Jessica CANCHIG https://orcid.org/0009-0001-6197-7929
Rinda KUSTINA✉ https://orcid.org/0009-0003-8118-4082
Mateusz GRYGORUK https://orcid.org/0000-0001-6465-9697

Warsaw University of Life Sciences – SGGW, Institute of Environmental Engineering, Poland

Developing an empirical model for assessment of total nitrogen inflow to rivers and lakes in the Biebrza river watershed, Poland

Introduction

In recent years, the discharge of nutrients has been identified as a significant cause of pollution in European water resources, affecting both drinking water and aquatic ecosystems (Grizzetti et al., 2017). While agriculture is a primary contributor to this issue, other sources, including human and industrial wastewater effluent, also contribute to the problem (Wuijts, Fraters, Boekhold & van Duijnen, 2022). However, crop production has been the largest contributor to the increase of the amount of nitrogen entering in the biospheric cycle (Sutton et al., 2011). Moreover, European vegetable production systems overused it, based on the experience of growers and technical advisors. It leads to nitrate leaching and eutrophication of surface water bodies (Thompson, Incrocci, van Ruijven & Massa, 2020), and the leaching of nitrogen from the surface into surface water bodies is a common form of diffuse nitrogen pollution that contaminates water resources (Almasri, 2007). Where the

main source of nitrogen is based on the use of fertilizers, with Europe and Asia being the main producers of nitrogen fertilizers (Llive et al., 2015). It is important to note that plants cannot take advantage of all the nitrogen present in the soil, which has significant implications. The amount of nitrogen that is used can vary from 25% to 85%, depending on the type of crop and the agricultural techniques used (Ávila & Sansores, 2003). Be that as it may, the deterioration of water quality is a global issue, and everyone contributing to pollution, including agriculture, industry, transport, and amenity managers, must be made aware of this fact (Knapp, 2005).

In addition, to maximize crop production, an excessive amount of nitrogenous fertilizer is often applied to the soil, significantly increasing nitrogen loss through stormwater, according to report by the Pan American Health Organization and the World Health Organization in 1980 (Ávila & Sansores, 2003). Due to all mentioned factors, the European Union (EU) has implemented ambitious water policies aimed at safeguarding and revitalizing aquatic ecosystems through the Water Framework Directive and the Marine Strategy Framework Directive (Grizzetti et al., 2021). Currently, over 50% of water bodies in the EU do not meet the ecological standards mandated by the Directive 2000/60/EC (so-called Water Framework Directive), with nutrient enrichment being a primary contributor to the degradation (Poikane et al., 2019). Likewise, chemical pollution (49%) is the main impact on surface water bodies, followed by altered habitats due to morphological changes (40%) and nutrient pollution (28%). This nutrient enrichment causes eutrophication, which in turn leads to the loss of aquatic biodiversity (Bednarek, Szklarek & Zalewski, 2014). What is more, excessive nutrient enrichment can be dangerous for human health, e.g. owing to toxic algal blooms, and can impair the use of water for drinking and bathing (*EEA Report*, 7/2018).

The case of Biebrza river is identical, where historically people used to cultivate this area for agricultural purposes. However, since 1960, these activities have increased significantly due to the discontinuation of traditional farming practices like hand mowing and livestock grazing. In fact, some parts of the wetlands were even drained to make way for farming (Sucholas, Molnár, Łuczaj & Poschlod, 2022). Also, it is important to mention that the Ramsar Convention has designated the Biebrza river as a wetland site of global significance (Dembicz, Kozub, Bobrowska & Dengler, 2020), it is the largest peatland complex in western and central Europe, and its peat soils have undergone degradation due to human activities (Razowska-Jaworek & Sadurski, 2014). The valuable ecosystems in this place are not just limited to natural peatlands, but also include large open semi-meadows resulting from extensive agriculture. When there are changes in water conditions or extensive agriculture is discontinued, the meadows and pastures undergo a transformation into tall herb vegetation and eventually, reed. In some parts of

the Biebrza river valley, this process leads to the succession of shrubs and forest over the non-forest ecosystems of peatlands (Świątek, Szporak, Chormański & Okruszko, 2008). Some of its tributaries have been modified to be drainage channels or to regulate their flow (Okruszko, 2005). When these soils are drained, they can release significant amounts of nitrogen and other nutrients into water bodies. Peat soils can store significant amounts of nutrients such as nitrogen. The cessation of agriculture has also allowed the growth of bushes and trees, further accelerating the degradation of peat soils and the release of mineral nitrogen into groundwater (Razowska-Jaworek & Sadurski, 2014).

Therefore, as nitrogen fertilizers are not fully absorbed by crops, it is important to describe an approach to control such loss before it enters the river basin, and thus substantially reduce the nitrogen load (Cai et al., 2014). Although several empirical models for modeling nutrients discharge to surface water exist (Naturstyrelsen, 2014), it has been observed that its application may not be universal, as the field observations may deviate significantly from the model predictions in certain cases. Mathematical methods are commonly used in connection with measured data for the evaluation of water quality (Pei-Yue, Hui & Jian-Hua, 2010). Even models based on the geographic information systems (GIS) have been developed to simulate nitrogen inputs in lakes such as the InVEST model (Yang et al., 2019). Although physical models have made significant progress in evaluating nitrogen load (Tan et al., 2023), their application is limited due to the need for extensive data and their accuracy (Duan et al., 2013). Hence, developing a scientifically based assessment model for the prediction of nitrogen is crucial for identifying sources of nitrogen and for preventing and controlling pollution in the watershed (Tan et al., 2023).

It is the same condition as for the rivers in the Biebrza watershed, lakes also have a pivotal role in contributing to the reduction of nutrients. The characteristics of lakes allow for an expansion of physical, chemical, and biological processes to predict the responses of large lakes based on the results from smaller ones (Schindler, 2012). Lakes are depositional environments, and the material that accumulates in the sediments of these water bodies can be deposits from outside the watershed (Downing et al., 2008). It can act as sinks for nutrients, absorbing and storing large amounts of nitrogen in their sediments and water. Additionally, they can provide a habitat for organisms that feed on nutrients, such as phytoplankton and zooplankton, thereby reducing the amount of nitrogen available in the water (Rekha, Raj, Aparna, Bindu & Anjaneyulu, 2005). Consequently, it is important to analyze the behavior of rivers and lakes in Biebrza. Therefore, there is a need to develop a location-specific empirical model that can effectively capture the site-specific factors that influence the nitrogen loss. In this study, we propose a modified empirical model using field observations to estimate the total

nitrogen load in the Biebrza catchment in the rivers and lakes. The modified model is expected to provide more accurate predictions by incorporating the site-specific factors that influence the nitrogen load and it will provide valuable information for better managing and protecting the water resources in the region.

