubmed. The search was limited to results in the English language. Besides recent publications, selected pioneering works and other relevant older publications were included.

Corresponding to the research objectives, the identified literature covers following topics: 1. Use of machine learning algorithms for estimating inland water quality parameters with a main focus on non-optically active parameter and on rivers; 2. General literature on the used algorithms; 3. General literature on river water quality and the investigated water quality parameters; 4. Hydrology, geology, and geography of the study area. Keywords were combined in different ways by the boolean operators AND and OR" and included "ammon*", "backpropagation", "Landsat*", "Lužnice", "machine learning", "neural net*", "nitrogen", "non-active", "non-optical*", "optically active" "phosphorus", "random forest", "remote sens*", "sentinel*", "South Bohemia", "support vector", "upper Vltava", and "water quality". Literature on specific questions that arose during the writing process was

identified by using the respectively relevant keywords in the previously enlisted literature repositories.

2.3 Data Sources and Dataset creation

2.3.1 Water Quality Data

Water quality data were provided by the Czech Hydrometeorological Institute (CHMI, 2022). The dataset includes a total of 62 sampling points (35 in the Lužnice River and its tributaries; 27 in the Vltava river and other of its tributaries), eleven of which could be used for model construction. It covers the period from January 2000 to December 2020. In includes water quality data from 9,217 samples, in total.
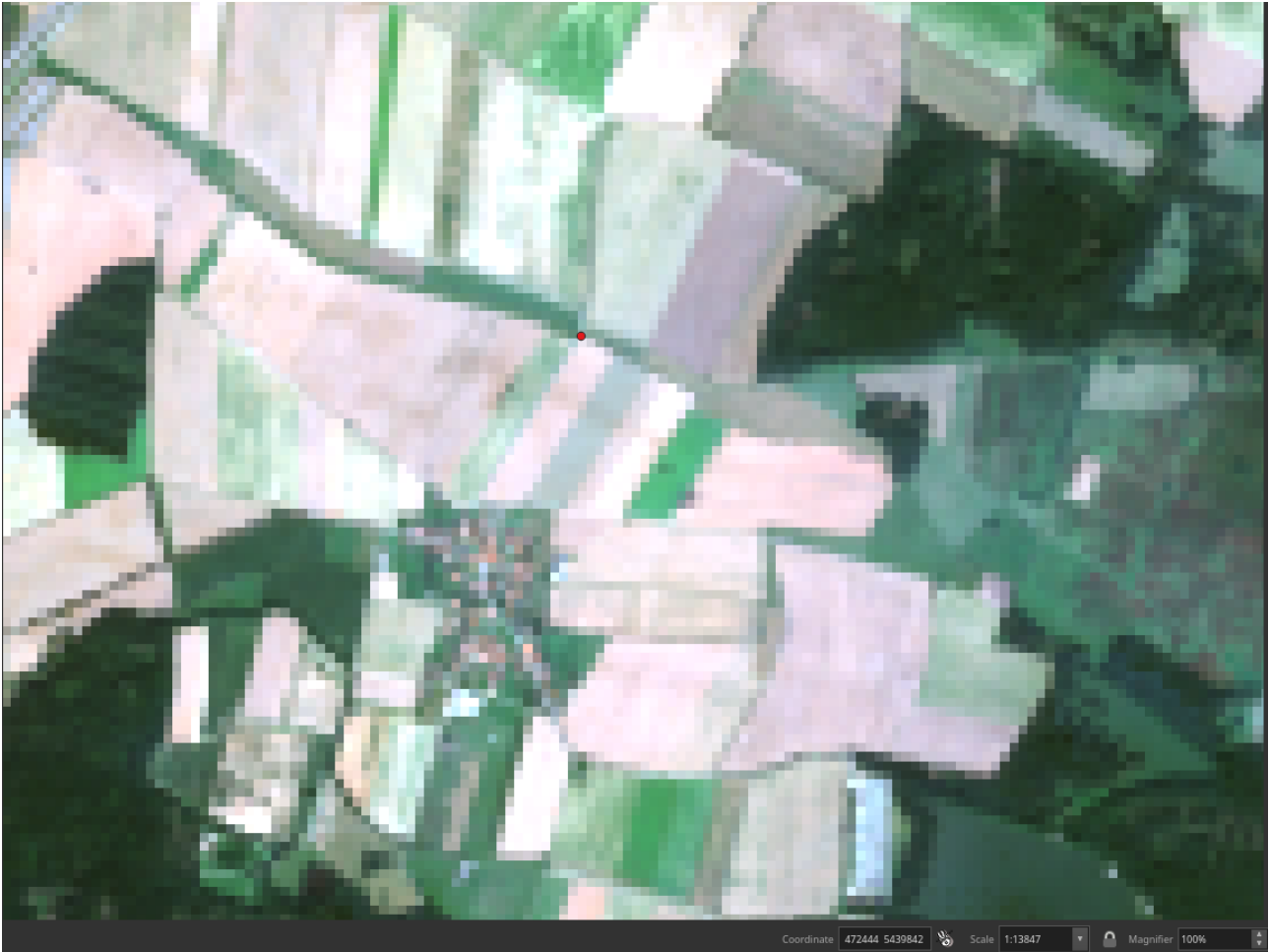
Figure 2: Point in QGIS with no apparent water in Sentinel-2 image (band combination 04-03-02)

Points were removed following visual inspection in QGIS (QGIS Development Team, 2009). Points were not removed from the dataset when the corresponding Sentinel-2 pixel from the HLS-S30 product was at least almost fully covered by water, and not hidden behind vegetation or objects. The satellite imagery for this procedure was loaded into the GIS via the Semi-automatic classification plugin (Congedo, 2023).

Only imagery with less than 5% cloud cover was considered. Decisions on whether to remove a point were made after inspecting the respective point with band combinations 04-03-02 (fig. no), representing natural color, and 08-11-04 (fig. no), suitable for identifying open water (Kaplan and Avdan, 2017). 51 points (82%) of the points had to be removed. Subsequently, remote sensing data from the remaining eleven point were obtained.
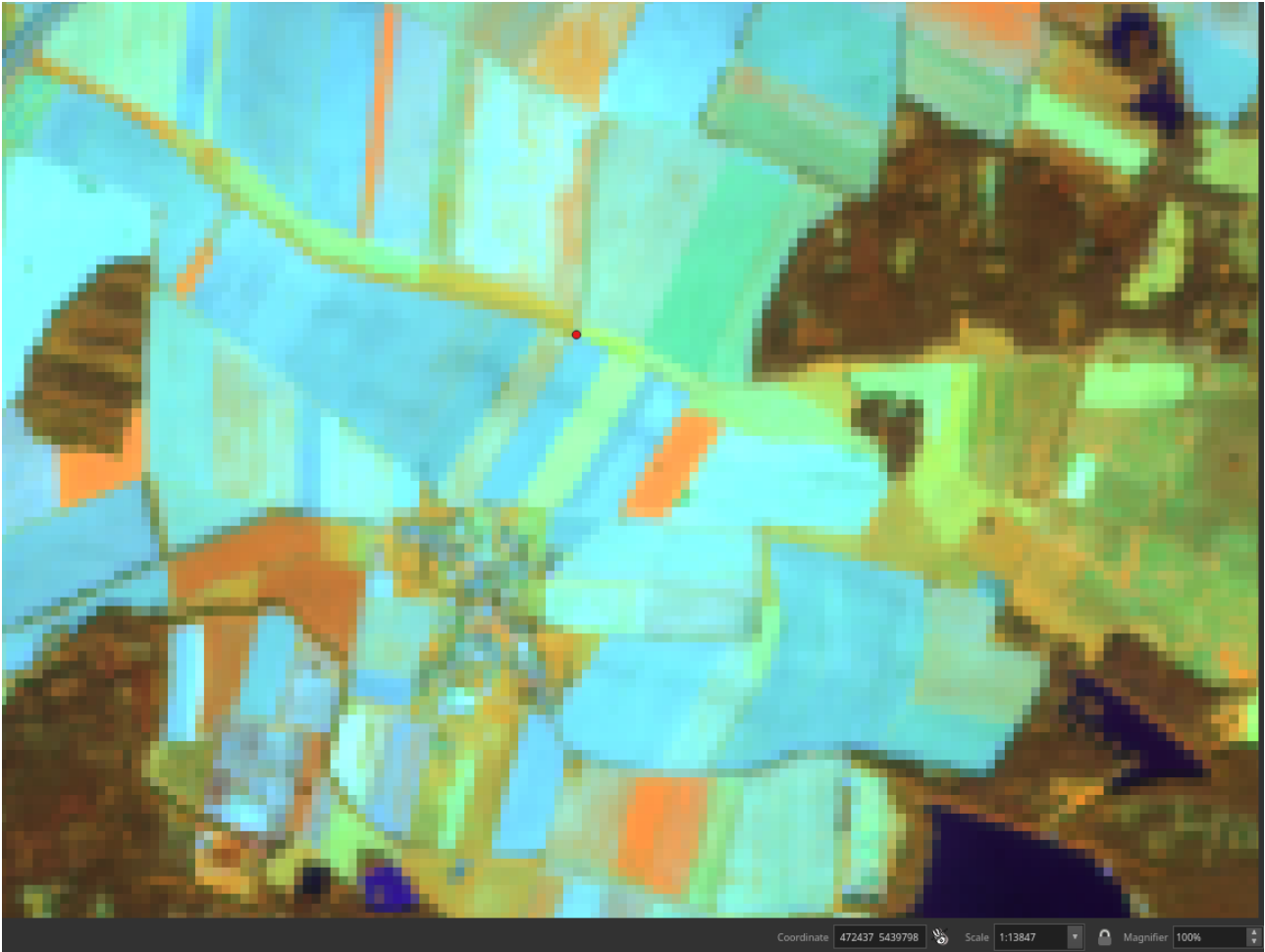
Figure 3: Same point, in QGIS, using different bands (band combination 08-11-04)

2.3.2 Remote Sensing Data

All remote sensing datasets were obtained using the JavaScript programming language in Google Earth Engine (GEE). The code is based on examples from the GEE documentation (Google, 2024b) and in parts written with the help of the large language model ChatGPT4 (OpenAI, 2023), as documented in the scripts. All relevant scripts can be found in the appendix.

Landast-5 surface reflectance data (Crawford et al., 2023; Google, 2024a) were used for the period from 01. January 2000 to 05. May 2008. Surface reflectance data from the HLS data product (Claverie et al., 2018; Ju et al., 2020) were used for the period from 12. April 2013 to 31. December 2020. Due to the lack of readily available satellite imagery in GEE, no data for the time range between 06. May 2012 and 11. April 2013 are used in the models.

Table 1: Wavelength ranges of used bands of LS5 and HLS L30

LS5 Band	Description	LS5 Wavelength (µm)	Corresponding HLS L30 Band	HLS L30 wavelength (approximation; µm)
Band 1	Blue	0.45-0.52	Band 2	0.45-0.51
Band 2	Green	0.52-0.60	Band 3	0.53-0.59
Band 3	Red	0.63-0.69	Band 4	0.64-0.67
Band 4	Near-Infrared (NIR)	0.76-0.90	Band 5	0.85-0.88
Band 5	Shortwave Infrared (SWIR) 1	1.55-1.75	Band 6	1.57-1.65
Band 7	Shortwave Infrared (SWIR) 2	2.08-2.35	Band 7	2.11-2.29

The exact HLS-product used for the models is called HLSL30: HLS-2 Landsat Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m (Masek et al., 2021). Data were also downloaded using the JavaScript programming language in Google Earth Engine. Cloud-covered points and measurements of poor quality were omitted based on the quality variables included in the product. The used bands were selected based on compatibility with the Landsat-5 product (see tab. 1), and on suitability for the modeling purpose. The fairly similar wavelength ranges of the used products allow using both as modeling input data without further spectral harmonization steps. LS-5 and HLS-data were joined based on the common bands (tab. 1) using the Python programming language. In the obtained dataset and in the further text, LS-5 band names are used to refer to both, LS-5 and the corresponding HLS bands.

2.3.2 Data merging and cleaning

Datasets were joined by appending each row from the water quality dataset with the respectively temporally closest row from the satellite dataset at the given point. Therefore, the absolute time difference in days was considered, i.e., the amount of days before *or* after the satellite passed a given point. Rows containing missing band values were removed. Rows with greater absolute time differences of more than 15 days were also removed from the dataset, resulting in a total of n = 835 remaining observations.

2.3 Data Analysis and Visualization

Data preparation, analysis and visualization were performed using the Python programming language in Jupyter notebooks in the VSCodium IDE from the miniconda distribution. Data preparation, statistical analysis, and ML model construction were performed using the libraries numpy (Harris et al., 2020), pandas (McKinney, 2010), statsmodels.api (Seabold and Perktold, 2010), Scikit-learn (Pedregosa et al., 2011), tensorflow(-gpu) (TensorFlow Developers, 2024), and the H_2O-AutoML library (LeDell and Poirier, 2020). Graphs were plotted using the additional libraries seaborn (Waskom, 2021), and matplotlib (Hunter, 2007). For data visualization, code from the python graph gallery was used and adapted to the purposes of the presented research (Holtz, 2024). Maps were created in QGIS. The satellite basemap from ESRI was loaded into QGIS via the SRTM-Dowloader (Duester, 2023) package for QGIS. Vector files of water bodies and catchment delineations were downloaded from the EEA (European Environment Agency., 2012). Further formatting of graphs and maps was performed using the vector graphics editor Inkscape (Inkscape Project, 2020). Parts of the codes were written with the help of Chat-GPT 4 and Chat-GPT 3.5 (OpenAI, 2023).

The input variables were normalized by applying the Z-score standardization (Pedregosa et al., 2011) algorithm (fig. 1), where z is the standardized value of a variable, x is its original value, μ the mean of the feature values, and σ is their standard deviation.

$$z = \frac{x - \mu}{\sigma}$$

Equation 1: Z-score standardization

Satellite measurements at a given point usually occurred on different days than water was sampled. Weighting columns were calculated in order to be able to penalize bigger time lags between a satellite revisit and water sampling. Weighting transformations were performed using an exponential function (eq. 2) in order to increase the punishment of greater time lags.

