

Response to Reviewer's Comments

This document contains copies of all comments of the reviewer 3 (*in italicized, blue text*) and our planned effort to address them (in normal, black text). Our proposed manuscript revisions are underlined.

Reviewer 3

Sheikholeslami and Hall, 2022 utilize machine learning approach to build a random forest model of nitrogen concentration in major river basins. They apply the model to global river basins to identify nitrogen concentration hot-spots and decadal increase in nitrogen concentration. The manuscript applies random forest approach for estimating riverine nitrogen concentration, but their findings are not filling a gap in the literature. Their finding that highest nitrogen concentrations are observed in United States, Europe, China, and India is not new given that these regions are agriculture dominated and utilize large amount of fertilizer and manure. The manuscript lacks a novel motivation, several important predictor variables are not considered, and the final analysis does not provide any new information.

Thank you very much for your comprehensive review and identification of key areas of improvement. We provide detailed response to your comments in the subsequent section. Below, we list your main concerns, and a summary of our proposed revisions to address them.

Overall, we believe that the reviewer has under-estimated the contribution in the paper. As outlined in this rebuttal document, the application of the random forest model can be considered to be state-of-the-art in a rapidly evolving field. The use of these techniques in the very challenging field of global water quality modelling is novel and has yielded predictive results that exceed other approaches in important respects. We think reviewer does not seem to have recognized (or has dismissed) the merit of the utilized methodology considering our comprehensive exercise in data processing, model selection, adaptation, and validation. We appreciate reviewer's suggestions on potential topics, and we provide responses to individual ones as following:

(1) Lack of a novel motivation: First, despite a plethora of powerful machine learning methods, ANN is the most popular method for water quality modelling. To fill this gap, we implemented random forest algorithm to examine its predictive performance when applied to large spatial scales. Second, our thorough search of the relevant literature indicated that machine learning methods, particularly random forests, have not been implemented at global scale to identify nitrogen hotspots and key drivers of nitrogen variability. Third, despite being successful in simulating and predicting surface water quality at catchment-scale, machine learning methods have not been utilized to provide spatially explicit (gridded) estimates of nitrogen levels. Based on our observation,

almost all machine learning models are lumped in space (see Table 1). This gap further motivated our development of the spatiotemporal random forest-based global model in the present study. Please note that these gaps have been discussed in Section 1.2. In the revised manuscript, we will improve the writing to reduce the reviewer's confusion that appear to have arisen due to lack of sufficient clarity in the original versions.

(2) Missing several important predictors: It would be helpful to have an indicator from the reviewer what important predictors they consider to have been missed out. As proposed in this rebuttal document, based on the reviewers' suggestion we are also planning to include additional variables, such as forest fraction, urban fraction, and hydrography data, to the structure of the model. We will also investigate their impact on model performance during the revision.

(3) Lack of new information: First, we restate that the model provided much higher spatial resolution than we were able to contain in our high-level summary of the results. Our model was used to predict nitrogen concentrations in large river basins globally, providing new information about dynamics of nitrogen concentrations in location with scarce/no observations. Furthermore, we provide NO_x-N concentrations (mg/l) worldwide (180°E–180°W; 90°S–90°N) at a spatial grain of 0.5-degree. The NO_x-N concentrations, mapped across the globe for 1992-2010, are available in a compressed GeoTiff file format in the WGS84 coordinate reference system (EPSG:4326 code). The developed stream nitrogen concentration maps have a wide array of potential applications in stream ecology, biodiversity research, conservation science, and stream and lake restoration ecology. For instance, the produced maps can be used to quantify the overall mass of nitrogen discharged into a specific lake or ocean body, enabling a deeper understanding of global-scale eutrophication. Furthermore, our estimates of nitrogen concentration can be used to verify new process-based models that predict nitrogen concentrations and transformations in inland waters worldwide. We encourage potential users of the described geo-dataset to contact the authors for future product updates. We will add this to the revised manuscript to better highlight the usefulness of our findings.

In addition, to improve the linkage between study objectives and results, we propose to present additional results to highlight several model capabilities. These include model validation using a new validation strategy, adding partial dependence plots, presenting the distribution of R² values, and further discussions on limitations of our approach along with recommendations for future studies.

Introduction:

1. The authors include “three critical observations” from the literature review of ML applications for studying water quality. What is the reason for including these observations?

