

## Responses to the reviewer 1

We greatly appreciate the reviewer's comments and suggestions. We have made a concerted (and hopefully successful) attempt to address the reviewer's concerns in the revised manuscript.

First, we have added time series evaluations of annual solid and nutrient loads from large U. S. rivers for the period ~1963-2000 against measurement-based estimates from the USGS National Water Quality Network.

Second, we have substantially expanded sensitivity/uncertainty analyses by perturbing all calibrated parameters and perturbing the fractions of dividing externally prescribed TN and TP inputs into different N and P forms, in addition to the previous manipulations of model components and inputs.

Third, we have modified analyzed years. For cross-watershed and global evaluations, annual results for the year 1990 are analyzed and presented. In parenthesis, ranges of using annual results for the years 1990-2000 are also provided throughout the manuscript.

Fourth, as the other reviewer 2 suggested, we have modified the model to simulate PON fluxes from terrestrial litter and soils, based on the simulated soil erosion fluxes and litter/soil N concentrations and have made associated modifications throughout the manuscript.

Finally, we have substantially extended the discussion of the uncertainties inherent in this initial development of the LM3-FANSY modeling framework and the need to pursue advances to provide a more comprehensive and consistent approach of modeling the coupled solid, N, C, and P cycles across the terrestrial and freshwater continuum. In accordance with this forward looking perspective on P, and the reviewer's suggestions, we have de-emphasized and strongly caveated our discussion of river N:P ratios.

In this response, the line numbers correspond to those in the original manuscript before making any modifications. Table, figure, and section numbers have been changed due to additions/removals of tables, figures, and sections. Unless noted, the table, figure, and section numbers correspond to the modified ones. The modified tables and figures are placed at the end of this response.

Sincerely,  
Dr. Minjin Lee  
Associate Research Scholar  
Princeton University, NOAA/Geophysical Fluid Dynamics Laboratory  
(on behalf of all co-authors)

## Referee: 1

*This paper presents the development of a coupled river (and lake) biogeochemical model for N and P on the global scale, based on the coupling of LM3, the freshwater module of FANSY. The P input and some of the N sources are taken from another data source (IMAGE-GNM).*

*The authors claim that the new model FANSY is incorporated in the Land Model (LM3). This coupling results in a more process-based representation of biochemistry model FANSY-LM3 which links the terrestrial and ocean biochemistry. The purpose of the LM3 model: “It captures processes including changes in vegetation functioning, plant-soil nitrogen cycling, and streamflow dynamics. The land model sheds light on the effects of these processes on atmospheric physics and chemistry and, conversely, the effects of changing climate and CO2 concentration on the land.” (from <https://www.gfdl.noaa.gov/land-model/>). It is clearly a goal of this coupled system (FANSY-LM3) to capture changes and streamflow dynamics. The FANSY model has indeed a lot of dynamics incorporated (and are described here), so this model is well suited to couple with the LM3 model. However, the results that are presented here are not dynamic, but are on one single time period. How can the reader discover whether FANSY is capable of producing reasonable results when it is used in combination with the LM3 model? If the purpose of the model only is to simulate the chosen time period, then I would not take this model, but the Global NEWS model. The advantage of the Global NEWS model is that it is published and well cited and the results are very transparent! The current presentation of the result does not convince me to start using a dynamic process-based model.*

First of all, we agree that time series evaluations would be a valuable addition to this model development study and have added time series analysis for large U. S. rivers where long time series of high quality data are available (see our responses below). We point out, however, that evaluation of model outputs from disparate globally-distributed systems under today’s climate and land use forms a core component of the development of Earth System Models used for applications of climate and land use changes, with the widely accepted premise that a model’s capacity to represent the range of outcomes across disparate systems spanning a range of climates and anthropogenic drivers has a bearing on its capacity to project changes over time.

To outline our thought process, we use a general guideline that building confidence in models used for global climate projections is a progressive process that rests on two pillars: 1) the construction of models from mechanistic principles likely to hold in novel environments expected under climate change, and 2) their ability to recreate features in the current climate and/or past climate changes (Randall et al., 2007). Our inclusion of a relatively comprehensive set of process-based dynamics attempts to advance the first confidence pillar (though we recognize that we are still only scratching the surface of the full complexity of the system and

have strengthened our discussion of the significant simplifications still required in accordance with the reviewer's later suggestions and those of reviewer 2). For the second pillar, we chose to focus on diverse solid and nutrient loads across disparate globally-distributed systems with different climates and other anthropogenic drivers. This focus on a chosen time spatial variation is common in validating new dynamic models added as a prognostic component of Earth System Models (e.g., Zhang et al., 2020), merged land-river model similarly to LM3-FANSY (e.g., He et al., 2011), and Earth System Model evaluation metrics more broadly (e.g., Dunne et al., 2020). It reflects the fact that the strongest current nutrient load constraints come from cross-system comparisons in today's climate.

We have clarified the underlying model evaluation rationale in the Methods section (Sect. 3.3). With the addition of the (admittedly limited) time series analysis and other additions suggested by the reviewers, we feel that our model evaluations are as thorough as any other published global river solid or nutrient modeling studies. We have furthermore noted in the Discussion the need for additional time varying constraints to build additional confidence in projected changes.

Secondly, the goal of FANSY is not to replace the Global NEWS model. Rather, the goal is to develop an initial module to be added as a prognostic component of an Earth System Model to link global terrestrial and ocean biogeochemistry. We anticipate that similar models will be developed for many global Earth System Models used for climate change projections in the coming years. This purpose motivated our efforts to advance the process resolution of the model relative to more empirical approaches, but we recognize that adding process-level constraints can degrade model fit. We thus feel that it is important to present both the overall skill and the skill relative to Global NEWS which, as the reviewer has stated, is a widely cited and pioneering community benchmark. We emphasize, however, that this is in no way an attempt to replace Global NEWS, but a necessary step toward our ultimate goal of integration of FANSY within an Earth System Model framework used for climate projections, and we have taken steps to clarify these points in the Introduction and Methods. We feel that this progressive approach is consistent with the goals of Geoscientific Model Development.

Lastly, as described several places in the manuscript, the merged land-freshwater framework of LM3-FANSY is built on the existing merged land-freshwater N framework of LM3-TAN, by integrating the new freshwater algae, solid, and P dynamics into it. Seasonal to centurial time series results of LM3-TAN have been extensively validated against observational estimates at both global and regional scales (Lee et al., 2014; 2016; 2018; 2019, 2021). The validated results have showed simulated vegetation and litter/soil N storage changes to significantly alter multi-decadal river N trends (Lee et al., 2019) and coastal ecosystem productivity and C exports (Liu et al., 2021), as well as seasonal to multi-year river N extremes (Lee et al., 2016; Lee et al., 2021).