Different weights were calculated based on different multipliers m_i (0.5, 1.0, 1.5, 2.0, 2.5), allowing greater flexibility and more appropriate weighting for each model. In the further text, natural numbers from 1 to 5 are used to refer to the weights, in ascending order.

$$w_i = e^{-m_i} \cdot \frac{\Delta t}{\Delta t_{max}}$$

Equation 2: Calculation of weights, based on time lag between sampling and revisit. w_i = weight; m_i = multiplier; Δt = absolute time difference (days)

PCA was performed on the standardized satellite band variables, obtaining two principal components that were later on tested as potential model input parameters. All models were tested with PCs in comparison to the standardized satellite bands as input features. Additionally, models were tested including the months (as categorical variable) as input feature. For each output feature in each model, the best available w_i was selected, using for-loops.

Both approaches of MLSR, forward selection and backward elimination were performed for all output variables. All possible combinations of polynomials from the first to the third order were modeled by using a for-loop. Time lag weight variables were iterated for each model. Subsequently, the respective model with the best fit (i.e., highest R^2) was identified automatically for each dependent variable. Exhaustive iterations over weight columns, and the best polynomial order (1-3) were performed, and the model summaries for the models with the highest R^2 were printed.

For PLSR, the standardized satellite band variables were used as inputs, with no further feature engineering. Models including or excluding the sampling month as input variable were tried. The models were constructed for all output variables at once. For each dependent variable, the model with the highest R^2 was selected and printed as output. For-loops were used to exhaustively iterate over all combinations of possible scaling methods (mean_centering, robust_scaling, or no scaling), and components components. Suitability of polynomials (up to the fourth order) as input variables was also assessed iteratively, but were not further considered after proving unsuitable. The final algorithm includes exhaustive iterations over all combinations of up to three excluded input features, of weight columns (including the option not to use any). The month-variable is part of the feature-space of this algorithm.

The modelled relationship between the input variables and a given output variable was considered statistically significant, if $p < 0.05$ (i.e., 95% level of significance).

Random forest regressors were constructed optionally with or without month (as categorical variable) as input variable. Iterations over all weight variables were performed to find the best one for each output variable. The minimum number of trees was set to 100. Incrementally, trees were added in steps of 20 additional trees, until further addition did not result in a significant decrease of the RMSE. The constructed RFR models predict all output features at once, but model evaluation was done for each water quality parameter individually. 100 bootstraps were applied.

Possible feature combinations for SVR-based estimation of the wq parameters was performed with a combined SVR-RFE algorithms, that assigned stability/importance estimates for the individual bands, and their simple polynomials of the second order (i.e., the squared observations of each individual band, and products of the observations of each possible combination of two individual bands). An adapted algorithm, including polynomials up to the third order polynomials was run for ANN feature selection. Possible feature combinations with different purposes were then constructed based on feature importance for each dependent variable. For instance, including features of the highest relevance, or covering a broad set of mid- to high-stability feature, or covering nonlinear dynamics.

SVR algorithms were trained by applying Bayes search to a search-space including pre-defined sets of input features, the weighting columns of the dataframe, kernel functions (rbf, linear, polynomial, and sigmoid), degree (if polynomial kernel), and the hyperparameters C, gamma, and epsilon. Bayes search was run with a varying amount of iterations (30-85), starting on a sample subset (n=400), and then generalized to the hole dataset, on which – if necessary – further hyperparameter optimization was performed. The Each iteration of Bayes search used 3-fold cross-validation. While 5-folds cross-validation would have increased model robustness, the increased computational cost was a constraint against it. Model training was done by gradually eliminating unsuitable feature combinations, and kernel function, and by gradually reducing (and sometimes enhancing) the search-space for the hyperparameters.

BP-ANN construction was first tried with hard-coded hyperparameters for all output variables at once, leading to mostly unsatisfactory results, and great computational efforts for hyperparameter tuning.

After, also mostly unsuccessful, computationally and temporally constrained attempts to implement a similar search- and training-approach as for the SVR-algorithms, an automated pipeline for BP-ANN construction and model selection, based on limited inputs, was implemented. For this approach, the automated machine learning (AutoML) Python-library H_2O was used. The included algorithm types were set to only to BP-ANN algorithms, encoded as "DeepLearning" in the H_2O package . Again, enforced by resource limitations, the maximum model number to be calculated was set to 50, with a respective maximum runtime of 2,500 seconds. A csv file with detailed results on model hyperparameters, and selected performance metrics (RMSE, R^2, MAPE) was returned.

Model evaluation for all types of algorithms was based on the performance metrics R^2, RMSE, and MAPE. Where applicable, information on weighting of the time lag, the polynomial degree, the best number of components (based on R^2). In the case of R^2, the R^2-score from the scikit-learn library was used. While it is equal to the regular coefficient of determination in the positive range, it can have negative values, indicating a worse fit than a model that always estimates the mean value of the target parameter (Pedregosa et al., 2011). Such a model would get the R^2-score 0.

More details on the methodology are described in the supplementary materials.

3. Results

3.1 Seasonal and spatial patterns

Most of the investigated water quality parameters averaged over the 62 initial input points show distinct seasonal patterns (Fig. 4). The biochemical oxygen demand is lowest during the winter months, with an average of slightly more than 3 mg L^{-1} between December and February. Highest values are measured between May and October. During this period, BOD5 is relatively stable.
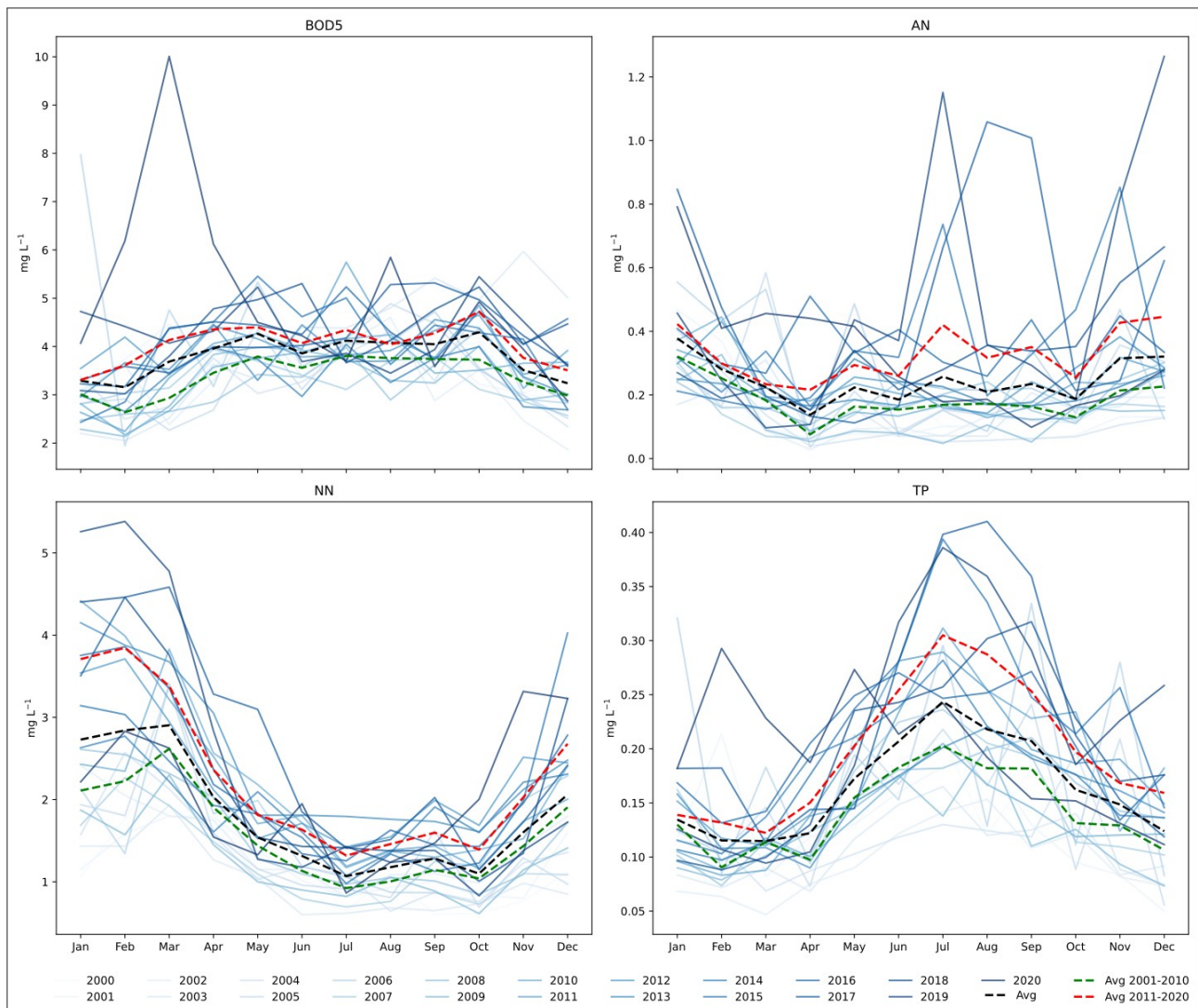

Figure 4: Monthly averages of water quality parameters across all sampling points (2000-2020), including decadal averages (2001-2010 and 2011-2020)

Seasonal dynamics of ammonia nitrogen are less pronounced than of the other measured parameters.

On average, there is only little seasonality, but in most years, a decline of AN concentration can be observed between December and April, followed by a strongly fluctuating increase over the rest of the year. The monthly average concentration (2000-2020) never exceeds 0.4 mg L^{-1}. However, there are considerable outliers, especially in recent years, where concentrations of more than 1 or even more than 1.2 mg L^{-1} are the mean concentration of the 62 sites at individual months in the second half of the year.

Nitrate nitrogen concentrations peak early in the year, between February and March. The 21-year average concentration in March is at almost 3 mg L^{-1}. After this, the concentration steeply drops to a minimum of slightly more than 1 mg L^{-1} in July. Until October it remains relatively stable and then gradually increases again to the annual maximum.

Total Phosphorus shows approximately the opposite pattern: On average, it has a distinctive peak in July (ca. 0.24 mg L^{-1}), from which it gradually drops to a minimum between February and March (ca. 0.12 mg L^{-1}). Between January and March, the concentration is relatively stable.

A notable increase of all four water quality parameters can be observed when comparing the 10-year averages of the decades between 2001 and 2010 on the one side and 2011 and 2020 on the other. This holds true for all four parameters in every single month of the year.