Material and methods

Site description

The Biebrza river is reputable as the most significant and invaluable riverside ecosystem in Europe. Its catchment area spans a vast expanse of approx. 7,250 km² with a length of 164 km in northeast Poland (Fig. 1), making it one of the region's most extensive catchments. Comprising a mix of agricultural land, forests, wetlands, and mires, agriculture dominates the land use of the catchment. Furthermore,
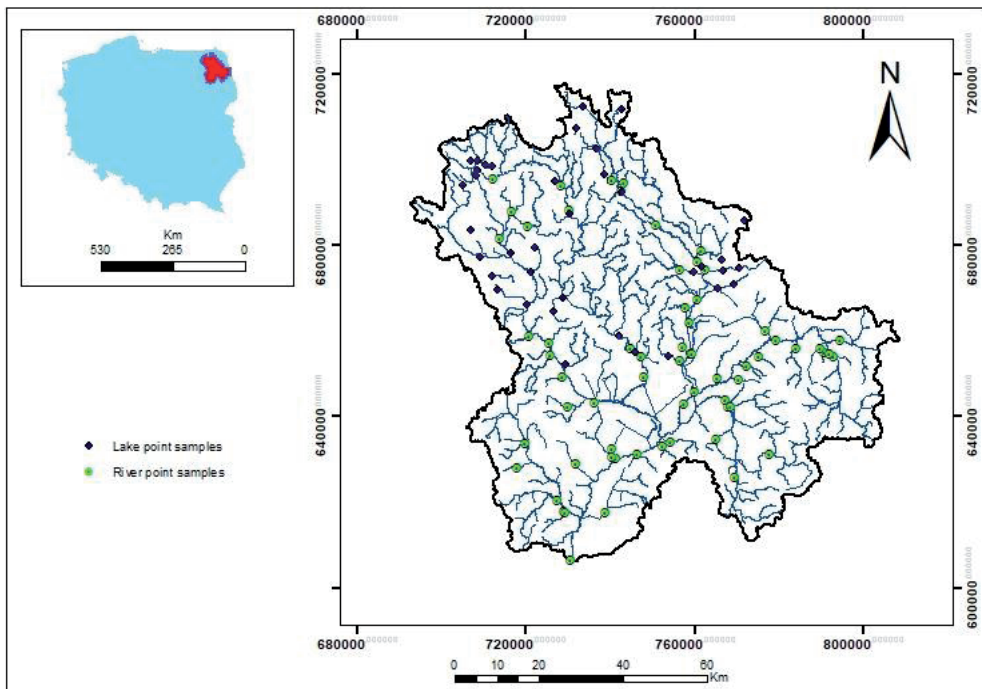

FIGURE 1. Location of study area in Biebrza catchment and the monitoring points in the rivers and lakes from 2005–2021

Source: own elaboration.

the catchment area houses several protected areas, such as the Biebrza National Park and the Narew National Park, that are integral to maintaining the ecological balance of the river and its surrounding environment (Wassen, Barendregt, Palczynski, de Smidt & de Mars, 1992). Nonetheless, environmental pressures, including land use changes, pollution, and climate change, can impact the water quality and ecological health in this area, warranting comprehensive monitoring and management of the catchment. The Biebrza river catchment is a major source of nutrients to the Baltic Sea, with significant inputs of nitrogen and phosphorus from agricultural and urban sources, resulting in eutrophication and reduced water quality (Xu et al., 2021).

According to Figure 2, both in the rivers and lakes, the highest load of nitrogen accumulates in the central part of the Biebrza river basin, represented by the red color. Given that the empirical model is mainly based on agricultural pollution, there are other parameters that were not considered and could have influenced these high concentrations, such as emissions from factories, vehicles, and burning of fuels, leaching from livestock manure, leachate percolation
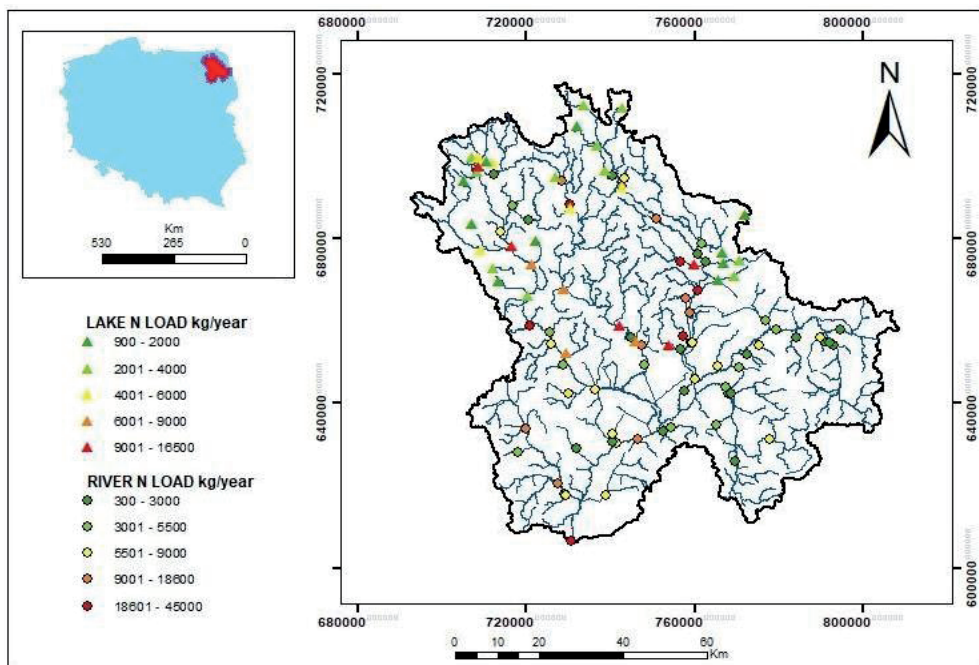

FIGURE 2. Location of study area in Biebrza catchment and the nitrogen load in the rivers and lakes from 2005–2021

Source: own elaboration.

from landfills, or discharge of wastewater (Silva, Cobelas & Gonzáles, 2017). Meanwhile, in the northern part of the basin, the nitrogen load varies between intermediate values. It is in these areas that agriculture covers most of the sub-catchment. However, in the southern part, most of the samples taken show low nitrogen values, where the agricultural area covers a smaller surface area of the sub-catchments in these areas.

Nitrogen load modeling

An empirical model is a mathematical representation of a real-world system, developed through observation or experimentation. Empirical models are used in scientific research and engineering to describe complex systems and make predictions about their behavior. The accuracy and the utility of an empirical model are dependent on the quality and quantity of data utilized in its development, as well as the assumption and simplifications made during the modeling process. Empirical models are iteratively refined and enhanced over time as new data becomes available and the underlying understanding of the modeled system evolves.