Dunne, J. P., et al.: The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics, JAME, 12(11), DOI:10.1029/2019MS002015, 2020.

Randall, D. A., et al: Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

He, B., Kanae, S., Oki, T., Hirabayashi, Y., Yamashiki, Y., and Takara, K.: Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework, Water Research, 45, 2573-2586, doi.org/10.1016/j.watres.2011.02.011, 2011.

Zhang, H. et al.: Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon, Glob Change Biol, 26: 2668–2685, <https://doi.org/10.1111/gcb.14994>, 2020.

***Also the sensitivity analyses are focused on the Pearson correlation coefficient of that same time period on a limited set of rivers. This approach does make it very hard to understand how the model is reacting on different inputs.***

The sensitivity analyses are based on both 1) the Pearson correlation coefficient (Table 4 in the original manuscript), which tests the impact of a parameter, input, or component change on the spatial agreement between the simulated and observed relative river loads, and 2) the amount of river solid and nutrient loads (Table 5 in the original manuscript), which tests the effect of a change on the magnitude of the river loads. We have emphasized this in the discussion of the load differences between the baseline and sensitivity simulations to show the effects of different inputs (e.g. lines 640-643, 651-658).

We have emphasized that the 69 rivers are the world's major rivers influenced by a broad range of climate and land use (lines 435-437). The 69 river basins cover 55% of global land area (excluding the Antarctic) and their broad global distribution and coverage is now illustrated in Fig. S11.

***It is not clear to me, why these 64 basins are chosen and whether they are a good representation of the global river basins.***

The total number of rivers considered for our analysis is 69 (not 64). As noted above, the 69 river basins cover 55% of global land area (excluding the Antarctic) across a broad range of climate

and land use. We have extended the description of how the 69 rivers were chosen later in this response.

***Also the use of different time periods to do this comparison is creating all kinds of uncertainties.***

We have modified the years for the analyses to consistently present values for the year 1990 and ranges for the years 1990-2000 throughout the manuscript. As suggested by the reviewer, this change was made to be consistent with the time limit of the model forcing.

***I would suggest to do a proper sensitivity analyses (not a one factor analysis but change all parameters at once) on the important model output parameters to show and understand the mechanistic behavior of the model.***

We have substantially expanded sensitivity/uncertainty analyses by changing all calibrated parameters and by changing the fractions of dividing the externally prescribed TN and TP inputs into different N and P species, in addition to the previous model component and input manipulations. This has increased the total number of sensitivity/uncertainty simulations from 22 to 87, requiring 4611 years of total sensitivity/uncertainty simulations. Discussion of the insights gained by these expanded simulations has also been extended considerably.

We do not feel that Monte Carlo experiments changing all parameters at once are appropriate at this initial development stage, because 1) it is challenging, if not impossible, to robustly attribute a response to a parameter or process when one is changing everything at once and 2) many of the simulations arising from such experiments proved implausible when compared with observations. We agree that such approaches are proper when constructing an ensemble of similarly plausible model permutations, but only after proper filtering for model fidelity and once the more basic single factor sensitivities are understood.

We feel that the significant expansion of single factor sensitivities/uncertainties and discussion made in response to the reviewer's comments and those of reviewer 2 have strengthened this model description paper considerably in this latter regard.

***The current validation of the model is not good enough and is not convincing.***

As emphasized in our responses to the reviewer's general comments above and in more detail below, we have included quantitative cross-watershed evaluations of solids, N, and P, in different forms (SS,  $\text{NO}_3^-$ ,  $\text{NH}_4^+$ , DIN, DON, TKN,  $\text{PO}_4^{3-}$ , DOP, and TP) and different units (yields, loads, and concentrations), across the 69 globally distributed rivers covering 55% of the land surface (excluding the Antarctic). We have augmented the evaluations further with time

series comparisons suggested by the reviewer and expansive sensitivity analysis. This frankly meets or exceeds the evaluations presented in comparable model description papers in GMD and elsewhere (e.g., Beusen et al., 2015, He et al., 2011, Pelletier, 2012). We also note that this is a model description paper, not just a model evaluation paper.

*Additional validation on long term data sets is needed to show that the dynamic patterns are well simulated.*

We have added time series evaluations of annual solid and nutrient loads from large U. S. rivers for the period ~1963-2000 against reported data from the USGS National Water Quality Network (NWQN, Lee, 2022). We still feel, however, that the global comparison of constituents and quantities across multiple systems, provides the strongest test of the model and the best initial basis for understanding the strengths and limitations of this model for global applications.

A description of the reported time series data has been added in Sect. 3.3 as follows:

**For time series evaluations, we used reported annual solid and nutrient loads across 8 stations in large U. S. rivers from the USGS National Water Quality Network (NWQN, Lee, 2022). We note that the robustness of evaluating simulated interannual variability against simple flow-weighted annual observations depends on the frequency and timing of chemical samplings (Lee et al., 2021). The reported annual loads from NWQN were estimated with the USGS’s latest load estimation method, WRTDS-K , for all analyzed rivers, except for the Columbia River for which only the REG method estimates are available (Lee et al., 2017). The WRTDS-K method was proved to be the most accurate annual load estimation method among methods studied by Lee et al. (2019).**

**The 8 stations are located in the 3 largest river basins in NWQN (6 stations in the Mississippi River Basin, 1 station in the St. Lawrence River Basin, and 1 station in the Columbia River Basin). The basin area of Yukon River station is larger than those of St. Lawrence and Columbia River stations, but the Yukon River station was excluded for this analysis, because data is only available beyond 2000. The corresponding station locations in the LM3-FANSY river network were identified by matching latitudes, longitudes, and basin areas of the reported and modeled rivers. The yearly data provided by NWQN is based on “a water year” defined as “the 12-month period from October 1 for any given year through September 30 of the following year”. The corresponding water yearly results of LM3-FANSY for the periods ~1963-2000 were used for this analysis.**

The time series evaluations have been added by introducing a new subsection of the results in line 630, with additions of new figures and a table, as follows:

### 4.3 Time series analysis

Correlations between the measurement-based vs. simulated annual river solid and nutrient load time series across the 8 stations for the period ~1963-2000 demonstrate that, while model results for individual rivers may be biased, the simulated solid and N loads in different N species covary with variations of the measurement-based loads (Table 6, Fig. 8, Fig. SI5). LM3-FANSY, however, has less capability of capturing interannual variability of the P loads. The prominent difference between the solid and N dynamics vs. the P dynamics in LM3-FANSY is the inputs from terrestrial litter and soils, which are found to be the largest input source to rivers (see sensitivity analysis in Sect. 4.4). The solid and N inputs from terrestrial litter and soils are simulated by LM3, while the corresponding P inputs are externally prescribed, because LM3 does not yet include terrestrial P dynamics. The lack of terrestrial P dynamics in LM3 is one of the most plausible reasons for the less capability of capturing the P load interannual variability (see Sect. 4.5 for further discussion).