Unlike in the previous decade, from 2011 to 2020, average BOD5 concentrations remain at levels of more than 4 mg L^{-1} from March to October, and do not drop below 3 mg L^{-1} for the rest of the year. A single extreme outlier was observed in March. when the average concentration at the 62 sampling points reached 10 mg L^{-1}. In the more recent decade, mean monthly AN concentrations exceeded 0.4 mg L^{-1} from November to January as well as in July. In the previous decade, the annual maximum monthly average was in January at around 0.3 mg L^{-1} with all other months showing substantially lower concentrations. Monthly mean AN in the later decade is constantly above 0.2 mg $^{-1}$, while it had been below that threshold for eight months (March – October) of the year in the previous decade. Mean nitrate nitrogen concentrations show an especially high increase in the first two months of the year. The decadal monthly maximum shifted from March to February and from a previous ~3 mg L^{-1} to almost 4 mg L^{-1}. The average minima are found in July and October in both decades, and increased from around 1 mg L^{-1} to more than 1.3 mg L^{-1}.
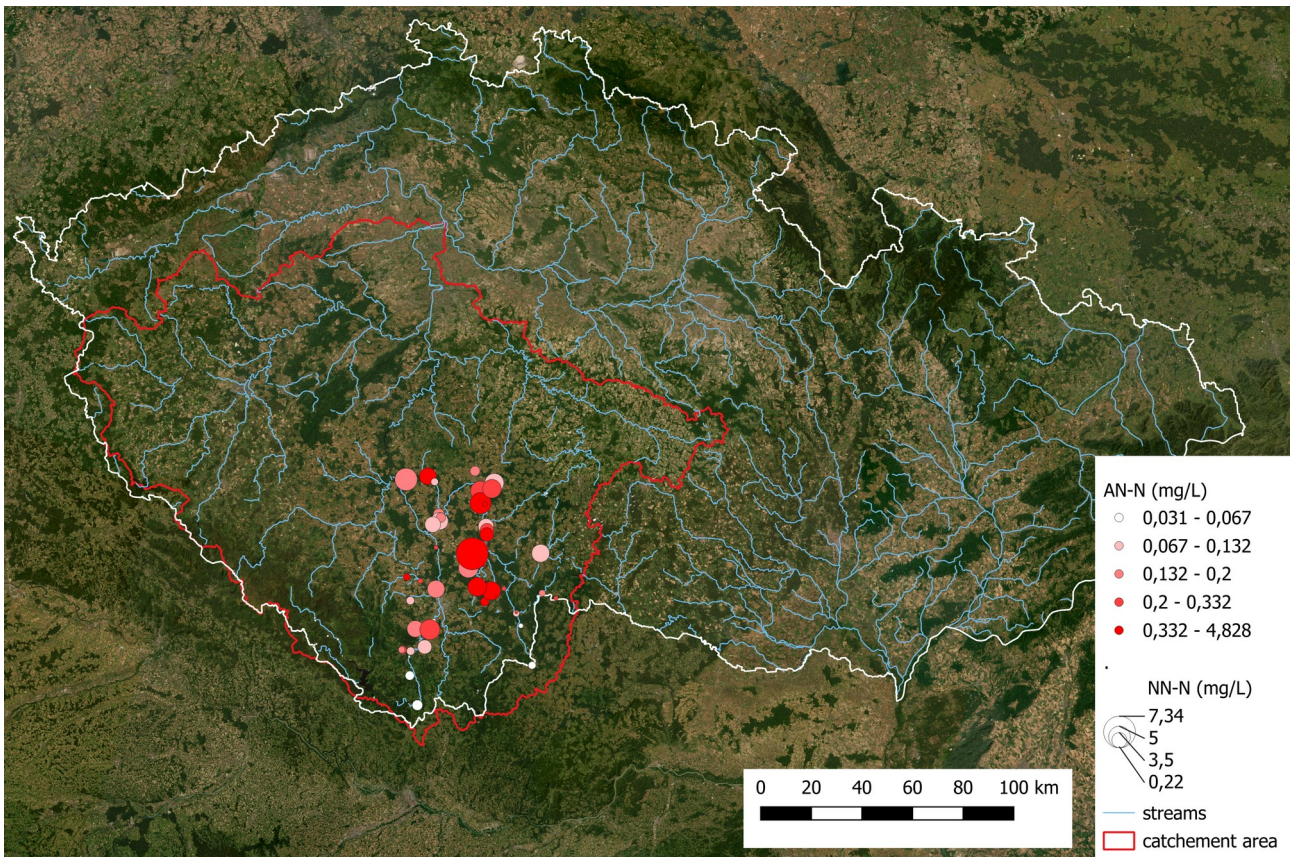
Figure 5: Reactive inorganic nitrogen (AN and NN) in the study area, considering all 62 sampling points from initial dataset
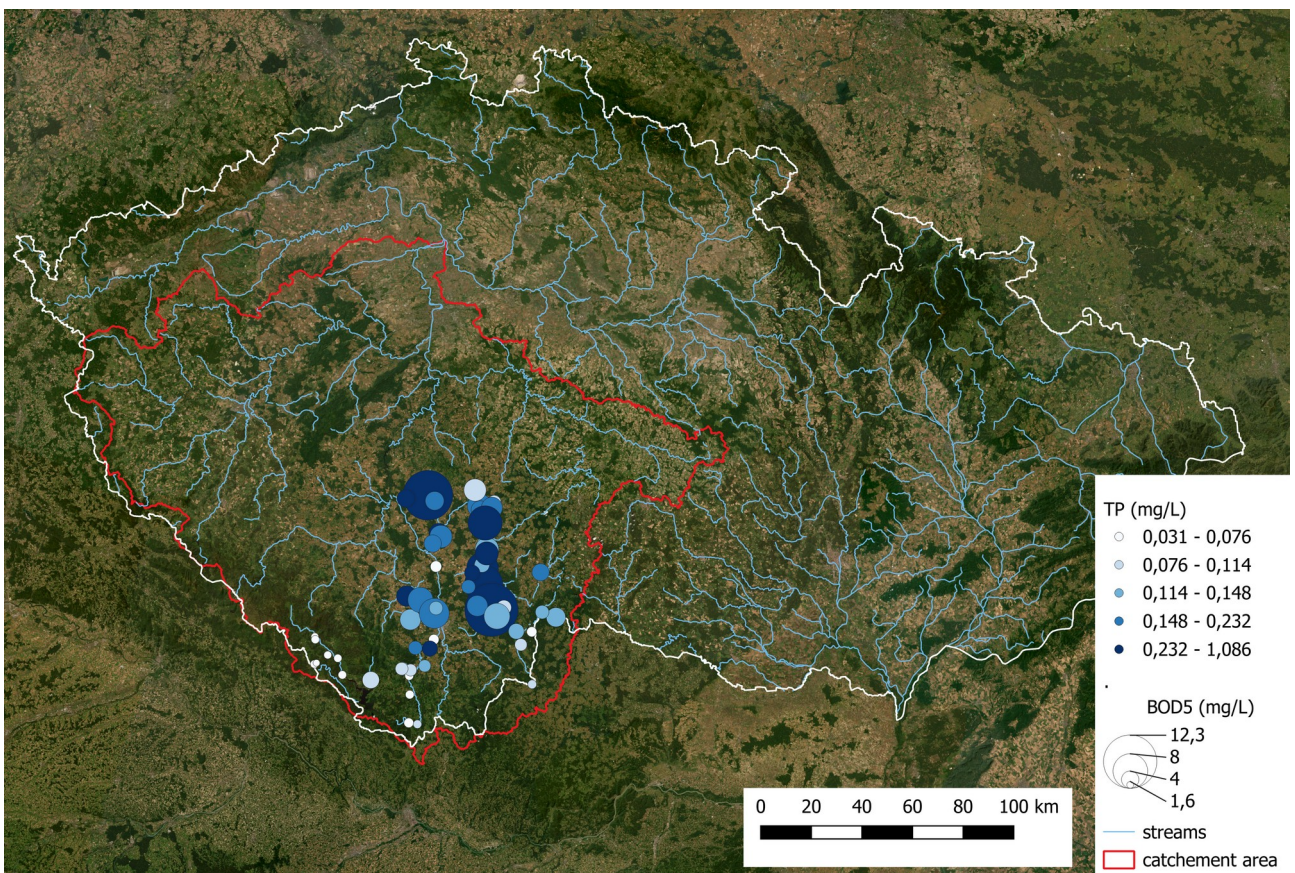

Figure 6: BOD5 and TP in study area (62 points)

The annual TP maximum in July increased by 50% from ~0.2 mg L^{-1} between 2001 and 2010 to ~0.3 L^{-1} between 2011 and 2020. While mean monthly TP concentrations only slightly increased for the months January and March, the increase in the remaining ten months is substantial or even dramatic.

Visual inspection of the mapped average concentrations of the four water quality parameters shows clear spatial patterns (fig. 5-6): All four water quality parameters appear in lower concentrations in the Šumava source region, while being more concentrated in the agricultural and urban areas more downstream. Furthermore, it appears that higher concentrations of all target parameters are found in stream segments of higher order. In-depth analysis could provide other relevant insights.

3.2 Spatial distribution of the water quality parameters

Averaged over the entire investigated period, the spatial distribution of all four water quality parameters behaves as hypothesized. An almost steady increase of all four, AN, NN, BOD5, and TP can be observed from the headwaters to the lower parts of the investigated areas.

It is important to keep in mind that the described temporal and spatial patterns refer to the dataset cover a wider range of water quality data than the models presented below, as the description of spatiotemporal patterns of the field measurements is not constrained by the availability of remote sensing data.

3.3 Model Preparation

None of the reflectance data from the bands in the combined LS5-HLS dataset were normally distributed. All of them were strongly skewed to the right. Often, local modes in higher ranges of the distributions were observed. While theoretically, the reflectance values should range between 0.00 and 1.00, the dataset includes few observations where 1.00 was exceeded. It was not further investigated whether such results are a consequence of data harmonization, satellite calibration, atmospheric correction, or randomly occurring artifacts. Instead, the observations were used as given in the datasets.
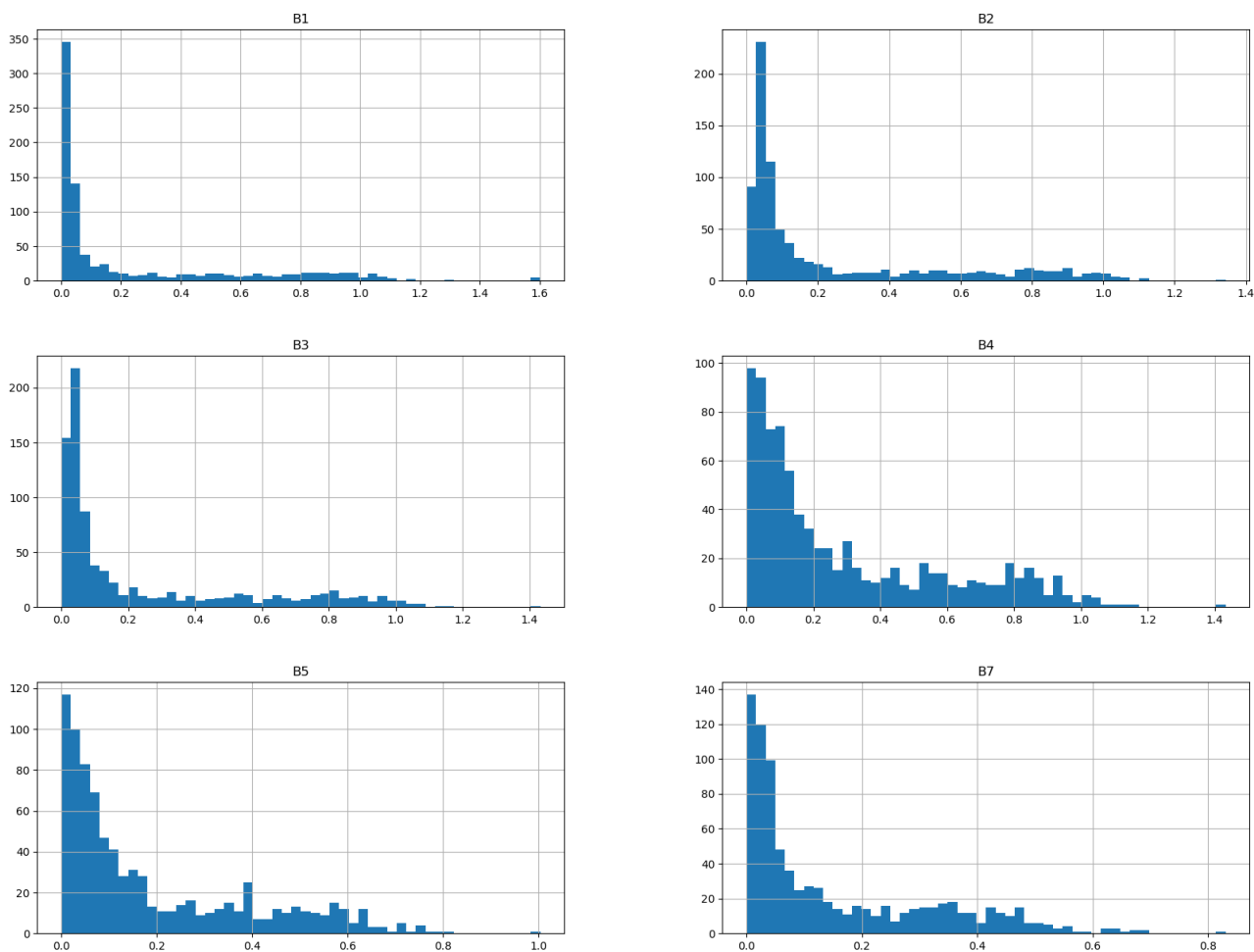
Figure 7: Distributions of surface reflectance data on the input bands of the combined LS5-HLS dataset

The non-normal distribution of all independent variables makes parametric tests unviable for first explorations of their suitability for estimating the concerning water quality parameters. Even after various transformations (quadratic, exponential, boxcox), none of the input variables was normally distributed, as shown by Shapiro-Wilk-tests.
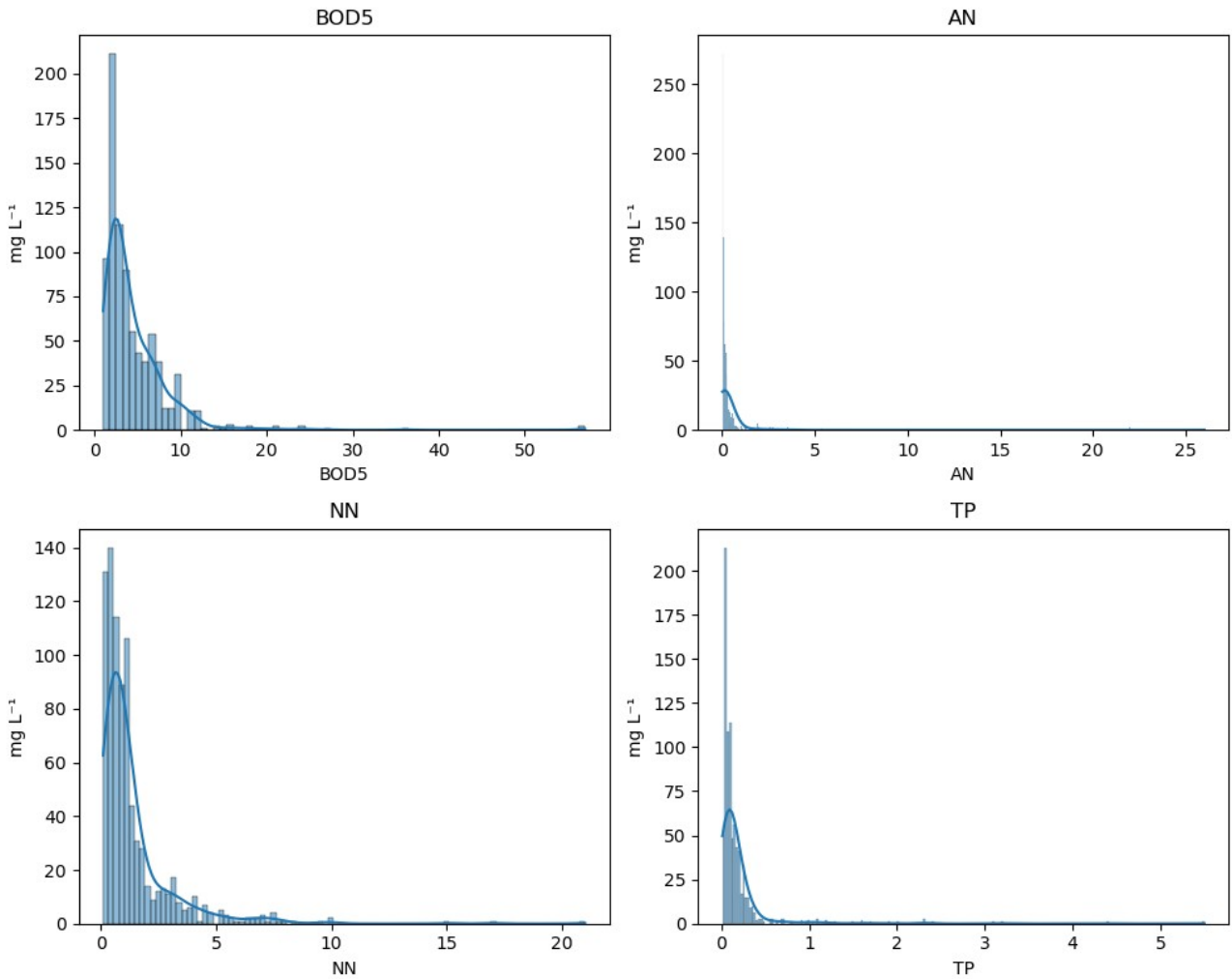
Figure 8: Distributions of the investigated water quality parameters

The distributions of all dependent variables are also strongly skewed to the right. The recorded levels of AN, NN, BOD5, and TP appear to resemble gamma-distributions, and could thus be transformed to normal distribution quite easily. But since the used algorithms do not depend on normal distribution, and the distributions of the dependent variables do not allow for parametric tests, this was not further investigated.