The total nitrogen loss was estimated using the empirical model provided by Naturstyrelsen (2014; Eq. 1):

$$N_{loss} = 1.124 \cdot \exp[-3.080 + 0.758 \cdot \ln(H) - 0.0030S + 0.0249D], \tag{1}$$

where N_{loss} is the nitrogen loss, measured in kg·ha^{-1}·year^{-1}, H is the annual run-off, measured in mm, S is the percentage of sandy soil, where the shapefile of global hydrologic soil groups (HYSOGs250m) for the curve number-based runoff modeling was used and processed in ArcTools, D is the percentage of agriculture, obtained from the TIFF Corine Land Cover – 100 m in 2018, the same that was converted into shapefile to be processed in ArcGIS.

Once the equation was applied, it was obtained an average nitrogen loss of 5.85 kg·ha^{-1}·year^{-1} and 6.2 kg·ha^{-1}·year^{-1} for rivers and lakes respectively, nitrogen total load is the total nitrogen loads in kg·year^{-1}, nitrogen total load can be calculated by nitrogen loss multiplied by the area of the sub-catchment of each motoring point, measured in ha.

Due to the lack of data on the discharge of the river, the formula of the rational method was applied for the calculation of the discharge average.

$$Q = C \cdot P \cdot A, \tag{2}$$

where Q is the average annual discharge, measured in $m^3 \cdot year^{-1}$, C is the runoff dimensionless coefficient obtained in 17 area Biebrza sub-basins, which data from the research by Venegas, Marcinkowski, Piniewski and Grygoruk (2022), P is the annual precipitation, measured in m, which was observed from 1951–2021 that were sourced by the input data set of meteorological data on rainfall totals were daily data from hydrological and meteorological monitoring (IMGW-PIB), which were summed up to individual calendar years.

These data were determined using the data of the characteristics of water balance components, raw water balance in water gauge catchments and within the boundaries of the Biebrza National Park and its buffer zone (Venegas et al., 2022), for each monitoring point. Variable A is the area of the sub-catchment determined for each sample point c. For this, the DEM image was downloaded in the STRM (Shuttle radar topographic mission). The outcomes derived from the application of the empirical model will be compared against the observational data obtained from the Regional Inspectorate of Environmental Conservation (*Wojewódzki Inspektorat Ochrony Środowiska*). The comparison will be predicted in the respective nitrogen load trends, the percentage deviation in nitrogen load between the nitrogen loads predicted by the empirical model and those obtained from the observational data, the mean nitrogen load values, and an array of statistical techniques such as the correlation coefficient, determination coefficient, and root mean square error. The original empirical model was calibrated by means of a trial-and-error process utilizing the square root method, based on data spanning the timeframe of 2005–2015 for rivers and 2008–2019 for lakes. The validation of the empirical model is carried out utilizing the data from 2016–2021 for rivers and 2020–2021 for lakes.

Results

Calculated total nitrogen loads

The calculations were derived from data obtained from 234 samples for rivers and 41 samples for lakes situated in the Biebrza river region. Total nitrogen load is expressed in $kg \cdot year^{-1}$, and it was calculated by Eq. (2) and multiplied by the data obtained from monitoring the water quality from the Regional Inspectorate of

Environmental Conservation. Nitrogen load observed is compared with the existing equation results. The values obtained from these calculations can be used to establish new equations that close the observed nitrogen loads results.

The results obtained from the use of the empirical model, its calibration and validation are shown in Figure 3. Nitrogen data taken in 2005 and 2015 were used to calibrate the model using linear regression. Figure 3a shows the observed nitrogen load and the calculated one using the empirical method Eq. (1), while Figure 3b shows the calculated nitrogen load and the calibrated model (Eq. 3). In Table 1, the coefficient of determination (R^2) has the same value, the calibrated model presents a steeper slope, and the data are also distributed better. Figures 3c and 3d show the data from 2016 to 2021, which were used to validate and determine the operation of the new empirical model, where the R^2 has a higher value, which implies a significant improvement. For the validated data (0.82), R^2 is higher than that calculated one for the calibrated data (0.81). Furthermore, the accuracy of the model improved
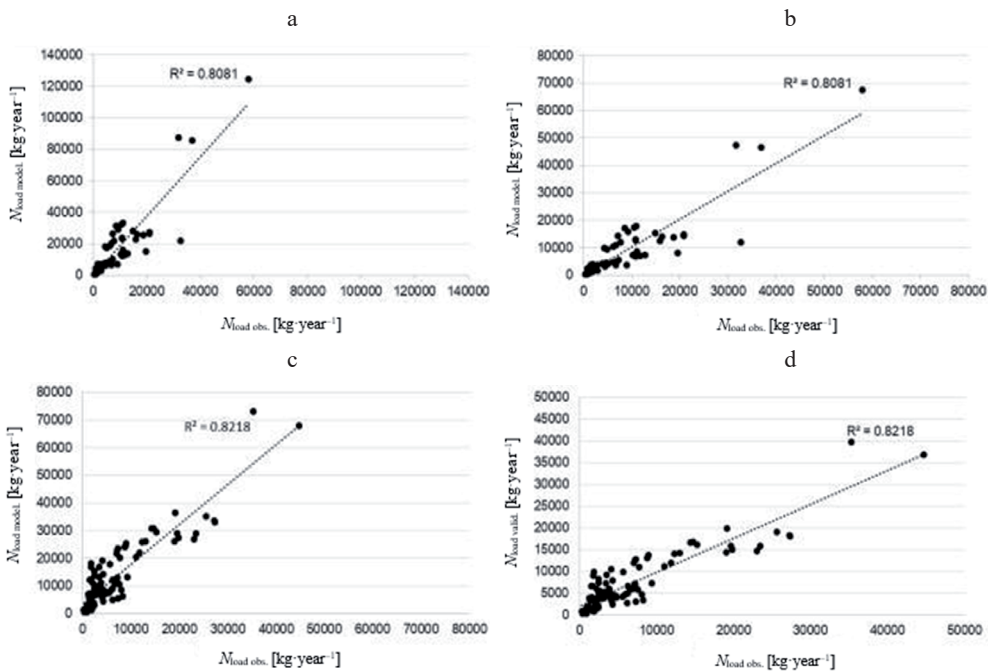

FIGURE 3. Linear regression of rivers monitoring points: a – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load modeling ($N_{\text{load model.}}$) in 2005–2015; b – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load calibrated ($N_{\text{load calib.}}$) in 2005–2015; c – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load modeling ($N_{\text{load model.}}$) in 2016–2021; d – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load validated ($N_{\text{load valid.}}$) in 2016–2021

Source: own elaboration.

significantly as the root mean square error (*RMSE*) value decreased for both calibration and validation (Table 1). Therefore, the model is accurately predicting real values, obtaining R^2 values close to 1 and where the *RMSE* value decreases, more than 30%. In addition, the data dispersion improved significantly, by approximately more than 50%, after the initial model calibration, indicating a clearer and more predictable relationship between them.