Prediction errors of the average river solid and nutrient loads for the periods ~1963-2000 across the 8 stations (Table 6) are comparable to the errors shown in the cross-watershed analysis (Table 4), as well as to the errors of cross-watershed analyses in previous global watershed modeling studies (Pelletier, 2012; Mayorga et al., 2010; Harrison et al., 2010; He et al., 2011). The cross-watershed emphasis reflects the intended global application of the model herein, but significant biases for individual rivers are a natural consequence of the prioritization. For example, the Mississippi River N loads in the LM3-FANSY simulation herein are significantly underestimated (Table 6) despite their minimal bias globally. Focused investigations of larger misfits will be pursued in future work to increase model robustness. Alternatively, for regional applications, tuning a single parameter for N dynamics or solid dynamics within the Mississippi River Basins allows LM3-FANSY to be calibrated for a specific river. For example, reducing the freshwater denitrification rate coefficient from  $0.15 \text{ day}^{-1}$  to  $0.05 \text{ day}^{-1}$  can reduce the errors from -62% to -15% for  $\text{NO}_3^-$  and from -26% to 8% for TN, while increasing the correlation coefficients from 0.7 to 0.75 for  $\text{NO}_3^-$  and from 0.63 to 0.67 for TN (Fig. 8). Reducing the free parameter of terrestrial soil erosion by half can reduce the error of SS from 157% to 27%.

In the Discussion, we have acknowledged the need to pursue advances across the terrestrial and freshwater continuum in LM3-FANSY as follows:

There is also a need to pursue advances to provide a more comprehensive and consistent approach of modeling the coupled N, C, and P cycles across the terrestrial and freshwater continuum of LM3-FANSY. Expansion of LM3 to include terrestrial P dynamics will be targeted to improve estimates of litter/soil P storage and fluxes to streams and rivers, generating mechanistically consistent estimates of N:P ratios of nutrient loads reaching

**coastal systems. Priority enhancements will also include integration of freshwater C and alkalinity dynamics with the current solid, algae, and nutrient dynamics of FANSY to simulate the impacts of river inputs on coastal C budgets and acidification.**

The following references have been newly added.

**Lee, C. J., Murphy, J. C., Crawford, C. G., and Deacon, J. R.: Methods for computing water-quality loads at sites in the U.S. Geological Survey National Water Quality Network (ver. 1.3, August 2021): U.S. Geological Survey Open-File Report 2017–1120, 20 p., <https://doi.org/10.3133/ofr20171120>., 2017.**

**Lee, C. J., Hirsch, R. M., and Crawford, C. G.: An evaluation of methods for computing annual water-quality loads: U.S. Geological Survey Scientific Investigations Report 2019–5084, 59 p., <https://doi.org/10.3133/sir20195084>., 2019.**

**Lee, C.: Nutrient and pesticide data collected from the USGS National Water Quality Network and previous networks, 1950-2021, U.S. Geological Survey, <https://doi.org/10.5066/P948Z0VZ>, 2022.**

**He, B., Kanae, S., Oki, T., Hirabayashi, Y., Yamashiki, Y., and Takara, K.: Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework, *Water Research*, 45, 2573-2586, [doi.org/10.1016/j.watres.2011.02.011](https://doi.org/10.1016/j.watres.2011.02.011), 2011.**

*I would advise to reject the current manuscript or major revisions under the condition that there will be a better sensitivity analyses including a more extensive validation over a longer time period.*

We have substantially extended sensitivity/uncertainty analyses and added time series evaluations of river solid and nutrient loads over long time periods ~1963-2000. We are confident that our documentation and evaluation now meets or exceeds those provided by other model development papers.



## *Major comments*

***1. It is very important to know if the calibration data and validation overlap. However, details about calibration data are absent in this paper .***

We have clarified that all data were included when evaluating and coarsely calibrating the model. The model evaluation thus tests the degree to which the model and forcing can explain the observed patterns across disparate watersheds and nutrient species, allowing for coarse calibration within the constraints of the model structure and parameters. Explicit separation of calibration and validation datasets is often omitted for global watershed model development. The examples include studies that evaluate simulated TN and TP concentrations (IMAGE-GNM, Beusent et al., 2015), nitrate loads and yields (He et al., 2011), and suspended solid yields (Pelletier, 2012).

We further note that LM3-FANSY has been evaluated against three independent datasets: the GEMS-GLORI world river discharge database (Meybeck and Ragu, 2012), Global NEWS (Mayorga et al., 2010), and, in the modified manuscript, time series data from the USGS National Water Quality Network (NWQN, Lee, 2022).

We now clarify these aspects, including the description of how the 69 rivers were chosen, in the Methods section (Sect. 3.3) as follows:

### **3.3 Comparisons of measurement-based and modeled estimates**

**For cross-watershed evaluations, we compare LM3-FANSY results of river SS, NO<sub>3</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, DIN (the sum of NO<sub>3</sub><sup>-</sup> and NH<sub>4</sub><sup>+</sup>), DON, TKN (the sum of NH<sub>4</sub><sup>+</sup>, DON, and PON), PO<sub>4</sub><sup>3-</sup>, DOP, and TP (the sum of PO<sub>4</sub>, DOP, PIP, and POP) yields (kg km<sup>-2</sup> yr<sup>-1</sup>), loads (kt yr<sup>-1</sup>), and concentrations (mg l<sup>-1</sup>) with measurement-based estimates from 69 world's major rivers (Table SII-9, the GEMS-GLORI world river discharge database, Meybeck and Ragu, 2012). The 69 river basins cover 55% of global land area (excluding the Antarctic) and are distributed globally across various climates and land use (Fig. SI1). The same comparisons are made by using the Global NEWS database against the measurement-based estimates, yet excluding a few unavailable rivers in Global NEWS (Table SI1, SI4, SI6-9).**