As stated in our response to previous comment, these three observations essentially motivated us to develop the proposed random forest-based global water quality model. To resolve this comment, we will make necessary changes in the revised manuscript to improve the quality and presentation of the Introduction section.

2. The two main objectives of the study related to identifying global nitrogen pollution hotspots and key drivers of nitrogen variability at a global scale are weak since various studies have examined these and it is not a novelty.

Politely disagree. To authors’ knowledge, there is no parallel in the literature in terms of time scale and spatial resolution with similar approach to identify nitrogen pollution hotspots and key drivers of nitrogen variability at a global scale. In fact, one essential purpose of the developed model is to enable water quality assessment and quantification of future scenarios (e.g., of livestock pasture extent) at a ‘global scale’. Though some instances of analysis of these questions exist (see, e.g., Mayorga et al., 2010; He et al., 2011; Beusen et al., 2016; etc.) they all have inevitable limitations, and none have used our proposed approach which we believe yields worthwhile results. Therefore, we believe that both objectives are strong enough to motivate our analyses. The reviewer argued that “*various studies have examined these*” we’d be more than happy if the reviewer helps us identify these studies, so we can cite them in the revised manuscript. For more discussion, please see our response to your first comment in this rebuttal document.

3. The Introduction currently has three section and is quite long. Condensing this section will improve the readability of the manuscript.

We will condense this section in the revised version, while keeping some important background information for the broad readership of the HESS as a multi-disciplinary journal.

Data and Methodology:

4. Lines 218-222: The authors mention reshaping all predictors to 0.5-degree resolution, but they have not described how precipitation, runoff, and other predictors from the entire upstream catchment/watershed is accounted for in predicting river nitrogen concentration. It appears that they only considered the contribution of the various predictors from the 0.5-degree grid. It is highly likely that only a fraction of this

grid falls in the upstream catchment and for several observations large fraction of the upstream catchment is not included in this grid. Thus, the estimation of contribution from various predictor variables is likely flawed.

Thank you, for this important comment, and we confirm that your interpretation about the spatial resolution of our study is all correct, i.e., the current variable used for predicting in-stream nitrogen concentrations only cover the properties within the grid cell of interest.

Regarding the catchment characteristics, this can be resolved, for example, by adding hydrography data delineating global river networks, though it will presumably add more complexity to the model. In the revision, we will add more variables to the list of predictors, including upstream characteristics, stream proximity (e.g., distance up to the stream) or log-transformed flow accumulation for better capturing spatial characteristics of watersheds. Previous studies have shown that these variables can be key drivers of water quality responses in rivers (see, e.g., Staponites et al., 2017; Lintern et al., 2017; Grabowski et al., 2016; Peterson et al., 2010). Particularly, they reported that accounting for the hydrological flow paths and flow accumulation through the landscape and coupling these processes with specific landscape features can improve model performance. We will also elaborate more on this with the supporting literature.

5. Adding latitude and longitude as predictor variables does not fully capture the spatial relationship between observations and predictors.

Agreed. But, as we extensively discussed in Section 3.1.1, two strategies have been proposed in the literature to account for spatial information: (1) using a hybrid modelling framework by embedding Kriging and Gaussian process modelling into the standard random forest method (Saha et al., 2021; Canion et al., 2019); and (2) using geographic information as the auxiliary inputs, for example, adding geographic coordinates (Behrens et al., 2018; Meng et al., 2018) or other spatial distances (Li et al., 2011; Wei et al., 2019) into the list of predictors. Since adding lat/lon as predictor variables to represent spatial relationship is common in data-driven modelling of Earth and environmental systems, we followed this strategy and incorporated lat/lon into our random forest model. During the revision, we will highlight the weakness of the second strategy in not fully capturing the spatial relationships between observations and predictors.

6. Additional land use predictors should be considered in the study for example, forest fraction, urban fraction.

Thank you for this suggestion. Please note that from our review of literature we observe that cropland fraction is one of the chief determinants of the nitrogen variability, as evident from our variable importance analysis (Fig 9). To address this comment, additional predictors such as forest fraction and urban fraction will be added to the list of predictors. Further, we will examine the relative importance of these variables when simulating nitrogen concentrations.

7. In addition to considering monthly precipitation, extreme precipitation variables and variables capturing dry spells should also be considered as they impact nitrogen concentration. For example, a long dry spell will result in high nitrogen concentrations however, if the dry spell is followed by a large precipitation event concentrations will drop. The monthly mean precipitation will ignore this temporal variability and thus not accurately predict nitrogen concentration.