TABLE 1. Statistical measures of the comparison of results obtained in calculations with the use of original empirical model and improved empirical model of the total nitrogen load for rivers

Specification	Statistical measure	Modeling	Calibration and validation
Calibration 2005–2015	coefficient of determination (R^2)	0.81	0.81
	root mean square error (*RMSE*)	14 743.05	5 022.38
	standard deviation (*SD*)	21 281.05	11 534.44
	Nash–Suthcliffe's model efficiency coefficient (*NSE*)	−1.14	0.75
Validation 2016–2021	coefficient of determination (R^2)	0.82	0.82
	root mean square error (*RMSE*)	8 847.55	3 342.31
	standard deviation (*SD*)	12 440.31	6 742.71
	Nash–Suthcliffe's model efficiency coefficient (*NSE*)	−0.29	0.82

Source: own elaboration.

Similarly, the values calculated with the initial model, without calibration, show a negative Nash–Suthcliffe's model efficiency coefficient (*NSE*), implying that the model is not suitable for predicting observed data. However, through calibration, this coefficient shows us that the model has a good fit and a robust predictive capacity.

Based on Figure 4b, watersheds that present a percent-age close to 33–66% show a more concentrated distribution of nitrogen load data, i.e., there is less variability in these data, and the distribution is symmetrical. However, the calculated nitrogen load for watersheds with a high percentage of agriculture shows widely distributed data, indicating high variability in the data. Regarding the data calculated for watersheds with a low percentage of agricultural area, the distribution of nitrogen load is asymmetrical and even contains outliers. Therefore, the empiri-
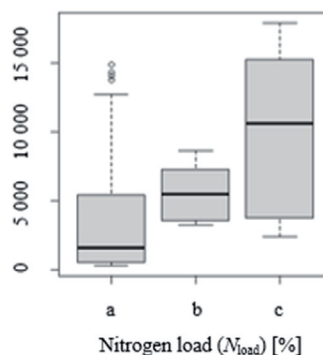

FIGURE 4. Distribution of nitrogen load (N_{load}) by percentage of agriculture: a – 0–33% of agriculture land; b – 33–66% of agriculture land; c – 66–100% of agriculture land

Source: own elaboration.

cal model fits the watersheds with an agriculture land ranging from the values already mentioned better.

The nitrogen load from observations and modeling for lakes is shown in Figure 5. The significant difference between the observed and modeled results indicates further development of the existing empirical models to accurately represent the studied area.
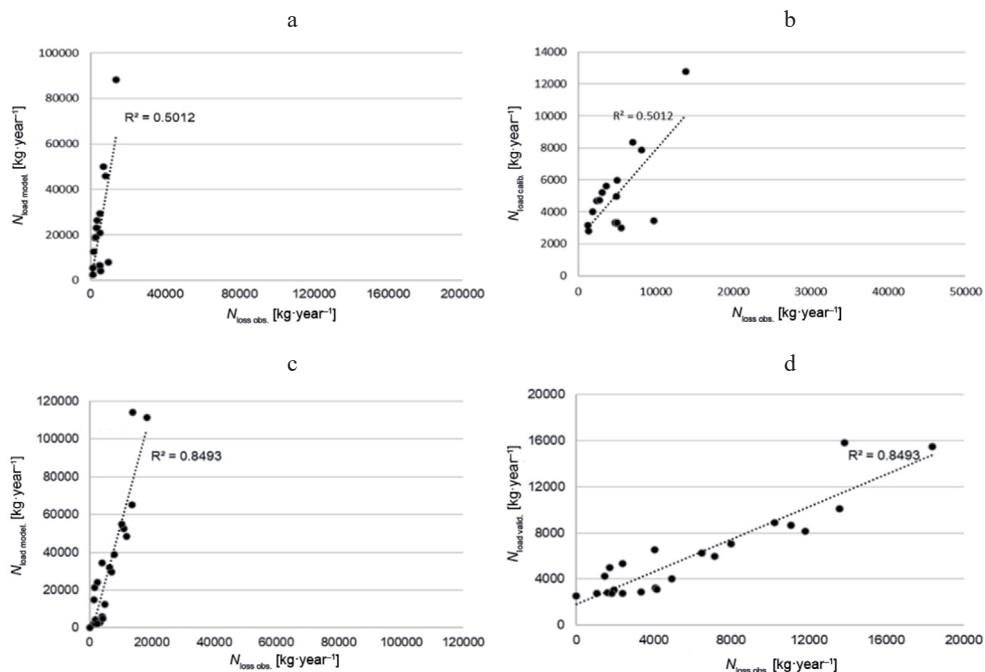

FIGURE 5. Linear regression of lakes monitoring points: a – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load modeling ($N_{\text{load model.}}$) in 2008–2019; b – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load calibrated ($N_{\text{load calib.}}$) in 2008–2019; c – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load modeling ($N_{\text{load model.}}$) in 2020–2021; d – nitrogen load observed ($N_{\text{load obs.}}$) and nitrogen load validated ($N_{\text{load valid.}}$) in 2020–2021

Source: own elaboration.

The correlation between the observed and modeled nitrogen load is depicted in Figure 5a, with applied calibration (2008–2019) and validation (2020–2021). The results show that both graphs have the same R^2 value, which is 0.5012 and 0.8493, respectively. The correlation quality can be determined by examining the R^2 value and correlation line. Based on the graph, the R^2 value indicates comparable results between the two figures, but the correlation line shows that Figure 5b performs better than Figure 5a, which can be attributed to the utilization of a new empirical model. In addition, Figure 5b

TABLE 2. Statistical measures of the comparison of results obtained in calculations with the use of original empirical model and improved empirical model of the total nitrogen load for lakes

Specification	Statistical measures	Modeling	Calibration and validation
Calibration 2008–2019	coefficient of determination (R^2)	0.5012	0.5012
	root mean square error (*RMSE*)	25 668.16	2 256.76
	standard deviation (*SD*)	21 427.88	2 491.61
	Nash–Suthcliffe's model efficiency coefficient (*NSE*)	–64.34	0.5
Validation 2020–2021	coefficient of determination (R^2)	0.8493	0.8493
	root mean square error (*RMSE*)	36 483.58	2055
	standard deviation (*SD*)	32 191.68	3 743.22
	Nash–Suthcliffe's model efficiency coefficient (*NSE*)	–54.43	0.82