**The 69 rivers were chosen by the following procedure. First, river basins with areas < 100,000 km<sup>2</sup>, about 10 grid cells in our 1 degree resolution, were excluded from the comparisons. Second, river locations were identified by matching latitudes, longitudes, and basin areas of the modeled and reported rivers, which are located either at river mouths or inlands. Third, rivers that were not properly represented within the LM3-FANSY river network were excluded. For example, the GEMS-GLORI database provides a Sanaga**

**River SS concentration at 3.8°N and 10.1°E with its basin area 119,300 km<sup>2</sup>. LM3-FANSY, however, does not capture the Sanaga River at a comparable location (3.5°N and 10.5°E), where the river has a much smaller basin area (5,607 km<sup>2</sup>). The Sanaga River was thus excluded. In total, 7 rivers (i.e., Sepik, Don, Huai, Uruguay, Burdekin, Pyasina, and Sanaga) were excluded for SS comparisons. None of the rivers were excluded for nutrient variables. When a river in the LM3-FANSY river network captures two merged small rivers in the GEMS-GLORI database, water discharge weighted mean concentration of the two rivers was used for analysis. When more than one data were given for a river, the data selected for the first line was used, since it was considered as “the most reliable and generally were obtained first hand by local engineers or scientists (Meybeck and Ragu, 1997)”. When more than one data were given for a river with different basin areas, the data monitored at the location with the largest basin area (i.e., nearest to the river mouth) was used.**

**Since hydraulic controls like damming, irrigation, and diversion affect many rivers, Meybeck and Ragu (2012) distinguish natural river discharges from actual, modified ones. LM3-FANSY does not resolve such hydraulic controls and thus, if available, the natural discharges of Meybeck and Ragu (2012) are used, when calculating loads and yields from their multi-year average concentrations. Comparisons of the model results with loads and yields calculated by using the actual discharges are also presented in Supplementary Information. We report the Pearson correlation coefficient (r) and Nash–Sutcliffe model efficiency coefficient (NSE) between the modeled and measurement-based estimates across the 69 rivers. We also report the prediction error computed as the difference between the modeled and measurement-based estimates of loads expressed as a percentage of the measurement-based load. For global evaluations, we compare LM3-FANSY results with reported global estimates from various references. For cross-watershed and global evaluations, annual results for the year 1990 are analyzed and presented. In parenthesis, ranges of using annual results for the years 1990-2000 are also provided throughout the manuscript.**

For the description of time series analysis, please refer to our response to the reviewer’s previous comment.

The following reference is newly added.

**Meybeck, M. and Ragu, A.: Presenting the GEMS-GLORI, a compendium of world river discharge to the oceans, *Freshwater Contamination*, 243, 3-14, 1997.**

*2. There should be an overview of the loads that flow into the river, the removal loads to the different sectors (sediments, air, ...) and the export to the coastal waters for N and P.*

We have added an overview of the loads in lines 493 and 507 respectively as follows:

**Globally, TN inputs to rivers in LM3-FANSY are 85 (85-91) TgN yr<sup>-1</sup>, of which about 59 (56-59)% are lost to the atmosphere and the other 41 (40-44)% are exported to the coastal ocean.**

**Globally, TP inputs to rivers in LM3-FANSY are 8 (8-9) TgN yr<sup>-1</sup>, of which about 9 (6-10)% are stored within freshwaters and 91 (90-94)% are exported to the coastal ocean.**

*3. The consistency of the P input compared to the N input is a concern. As long as this is not clear, I would suggest to avoid to calculate and present the N:P ratio.*

We agree that the inconsistency of the P and N inputs is a concern that should be addressed in future work. We have highlighted this as a development priority in the Discussion (shown in our response to the reviewer's earlier comment).

We also agree that, given the inconsistency pointed out by the reviewer, and the generally modest skill achieved, we had placed too much emphasis on the N:P ratio in the initial draft. This conclusion was reinforced by the finding that the modest skill achieved was highly sensitive to outliers and subtle recalibration. We thus have removed Fig. 5, Fig. 9, and Sect. 4.3 in the original manuscript, and have limited ourselves to making the following statements and including one supplementary figure (Fig. SI4) to support it (lines 518-526):

**While the simulations are reasonably successful in capturing cross ecosystem differences in both N and P species, variations in their ratio prove more challenging. The model captures the mean ratio, but little of the variation (Fig. SI4). One factor that likely contributes to this is the inconsistency between the N inputs to rivers, the majority of which (~92%) were simulated within LM3, and the P inputs, all of which were drawn from another model (IMAGE-GNM). Continued refinement is thus needed to reliably capture N:P ratio variations in rivers and their subsequent water quality implications.**

4. *In line 438-440 it is clearly states that FANSY does not simulate the processes in reservoirs. This should be discussed including what this means when this model is used in “unprecedented scenarios”.*

The associated discussion was previously in lines 673-676. We now have extended it within the newly introduced Sect. 4.5, which provides a consolidated discussion of model uncertainties and priority developments. The text is as follows:

**Uncertainties associated with sediment dynamics and bed load transport are compounded by the relatively simple representation of lakes, and the exclusion of anthropogenic hydraulic controls like damming, irrigation, and diversion that affect many rivers. For model evaluation, if available, we used the natural water discharges of Meybeck and Ragu (2012) when calculating loads and yields from their multi-year average concentrations (see Sect. 3.3). Large dams or reservoirs, however, have been shown to impound solids and nutrients to substantially decrease their loadings to rivers (Vorosmarty et al., 2003). Thus, despite the relatively low global river SS loads in this first implementation of LM3-FANSY, the lack of such sediment trapping may have induced overestimations of solid and nutrient loads from river basins including large dams or reservoirs. As a representative example, the Colorado River Basin is known for nearly complete trapping of solids due to large reservoir construction and flow diversion (Vorosmarty et al., 2003). LM3-FANSY does not capture such an extreme trapping. As a result, the Colorado River SS load simulated by LM3-FANSY ( $99,235 \text{ kt yr}^{-1}$ ) is more consistent with the corresponding load calculated by using the “natural” water discharge of Meybeck and Ragu (2012) ( $120,010 \text{ kt yr}^{-1}$ ) than with the load calculated by using the “actual” water discharge ( $649 \text{ kt yr}^{-1}$ ). Although use of the actual water discharges is found to not significantly alter the cross-watershed evaluations (Fig. SI6), such anthropogenic hydraulic controls are expected to further increase in the future (Seitzinger et al., 2010). It will be thus important to consider the effects of such controls in future work.**

5. *It is not appropriate to use the measurement data unevenly collected during for 1970s-1990s to validate model results for 1982-2010, especially because there are a lot of published global measurement data available for validation for the period 1982-2010. This introduced high temporal uncertainty may be larger than the model uncertainty itself. Substantial work is needed to re-do the model validation to eliminate this uncertainty.*

The GEMS-GLORI database has been used to evaluate global model results for various years including 1990 (IMAGE-GNM, Beusen et al., 2015), 1995 (He et al., 2011), and 2000 (Global NEWS 2, Mayorga et al., 2010). We previously used the LM3-FANSY results for the year 1990

for cross-watershed and global evaluations and additionally provided the ranges by using the annual results for the years 1982-2010.