Due to the non-normal data distributions – even after transformations –, it is not admissible to estimate basic interactions between dependent and independent variables with Pearson's correlation.

Therefore, a first estimate of relevant interactions of reflectance at given satellite bands and concentrations of AN, NN, BOD5, and TP was made, based on Spearman's rank correlation. The satellite reflectance data show strong positive intercolinearity between all bands.

The weakest association, found between bands 1 and 5, is still very high at $\rho = 0.84$, while the strongest one ($\rho = 0.99$) almost reaches the theoretical maximum.

All of the investigated water quality parameters also positively correlated with each other. The weakest rank correlation is found between NN and TP ($\rho = 0.12$) and the highest one between BOD5 and TP ($\rho = 0.77$).
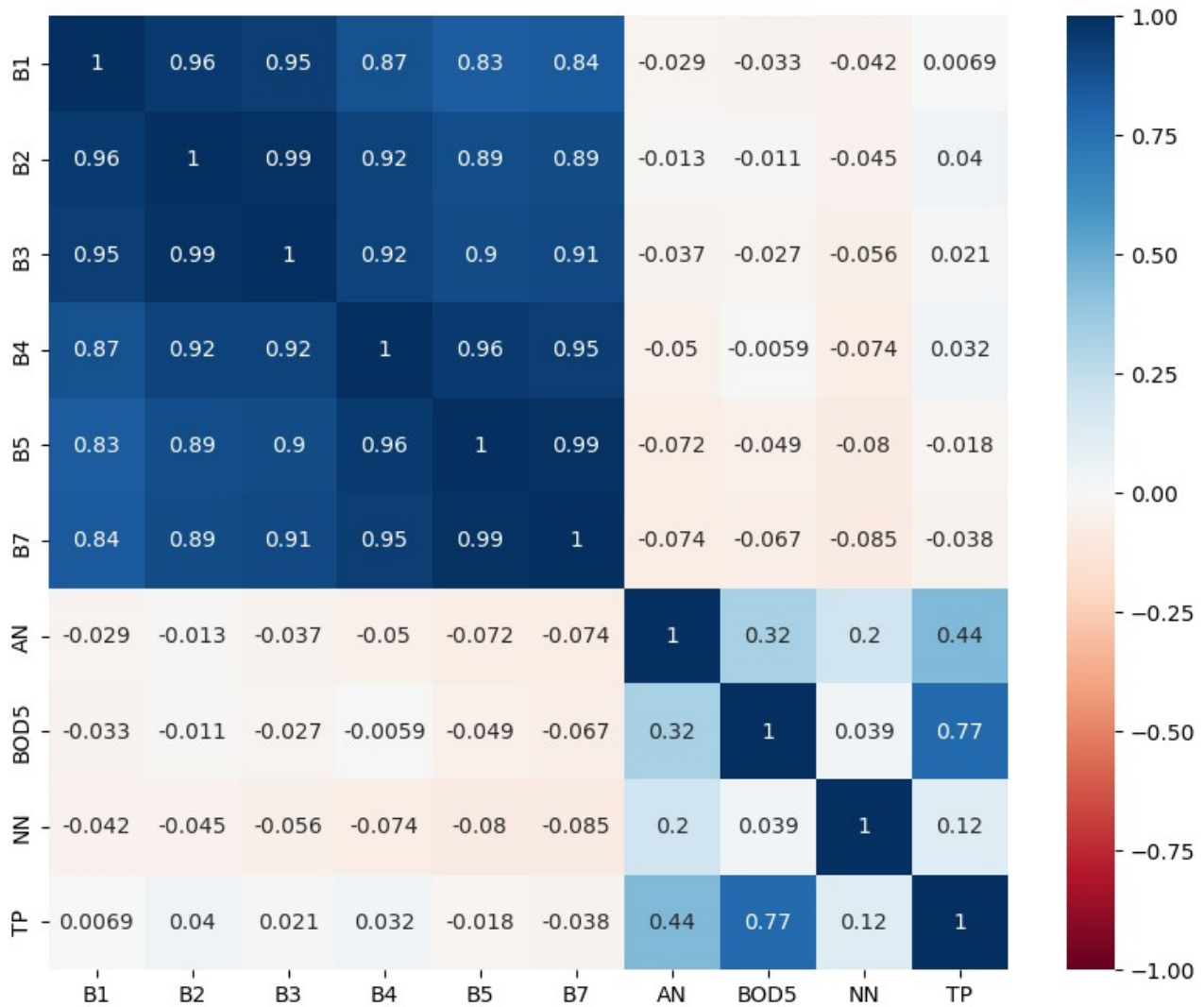
Figure 9: Correlation matrix of input (satellite bands) and output (water quality parameters) features using Spearman's ρ

Spearman's rank correlation coefficients between the bands on the one side and the water quality parameters on the other, are mostly relatively weak and negative. The greatest relationship is found between B7 and NN ($\rho = -0.085$). Positive coefficients are found only found between the bands 2, 3, and especially 4 ($\rho = 0.032$) with TP.

The identified rank correlations between surface reflectance on the given bandwidths on the one side, and the levels of the investigated water quality parameters on the other side suggest that there might be more complex underlying patterns, allowing to interfere between these data.

3.4 Model Performance

These patterns were captured by the different algorithms with varying success (Tab. 2).

Table 2: Model performances on testing dataset for each water quality parameter (best model of each type based on R^2; For BP-ANN: No weight variables tested)

WQP	Model	R^2	RMSE	MAPE (%)	w_i
AN	MSLR (ns)	0.005	0.585	488.61	5
	PLSR	0.012	0.918	567.22	-
	SVR	0.078	0.500	376.48	5
	RFR	0.156	1.180	319.36	4
	BP-ANN	0.010	1.804	288.95	-
NN	MSLR (***)	0.212	1.768	141.68	-
	PLSR	0.112	2.450	138.97	1
	SVR	0.091	2.472	169.84	1
	RFR	0.555	1.124	80.01	5
	BP-ANN	0.180	1.354	112.69	-
BOD5	MSLR (***)	0.035	2.747	68.11	-
	PLSR	0.061	3.180	75.68	1
	SVR	0.042	3.218	52..3	4
	RFR	0.462	2.667	46.15	3
	BP-ANN	0.040	4.370	72.31	-
TP	MSLR (***)	0.041	0.431	128.72	-
	PLSR	0.004	0.229	127.71	1
	SVR	0.049	0.224	98.84	5
	RFR	0.390	0.257	91.24	5
	BP-ANN	0.041	0.397	105.92	-

The MLSR models indicate no significant linear relationship (p < 0.05) between the input band variables and Ammonia Nitrogen, regardless of weighting, polynomial degree of input variables, or type of the MSLR (forward selection vs. backward elimination). Very little variance is explained by the model ($R^2 = 0.005$), and errors are high (RMSE = 0.585 mg L^{-1}; MAPE = 488.61%). Most variance of AN is explained by RFR ($R^2 = 0.156$) and SVR ($R^2 = 0.078$) models. All other algorithms perform very poorly on this target variable. Even these comparatively well-performing models predicted AN with an RMSE of 0.500 mg L^{-1} (RFR) to 1.180 mg L^{-1} (SVR). Given the monthly mean range of ca. 0.2-0.4 mg L^{-1}, a root mean square error this high cannot be satisfactory. PLSR ($R^2 = 0.012$) and BP-ANN, both explain only an almost negligible fraction of the variance. The relative differences of estimated and measured AN concentrations are high for all models, ranging from 288.95% (BP-ANN) to 567.22% (PLSR). It is interesting to note that models with higher penalties for time lags between sampling and satellite revisit tend to perform better than with lower penalties.
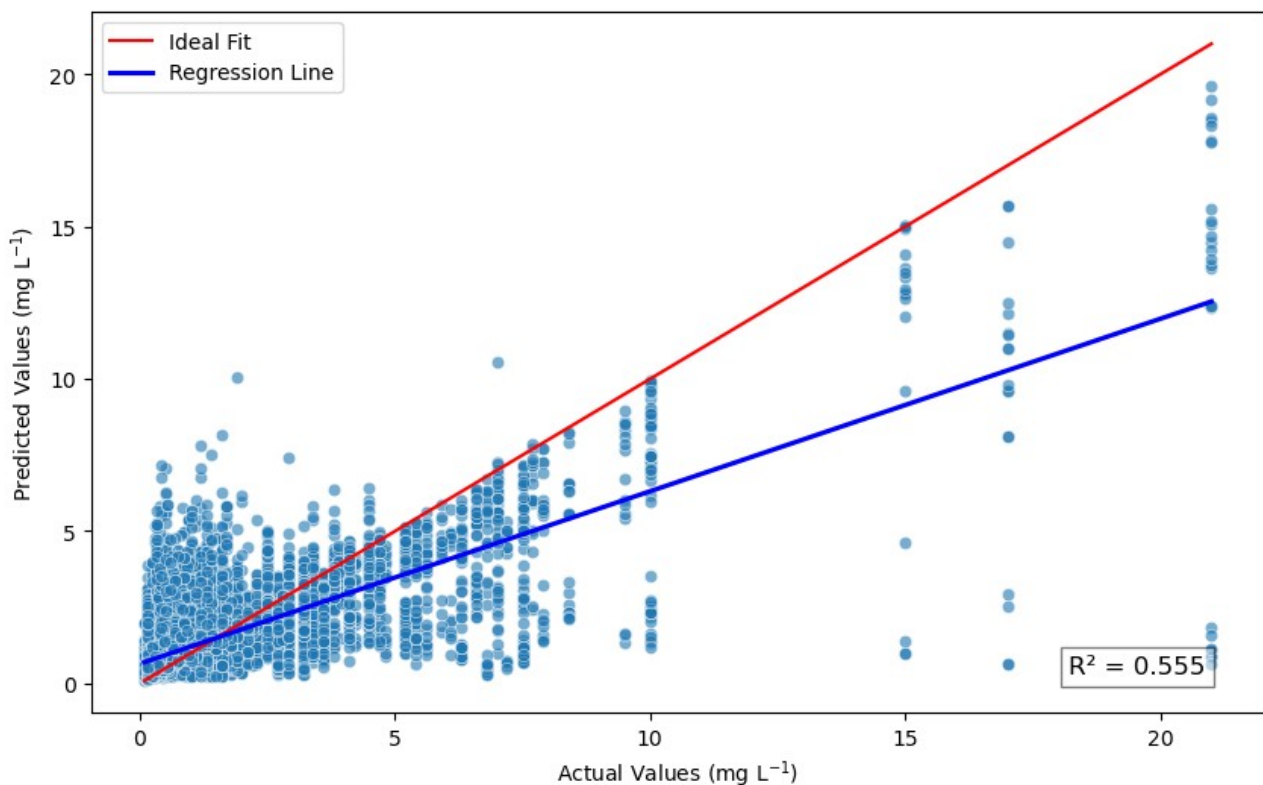

Figure 10: Predictions vs. actual values of test-dataset of RFR for NN estimation

Estimation of nitrate-N, the other nitrogen fraction among the dependent variables, works substantially better. The linear MSLR-model can explain a considerable fraction of the variance of NN ($R^2 = 0.212$). Apart from SVR ($R^2 = 0.091$), all types of algorithms return $R^2 > 0.10$. The highest fraction of variance ($R^2 = 0.555$) is explained by a random forest regressor (fig.7), which also returns the lowest mean errors (RMSE = 1.124; MAPE = 80.01). PLSR ($R^2 = 0.112$), and BP-ANN (0.180) explain less of the variance than MSLR. Due to mean observations an order of magnitude higher than those of AN, also the RMSE values in mg L^{-1} are higher, while the relative deviation of estimated from measured concentrations, expressed by the MAPE, is a lot lower (AN: 288.95% - 567.22%; NN: 77.52.95% - 169.84.%). Overall, all types of models return useful outputs for estimating nitrate nitrogen.

BOD5-estimates are less powerful than those of NN, but still a lot more valuable than those of AN. The multiple stepwise linear regression indicates a highly significant relationship with $R^2 = 0.035$ (RMSE = 2.747 mg L^{-1}). The RFR model performs the best on all investigated metrics ($R^2 = 0.462$; RMSE = 2.557 mg L^{-1}; MAPE = 46.15). The remaining models also explain a greater fraction of the variance of BOD5, but their RMSE values indicate lower accuracy than for MSLR (see tab. 2). Over all selected models, the deviation of predicted values from measured values stayed well below 100% (MAPE: 46.15% - 75.68%), indicating considerably high accuracy, compared to other dependent variables. Despite the relatively low fraction of variance explained by MSLR, PLSR, SVR, and BP-ANN, the results still show a clear response between satellite bands and BOD5, even though it might be weak, and hard to capture in a model.

Quite similar results as for BOD5 were obtained for TP. The MSLR showed a highly significant (p <) relationship with the predictors and TP, explaining a small part of the variance ($R^2 = 0.041$). Lower errors at the same R^2 were achieved by the BP-ANN (see tab. 2). While having somewhat lower errors than MLSR, partial least square regression ($R^2 = 0.004$) can hardly explain more variance than a constant model, always assuming the average value of TP, suggesting that it is not ideal for predicting this water quality parameter. SVR, in contrast, can explain a small, yet non-negligible part of the variance ($R^2 = 0.049$), while also returning substantially lower errors (see tab. 2). At similar errors, but clearly higher explained variance, SVR was only outperformed by RFR ($R^2 = 0.390$).

44

The main results of the model evaluation are: Estimating AN remains challenging, as none of the used algorithms was able to explain a high part of its variance, while all of them produce high errors when estimating AN concentrations. In contrast, NN estimates are comparably accurate. Various types of models can return valuable information, promising at least some degree of suitability for NN prediction, and promising to be a good base for further optimization.

The results for BOD5- and TP-estimation indicate a relevant response between the used predictors and the target variables. Yet, the models mostly fail to accurately capture the variance of BOD5, and TP. The comparably low relative error of BOD5 estimates, as expressed by MAPE, indicates the possibility of relatively accurate predictions. Yet, the minimum MAPE vales of almost 50% are still far from ideal.

One of the most important results is the capital effectivity of RFR to predict any of the target parameters from the dataset, as compared to all other tested algorithms. In fact, it can be considered the only used algorithm providing solid estimates of NN, BOD5, and TP, while still being very limited for AN.