Source: own elaboration.

exhibits better results, other outcomes obtained from utilizing the newly calibrated and validated model are displayed in Figure 5d. Moreover, the Nash–Suthcliffe's values indicate a nitrogen load by the existing empirical model and observed for 2008–2019 and 2020–2021 is –64.34 and –54.43, respectively. While the new empirical model and observed is 0.5 for period 2008–2019 and 0.82 for 2020–2021, which are close to 1 that indicates the new empirical model has suitable results to the observed load.
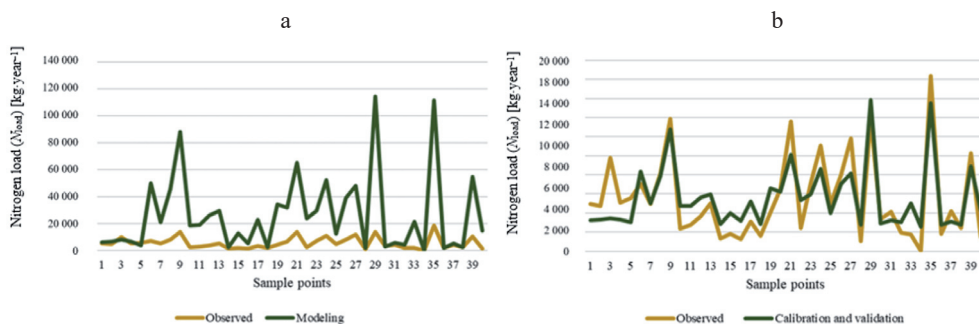

FIGURE 6. Sample points in the lake in 2008–2020: a – nitrogen load observed and modeling; b – nitrogen load observed, calibration, and validation

Source: own elaboration.

Figure 6a depicts the nitrogen load results through both observations and empirical models for each sample point within the lake. It is apparent that the calculated nitrogen load values from the empirical model are higher compared to the observation results. On the other hand, Figure 6b demonstrates outcomes from the same source, but it is compared to the results from the newly calibrated empirical model.

By using the new empirical model, the nitrogen results approach the observation results. This is further supported by the *RMSE* calculations in Table 2 which indicate the decreasing of the *RMSE* 92% for 2008–2019 and 95% decrease for 2020–2021 from the existing empirical model results. The determination of data approximation is assessed through the analysis of the R^2 and the *RMSE* values. According to Kutner, Nachtsheim, Neter and Li (2005), and Hair, Black, Babin, Anderson and Tatham (2019), that the coefficient of determination, or the R^2, is a commonly used matrix to assess the goodness of fit of a regression model to the data. Measures of the R^2 for the proportion of the variance in the dependent variable that can be explained by independent variables in the model, with values ranging from 0 to 1. The high value of the R^2 indicates a good fit of the model to the observed data, with more variance explained by the independent variables. Whereas it is important to consider the limitations or potential errors with the new equation when applied to lakes in the Biebrza catchment. The equation consistently yields results that are 1.3% lower than the actual values, which indicates a systematic discrepancy in the estimation.

Calibration and validation of the empirical model

Although the results of the original empirical equation showed values that follow a trend equal to those observed, the difference between them is huge. Therefore, it was necessary to calibrate said equation to adapt it to the conditions presented by the Biebrza catchment in rivers and lakes. For which, a linear regression was performed in RStudio and using the value of the intercept coefficient, the equation was adjusted. Resulting in the original equation divided by the established values in the case of the rivers.

$$N_{loss} = 1.124 \cdot \exp(-3.080 + 0.758 \cdot \ln(H) - 0.0030S + 0.0249D) / 1.845 + 70.38. \quad (3)$$

In addition, the results of the original equation for the lakes around Biebrza catchment need to be developed to approach the results of observation. Therefore, a better equation should be used in this case (Eq. 4):

$$N_{loss} = 1.124 \cdot \exp(-3.080 + 0.758 \cdot \ln(H) - 0.0030S + 0.0249D) / 8.6 + 2,541. \quad (4)$$

Likewise, these new equations were validated using the water quality data for the years 2016 to 2021 (Eq. 3) for rivers and 2020–2021 for lakes (Eq. 4). Obtaining values that fit better and with greater precision (Tables 1 and 2). It is important to emphasize that these values were not used for calibration.

Discussion

Agriculture is one of the factors contributing to nitrogen load in the Biebrza river watershed. Therefore, tools are needed to identify factors contributing to nitrogen load to establish sustainable guidelines for minimizing the nitrogen impact. Thus, the use of this empirical model will serve as a tool to implement guidelines that can be used by farmers in the surrounding areas and by authorities responsible for establishing agricultural policies in the area. Fertilizers play a crucial role in agriculture, but they have negative effects on the water bodies, such as eutrophication. Therefore, they should be applied considering optimal economic and environmental performance (Berger et al., 2020).

Based on the equation used (Eq. 1), the results calculated differ from the observation results. Consequently, a new equation that can be tailored to approach the total nitrogen derived from observations is necessary. According to Figures 3 and 5, the average nitrogen load from observed data is approx. 7,000 and approx. 5,000 kg·year^{-1}, while the existing model is approx. 13,000 kg·year^{-1} and approx. 26,000 kg·year^{-1} (Eq. 1), and the new empirical model is approx. 7,000 kg·year^{-1} and approx. 5,000 kg·year^{-1} (Eqs 3 and 4) for the river and lakes respectively. The nitrogen load from the existing empirical model (Eq. 1) has increased almost twice in the river and by five times in the lakes the actual result, while the new empirical model (Eqs 3 and 4) shows a difference of less than 4% from actual result. To establish the new equation, a linear regression analysis was used, which is a statistical technique that allows the analysis to model the relationship between a dependent variable and independent variables. Furthermore, it evaluates the strength of the association between the observed and modeling data. It assumes that the relationship between the variables can be described by a linear equation and seeks to find the best-fitting line summarizes this relationship (Aloui, Jammazi & Nguyen, 2014).