We now have modified years for the analyses as 1990-2000 for the cross-watershed and global evaluations, the description of which is summarized in line 442 as follows:

**For cross-watershed and global evaluations, annual results for the year 1990 are analyzed and presented. In parenthesis, ranges of using annual results for the years 1990-2000 are also provided throughout the manuscript.**

### *Specific comments*

*Line 22-23: How you do evaluate which are “better global estimates of nutrient inputs to rivers”? As far as I know, there is no way to validate nutrient inputs to rivers, and it’s hard to say which is a better estimate.*

We have modified the sentence as follow:

**Analyses of the model results and sensitivities to components, parameters, and inputs suggest that fluxes from terrestrial litter and soils, wastewater, and weathering are the most critical inputs to the fidelity of simulated river nutrient loads to the observation-based estimates.**

*Line 32-38: This sentence is too long, and extends for 7 lines. Please re-organize it into several short sentences to increase readability.*

We have modified “, which in turn have been linked to myriad consequences, including”  
to  
“. The consequences include”.

*Line 40: “a comprehensive model” is not clear. In terms of what aspect?*

We have modified “a comprehensive freshwater biogeochemistry model”  
to

“a freshwater biogeochemistry model that captures intertwined algae, nutrient, and solid dynamics”.

***Line 48 and 52: Term use: NO<sub>3</sub>- instead of “NO<sub>3</sub>”, same for NH<sub>4</sub> and PO<sub>4</sub>, same for everywhere in the text.***

We have made the suggested modifications throughout the manuscript.

***Line 43-45: “In particular, inorganic nutrients, which are characterized by higher bioavailability than organic forms (Sipler and Bronk, 2004), have been recognized as critical drivers of algal blooms (including non-HABs) and hypoxic events (Kemp et al., 2005).” It is “excessive inorganic nutrients” rather than simply “inorganic nutrients” that may drive algal blooms and hypoxic blooms.***

We have modified “inorganic nutrients”  
to  
“excessive inorganic nutrients”.

***Line 45: “Meanwhile” should be “Besides”***

We have modified “Meanwhile”  
to  
“Besides”.

***Line 51-53: “Furthermore, nutrient and algae dynamics are strongly linked with solid dynamics through phosphate (PO<sub>4</sub>) sorption/desorption interactions with solid particles (McGechan and Lewis, 2002) and algae growth reduction due to light shading by suspended solids (SS) (Dio Toro, 1978) ”: There are more processes that link nutrient and algae dynamics with solid dynamics, and what you list here are only two of them.***

We have added “for example” between “dynamics” and “through”.

*Line 55-57: “Projecting global freshwater biogeochemistry changes requires process-based models that are robust under unprecedented conditions expected in the next century.” This paper does not involve any future projections, but “unprecedented conditions expected in the next century” and similar statements have appeared many times in this paper. This paper does not show that this model can handle “unprecedented” situations. Therefore, I would advise to remove this kind of sentences.*

The term “unprecedented” was used only twice throughout the manuscript: in abstract and line 57.

Lines 57-70 provide background for this study. The paragraph indicates a generally accepted rationale (e.g., Randall et al., 2007) that future projections into novel environments will be more reliable, as they rest on process-level understanding that is likely to hold in these novel environments, as opposed to extrapolation beyond the bounds of empirical relationships based on historical data. This broad goal sets the context for FANSY development, but does not claim that it has been achieved. We thus feel that it is appropriate here, but have added extensive discussion of areas: this first version of LM3-FANSY should be improved in pursuit of this goal (Sect. 4.5).

The statement in abstract lines 24-27 also describes the purpose of FANSY development to be added as a prognostic component of Earth System Models which is commonly used for future projections.

The extensive discussion in Sect. 4.5 includes:

**There is also significant room for further model development and improvement. An improved representation of lakes (e.g., vertical layering) is necessary to better resolve algal processes and associated transformations between inorganic and organic phases. Modeling large lakes with the ocean component of GFDL’s Earth System Model (Adcroft et al., 2019) is one of our priority developments, particularly given the importance of algae as a control on the relative proportions of inorganic vs. organic nutrients in freshwaters. An initial configuration for the U. S. Great Lakes is currently under development.**

**There is also a need to pursue advances to provide a more comprehensive and consistent approach of modeling the coupled N, C, and P cycles across the terrestrial and freshwater continuum of LM3-FANSY. Expansion of LM3 to include terrestrial P dynamics will be targeted to improve estimates of litter/soil P storage and fluxes to streams and rivers, generating mechanistically consistent estimates of N:P ratios of nutrient loads reaching coastal systems. Priority enhancements will also include integration of freshwater C and**

alkalinity dynamics with the current solid, algae, and nutrient dynamics of FANSY to simulate the impacts of river inputs on coastal C budgets and acidification.

The current version of LM3-FANSY simulates denitrification emissions from freshwaters, but does not include processes that explicitly separate N<sub>2</sub>O and N<sub>2</sub> emissions from the freshwater denitrification. As freshwater N<sub>2</sub>O emissions have been recognized as an increasingly important greenhouse gas source (Yao et al., 2020), it will be important to differentiate N<sub>2</sub> and N<sub>2</sub>O emission processes in future work.

Although LM3-FANSY is capable of producing river solids and nutrients, in various forms and units, some disagreements between the modeled and measurement-based estimates remain. Many observational studies have noted the uncertainties associated with measurement methods, location, and frequency that likely contribute to these disagreements. Additional time varying constraints are also needed to build additional confidence in projected changes. Finally, all of these model improvement efforts will be greatly facilitated by implementing LM3-FANSY at higher resolution capturing the larger number of rivers and by extensive river measurements across the world, with a better assessment of uncertainties.

The following references are newly added.

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al.: The GFDL global ocean and sea ice model OM4.0: Model description and simulation features, *J. Adv. Model. Earth Syst.*, **11**, 3167–3211, <https://doi.org/10.1029/2019MS001726>, 2019.