In contrast, the MLSR algorithms tended to produce the weakest predictions, with the exceptions of NN, where MLSR explains the second-highest fraction of variance among the tested models, and for TP, where its R^2 value is ten times higher than that of PLSR. However, the MLSR outcomes also indicate that some significant relationships can be captured between the input variables on the one side, and NN, BOD5, or TP on the other side, even by linear models.

For all sorts of algorithms, and all investigated water-quality parameters, finding appropriate weights for the absolute time-differences between water sampling and satellite revisit at a given point is not a trivial task. This is underlined by the wide range of weights used in the respectively best models, with no apparent clear pattern.

4. Discussion

4.1 Spatiotemporal patterns of Water Quality Parameters

AN. Relatively unobstrusive seasonality: AN largely determined by plant uptake in growing season. Effect maybe neutralized by elevated inputs, or by changed patters of ammonia oxidation. Increase likely due to higher agri input.

NN: Seasonality matches observations from East-Polish Supraśl River (Skorbiłowicz and Ofman, 2014) and from Latvian river sites (Tsirkunov et al., 1992). Explanation: Growing season: More N in primary producers. Increase likely due to higher agri input, and not due to lower uptake or the like. Solubility of nitrates increases with temperature, but direct climate change contribution low. After decreasing agricultural nitrate loads between 2000 and 2014 in the catchment of the Rímov reservoir, an increase until 2020 was reported. In contrast, nitrate loads from settlements showed a low, but steady decrease between 2000 and 2020, while no significant changes were reported for nitrate loads from forests. Maybe different pattern further downstream? Especially

TP Seasonality: Hgher solubility in summer. Increase: Higher P-input. It is yet to be investigated, if effects of climate change contribute to P-increase on this local level. A likely explanation is increased use of phosphate fertilizers. Also accumulation in reservoirs might be a factor. Contribution of point sources? Easier to control. But improving WWTP efficiencies. Rímov reservoir: Insignificant, and almost negligible decrease from forests, little change from agricultural, slow increase from settlement. Overall, very slight increase. Loads from all systems peak in summer, indicating that higher mobility and loads might also contribute to seasonal patterns. Observed overall increase corresponds to observations presented in previous chapter. Increase in temperature was found.

Weaker seasonality of BOD5: Microbial activity in summer increased due to hugher temp. and P-availability (assuming P-limitation of the system). But: Also, lower O_2-availability, due to lower solubility, so faster O_2 depletion. Also, BOD_5 tested under lab conditions, so temperature during respiration is same and growth promoted again in lab. Increase due to eutrophication, and like for P: climate change effects would be interesting to investigate.

Might have direct effects due to higher microbial activity at higher temperatures. Due to usually high O_2 saturation in rivers, lower solubility might not have a contrary effect.

Slapy reservoir: Slight temperature increase, with decade 2010-2019 as warmest decade in past 60 years. Increased O_2 saturation. Similar levels of nitrate, and in some months lower levels of ammonium . Increased DOC. Strong increase of DON. Very similar TP (Kopáček et al., 2021).

Seasonality: Identified patterns correspond to those reported in literature. Increase: Partly supported by other published results, but partly in contrast to them. Maybe: addition of points with higher trophic status later included to dataset? Refinement of methodology required. Also, literature usually focuses on individual basins, and not on whole area, so further investigation is needed. After further progress in methodology, ML models could be a good way for quantifying this development in a good temporal and spatial resolution.

Streamflow decline observed (Vystavna et al., 2023).

Spatial: Downstream increase corresponds to land use gradient: Lowest inputs from forests also reported by (Vystavna et al., 2023).

The hypothesis of decadal decrease cannot be accepted. The results clearly indicate an increase of all four water quality parameters between the two compared decades. However, due to methodological limitations, and inconsistencies with the findings reported in published research, a refined approach of investigation is required.

Spatially, a downstream increase of all four parameters was observed, matching the initially stated hypothesis. Two factors might contribute to that: Faster flor rates in the upstream area allow faster transport of input contaminants and other substances, giving less time for accumulation (Schwoerbel and Brendelberger, 2022). But more important might even be the land use gradient described earlier that is closely associated to the "continuous gradient of pollution and habitat degradation" (Horka et al., 2023, p. 2) in the area.

4.2 Model Evaluation

Apart from some RFR models, the accuracy of all predictions is fairly low, as indicated by the test metrics. Other model types can at least still explain considerable amounts of variance of nitrate nitrogen. Furthermore, estimations of BOD5 differ on average by less than 80% (MAPE < 80%) from the actually observed values, also indicating some relative accuracy. This lower mean absolute percentage error of BOD5 estimates, compared to the remaining dependent variables, might be due to a narrower range or smaller variance of BOD5. For now, the hypothesis of accurate ML-based estimates in the given research area cannot be accepted, but the results point toward a promising direction.

Especially estimating AN turned out to be difficult. Two types of algorithms, SVR ($R^2 =$ 0.08), and RFR ($R^2 = 0.16$), explained similar levels of AN variance in the upper Vltava, as the equivalent models by Li et al. (2022) explain for the tropical Nandu river in China (R^2 (SVR) = 0.07; R^2 (RFR) = 0.24). However, the poor performance of the BP-ANN in the present thesis ($R^2 = 0.10$) is in stark contrast to the ANN presented in the Chinese study ($R^2 = 0.44$). Still, also for Li et al. (2022) AN turned out to be the variable hardest to estimate. Despite explaining a substantial share of AN variance, also the estimates by Li et al. show large deviations from the actual values, with MAPE ranging from 274% to 318%.

While the errors might still be relatively large and the explained fractions of AN variance by the different models presented in the previous chapter might be low, they are far from negligible. Future improvements of ML methodology application are required to allow for robust and accurate estimates of AN.

The comparably good estimates if NN are in correspondence to those presented in literature. SVR ($R^2 = 0.29$) and NN ($R^2 = 0.29$) models in a study by Sagan et al. (2020) could to explain considerable parts of the variance of nitrate in the water. Their PLSR model ($R^2 = 0.26$) also outperformed the one presented in this thesis ($R^2 = 0.11$). Other similar studies do usually not focus on nitrate (nitrogen). Anyhow, comparing the presented NN estimates to such of TN/TIN can still provide valuable insights, as river N generally mostly consists of NN (Boyd, 2015).

Gao reported R^2 0.31 (SVR) – 0.45 (XGB). XGB was generally among best algorithms for most variables. Li: RF performed best with $R^2 = 0.49$. SVR had 0.2. MAPE ranging from 33.11-43.59%. Maybe narrower distribution of values? Overall: Explained variance in my models can compete. Li had ANN with $R^2 = 0.45$. Gao was not so good in TN estimation with best model XGB $R^2 = 0.45$, but had strong SVR with 0.35%. According to the coefficient of determination, RFR models by Gao et al. (2024; $R^2 = 0.42$) and Li et al. (2022; $R^2 = 0.49$) performed slightly weaker than the one from the upper Vltava ($R^2 = 0.56$). However, at 33.53%, also the MAPE reported by Li et al. is a lot lower. Both studies also presented relatively well-performing ANNs and gradient boosting algorithms. It can be concluded that besides the present work, various publications support the observation that ML algorithms have considerable predictive power for AN- or TN-estimation from satellite imagery.

BOD5 has largely been neglected in similar studies. The remarkably low MAPE values, the large fraction of variance explained by RFR, and the small, but still obvious responses shown by other models are strong indicators of the suitability of ML-approaches for also estimating this parameter from RS-imagery. This, however, demands more in-depth research, ideally from the upper Vltava and other riverine systems around the world.

Estimating TP in the upper Vltava turned out to be somewhat more challenging than BOD5, and a lot more than AN. This is in line with the results by Sagan et al. (2020) who failed to explain any variance of TP ($R^2 = 0.00$) with SVR or ANN ($R^2 = 0.00$) algorithms. While clearly outperforming the best calculated PLSR model on the upper Vltava ($R^2 = 0.00$), the one by Sagan et al. still predicted TP very poorly ($R^2 = 0.02$). The explained TP variance by SVR (0.05) and RFR (0.43) in the upper Vltava are higher by orders of magnitude, and might represent promising starting points for the development of more accurate ML algorithms for this geographic realm. Substantially stronger model performances for TP estimations in commensurable systems have been presented by Li et al. (2022), who obtained an SVR model with $R^2 = 0.59$, and an ANN with $R^2 = 0.67$, both of which came with MAPE values around 50%. This level accuracy remains hard to reach for the upper Vltava. The models presented by Gao et al. had a somewhat weaker performance, with SVR having the lowest ($R^2 = 0.2$), and RFR the highest ($R^2 = 0.39$) coefficient of determination.

The huge variability of model outputs from the upper Vltava, and their partly poor performances when compared to other research results points to a demand of further refinement of ML methodologies. The high R^2 of the RFR model clearly indicates suitability of such models for TP estimation in the area.

From a broader perspective, it can be observed that model performances from other studies are often better, and more stable across different model types. But there are also exceptions to this, and not even uniform patterns on which variables are hardest to predict. However, there seems to be a tendency, that various approaches for NN or TN prediction tends to perform very well, while AN prediction remains challenging.

The approach of neural network construction in the present thesis turned out to likely be insufficient, as ANNs in published research generally performed a lot better, than those used for the upper Vltava. The Azto ML-approach for ANN-construction might be useful for getting a first orientation for the algorithm and the behaviors of the dependent variables, but it might be insufficient for actually identifying suitable ANN models, or at least, requires more control by the modeler.

4.3 Limitations and potential for future model improvements

The brief analysis of spatial and temporal patterns can only give a rough idea of them, but due to massive temporal constraints, it comes with some methodological flaws. The monthly averages, the decadal changes, and the mapping are all based on averages from all sampling locations. While this increases the sample size for these calculations and corresponding visualizations, the results cannot easily be generalized, and must not be overvalued: Since samplings have not been taken from all sites since January 2000, but some sites were added later, there might be some introduced bias. This might for instance be because stations which come into service later, might have artificially increased mean levels of the four parameters, due to a temporal increase between the two decades. Or maybe, sites with higher levels of one or more of the parameters have been introduced later, leading to only seemingly increased concentrations, so that the proclaimed increases might actually be lower than described. Neither possibility has been tested, due to temporal limitations.

Work remains to be done to eliminate these shortcomings and reduce the bias. Only this could enable reliable detailed conclusions on the spatiotemporal distribution and development of AN, NN, BOD5, and TP. However, despite these limitations, the simplified analysis approach satisfies the requirements of the performed further investigations.

Additional valuable spatiotemporal information, some of which could also be useful for the model, could be obtained by methods such as time-series analysis, allowing to isolate seasonal patterns from long-term trends in water quality development (Halliday et al., 2012). Another valuable approach could be catchment modeling based on digital elevation models, providing reliable estimates of catchment area (Kwast and Menke, 2022) and river length until a certain point (e.g., sampling locality).

As already stated, also the ML models themselves come with limitations. Some of these are necessary consequences of the nature of the ML methods. Others are more related to modifiable parts of the learning systems, and thus represent potentials for future improvements.

First of all, ML not able to explain mechanisms of interactions between water quality parameters and surface reflectance, thus leaving them as a black box (Cao et al., 2020; Gao et al., 2024). Their interpretability is very limited and they cannot shed light on deeper underlying patterns. While to some degree, we have to just live with that fact for now, increased interpretability and more detailed interpretations would be possible in some cases. Some of the possible methods can be useful for future improvements of the presented approach.

One easy approach could for instance be, to include feature importance to the output from random forest models. This can not only increase their interpretability by giving insights to the strength of connections between input features and target variables, but can also be valuable for feature selection in other models (Pedregosa et al., 2011). Another strategy specifically for random forests could be to visualize a simplified decision tree that represents the average decision-making process of the forest, in a so-called born-again tree (Vidal and Schiffer, 2020). This approach could represent the unmanageable complexity of an RFR model in the more comprehensible form of a single, condensed regression tree, and thus allow an approximate interpretation of the decision procedures performed by the tree.

However, even the born-again tree can be sufficiently complex to be of only limited interpretability.

A relatively simple and highly effective improvement of interpretability could further be using additional metrics for model evaluation across all models, such as MSE, RMPSE or MAE,/MAD as it has been done in previous studies (Gao et al., 2024; Guo et al., 2021; Li et al., 2022; Sagan et al., 2020). Still simple and even more informative would be further visualization of model performances, by including different types of visualization, and – more importantly – by visually comparing predictions of all selected models to the respective actual values of the respective dependent variable in the test dataset, as it was done in the study by Li et al. (2022). The information obtained by this visualization might even be valuable for optimizing the models themselves. For instance, model adaptions could be implemented to try to mitigate the overestimation of NN in the upper concentration ranges, and the underestimation in the very low range.

Besides aiming at higher interpretability, the architecture of the models can also changed in a way that allow for greater robustness, and higher accuracy. An obvious approach for increased robustness could be to add more folds to cross-validation or more iterations to bootstrapping. This strategy is, however, limited by computational constraints, and it should be kept in mind that at some point, the added robustness from additional folds or bootstraps gets marginal, while still substantially increasing computational cost.

Model predictions could be improved by refining the strategy of weighting individual observations. The entire range of weighting variables for penalizing the time lag between sampling at satellite revisit was used in the models. This indicates that both, a wider and denser set of weighting variables, or even a more dynamic weighting approach might benefit model performances. It is furthermore interesting, that there was no uniform pattern regarding preferred weights and dependent variables, which could indicate greater stability of a wq-parameter, if the penalty was lower, and larger fluctuations at greater preferred penalties.

More exhaustive strategies of identifying the best hyperparamters might further be helpful. Even more important might be to optimize the approach to input-features.

More individualized feature selection could be implemented for each model and output variable, in order to keep a balance between the amount of variables containing valuable information, and dimensionality. This can also include polynomial features, obtained by multiplication of individual input bands. The feature selection engineering and selection approach used for SVR is a step into that direction, yet it may have failed to identify the ideal set of features for each output variable. Also, there was no systematic approach to identifying the best number of input variables. For-loops, or preferably search mechanisms such as Bayesian search, grid search, or maybe even genetic algorithms (Yang, 2021) could be (further) implemented. While promising powerful improvements, sophisticated strategies of this kind are nevertheless far beyond the scope of this thesis.