Nitrogen load calculations based on observations involve determining the catchment area for each sample, knowing the precipitation and runoff around the Biebrza watershed, and multiplication of these values by the nitrogen content in each observation. This leads to the calculation of nitrogen load in kilogram per year, which serves as a reference point for the observed data. On the other hand, nitrogen load calculations can be done using an equation (Eq. 1), which does not require direct field observations but only needs precipitation in the catchment area, runoff, and samples from the area. In the case of the rivers 11 sub-basins were used to modify the original empirical model (Eq. 3) to establish an equation that fits the conditions of the Biebrza area. Furthermore, the determination of data approximation is assessed through the analysis of R^2 and *RMSE* values. According to Kutner

etal. (2005), and Hair et al. (2019), that the coefficient of determination, or the R^2, is a commonly used matrix to assess the goodness of fit of a regression model to the data. Measures of the R^2 for the proportion of the variance in the dependent variable that can be explained by independent variables in the model, with values ranging from 0 to 1. The high value of R^2 indicates a good fit of the model to the data, with more variance explained by the independent variables. To verify its efficiency, R^2 was first determined and showed no variation with respect to observed values compared to the modified model during the years 2005 to 2015. Validation performed for the following years, 2016 to 2021, showed an improvement in this value. Therefore, to ensure more accurate comparison between observed data, existing model, and new empirical model, an analysis using the Nash–Suthcliffe's coefficient which is a statistical measure of goodness-of-fit between observed and modeled values in hydrological and environmental modeling. It is calculated as the ratio of the sum of squared differences between observed and modeled values to the variance of the observed values. According to Nash and Sutcliffe (1970), as cited in Legates and McCabe Jr. (1999), defined the coefficient of efficiency which ranges from minus infinity to 1.0, with higher values indicating better agreement (Legates & McCabe Jr., 1999). In addition, the statistical values of the *RMSE*, and standard deviation show statistically acceptable results compared to the modified equation that fits better with the conditions of this basin, indicating that this model can be used to predict nitrogen load in the Biebrza river and for decision-making regarding the reduction of this nutrient.

The nitrogen from agricultural runoff is the main source of nutrient input into a river basin (Xia et al., 2020). Leading to environmental problems, in addition, the efficiency of nitrogen use in agriculture is very low and the load of this nutrient occurs mainly towards water bodies (Salvador-Castillo et al., 2021). Similarly, the accumulation of this type of nutrients leads to a process of eutrophication in water resources (Toro Gallego, 2019). Therefore, the agricultural percentage that covers a watershed is a primary factor in determining nitrogen load (Boyer, Goodale, Jaworski & Howarth, 2002). Thus, through the box plot (Fig. 3), it was observed that while the model works for the entire watershed, this model adapts better to sub-basins the agricultural percentage of which varies between 33% and 66%. That is, if stakeholders want to know the behavior of nitrogen load in the Biebrza river, applying this model (Eq. 3) will provide adequate data for decision-making in nitrogen reduction. However, more accurate and better-adjusted data to reality will be obtained if this equation is applied to those sub-basins which agricultural area covers the afore-mentioned range.

However, it is essential to note that although, based on statistical calculations, this model allows for fairly accurate prediction of nitrogen load, it has its limitations. Firstly, since it is based on historical water quality data, there are several gaps in different times of the year that prevent a more precise determination of nitrogen behavior during these times. Likewise, the accuracy of this model also depends on the quality and precision with which such samples were taken. During the analysis of water quality data, outliers were observed, which could be due to a specific event that caused an increase or decrease outside the range of all data, or to an inadequate sampling, as well as some sub-catchments not being sampled periodically, which prevented an analysis of those areas. The outlier values (Figs 3a and 3b) that use for calibration of the empirical model represent the highest value of each year (2005–2015) of Zelwianka (Zalewianka) – Mazurki sub-catchment. Therefore, these values were considered, when calibrating the empirical model. In addition, extreme weather conditions or unpredictable events that could affect nutrient load values were not considered. However, despite these limitations, this model can be used for the conditions of the Biebrza river, to make decisions to mitigate the amount of nitrogen affecting this watershed.

The calculation of nitrogen loss developed by (Naturstyrelsen, 2014) present in (Eq. 1) shows values that do not match with the nitrogen load observed. Hence, to be used for the conditions of the Biebrza basin and its lakes, the model was calibrated (Eqs 3 and 4) using a linear regression analysis. Once the new conditions were established, this new empirical model allowed to have more accurate and precise data compared to the original values. Therefore, the original empirical model presents a versatility that allows it to be adapted to the characteristics of various hydrographic basins. As is the case of the Ryck river in Germany, where this model was calibrated to meet the condition of this basin and it could be used as a key for the implementation of water buffer zones for the mitigation of the nutrient runoff (Trehan, Wichtmann & Grygoruk, 2022). Similarly, there are more complex models that forecast Nitrogen load, for instance Global NEWS, a Phyton framework and a global, multi--element and multi-form model of nutrient export by rivers (Mayorga et al., 2010), nonetheless, its application can become complicated if it is used by people who do not have the skills to apply the model in mentioned software.

Knowing the nitrogen load will allow determining the amount of fertilizer that should be applied to a crop will also help identify which sub-catchments are most affected, to minimize nitrogen load in the most affected areas. Fertilization aims to maximize yield using minimal fertilizer. There are different methods to determine this factor, ranging from recommendations to mathematical models that establish

dosage and yield, using chemical analysis, climatic characteristics, and crop management techniques as the main information (Quiroga & Bono, 2012). Due to the high economic costs, chemical analysis can become a disadvantage when determining fertilization guidelines. Therefore, by applying the previously established model, the amount of nitrogen load can be calculated in a simple way.

The new proposed empirical model allows determining the Nitrogen load, adapted to the conditions of the Biebrza River basin. As well as showing the adaptability of the original model towards new basins.

Conclusions

The assessment of nitrogen in the Biebrza catchment for agriculture is the crucial issue. That is the reason why it is important to have information related to the inflow of nitrogen in the area, to mitigate pollution processes. This research presents that the existing model for nitrogen load shows significant results, however, in order to be applied in the Biebrza basin it needs to be developed into a more advanced empirical model. By the analysis regression and some statistical analysis, the prediction result of a new empirical model was given in the study case and it is more suitable with the nitrogen load observed. Nitrogen load for the existing model is approx. 13,000 kg·year^{-1} and approx. 26,000 kg·year^{-1}, whereas, by the new empirical model is approx. 7,000 kg·year^{-1} and approx. 5,000 kg·year^{-1} for rivers and lakes, respectively. Therefore, it can be used for future research in the Biebrza river or water bodies with similar characteristics, such as precipitation and agricultural land. However, it is important to note that if more accurate values with a lower range of error are desired, the model should be optimized for each sub-watershed, considering the respective factors that affect the Nitrogen load. Thus, this model can be used as a guide in decision-making regarding the management and protection of a river, such as establishing sustainable agricultural measures that prevent nitrogen load and enabling the establishment of public policies for the protection of water resources.

Acknowledgements

References

Almasri, M. N. (2007). Nitrate contamination of groundwater: a conceptual management framework. *Environmental Impact Assessment Review*, *27* (3), 220–242. https://doi.org/10.1016/j.eiar.2006.11.002

Aloui, C., Jammazi, R. & Nguyen, D. K. (2014). A wavelet analysis of the relationship between oil prices and stock prices. *Journal of Financial Economics*, *111* (1), 1–13.