Yao, Y., Tian, H., Shi, H. et al.: Increased global nitrous oxide emissions from streams and rivers in the Anthropocene, *Nat. Clim. Chang.*, **10**, 138–142, <https://doi.org/10.1038/s41558-019-0665-8>, 2020.

*Line 59: “Prior applications of process-based freshwater biogeochemistry models, such as ..., have generally been limited to small watersheds”. This is not true. Many process-based models have been developed to simulate freshwater biogeochemistry in inland waters more than “small watersheds”, such as DLEM (Tian et al., 2020; Yao et al., 2020; Bian et al., 2022), LOAC (Akbarzadeh et al., 2019), INCA(Whitehead et al., 1998a; Whitehead et al., 1998b; Wade et al., 2002), IMAGE-DGNM (Vilmin et al., 2020; Vilmin et al., 2022), for large river basins and global scale. A better understanding and summary of processed freshwater biogeochemistry models are needed.*

We agree that the statement was too general as written and, as a result, inaccurate. We have rewritten it to acknowledge that process-based models resolving N or P cycle alone have been



widely applied across scales. Applications of models that simulate intertwined algae, multi-nutrient and/or solid dynamics, such as RIVE and QUAL2K, however, have largely been limited to focused studies in smaller watersheds.

We have modified and added statements to clarify our points as follows:

**Process-based freshwater biogeochemistry models of N cycle alone (LM3-TAN, Lee et al., 2014; DLEM, Tian et al., 2020; IMAGE-DGNM, Vilmin et al., 2020; INCA, Wade et al., 2002) or P cycle alone (Bian et al., 2022; Vilmin et al., 2022) have been widely applied across scales. However, applications of models that capture coupled algae, multi-nutrient, and/or solid cycles, such as RIVE (Billen et al., 1994) and QUAL2K (Pelletier et al., 2006), have generally been limited to relatively small watersheds.**

The following references are newly added.

**Bian, Z., Pan, S., Wang, Z., Yao, Y., Xu, R., Shi, H., Kalin, L., Anderson, C., Justic, D., Lohrenz, S., and Tian, H.: A Century-Long Trajectory of Phosphorus Loading and Export From Mississippi River Basin to the Gulf of Mexico: Contributions of Multiple Environmental Changes, *Global Biogeochem. Cycles*, 36, e2022GB007347, 2022.**

**Tian, H., Xu, R., Pan, S., Yao, Y., Bian, Z., Cai, W.-J., Hopkinson, C.S., Justic, D., Lohrenz, S., Lu, C., Ren, W., and Yang, J.: Long-term trajectory of nitrogen loading and delivery from Mississippi River Basin to the Gulf of Mexico, *Global Biogeochem. Cycles* 34, e2019GB006475, 2020.**

**Vilmin, L., Bouwman, A. F. , Beusen, A. H. W., van Hoek, W. J., and Mogollón, J. M.: Past anthropogenic activities offset dissolved inorganic phosphorus retention in the Mississippi River basin, *Biogeochemistry*, 161, 157-169, 2022.**

**Vilmin, L., Mogollón, J. M., Beusen, A. H. W., van Hoek, W. J., Liu, X., Middelburg, J. J., and Bouwman, A. F.: Modeling process-based biogeochemical dynamics in surface fresh waters of large watersheds with the IMAGE-DGNM framework, *J. Adv. Model. Earth Syst.*, 12, 1-19, 2020.**

**Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., Butterfield, D., Rankinen, K., and Lepisto, A.: A nitrogen model for European catchments: INCA, new model structure and equations, *Hydrol. Earth Syst. Sci.*, 6, 559-582, 2002.**

***Line 69-70: Indeed IMAGE-GNM does not differentiate the different forms, but therefore the IMAGE-DGNM is developed (Vilmin et al., 2020; Vilmin et al., 2022).***

In lines 62-70, we are discussing global applications. In lines 58-59 where regional applications are discussed, we have added the IMAGE-DGNM references.

***Line 73-75: “IMAGE-GNM takes a mass balance approach to calculate soil nutrient budgets, which at times rests on simple scaling without potential dynamical feedbacks (e.g., an estimation of litter from floodplains to rivers as 50% of total net primary production (Beusen et al., 2015)).” Firstly, floodplain vegetation is not calculated as an item in the soil budget in IMAGE-GNM but as a direct source to surface water. Secondly, the use of 50% of net primary production (NPP) for wetlands and floodplains can be uncertain, but the area of floodplain and wetlands due to hydrology, as well as the simulated NPP are both spatially explicit and time-dependent. It is therefore not correct to say that calculation of nutrients from floodplain vegetation to rivers in IMAGE-GNM is just simplified without dynamic feedback. Not to mention that for all the other soil and water budget, IMAGE-GNM simulates them in a mass-balanced and dynamic way.***

We apologize for the confusion. What we meant has been clarified as follow:

**IMAGE-GNM does not explicitly simulate terrestrial dynamics, such as vegetation growth, leaf fall, natural and fire-induced mortality, and soil microbial processes, but takes a mass balance approach to calculate dynamic soil nutrient budgets (Beusen et al., 2015).**

***Line 86-89: A sentence too long to read.***

We have modified “dynamics linking”  
to  
“dynamics. LM3-FANSY is aimed at linking ”.

***Line 90: “performance assessment against measurement-based global and regional estimates across world major rivers”. Ambiguous. The global result cannot be compared with “major river (s)” to assess model performance.***

We have clarified it as “performance assessment against measurement-based estimates of nutrient and suspended solid loads across major world rivers”.

*Line 119-123: These sentences are unclear: First sentences states that sewage and aquaculture are from prescribed sources. Last line claims all P input is not from LM3? Please make more clear.*

We have clarified the sentences as follow:

**The freshwater component FANSY receives N, P, and solids in multiple forms either from LM3 or from prescribed inputs. It then simulates biogeochemical transformations and transport of each form of the nutrients and solids within streams, rivers, and lakes. The N and P inputs to rivers are specified in Sect. 3.1.**

For clarification, we have extended the input description in Lines 402-403 as follows:

**Solid and nutrient inputs from terrestrial systems and from the atmosphere to rivers are either simulated by LM3-FANSY or provided by Beusen et al. (2015) (Table 3). For N, all inputs were simulated by LM3-FANSY except aquaculture, wastewater, and atmospheric deposition, which were provided by Beusen et al. (2015). For P, which is not currently included in LM3, all inputs were provided by Beusen et al. (2015).**