The evaluation – and massive removal – of measurement points evokes the question if the used satellite imagery is appropriate. Instead, sacrificing a wide time span, and all Landsat imagery in favor of Sentinel-2 data might be a better approach. Their spatial resolution of up to 15 m, would allow the inclusion of a much greater amount of points. Combined with the better spectral resolution, this could also improve the model performance. A comparison of predictions of HLS and S-2 data, could bring clarity. But, as previously elaborated, relying only on Sentinel-2 data has also has a great limitation: I can only provide estimates from June 2015 onward. Models based on these data can thus only cover a short time range, which makes them unsuitable for many all applications that require long-term estimates of the development of the investigated parameters.

Not only might different remote sensing data allow better models. After understanding which band combinations from which instrument allow the best predictions of each concerning water quality parameter, it might also be of interest to consider other freely available predictors. Combining the remote sensing-based approach with other freely available predictors, such as meteorological data, hydrogeomorphological data (e.g., catchment area at sampling locality), or data on adjacent landcover and landuse could enhance the performance of some of the algorithms significantly.

Besides different input data, also different models should be tried in future research. Plenty of studies have shown, for instance, that gradient boosting algorithms have outperformed other methods for estimation of various non-optically active water quality parameters (Gao et al., 2024; Li et al., 2022).

The pursuit for finding the most appropriate model for predicting each of them should thus also focus on gradient boosting methods, but also other algorithms, like simple regression trees, have performed surprisingly well in some cases (Li et al., 2022), and should thus also be considered.

From the beginning to the end, parts of our learning systems could be adapted with potential benefits. Even if the models were calculated using ideal input datasets, ideal feature engineering and selection, and ideal hyperparameter selection, at least one further step could be optimized: The performed model-selection strategy, focusing almost exclusively on R^2 is very limited. Given the limited resources and the modest scope of this thesis, focusing on R^2 as main target metric is a justified, yet not ideal approach.

The availability of better models will also allow mapping of spatial distribution and temporal developments of estimated water quality parameters, opening up a new, rapidly-availble source of valuable information for water quality management. This can provide unprecedented insights on behavior of the distribution of the mapped parameters, and help identifying and managing the influence of various point- and diffuse sources of nitrogen and phosphorus. So far, however, the models have limitations due to which such efforts are unlikely to produce very valuable knowledge.

The shortcomings of the models presented in this thesis demonstrate the challenges of estimating non-optically active wq params based on optical data. They further show the demand for further refinement of methodologies. The comparison to published data does not that there are slight methodological deficiencies in the present work, that call for optimization; it also shows that the usually narrower, and faster-flowing upper reaches of rivers come with additional particular challenges for RS-ML-based estimation of non-optically active water quality parameters. But despite undeniable weaknesses of the presented results, they still clearly show that even for such systems, machine learning methods have the potential to be powerful tools for estimating NN, BOD5, TP, and even AN concentrations in the upper Vltava and similar riverine systems.

For direct application in water quality management, the presented results are still insufficient.

But they can serve as a good starting point for developing robust models that can reliably provide accurate estimates of AN, NN, BOD5, and TP, and thus become an invaluable and unprecedented resource for water quality management and similar practical applications.

5. Conclusions and Outlook

Management of river water quality is important for mitigating eutrophication. After the water quality improvements since the 1990s, this trend might have turned in the upper Vltava catchment after 2010: Concentrations of the four optically inactive water quality parameters ammonia nitrogen, nitrate nitrogen, five-day biochemical oxygen demand, and total phosphorus might have increased, as results presented in this thesis suggest. All four parameters follow a land-use and ecosystem degradation gradient in the upper Vltava, increasing in concentration from the headwaters to the lower parts of the research area.

In this study, five different types of algorithms (multiple stepwise linear regression, partial least squares regression, support vector regression, random forest regressor, and backpropagation artificial neural network) were performed on harmonized Landsat-Sentinel-2-data (HLS), in order to predict AN, NN, BOD5, and TP. Such machine learning methods can be useful tools for estimating the four named parameters, but they still bear a lot of challenges, especially given the often narrow streambed. Further improvements at all steps – from input data selection, over model optimization, to model selection – are required for ensuring accurate and robust predictions. The presented results indicate that this goal can be achieved. In most cases, machine learning algorithms clearly outperformed simpler linear models. Especially the method of the random forest regressor can often estimate a big fraction of variance, while also producing comparatively low errors.

Of all four water quality parameters, NN was most effectively-predicted, with $R^2 = 0.555$ by a random forest regressor, whereas AN is the most challenging with a maximum $R^2 = 0.156$, also by a random forest regressor. BOD5 and TP prediction also remains challenging, but some models returned relatively strong metrics, indicating great potential for sound predictions of these parameters as well.

Optimization of used input data, feature engineering, feature selection, model architectures, hyperparameter optimization and model selection will allow broad application of ML methods for river water quality estimations, even in upper reaches of the Vltava and other major European rivers. Reaching this goal requires big further efforts.

6. References

Abdi, H., 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). WIREs Computational Statistics 2, 97–106. https://doi.org/10.1002/wics.51

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Ahmed, M., Mumtaz, R., Anwar, Z., 2022. An Enhanced Water Quality Index for Water Quality Monitoring Using Remote Sensing and Machine Learning. Applied Sciences 12, 12787. https://doi.org/10.3390/app122412787

Akbarzadeh, Z., Maavara, T., Slowinski, S., Van Cappellen, P., 2019. Effects of Damming on River Nitrogen Fluxes: A Global Analysis. Global Biogeochemical Cycles 33, 1339–1357. https://doi.org/10.1029/2019GB006222

Amari, S., 1993. Backpropagation and stochastic gradient descent method. Neurocomputing 5, 185–196. https://doi.org/10.1016/0925-2312(93)90006-O

Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. Estuaries 25, 704–726. https://doi.org/10.1007/BF02804901

Anding, D., Kauth, R., 1970. Estimation of sea surface temperature from space. Remote Sensing of Environment 1, 217–220. https://doi.org/10.1016/S0034-4257(70)80002-5

Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., de Freitas, N., 2016. Learning to learn by gradient descent by gradient descent. https://doi.org/10.48550/arXiv.1606.04474

Baburek, J., Pertoldová, J., Verner, K., Jiřička, J., 2013. Guide to the geology of the Šumava Mts. Administration of the Šumava National Park and Protected Landcape Area, Vimperk.

Bakhteev, O.Y., Strijov, V.V., 2020. Comprehensive analysis of gradient-based hyperparameter optimization algorithms. Ann Oper Res 289, 51–65. https://doi.org/10.1007/s10479-019-03286-z

Bardenet, R., Brendel, M., Kégl, B., Sebag, M., 2013. Collaborative hyperparameter tuning, in: Proceedings of the 30th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 199–207.

Basak, D., Pal, S., Patranabis, D.C., others, 2007. Support vector regression. Neural Information Processing-Letters and Reviews 11, 203–224.

Basu, N.B., Van Meter, K.J., Byrnes, D.K., Van Cappellen, P., Brouwer, R., Jacobsen, B.H., Jarsjö, J., Rudolph, D.L., Cunha, M.C., Nelson, N., Bhattacharya, R., Destouni, G., Olsen, S.B., 2022. Managing nitrogen legacies to accelerate water quality improvement. Nat. Geosci. 15, 97–105. https://doi.org/10.1038/s41561-021-00889-9

Baxter, R.M., 1977. Environmental Effects of Dams and Impoundments. Annual Review of Ecology and Systematics 8, 255–283. https://doi.org/10.1146/annurev.es.08.110177.001351

Beaulieu, J.J., DelSontro, T., Downing, J.A., 2019. Eutrophication will increase methane emissions from lakes and impoundments during the 21st century. Nat Commun 10, 1375. https://doi.org/10.1038/s41467-019-09100-5

Bellman, R., 1984. Dynamic programming. Princeton Univ. Pr, Princeton, NJ.

Berry, M.W., Mohamed, A., Yap, B.W. (Eds.), 2020. Supervised and Unsupervised Learning for Data Science, Unsupervised and Semi-Supervised Learning. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-22475-2

Beusen, A.H.W., Dekkers, A.L.M., Bouwman, A.F., Ludwig, W., Harrison, J., 2005. Estimation of global river transport of sediments and associated particulate C, N, and P. Global Biogeochemical Cycles 19. https://doi.org/10.1029/2005GB002453

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bohnet, I.C., Janeckova Molnarova, K., van den Brink, A., Beilin, R., Sklenicka, P., 2022. How cultural heritage can support sustainable landscape development: The case of Třeboň Basin, Czech Republic. Landscape and Urban Planning 226, 104492. https://doi.org/10.1016/j.landurbplan.2022.104492

Boyd, C.E., 2015. Water Quality: An Introduction. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-17446-4

Boyer, E.W., Howarth, R.W. (Eds.), 2002. The Nitrogen Cycle at Regional to Global Scales. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-3405-9

Breaux, H.J., 1967. On stepwise multiple linear regression (Master's Thesis). University of Delaware.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brush, M.J., Giani, M., Totti, C., Testa, J.M., Faganeli, J., Ogrinc, N., Kemp, W.M., Umani, S.F., 2020. Eutrophication, Harmful Algae, Oxygen Depletion, and Acidification, in: Coastal Ecosystems in Transition. American Geophysical Union (AGU), pp. 75–104. https://doi.org/10.1002/9781119543626.ch5

Burges, C.J.C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121–167. https://doi.org/10.1023/A:1009715923555

Cai, W.-J., Hu, X., Huang, W.-J., Murrell, M.C., Lehrter, J.C., Lohrenz, S.E., Chou, W.-C., Zhai, W., Hollibaugh, J.T., Wang, Y., Zhao, P., Guo, X., Gundersen, K., Dai, M., Gong, G.-C., 2011. Acidification of subsurface coastal waters enhanced by eutrophication. Nature Geosci 4, 766–770. https://doi.org/10.1038/ngeo1297

Canfield, D.E., Glazer, A.N., Falkowski, P.G., 2010. The Evolution and Future of Earth's Nitrogen Cycle. Science 330, 192–196. https://doi.org/10.1126/science.1186120

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sensing of Environment 248, 111974. https://doi.org/10.1016/j.rse.2020.111974

Carpenter, S.R., Bennett, E.M., 2011. Reconsideration of the planetary boundary for phosphorus. Environ. Res. Lett. 6, 014009. https://doi.org/10.1088/1748-9326/6/1/014009

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen. Ecological Applications 8, 559–568. https://doi.org/10.2307/2641247

Carvalho, D.V., Pereira, E.M., Cardoso, J.S., 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics 8, 832. https://doi.org/10.3390/electronics8080832

CHMI, 2022. Non-optically active water quality parameters from Povodí Vltavy (unpublished data).

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

Congedo, L., 2023. Semi-Automatic Classification Plugin for QGIS.

Conley, D.J., Paerl, H.W., Howarth, R.W., Boesch, D.F., Seitzinger, S.P., Havens, K.E., Lancelot, C., Likens, G.E., 2009. Controlling Eutrophication: Nitrogen and Phosphorus. Science 323, 1014–1015. https://doi.org/10.1126/science.1167755

Crawford, C.J., Roy, D.P., Arab, S., Barnes, C., Vermote, E., Hulley, G., Gerace, A., Choate, M., Engebretson, C., Micijevic, E., Schmidt, G., Anderson, C., Anderson, M., Bouchard, M., Cook, B., Dittmeier, R., Howard, D., Jenkerson, C., Kim, M., Kleyians, T., Maiersperger, T., Mueller, C., Neigh, C., Owen, L., Page, B., Pahlevan, N., Rengarajan, R., Roger, J.-C., Sayler, K., Scaramuzza, P., Skakun, S., Yan, L., Zhang, H.K., Zhu, Z., Zahn, S., 2023. The 50-year Landsat collection 2 archive. Science of Remote Sensing 8, 100103. https://doi.org/10.1016/j.srs.2023.100103

Czech Statistical Office, 2023. Population of Municipalities - 1 January 2023 [WWW Document]. Population of Municipalities - 1 January 2023. URL https://www.czso.cz/csu/czso/population-of-municipalities-1-january-2023 (accessed 3.17.24).

De Mol, C., De Vito, E., Rosasco, L., 2009. Elastic-net regularization in learning theory. Journal of Complexity 25, 201–230. https://doi.org/10.1016/j.jco.2009.01.002

de Moraes, K.R., Souza, A.T., Bartoň, D., Blabolil, P., Muška, M., Prchalová, M., Randák, T., Říha, M., Vašek, M., Turek, J., Tušer, M., Žlábek, V., Kubečka, J., 2023. Can a Protected Area Help Improve Fish Populations under Heavy Recreation Fishing? Water 15, 632. https://doi.org/10.3390/w15040632

Delzer, G.C., McKenzie, S.W., 2003. Chapter A7. Section 7.0. Five-Day Biochemical Oxygen Demand. https://doi.org/10.3133/twri09A7.0

Dey, J., Vijay, R., 2021. A critical and intensive review on assessment of water quality parameters through geospatial techniques. Environ Sci Pollut Res 28, 41612–41626. https://doi.org/10.1007/s11356-021-14726-4

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., 2009. Eutrophication of U.S. Freshwaters: Analysis of Potential Economic Damages. Environ. Sci. Technol. 43, 12–19. https://doi.org/10.1021/es801217q

Dong, L., Gong, C., Huai, H., Wu, E., Lu, Z., Hu, Y., Li, L., Yang, Z., 2023. Retrieval of Water Quality Parameters in Dianshan Lake Based on Sentinel-2 MSI Imagery and Machine Learning: Algorithm Evaluation and Spatiotemporal Change Research. Remote Sens. 15, 5001. https://doi.org/10.3390/rs15205001

Duester, H., 2023. SRTM Downloader Plugin for QGIS.

Egbert, J., Plonsky, L., 2020. Bootstrapping Techniques, in: Paquot, M., Gries, S.Th. (Eds.), A Practical Handbook of Corpus Linguistics. Springer International Publishing, Cham, pp. 593–610. https://doi.org/10.1007/978-3-030-46216-1_24

El Naqa, I., Murphy, M.J., 2015. What Is Machine Learning?, in: El Naqa, I., Li, R., Murphy, M.J. (Eds.), Machine Learning in Radiation Oncology: Theory and Applications. Springer International Publishing, Cham, pp. 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

European Environment Agency., 2012. EEA Catchments and Rivers Network System ECRINS v1.1: rationales, building and improving for widening uses to Water Accounts and WISE applications. Publications Office, LU.