Ávila, J. P. & Sansores, A. C. (2003). Fuentes principales de nitrógeno de nitratos en aguas subterráneas [Main sources of nitrogen from nitrates in groundwater]. *Ingeniería*, *7* (2), 47–54.

Bednarek, A., Szklarek, S. & Zalewski, M. (2014). Nitrogen pollution removal from areas of intensive farming – Comparison of various denitrification biotechnologies. *Ecohydrology & Hydrobiology*, *14* (2), 132–141. https://doi.org/10.1016/j.ecohyd.2014.01.005

Berger, K., Verrelst, J., Féret, J. B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W. & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, *242*, 111758. https://doi.org/10.1016/j.rse.2020.111758

Boyer, E., Goodale, C., Jaworski, N. & Howarth, R. (2002). Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern USA. *Biogeochemistry*, *57*, 137–169. https://doi.org/10.1023/A:1015709302073

Cai, D., Wu, Z., Jiang, J., Wu, Y., Feng, H., Brown, I. G., Chu, P. K. & Yu, Z. (2014). Controlling nitrogen migration through micro-nano networks. *Scientific Reports*, *4* (1), 3665. https://doi.org/10.1038/srep03665

Dembicz, I., Kozub, Ł., Bobrowska, I. & Dengler, J. (2020). Photo Story – Grasslands of the mineral islands in the Biebrza National Park, Poland. *Palaearctic Grasslands – Journal of the Eurasian Dry Grassland Group*, *47*, 43–51. https://doi.org/10.21570/EDGG.PG.47.43-51

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. OJ L 327, 22.12.2000.

Downing, J. A., Cole, J. J., Middelburg, J. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Prairie, Y. T. & Laube, K. A. (2008). Sediment organic carbon burial in agriculturally eutrophic impoundments over the last century. *Global Biogeochemical Cycles*, *22* (1), 1–10. https://doi.org/10.1029/2006GB002854

Duan, W., Takara, K., He, B., Luo, P., Nover, D. & Yamashiki, Y. (2013). Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. *Science of The Total Environment*, *461*, 499–508. https://doi.org/10.1016/j.scitotenv.2013.05.022

European waters – assessment of status and pressures 2018. EEA Report, *7*. https://doi.org/10.2800/303664

Grizzetti, B., Pistocchi, A., Liquete, C., Udias, A., Bouraoui, F. & van de Bund, W. (2017). Human pressures and ecological status of European rivers. *Scientific Reports*, *7* (1), 1–11. https://doi.org/10.1038/s41598-017-00324-3

Grizzetti, B., Vigiak, O., Udias, A., Aloe, A., Zanni, M., Bouraoui, F., Pistocchi, A., Dorati, C., Friedland, R., De Roo, A., Benitez Sanz, C., Leip, A. & Bielza, M. (2021). How EU policies could reduce nutrient pollution in European inland and coastal waters. *Global Environmental Change*, *69*, 102281. https://doi.org/10.1016/j.gloenvcha.2021.102281

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2018). *Multivariate data analysis* (8th ed.). Boston: Cengage Learning.

Knapp, M. F. (2005). Diffuse pollution threats to groundwater: A UK water company perspective. *Quarterly Journal of Engineering Geology and Hydrogeology*, *38* (1), 39–51. https://doi.org/10.1144/1470-9236/04-015

Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York: McGraw-Hill.

Legates, D. R. & McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35* (1), 233–241. https://doi.org/10.1029/1998WR900018

Llive, F., Cadillo Benalcazar, J. J., Liger, B., Rosero Asqui, G., Fraga Ramos, E. & Ramos Martín, J. (2015). Vulnerabilidad y dependencia internacional de fertilizantes en el Ecuador [Vulnerability and international dependence on fertilizers in Ecuador]. *Technology Magazine – ESPOL*, *29* (2), 1–28.

Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H. W., Bouwman, A. F., Fekete, B. M., Kroeze, C. & Van Drecht, G. (2010). Global Nutrient Export from WaterSheds 2 (NEWS 2): model development and implementation. *Environmental Modelling & Software*, *25* (7), 837–853. https://doi.org/10.1016/j.envsoft.2010.01.007

Nash, J. E. & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I – a discussion of principles. *Journal of Hydrology*, *27* (3), 282–290.

Naturstyrelsen (2014). *The Danish Nature Agency's Guidelines for Nitrogen Calculations*. Copenhagen: Ministry of Environment of Denmark. https://mst.dk/media/121898/kvaelstofberegvejledningmaj2014.pdf [accessed: 26.04.2023].

Okruszko, T. (2005). *Kryteria hydrologiczne w ochronie mokradeł* [*Hydrological criteria in the protection of wetlands*]. Warszawa: Wydawnictwo SGGW.

Pei-Yue, L., Hui, Q. & Jian-Hua, W. (2010). Groundwater Quality Assessment Based on Improved Water Quality Index in Pengyang County, Ningxia, Northwest China. *Journal of Chemistry*, *7* (S1), S209–S216. https://doi.org/10.1155/2010/451304

Poikane, S., Kelly, M. G., Herrero, F. S., Pitt, J. A., Jarvie, H. P., Claussen, U., Leujak, W., Solheim, A. L., Teixeira, H. & Phillips, G. (2019). Nutrient criteria for surface waters under the European Water Framework Directive: Current state-of-the-art, challenges and future outlook. *Science of The Total Environment*, *695*, 133888. https://doi.org/10.1016/j.scitotenv.2019.133888

Quiroga, A. & Bono, A. (2012). *Manual de fertilidad y evaluación de suelos* [*Soil fertility and evaluation manual*]. Buenos Aires: Instituto Nacional de Tecnología Agropecuaria. Retrieved from: https://aulavirtual.agro.unlp.edu.ar/pluginfile.php/75200/mod_resource/content/0/Quiroga%20 et%20al%202007.pdf [accessed: 26.04.2023].

Razowska-Jaworek, L. & Sadurski, A. (2014). *Nitrates in groundwater*. Abingdon-on-Thames: Routledge.