**Beusen et al. (2015) provided five-year interval data for the period 1900-2000 at 0.5 degree resolution. The data were regridded to our 1 degree resolution by summing up the values given in kg yr<sup>-1</sup> and linearly interpolated across the five-year intervals. Beusen et al. (2015)'s wastewater N and P inputs were from Morée et al. (2013)'s urban waste N and P discharge estimates to surface waters. Beusen et al. (2015) calculated aquaculture N and P inputs using Bouwman et al. (2013)'s finfish and Bouwman et al. (2011)'s shellfish data. For atmospheric N deposition inputs, Beusen et al. (2015)'s input for the year 2000 was from Dentener et al. (2006)'s ensemble of reactive-transport models and those for the years before 2000 were made by scaling the deposition with Bouwman et al. (2013)'s ammonia emissions. Beusen et al. (2015)'s surface runoff P inputs include those leached from soil P budgets (i.e., the sum of fertilizer and animal manure minus crop and grass withdrawal) and those driven by soil erosion estimates based on Cerdan et al. (2010). Beusen et al. (2015)'s P inputs of litter from floodplains were estimated as 50% of total NPP with a C:P ratio of 1200. Beusen et al. (2015)'s weathering P inputs were computed based on Hartmann et al. (2014)'s chemical weathering P release estimates.**

The following references are newly added.

**Bouwman, A. F., Beusen, A. H. W., Overbeek, C. C., Bureau, D. P., Pawlowski, M., and Glibert, P. M.: Hindcasts and future projections of global inland and coastal nitrogen and phosphorus loads due to finfish aquaculture, Rev. Fish Sci., 21, 112–156, doi:10.1080/10641262.2013.790340, 2013.**

**Bouwman, A. F., Klein Goldewijk, K., Van der Hoek, K. W., Beusen, A. H. W., Van Vuuren, D. P., et al.: Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period, P. Natl. Acad. Sci. USA, 110, 20882–20887, doi:10.1073/pnas.1012878108, 2013.**

**Bouwman, A. F., Pawlowski, M., Liu, C., Beusen, A. H. W., Shumway, S. E., Glibert, P. M., and Overbeek, C. C.: Global hindcasts and future projections of coastal nitrogen and phosphorus loads due to shellfish and seaweed aquaculture, Rev. Fish Sci., 19, 331–357, doi:10.1080/10641262.2011.603849, 2011.**

**Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., et al.: Rates and spatial variations of soil erosion in Europe: A study based on erosion plot data, Geomorphology, 122, 167–177, doi:10.1016/j.geomorph.2010.06.011, 2010.**

**Dentener, F., Stevenson, D., Ellingsen, K., Noije, T. v., Schultz, M., et al.: The global atmospheric environment for the next generation, Environ. Sci. Technol., 40, 3586–3594, 2006.**

**Hartmann, J., Moosdorf, N., Lauerwald, R., Hinderer, M., and West, A. J.: Global chemical weathering and associated P-release – The role of lithology, temperature and soil properties, Chem. Geol., 363, 145–163, doi:10.1016/j.chemgeo.2013.10.025, 2014.**

**Morée, A. L., Beusen, A. H. W., Bouwman, A. F., and Willems, W. J.: Exploring global nitrogen and phosphorus flows in urban wastes during the twentieth century, Global Biogeochem. Cy., 27, 1–11, doi:10.1002/gbc.20072, 2013.**

*Line 132-133: “Algae chlorophyll a (CHL), algae C (Ca), algae P (Pa), and algae dry matter (Da) are diagnosed from algal N (Na) and, in the case of CHL, nutrient and light conditions”, not clear. It’s better to separate it into two sentences and describe the two situations.*

We have modified the sentence as follow:

Algae C ( $C_{al}$ ) and P ( $P_{al}$ ) are diagnosed from algal N ( $N_{al}$ ) assuming the Redfield C:N:P ratio (Chapra, 1997; Redfield et al., 1963). Chlorophyll a (CHL) is derived using the photoacclimation model of Geider et al. (1997) to predict a CHL-to-C ratio ( $r_{CHLC}$ ) and calculate CHL from  $r_{CHLC}$  and  $C_{al}$ .

*Line 135-138: “The 5 prognostic P variables include the same organic forms as for N (dissolved organic P (DOP), particulate organic P (POP), and sedimentary organic P (SedP)), but includes dissolved and particulate inorganic forms (PO<sub>4</sub> and PIP, Sec. 2.2.4) rather than the oxidation state distinction as done for N. ” This sentence is too long and not so readable. Consider improving it. Perhaps add “variables” at the end of this sentence, after “N”.*

We have modified the sentence as follow:

**The 5 prognostic P variables include the same organic forms as for N variables (dissolved organic P (DOP), particulate organic P (POP), and sedimentary organic P (SedP), Sect. 2.2.4). The other two variables are dissolved and particulate inorganic forms (PO<sub>4</sub><sup>3-</sup> and PIP), rather than distinguished by the oxidation state as done for N variables.**

*Line 140: Ambiguous expression of “variable processes”: do you mean processes which are variable, or variable’s processes, or variables and processes?*

We clarified it as “a detailed description of each variable and associated processes”.

*Line 4 in Table 1: From Line 128 in the text, I think you mean Sed is the abbreviation for “bottom sediment”. But in Table 1, it is not clear and differs from the text. Modification is needed to ensure the term use is consistent and clear.*

We have replaced the “sediment solids” with “benthic sediment inorganic solids” in Table 1 and throughout the manuscript.

*Table 1 and text: This is a biogeochemistry model which includes quite some biogeochemical processes and different chemical forms. Therefore, it is not appropriate to use the common*

*symbols of Na, Ca and Pa, as new abbreviations of the terms in this study. Please consider revising these symbols to ensure clarity and avoid confusion.*

We have replaced the symbols to  $N_{al}$ ,  $C_{al}$ , and  $P_{al}$ .

*Figure 1: (1) The pathway of Algae dry matter to SS is quite strange. Do you mean that this part of organic matter will go to the pool of SS in water or finally Sed in sediment? Then this results in permanent additive pools and the pathway, which will only increase. This means that you may have partitioned the algae organic matter directly into one part which will be degraded, and the other refractory part which will not participate in biogeochemical reactions. The refractory part after partitioning will always increase the SS (water) and Sed pools (bottom sediment) pools, but will not add to the total PON and POP pools, which is quite strange. Or another possible assumption for this is that the dead algae are counted as SS which only participates in the physical processes, while in fact the dead algae already become  $NH_4^+$  (see your pathway of another algae mortality), which should not add to the SS pool.*

*Besides, you have SedN and SedP specially for N- and P-contained sediment. The interactions seem somewhat unclear and not self-explanatory.*

We were able to remove “the pathway of algae dry matter to SS” after adopting the advice of reviewer 2 to link SS and PON fluxes from land to rivers. Please see the modified Fig. 1.