Thanks for sharing the interesting idea. Perhaps you missed it, but we have already discussed the impact of extreme and/or prolonged hydroclimatic events on nutrient concentrations in Section 4.3 (line 454-464). We believe quantifying the impact of extreme events on nitrogen concentrations using machine learning is an interesting topic, which we'd like to explore in future. In the Discussion section of the revised manuscript, we will mention this as an important recommendation for future research.

8. Table 2 should be replaced with a table containing the final 17 predictor variables. In its current format the table 2 does not list the four time and space predictors and lists the livestock predictors in a single row that makes it appear as a single predictor not five different predictors.

Thanks for this comment. We will update Table 2 in the revised manuscript based on reviewer's suggestion.

Results:

9. In the Results section, the authors have not compared their findings with any of studies listed in Table 1. The authors should discuss the similarity and differences between their findings and that of others using similar model building approach.

As we mentioned in our response to your second comment, there is no parallel in the literature with similar approach. None of the studies listed in Table 1 systematically investigated the key drivers of nitrogen variability, which was one of the main objectives of our study. Additionally, none of them provided spatially grided estimates of nitrogen concentration, as all of them were lumped in space. However, we will add some evaluation of the closest relevant studies in the revised manuscript to compare our findings and that of others.

We want to highlight that our paper has been sufficiently contextualized, particularly in the Introduction section by (i) placing our research topic within its larger setting, (ii) providing important perspective by citing similar examples or relevant background, (iii) explaining what historical circumstances led up to the topic we are discussing, (iv) citing other scholars who have recently contributed to the field, and (v) exploring how our analysis fits into a larger discussion about the field. Indeed, we agree that the regional observations would benefit from further contextualization in relation to previous regional studies, which we will do in the revised manuscript.

10. In addition to the annual concentrations (Figure 6), authors should also analyze and discuss monthly or seasonal maximum concentration.

Thanks for pointing this out. We will present more results on monthly/seasonal variability of maximum concentrations in the Results section of the revised manuscript.

11. The authors had developed spatial plots of nitrogen concentration for every month between 1992-2010, then why did they limit the analysis of change in nitrogen concentration to a decadal scale difference only. For large river basins, they can perform trend analysis. This will provide more useful information for policy makers than decadal scale difference.

Thanks for this comment, we strongly believe that decadal analysis is also helpful to identify long-term trends. While we decided to focus this study on the decadal scale, we have noted trend analysis for the selected large river basins as a possible option, and we would like to explore its usefulness in future studies.

12. Line 387 – 389: What factors contributed to the decline in nitrogen concentrations? Just merrily stating decline is not enough and the drivers behind this observation should be discussed especially given that few regions with highest nitrogen concentration (India and South Korea) also have the largest decline.

Thank you for raising this point, and we confirm the reviewer's interpretation that significant decline in nitrogen concentrations has been occurred in a few basins with highest nitrogen concentrations. Our focus, however, was not to explore which factors were mainly responsible for decline in nitrogen level in some regions. We therefore think a thorough analysis of this question is beyond the scope of the present study. There might be several reasons for this observation, such as dietary behavior change, improved nitrogen fertilizer management, increasing efficiency of crop production, hydroclimatic regime shifts, change in upstream

management, etc. We will elaborate more on the possible causes of the observed decline in nitrogen concentrations with the supporting literature.

13. The fact that month of year (MOY) is significantly more important than precipitation, runoff, and temperature seems concerning. If precipitation, runoff, and/or temperature, were selected as predictor variables it would have a direct physical meaning. For example, increase/decrease in precipitation can decrease/increase the nitrogen concentration.

We are unclear on the interpretation of your comment. Please note that this variable has been included in our model to represent ‘distance’ in the time domain, particularly to capture the seasonality effect. Based on our factor importance analysis, therefore, the most important covariate for predicting monthly nitrogen concentration given the utilized datasets is: time. One possible reason is that the data used in the model has a strong seasonality that makes the variable “month of year” highly important. This can be justified because generally there is a considerable seasonal variability in major factors influencing nitrogen level, such as vegetation, land-use change, hydro-climatic parameters, and farming activities, which strongly influence constituents’ concentrations in different seasons (see, e.g., Pejman et al., 2009; Shabalala et al., 2013; Xu et al., 2019; etc.). Consequently, it is expected to observe seasonal variation of water quality in many regions of the world. Furthermore, to avoid confusion, we have to clarify that the importance of this variable should not be miss-understood as a strong trend in the sense that the monthly nitrogen concentrations increase over time or the like.