Foley, B., Jones, I.D., Maberly, S.C., Rippey, B., 2012. Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication. Freshwater Biology 57, 278–289. https://doi.org/10.1111/j.1365-2427.2011.02662.x

Fu, B., Lao, Z., Liang, Y., Sun, J., He, X., Deng, T., He, W., Fan, D., Gao, E., Hou, Q., 2022. Evaluating optically and non-optically active water quality and its response relationship to hydro-meteorology using multi-source data in Poyang Lake, China. Ecological Indicators 145, 109675. https://doi.org/10.1016/j.ecolind.2022.109675

Gao, L., Shangguan, Y., Sun, Z., Shen, Q., Shi, Z., 2024. Estimation of Non-Optically Active Water Quality Parameters in Zhejiang Province Based on Machine Learning. Remote Sens. 16, 514. https://doi.org/10.3390/rs16030514

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9

Gholizadeh, M., Melesse, A., Reddi, L., 2016. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. Sensors 16, 1298. https://doi.org/10.3390/s16081298

Google, 2024a. USGS Landsat 5 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog [WWW Document]. Google for Developers. URL https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT05_C02_T1_L2 (accessed 2.22.24).

Google, 2024b. JavaScript and Python Guides | Google Earth Engine | Google for Developers [WWW Document]. URL https://developers.google.com/earth-engine/guides (accessed 3.14.24).

Grossman, Y.L., Ustin, S.L., Jacquemoud, S., Sanderson, E.W., Schmuck, G., Verdebout, J., 1996. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. Remote Sensing of Environment 56, 182–193. https://doi.org/10.1016/0034-4257(95)00235-9

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451, 293–296. https://doi.org/10.1038/nature06592

Guo, H., Huang, J.J., Chen, B., Guo, X., Singh, V.P., 2021. A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. Int. J. Remote Sens. 42, 1841–1866. https://doi.org/10.1080/01431161.2020.1846222

Halliday, S.J., Wade, A.J., Skeffington, R.A., Neal, C., Reynolds, B., Rowland, P., Neal, M., Norris, D., 2012. An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. Science of The Total Environment, Climate Change and Macronutrient Cycling along the Atmospheric, Terrestrial, Freshwater and Estuarine Continuum - A Special Issue dedicated to Professor Colin Neal 434, 186–200. https://doi.org/10.1016/j.scitotenv.2011.10.052

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M.H., Brett, M., Haldane, A., Del Río, J.F., Wiebe, M., Peterson, P.,

Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hassine, K., Erbad, A., Hamila, R., 2019. Important Complexity Reduction of Random Forest in Multi-Classification Problem, in: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). Presented at the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 226–231. https://doi.org/10.1109/IWCMC.2019.8766544

Holtz, Y., 2024. Python Graph Gallery [WWW Document]. The Python Graph Gallery. URL https://python-graph-gallery.com/ (accessed 3.18.24).

Horka, P., Musilova, Z., Holubova, K., Jandova, K., Kukla, J., Rutkayova, J., Jones, J.I., 2023. Anthropogenic nutrient loading affects both individual species and the trophic structure of river fish communities. Frontiers in Ecology and Evolution 10. https://doi.org/10.3389/fevo.2022.1076451

Houser, J.N., Richardson, W.B., 2010. Nitrogen and phosphorus in the Upper Mississippi River: transport, processing, and effects on the river ecosystem. Hydrobiologia 640, 71–88. https://doi.org/10.1007/s10750-009-0067-4

Howarth, R., Chan, F., Conley, D.J., Garnier, J., Doney, S.C., Marino, R., Billen, G., 2011. Coupled biogeochemical cycles: eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. Frontiers in Ecology and the Environment 9, 18–26. https://doi.org/10.1890/100008

Howarth, R.W., Marino, R., 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. Limnology and Oceanography 51, 364–376. https://doi.org/10.4319/lo.2006.51.1_part_2.0364

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Inkscape Project, 2020. Inkscape.

James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., 2023. Linear Model Selection and Regularization, in: James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J. (Eds.), An Introduction to Statistical Learning: With Applications in Python, Springer Texts in Statistics. Springer International Publishing, Cham, pp. 229–288. https://doi.org/10.1007/978-3-031-38747-0_6

Ju, J., Neigh, C., Claverie, M., Skakun, S., Roger, J.-C., Vermote, E., Dungan, J., 2020. Harmonized Landsat Sentinel-2 (HLS) Product User Guide, 2.0. ed.

Kakade, A., Salama, E.-S., Han, H., Zheng, Y., Kulshrestha, S., Jalalah, M., Harraz, F.A., Alsareii, S.A., Li, X., 2021. World eutrophic pollution of lake and river: Biotreatment potential and future perspectives. Environmental Technology & Innovation 23, 101604. https://doi.org/10.1016/j.eti.2021.101604

Kaplan, G., Avdan, U., 2017. Object-based water body extraction model using Sentinel-2 satellite imagery. European Journal of Remote Sensing.

Keith, T.Z., 2019. Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. Routledge.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1137–1143.

Köhler, J., Gelbrecht, J., 1998. Interactions between phytoplankton dynamics and nutrient supply along the lowland river Spree, Germany. SIL Proceedings, 1922-2010 26, 1045–1049. https://doi.org/10.1080/03680770.1995.11900880

Kondratyev, K.Ya., Pozdnyakov, D.V., Pettersson, L.H., 1998. Water quality remote sensing in the visible spectrum. International Journal of Remote Sensing 19, 957–979. https://doi.org/10.1080/014311698215810

Kopáček, J., Hejzlar, J., Porcal, P., Znachor, P., 2021. Biogeochemical causes of sixty-year trends and seasonal variations of river water properties in a large European basin. Biogeochemistry 154, 81–98. https://doi.org/10.1007/s10533-021-00800-z

Kopáček, J., Hejzlar, J., Posch, M., 2013. Factors controlling the export of nitrogen from agricultural land in a large central european catchment during 1900-2010. Environmental Science and Technology 47, 6400–6407. https://doi.org/10.1021/es400181m

Kupssinsku, L.S., Guimaraes, T.T., de Freitas, R., de Souza, E.M., Rossa, P., Marques Jr, A., Veronez, M.R., Gonzaga Jr, L., Cazarin, C.L., Mauad, F.F., 2019. Prediction of chlorophyll-a and suspended solids through remote sensing and artificial neural networks, in: 2019 13TH INTERNATIONAL CONFERENCE ON SENSING TECHNOLOGY (ICST), International Conference on Sensing Technology. Presented at the 13th International Conference on Sensing Technology (ICST), IEEE, New York. https://doi.org/10.1109/icst46873.2019.9047682

Kwast, H. van der, Menke, K., 2022. QGIS for hydrological applications: recipes for catchment hydrology and water management, Second edition. ed. Locate Press, Penticton, BC.

Latt, Z.Z., Wittenberg, H., 2014. Improving Flood Forecasting in a Developing Country: A Comparative Study of Stepwise Multiple Linear Regression and Artificial Neural Network. Water Resour Manage 28, 2109–2128. https://doi.org/10.1007/s11269-014-0600-8

Le, C., Zha, Y., Li, Y., Sun, D., Lu, H., Yin, B., 2010. Eutrophication of Lake Waters in China: Cost, Causes, and Control. Environmental Management 45, 662–668. https://doi.org/10.1007/s00267-010-9440-3

LeDell, E., Poirier, S., 2020. H2O AutoML: Scalable Automatic Machine Learning.

Leggesse, E.S., Zimale, F.A., Sultan, D., Enku, T., Srinivasan, R., Tilahun, S.A., 2023. Predicting Optical Water Quality Indicators from Remote Sensing Using Machine Learning Algorithms in Tropical Highlands of Ethiopia. Hydrology 10, 110. https://doi.org/10.3390/hydrology10050110

Li, D., Liu, S., 2019. Chapter 7 - Detection of River Water Quality, in: Li, D., Liu, S. (Eds.), Water Quality Monitoring and Management. Academic Press, pp. 211–220. https://doi.org/10.1016/B978-0-12-811330-1.00007-7

Li, N., Ning, Z., Chen, M., Wu, D., Hao, C., Zhang, D., Bai, R., Liu, H., Chen, X., Li, W., Zhang, W., Chen, Y., Li, Q., Zhang, L., 2022. Satellite and Machine Learning Monitoring of Optically Inactive Water Quality Variability in a Tropical River. Remote Sensing 14, 5466. https://doi.org/10.3390/rs14215466

Lin, K., Zhu, Y., Zhang, Y., Lin, H., 2019. Determination of ammonia nitrogen in natural waters: Recent advances and applications. Trends in Environmental Analytical Chemistry 24, e00073. https://doi.org/10.1016/j.teac.2019.e00073

Liu, Y., Heuvelink, G.B.M., Bai, Z., He, P., Xu, X., Ding, W., Huang, S., 2021. Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear

regression. Field Crops Research 264, 108098. https://doi.org/10.1016/j.fcr.2021.108098

Ljumović, M., Klar, M., 2015. Estimating expected error rates of random forest classifiers: A comparison of cross-validation and bootstrap, in: 2015 4th Mediterranean Conference on Embedded Computing (MECO). Presented at the 2015 4th Mediterranean Conference on Embedded Computing (MECO), pp. 212–215. https://doi.org/10.1109/MECO.2015.7181905

Magri, S., Ottaviani, E., Prampolini, E., Besio, G., Fabiano, B., Federici, B., 2023. Application of machine learning techniques to derive sea water turbidity from Sentinel-2 imagery. Remote Sens. Appl.-Soc. Environ. 30, 100951. https://doi.org/10.1016/j.rsase.2023.100951

Maier, P.M., Keller, S., 2018. Machine Learning Regression on Hyperspectral Data to Estimate Multiple Water Parameters, in: 2018 9TH WORKSHOP ON HYPERSPECTRAL IMAGE AND SIGNAL PROCESSING: EVOLUTION IN REMOTE SENSING (WHISPERS), Workshop on Hyperspectral Image and Signal Processing. Presented at the 9th Workshop on Hyperspectral Image and Signal Processing - Evolution in Remote Sensing (WHISPERS), IEEE, New York.

Mainstone, C.P., Parr, W., 2002. Phosphorus in rivers — ecology and management. Science of The Total Environment, Water quality functioning of lowland permeable catchments:inferences from an intensive study of the River Kennet and upper River Thames 282–283, 25–47. https://doi.org/10.1016/S0048-9697(01)00937-8

Masek, J., Ju, J., Roger, J.-C., Skakun, S., Vermote, E., Claverie, M., Dungan, J., Yin, Z., Freitag, B., Justice, C., 2021. HLS Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m v2.0. https://doi.org/10.5067/HLS/HLSL30.002

Maúre, E. de R., Terauchi, G., Ishizaka, J., Clinton, N., DeWitt, M., 2021. Globally consistent assessment of coastal eutrophication. Nat Commun 12, 6142. https://doi.org/10.1038/s41467-021-26391-9

McKinney, W., 2010. Data Structures for Statistical Computing in Python. Presented at the Python in Science Conference, Austin, Texas, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Mentlík, P., 2016. Bohemian ForestBohemian Forest: Landscape and People on the Frontier, in: Pánek, T., Hradecký, J. (Eds.), Landscapes and Landforms of the Czech Republic. Springer International Publishing, Cham, pp. 87–100. https://doi.org/10.1007/978-3-319-27537-6_8

Meybeck, M., 1982. Carbon, Nitrogen, and Phosphorus Transport by World Rivers. Am. J. Sci. 282. https://doi.org/10.2475/ajs.282.4.401

Minsky, M., 1961. Steps Toward Artificial Intelligence. Proceedings of the IRE 49, 8–30.

Mooney, C.Z., Duval, R.D., 2006. Bootstrapping: a nonparametric approach to statistical inference, Nachdr. ed, Quantitative applications in the social sciences. Sage Publ, Newbury Park, Calif.

Myers, J.S., Miller, R.L., 2005. Optical Airborne Remote Sensing, in: Miller, R.L., Del Castillo, C.E., Mckee, B.A. (Eds.), Remote Sensing of Coastal Aquatic Environments: Technologies, Techniques and Applications. Springer Netherlands, Dordrecht, pp. 51–67. https://doi.org/10.1007/978-1-4020-3100-7_3

Ng, A.Y., 2004. Feature selection, L1 vs. L2 regularization, and rotational invariance, in: Proceedings of the Twenty-First International Conference on Machine Learning,

ICML '04. Association for Computing Machinery, New York, NY, USA, p. 78. https://doi.org/10.1145/1015330.1015435

Niu, C., Tan, K., Jia, X., Wang, X., 2021. Deep learning based regression for optically inactive inland water quality parameter estimation using airborne hyperspectral imagery. Environmental Pollution 286, 117534. https://doi.org/10.1016/j.envpol.2021.117534

OpenAI, 2023. ChatGPT-4.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Penn, M.R., Pauer, J.J., Mihelcic, J.R., 2009. Biochemical oxygen demand. Environmental and ecological chemistry 2, 278–297.