Canchig, J., Kustina, R., Grygoruk, M. (2023). Developing an empirical model for assessment of total nitrogen inflow to rivers and lakes in the Biebrza river watershed, Poland. *Sci. Rev. Eng. Env. Sci.*, *32* (3), 201–220. DOI 10.22630/srees.4886

219

Rekha, P., Raj, D. S. S., Aparna, C., Bindu, V. H., & Anjaneyulu, Y. (2005). Bioremediation of contaminated lake sediments and evaluation of maturity indicies as indicators of compost stability. *International Journal of Environmental Research and Public Health*, *2* (2), 251–262. https://doi.org/10.3390/ijerph2005020008

Salvador-Castillo, J. M., Bolaños-González, M. A., Palacios-Vélez, E., Palacios-Sánchez, L. A., López-Pérez, A., Muñoz-Pérez, J. M., Salvador-Castillo, J. M., Bolaños-González, M. A., Palacios-Vélez, E., Palacios-Sánchez, L. A., López-Pérez, A. & Muñoz-Pérez, J. M. (2021). Estimación de la fracción de cobertura vegetal y contenido de nitrógeno del dosel en maíz mediante sensores remotos [Estimation of fractional vegetation cover and canopy nitrogen content in corn by remote sensing]. *Terra Latinoamericana*, *39*, 1–11. https://doi.org/10.28940/terra.v39i0.899

Schindler, D. W. (2012). The dilemma of controlling cultural eutrophication of lakes. *Proceedings of the Royal Society B: Biological Sciences*, *279* (1746), 4322–4333. https://doi.org/10.1098/rspb.2012.1032

Silva, A. R. E., Cobelas, M. Á. & González, E. M. (2017). Impactos del nitrógeno agrícola en los ecosistemas acuáticos: *Ecosistemas*, *26* (1), 37–44. https://doi.org/10.7818/ECOS.2017.26-1.06

Sucholas, J., Molnár, Z., Łuczaj, Ł. & Poschlod, P. (2022). Local traditional ecological knowledge about hay management practices in wetlands of the Biebrza Valley, Poland. *Journal of Ethnobiology and Ethnomedicine*, *18* (1), 1–42. https://doi.org/10.1186/s13002-022-00509-9

Sutton, M. A., Howard, C. M., Erisman, J. W., Billen, G., Bleeker, A., Grennfelt, P., Grinsven, H. van & Grizzetti, B. (2011). *The European nitrogen assessment: sources, effects and policy perspectives*. Cambridge: Cambridge University Press.

Świątek, D., Szporak, S., Chormański, J. & Okruszko, T. (2008). Hydrodynamic model of the Lower Biebrza River flow – A tool for assessing the hydrologic vulnerability of a floodplain to management practices. *Ecohydrology & Hydrobiology*, *8* (2), 331–337. https://doi.org/10.2478/v10104-009-0026-8

Tan, S., Xie, D., Ni, J., Chen, L., Ni, C., Ye, W., Zhao, G., Shao, J. & Chen, F. (2023). Output characteristics and driving factors of non-point source nitrogen (N) and phosphorus (P) in the Three Gorges reservoir area (TGRA) based on migration process: 1995–2020. *Science of The Total Environment*, *875*, 162543. https://doi.org/10.1016/j.scitotenv.2023.162543

Thompson, R. B., Incrocci, L., Ruijven, J. van & Massa, D. (2020). Reducing contamination of water bodies from European vegetable production systems. *Agricultural Water Management*, *240*, 106258. https://doi.org/10.1016/j.agwat.2020.106258

Toro Gallego, L. (2019). *Aplicación de metodologías para la estimación de la eutrofización en emblases tropicales y selección del índice de estado trópico mas adecuado para le embalse Peñol-Guatapé, Colombia* [*Application of methodologies for the estimation of eutrophication in tropical reservoirs and selection of the most appropriate tropical state index for the Peñol-Guatapé reservoir, Colombia*] (engineering thesis). Medellin: Universidad de Antioquia. Retrieved from: https://bibliotecadigital.udea.edu.co/ handle/10495/12104 [accessed: 26.04.2023].

Trehan, M., Wichtmann, W. & Grygoruk, M. (2022). Assessment of nutrient loads into the Ryck river and options for their reduction. *Water*, *14* (13), 2055. https://doi.org/10.3390/w14132055

Venegas, N., Marcinkowski, P., Piniewski, M. & Grygoruk, M. (2022). *Charakterystyka składników bilansu wodnego, surowy bilans wodny w zlewniach wodowskazowych oraz w granicach Biebrzańskiego Parku Narodowego i jego otuliny* [*Characteristics of water balance components, raw water balance in water gauge catchments and within the boundaries of the Biebrza National Park and its buffer zone*]. Goniądz: Biebrza National Park.

Wassen, M. J., Barendregt, A., Palczynski, A., Smidt, J. T. de & Mars, H. de (1992). Hydro-ecological analysis of the Biebrza mire (Poland). *Wetlands Ecology and Management*, *2*, 119–134.

Wuijts, S., Fraters, D., Boekhold, S. & Van Duijnen, R. (2022). *Monitoring of nitrogen in water in the EU – legal framework, effects of nitrate, design principles, effectiveness and future developments*. Brussels: European Parliament.

Xia, Y., Zhang, M., Tsang, D. C. W., Geng, N., Lu, D., Zhu, L., Igalavithana, A. D., Dissanayake, P. D., Rinklebe, J., Yang, X. & Ok, Y. S. (2020). Recent advances in control technologies for non-point source pollution with nitrogen and phosphorous from agricultural runoff: Current practices and future prospects. *Applied Biological Chemistry*, *63* (1), 1–13. https://doi.org/10.1186/s13765-020-0493-6

Xu, P., Zhu, J., Wang, H., Shi, L., Zhuang, Y., Fu, Q., Chen, J., Hu, H. & Huang, Q. (2021). Regulation of soil aggregate size under different fertilizations on dissolved organic matter, cellobiose hydrolyzing microbial community and their roles in organic matter mineralization. *Science of The Total Environment*, *755*, 142595. https://doi.org/10.1016/j.scitotenv.2020.142595

Yang, X., Ji, G., Wang, C., Zuo, J., Yang, H., Xu, J. & Chen, R. (2019). Modeling nitrogen and phosphorus export with InVEST model in Bosten Lake basin of Northwest China. *PLOS ONE*, *14* (7), e0220299. https://doi.org/10.1371/journal.pone.0220299

Summary

Developing an empirical model for assessment of total nitrogen inflow to rivers and lakes in the Biebrza river watershed, Poland. Nitrogen load is crucial for its application in various fields such as agriculture and improving water quality control for authorities responsible for establishing agricultural policies in the area. The calculation of nitrogen load using existing equations is not applicable for all types of rivers, thus requiring the development of a new equation that can be applied to lakes and rivers in the Biebrza river catchment. To determine the new equation, extensive mapping of the catchment area was conducted, which was adjusted to precipitation and runoff in the area, allowing the observed results to be compared. Based on several analyses, the new equation has better accuracy, *RMSE* of the new model-based estimation decreased by 65.9% in 2005–2015 and 62.2% in 2016–2021 for river and 92% in 2008–2019 and 95% in 2020–2021 for lakes. Therefore, the application of the new calibrated empirical model provides results close to the real values and it can be used in the Biebrza river basin to simulate the total nitrogen runoff.