14. What is the physical meaning behind cumulative month (CM) variable being identified as the second most influential predictor?

We should again highlight that this variable has been included in our model to capture the long-term trends. Furthermore, it somehow can compensate for biogeochemical legacy and the long travel time between N input and riverine N export signals. The covariate CM allows the random forest model to fit different spatial patterns for each month underpinning that the observed nitrogen level is different from month to month. To address this comment, we will add more details to the relevant discussions in Section 4.3 (High importance factors influencing predictions of nitrogen levels).

15. Figure 9 - Why is the relative importance of 3-15 predictors almost same?

We believe that for the top 7 important variables the ranking is distinguishable. However, for the rest of predictors, i.e., 8th to 15th, we agree that the difference in variable importance values is not significant. It means that randomly permuting the values of these predictors resulted in quite the same change in prediction accuracy. In other words, the importance of these variables cannot be robustly ranked, even though they are all influential. As mentioned in the manuscript (Section 5), to comprehensively analyze how various factors influence model output variability a more advanced approach is required. Global sensitivity analysis methods are suitable candidates in this regard.

16. Did the most influential predictors vary over space and/or time?

Of course. Considering the spatio-temporal variability in both target and predictor variables, we can assert that ‘factor sensitivity’ also varies over time and/or space. However, the inherent feature of the random forest algorithm for evaluating variables' predicting strength (expressed as variable importance ranking) can only measure variable importance for all input data using the whole space-time regression matrix. Thus, it cannot provide spatial-temporal characteristic of the factor sensitivity.

Assessing how the impact of predictors and their interactions varies in both space and time requires a more systematic approach, such as using advanced global sensitivity analysis methods, which we'd like to explore in future studies. To clarify this, we will add a brief discussion on spatio-temporal sensitivity analysis of random forest model and will elaborate more on this issue with the supporting literature.

Reference

- Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., & MacMillan, R. A. (2018). Spatial modelling with Euclidean distance fields and machine learning. *European journal of soil science*, 69(5), 757-770.
- Beusen, A. H. W., Van Beek, L. P. H., Bouwman, A. F., Mogollón, J. M., and Middelburg, J. J.: Coupling global models for hydrology and nutrient loading to simulate nitrogen and phosphorus retention in surface water—description of IMAGE–GNM and analysis of performance, *Geoscientific Model Development*, 8(12), 4045-4067.
- Canion, A., McCloud, L., & Dobberfuhr, D. (2019). Predictive modeling of elevated groundwater nitrate in a karstic spring-contributing area using random forests and regression-kriging. *Environmental Earth Sciences*, 78(9), 1-11.
- He, B., Kanae, S., Oki, T., Hirabayashi, Y., Yamashiki, Y., & Takara, K. (2011). Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework. *Water research*, 45(8), 2573-2586.
- Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, 26(12), 1647-1659.

Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H., Bouwman, A. F., ... and Van Drecht, G.: Global nutrient export from WaterSheds 2 (NEWS 2): model development and implementation, *Environmental Modelling & Software*, 25(7), 837-853.

Meng, X., Hand, J. L., Schichtel, B. A., & Liu, Y. (2018). Space-time trends of PM_{2.5} constituents in the conterminous United States estimated by a machine learning approach, 2005–2015. *Environment international*, 121, 1137-1147.

Pejman, A. H., Bidhendi, G. R., Karbassi, A. R., Mehrdadi, N., & Bidhendi, M. E. (2009). Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *International Journal of Environmental Science & Technology*, 6(3), 467-476.

Saha, A., Basu, S., & Datta, A. (2021). Random forests for spatially dependent data. *Journal of the American Statistical Association*, 1-19.

Shabalala, A. N., Combrinck, L., & McCrindle, R. (2013). Effect of farming activities on seasonal variation of water quality of Bonsma Dam, KwaZulu-Natal. *South African Journal of Science*, 109(7), 1-7.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM_{2.5} concentrations across China using the space-time random forest approach. *Remote Sensing of Environment*, 231, 111221.

Xu, G., Li, P., Lu, K., Tantai, Z., Zhang, J., Ren, Z., ... & Cheng, Y. (2019). Seasonal changes in water quality and its main influencing factors in the Dan River basin. *Catena*, 173, 131-140.