Peshawa, J.M.A., Rezhna, H.F., 2014. Data Normalization and Standardization: A Technical Report. https://doi.org/10.13140/RG.2.2.28948.04489

Peterson, K.T., Sagan, V., Sloan, J.J., 2020. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience & Remote Sensing 57, 510–525. https://doi.org/10.1080/15481603.2020.1738061

Piotrowski, A.P., Napiorkowski, J.J., 2013. A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. Journal of Hydrology 476, 97–111. https://doi.org/10.1016/j.jhydrol.2012.10.019

Pitcher, G.C., Aguirre-Velarde, A., Breitburg, D., Cardich, J., Carstensen, J., Conley, D.J., Dewitte, B., Engel, A., Espinoza-Morriberón, D., Flores, G., Garçon, V., Graco, M., Grégoire, M., Gutiérrez, D., Hernandez-Ayon, J.M., Huang, H.-H.M., Isensee, K., Jacinto, M.E., Levin, L., Lorenzo, A., Machu, E., Merma, L., Montes, I., SWA, N., Paulmier, A., Roman, M., Rose, K., Hood, R., Rabalais, N.N., Salvanes, A.G.V., Salvatteci, R., Sánchez, S., Sifeddine, A., Tall, A.W., Plas, A.K.V.D., Yasuhara, M., Zhang, J., Zhu, Z.Y., 2021. System controls of coastal and open ocean oxygen depletion. Progress in Oceanography 197. https://doi.org/10.1016/j.pocean.2021.102613

Pretty, J.N., Mason, C.F., Nedwell, D.B., Hine, R.E., Leaf, S., Dils, R., 2003. Environmental Costs of Freshwater Eutrophication in England and Wales. Environ. Sci. Technol. 37, 201–208. https://doi.org/10.1021/es020793k

Pu, F., Ding, C., Chao, Z., Yu, Y., Xu, X., 2019. Water-Quality Classification of Inland Lakes Using Landsat8 Images by Convolutional Neural Networks. Remote Sensing 11, 1674. https://doi.org/10.3390/rs11141674

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation.

Ramachandran, P., Zoph, B., Le, Q.V., 2017. Searching for Activation Functions. https://doi.org/10.48550/arXiv.1710.05941

Razon, L.F., 2018. Reactive nitrogen: A perspective on its global impact and prospects for its sustainable production. Sustainable Production and Consumption 15, 35–48. https://doi.org/10.1016/j.spc.2018.04.003

Records, R.M., Wohl, E., Arabi, M., 2016. Phosphorus in the river corridor. Earth-Science Reviews 158, 65–88. https://doi.org/10.1016/j.earscirev.2016.04.010

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Richards, J.A., 2013. Sources and Characteristics of Remote Sensing Image Data, in: Richards, J.A. (Ed.), Remote Sensing Digital Image Analysis: An Introduction. Springer, Berlin, Heidelberg, pp. 1–26. https://doi.org/10.1007/978-3-642-30062-2_1

Riebeek, Holly, 2013. Historic Landsat 5 Mission Ends | Landsat Science [WWW Document]. URL https://landsat.gsfc.nasa.gov/article/historic-landsat-5-mission-ends/ (accessed 3.27.24).

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. Nature 461, 472–475. https://doi.org/10.1038/461472a

Rodríguez, J.D., Pérez, A., Lozano, J.A., 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 569–575. https://doi.org/10.1109/TPAMI.2009.187

Rosendorf, P., Vyskoč, P., Prchalová, H., Fiala, D., 2016. Estimated contribution of selected non-point pollution sources to the phosphorus and nitrogen loads in water bodies of the Vltava river basin. Soil and Water Research 11, 196–204. https://doi.org/10.17221/15/2015-SWR

Roy, D.P., Li, Z., Zhang, H.K., 2017. Adjustment of Sentinel-2 Multi-Spectral Instrument (MSI) Red-Edge Band Reflectance to Nadir BRDF Adjusted Reflectance (NBAR) and Quantification of Red-Edge Band BRDF Effects. Remote Sensing 9, 1325. https://doi.org/10.3390/rs9121325

Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sensing of Environment 176, 255–271. https://doi.org/10.1016/j.rse.2016.01.023

Ruescas, A.B., Mateo-Garcia, G., Camps-Valls, G., Hieronymi, M., 2018. Retrieval of Case 2 Water Quality Parameters with Machine Learning, in: IGARSS 2018 - 2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, IEEE International Symposium on Geoscience and Remote Sensing IGARSS. Presented at the 38th IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, New York, pp. 124–127.

Růžička, M., Hejzlar, J., Jarošík, J., 2005. Modelling the Hydrodynamics and Water Quality of the Vltava River Reservoir Cascade, in: Informatics for Environmental Protection-Networking Environmental Information. Masaryk University Brno.

Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Maalouf, S., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Reviews 205, 103187. https://doi.org/10.1016/j.earscirev.2020.103187

Sagi, O., Rokach, L., 2020. Explainable decision forest: Transforming a decision forest into an interpretable tree. Information Fusion 61, 124–138. https://doi.org/10.1016/j.inffus.2020.03.013

Santos, C.F.G.D., Papa, J.P., 2022. Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. ACM Comput. Surv. 54, 213:1-213:25. https://doi.org/10.1145/3510413

Schindler, D.W., 1977. Evolution of phosphorus limitation in lakes. Science 195, 260–262. https://doi.org/10.1126/science.195.4275.260

Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., Beaty, K.G., Lyng, M., Kasian, S.E.M., 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. Proceedings of the National Academy of Sciences of the United States of America 105, 11254–11258. https://doi.org/10.1073/pnas.0805108105

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schwoerbel, J., Brendelberger, H., 2022. Einführung in die Limnologie: Stoffhaushalt - Lebensgemeinschaften - Technologie. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-63334-2

Seabold, S., Perktold, J., 2010. statsmodels: Econometric and statistical modeling with Python.

Selman, M., Greenhalgh, S., 2009. Eutrophication: Sources and Drivers of Nutrient Pollution.

Sharma, Sagar, Sharma, Simone, Athaiya, A., 2017. Activation functions in neural networks. Towards Data Sci 6, 310–316.

Shuiwang, D., Shen, Z., Hongyu, H., 2000. Transport of dissolved inorganic nitrogen from the major rivers to estuaries in China. Nutrient Cycling in Agroecosystems 57, 13–22. https://doi.org/10.1023/A:1009896032188

Singhal, G., Bansod, B., Mathew, L., Goswami, J., Choudhury, B.U., Raju, P.L.N., 2019. Chlorophyll estimation using multi-spectral unmanned aerial system based on machine learning techniques. Remote Sens. Appl.-Soc. Environ. 15, 100235. https://doi.org/10.1016/j.rsase.2019.100235

Skorbiłowicz, M., Ofman, P., 2014. Seasonal changes of nitrogen and phosphorus concentration in supraśl river. Journal of Ecological Engineering 15, 26–31. https://doi.org/10.12911/22998993.1084172

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environ Sci & Pollut Res 10, 126–139. https://doi.org/10.1065/espr2002.12.142

Smith, V.H., Joye, S.B., Howarth, R.W., 2006. Eutrophication of freshwater and marine ecosystems. Limnology and Oceanography 51, 351–355. https://doi.org/10.4319/lo.2006.51.1_part_2.0351

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing 14, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Soana, E., Gervasio, M.P., Granata, T., Colombo, D., Castaldelli, G., 2024. Climate change impacts on eutrophication in the Po River (Italy): Temperature-mediated reduction in nitrogen export but no effect on phosphorus. Journal of Environmental Sciences (China) 143, 148–163. https://doi.org/10.1016/j.jes.2023.07.008

Soballe, D.M., Kimmel, B.L., 1987. A Large-Scale Comparison of Factors Influencing Phytoplankton Abundance in Rivers, Lakes, and Impoundments. Ecology 68, 1943–1954. https://doi.org/10.2307/1939885

Tan, Z., Ren, J., Li, S., Li, W., Zhang, R., Sun, T., 2023. Inversion of Nutrient Concentrations Using Machine Learning and Influencing Factors in Minjiang River. Water 15, 1398. https://doi.org/10.3390/w15071398

TensorFlow Developers, 2024. TensorFlow. https://doi.org/10.5281/ZENODO.4724125

Tesfaye, M., Souza, A.T., Soukalová, K., Šmejkal, M., Hejzlar, J., Prchalová, M., Říha, M., Muška, M., Vašek, M., Frouzová, J., Blabolil, P., Boukal, D.S., Kubečka, J., 2023. Somatic growth of pikeperch (*Stizostedion lucioperca*) in relation to variation in temperature and eutrophication in a Central Europe Lake. Fisheries Research 267, 106824. https://doi.org/10.1016/j.fishres.2023.106824

Tesoriero, A.J., Robertson, D.M., Green, C.T., Böhlke, J.K., Harvey, J.W., Qi, S.L., 2024. Prioritizing river basins for nutrient studies. Environmental Monitoring and Assessment 196. https://doi.org/10.1007/s10661-023-12266-7

Tian, D., Zhao, X., Gao, L., Liang, Z., Yang, Z., Zhang, P., Wu, Q., Ren, K., Li, R., Yang, C., Li, S., Wang, M., He, Z., Zhang, Z., Chen, J., 2024. Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, South China. Environmental Pollution 342. https://doi.org/10.1016/j.envpol.2023.123104

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tobias, R.D., 1995. An introduction to partial least squares regression, in: Proceedings of the Twentieth Annual SAS Users Group International Conference. Citeseer, pp. 1250–1257.

Topp, S.N., Pavelsky, T.M., Jensen, D., Simard, M., Ross, M.R.V., 2020. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. Water 12, 169. https://doi.org/10.3390/w12010169

Tsamardinos, I., Greasidou, E., Borboudakis, G., 2018. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. Mach Learn 107, 1895–1922. https://doi.org/10.1007/s10994-018-5714-4

Tsirkunov, V.V., Nikanorov, A.M., Laznik, M.M., Dongwei, Z., 1992. Analysis of long-term and seasonal river water quality changes in Latvia. Water Research 26, 1203–1216. https://doi.org/10.1016/0043-1354(92)90181-3

USGS, 2012. Landsat Data Continuity Mission.

Vidal, T., Schiffer, M., 2020. Born-Again Tree Ensembles, in: Proceedings of the 37th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 9743–9753.

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H., Tilman, D.G., 1997. Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. Ecological Applications 7, 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Vymazal, J., 2007. Removal of nutrients in various types of constructed wetlands. Science of The Total Environment, Contaminants in Natural and Constructed Wetlands: Pollutant Dynamics and Control 380, 48–65. https://doi.org/10.1016/j.scitotenv.2006.09.014

Vystavna, Y., Hejzlar, J., Kopáček, J., 2017. Long-term trends of phosphorus concentrations in an artificial lake: Socio-economic and climate drivers. PLOS ONE 12, e0186917. https://doi.org/10.1371/journal.pone.0186917

Vystavna, Y., Paule-Mercado, M.C., Schmidt, S.I., Hejzlar, J., Porcal, P., Matiatos, I., 2023. Nutrient dynamics in temperate European catchments of different land use under changing climate. Journal of Hydrology: Regional Studies 45, 101288. https://doi.org/10.1016/j.ejrh.2022.101288

Wang, H., García Molinos, J., Heino, J., Zhang, H., Zhang, P., Xu, J., 2021. Eutrophication causes invertebrate biodiversity loss and decreases cross-taxon congruence across anthropogenically-disturbed lakes. Environment International 153, 106494. https://doi.org/10.1016/j.envint.2021.106494

Waskom, M., 2021. seaborn: statistical data visualization. JOSS 6, 3021. https://doi.org/10.21105/joss.03021

Whitehead, P.G., Wilby, R.L., Battarbee, R.W., Kernen, M., Wade, A.J., 2009. A review of the potential impacts of climate change on surface water quality. Hydrological Sciences Journal 54, 101–123. https://doi.org/10.1623/hysj.54.1.101

Withers, P.J.A., Jarvie, H.P., 2008. Delivery and cycling of phosphorus in rivers: A review. Science of The Total Environment 400, 379–395. https://doi.org/10.1016/j.scitotenv.2008.08.002

Wrigley, R.C., Horne, A.J., 1974. Remote sensing and lake eutrophication. Nature 250, 213–214. https://doi.org/10.1038/250213a0

Wu, H., Chen, J., Xu, J., Zeng, G., Sang, L., Liu, Q., Yin, Z., Dai, J., Yin, D., Liang, J., Ye, S., 2019. Effects of dam construction on biodiversity: A review. Journal of Cleaner Production 221, 480–489. https://doi.org/10.1016/j.jclepro.2019.03.001

Xia, X., Zhang, S., Li, S., Zhang, Liwei, Wang, G., Zhang, Ling, Wang, J., Li, Z., 2018. The cycle of nitrogen in river systems: sources, transformation, and flux. Environ. Sci.: Processes Impacts 20, 863–891. https://doi.org/10.1039/C8EM00042E

Yadav, S., Shukla, S., 2016. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification, in: 2016 IEEE 6th International Conference on Advanced Computing (IACC). Presented at the 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 78–83. https://doi.org/10.1109/IACC.2016.25

Yan, D., Wang, K., Qin, T., Weng, B., Wang, H., Bi, W., Li, X., Li, M., Lv, Z., Liu, F., He, S., Ma, J., Shen, Z., Wang, J., Bai, H., Man, Z., Sun, C., Liu, M., Shi, X., Jing, L., Sun, R., Cao, S., Hao, C., Wang, L., Pei, M., Dorjsuren, B., Gedefaw, M., Girma, A., Abiyu, A., 2019. A data set of global river networks and corresponding water resources zones divisions. Sci Data 6, 219. https://doi.org/10.1038/s41597-019-0243-y

Yang, X.-S., 2021. Chapter 6 - Genetic Algorithms, in: Yang, X.-S. (Ed.), Nature-Inspired Optimization Algorithms (Second Edition). Academic Press, pp. 91–100. https://doi.org/10.1016/B978-0-12-821986-7.00013-5

Zhan, X., Liang, X., Xu, G., Zhou, L., 2013. Influence of plant root morphology and tissue composition on phenanthrene uptake: Stepwise multiple linear regression analysis. Environmental Pollution 179, 294–300. https://doi.org/10.1016/j.envpol.2013.04.033

Zhang, H., Lin, C., Lei, P., Shan, B., Zhao, Y., 2015. Evaluation of river eutrophication of the Haihe River Basin. Huanjing Kexue Xuebao/Acta Scientiae Circumstantiae 35, 2336–2344. https://doi.org/10.13671/j.hjkxxb.2015.0025

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L., 2022. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health 1, 107–116. https://doi.org/10.1016/j.eehl.2022.06.001

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Zou, J., Han, Y., So, S.-S., 2009. Overview of Artificial Neural Networks, in: Livingstone, D.J. (Ed.), Artificial Neural Networks: Methods and Applications. Humana Press, Totowa, NJ, pp. 14–22. https://doi.org/10.1007/978-1-60327-101-1_2

CZECH UNIVERSITY OF LIFE SCIENCESPRAGUE

FACULTY OF ENVIRONMENTSL SCIENCES

DIPLOMA THESIS

2024   Jonas Niese