We now simulate PON fluxes from terrestrial litter and soils to rivers based on the simulated soil erosion fluxes and litter/soil N concentrations, adapting an approach of previous studies that simulate POC fluxes based on soil erosion fluxes and soil C concentrations (Tian et al., 2015; Zhang et al., 2022). We also now use the simulated PON fluxes to calculate inorganic soil fluxes from the total soil erosion fluxes, so that the eroded soil and PON inputs to rivers are self consistent. This now allows us to use inorganic suspended solids (ISS) in freshwaters as our prognostic variable, and derive total suspended solids (SS, i.e., the sum of ISS and POM) from the combination of freshwater ISS and PON. Accordingly, Sed is now benthic sediment “inorganic” solids, differently from “organic” nutrients of SedN and SedP.

For details, please see our response to the reviewer 2.

The following references are newly added.

**Tian, H., Yang, Q., Najjar, R. G., Ren, W., Friedrichs, M. A. M., Hopkinson, C. S., and Pan, S.: Anthropogenic and climatic influences on carbon fluxes from eastern North**

America to the Atlantic Ocean: A process-based modeling study, *J. Geophys. Res. Biogeosci.*, **120**, 757– 772, 2015.

Zhang, H., Lauerwald, R., Regnier, P., Ciais, P., Van Oost, K., Naipal, V., Guenet, B., and Yuan, W.: Estimating the lateral transfer of organic carbon through the European river network using a land surface model, *Earth Syst. Dynam.*, **13**, 1119–1144, <https://doi.org/10.5194/esd-13-1119-2022>, 2022.

*(2) algae growth and mortality are not the common definitions themselves (see Line 216-217), which calls for earlier explanations or notes to inform the readers from the very beginning. Otherwise this is too confusing.*

We have clarified the statement as follows:

**Algae dynamics are governed by the balance of net growth (i.e., gross growth – respiration) and generalized mortality (i.e., non-predatory mortality + grazing + settling + excretion). The net growth is the difference between gross photosynthesis and respiration. The mortality may include contributions from grazing, excretion, viruses, cell death, and excretion, though these diverse contributions are ultimately parameterized as a simple density-dependent loss term (Dunne et al., 2005). For a batch river and lake system, a mass balance for algae is written as:**

$$\frac{dSB_{N_{al}}}{dt} = \mu(I_{av}, T, N, P) \cdot N_{al} - m(T), \quad (12)$$

where  $N_{al}$  is algae N (kgN) and  $\mu(I_{av}, T, N, P)$  is algae net growth rate ( $s^{-1}$ ) as a function of euphotic zone averaged irradiance  $I_{av}$ ,  $T$ ,  $N$ , and  $P$ .

*Line 156 – 157: Equation 1 and 2. The parameter  $R_i$  is used in both equations with the same sign. Is that correct? Is  $R_i$  not part of  $F_{in}$  or  $F_{out}$ ? Can you explain this better?*

As described in lines 152-154,  $R_i$  is settling/resuspension dynamics and/or biogeochemical reactions for variable  $i$ , which are detailed throughout the sections 2.2.1-2.2.4. For example, if  $i$  is PON,  $R_i$  is Eq. (29) in the original manuscript (now the Eq numbers have been changed). If  $i$  is SedN,  $R_i$  is Eq. (30).

Inputs and outputs via river flows ( $F_{in}$  and  $F_{out}$ ) are only applicable for water column species in Eq. (1). Because Eq. (2) is only for benthic sediment species, it does not have  $F_{in}$  and  $F_{out}$  terms.

We have clarified this as follows:

**In each river reach or lake, for each prognostic variable  $i$ , Settling/resuspension dynamics and/or Biogeochemical reactions ( $SB_i$ ) are calculated according to the process-based formulations described in the following subsections 2.2.1-2.2.4. For example, if  $i$  is PON,  $SB_i$  is Eq. (28). If  $i$  is SedN,  $SB_i$  is Eq. (29). A general mass balance for a variable  $i$  in a river reach or lake at each computation time step (30 minutes in this study) is written as:**

$$\frac{dX_i}{dt} = F_i^{\text{in}} - F_i^{\text{out}} + I_i + SB_i, \quad \text{if } i \text{ is a river/lake water column variable,} \\ \text{i.e., } i = \text{all variables except Sed, SedN, and SedP} \quad (1)$$

$$\frac{dX_i}{dt} = SB_i, \quad \text{if } i \text{ is a river/lake benthic sediment variable,} \\ \text{i.e., } i = \text{Sed, SedN, or SedP} \quad (2)$$

where  $i$  is a prognostic variable listed in Table 1,  $X_i$  is the amount of species  $i$  (kg),  $F_i^{\text{in}}$  and  $F_i^{\text{out}}$  are inflow and outflow of the variable  $i$  ( $\text{kg s}^{-1}$ ),  $I_i$  is inputs of the variable  $i$  from terrestrial systems and the atmosphere ( $\text{kg s}^{-1}$ ), and  $SB_i$  is settling/resuspension dynamics and/or biogeochemical reactions of the variable  $i$  ( $\text{kg s}^{-1}$ ).

We have also consistently modified the notation of the equations in the sections 2.2.1-2.2.4 (Please notice that we have changed  $N_{\text{al}}$  to  $SB_{N_{\text{al}}}$  in Eq. (12) in the previous response.

*Line 167: Which rainfall is meant here? Annual, monthly, weekly, daily? Clarify.*

This is the simulated 3 hourly rainfall. The details can be found in the description of model forcing (lines 394-396).

*Line 170:  $C_1$  is calibrated. How and on what? Please elaborate on this.*

Yes. We have modified the statement in lines 170-172 as follows:

**$C_1$  is a single global value scale parameter (Pelletier, 2012) and coarsely calibrated to reduce prediction errors of SS loads across world major rivers (see Sect. 2.2.5 for details in model calibration).**

We have also introduced a new subsection “2.2.5 Model calibration” to describe details about the LM3-FANSY calibration.



### 2.2.5 Model calibration

Because many of the reported parameters required to simulate the coupled algae, nutrient, and solid dynamics within LM3-FANSY vary widely, it is difficult to assign a single global value for each parameter. Informed by the parameter sensitivity analysis herein (Sect. 4.4), our approach was to coarsely calibrate a limited set of uncertain yet highly influential parameters within their broad observed ranges to reduce prediction errors of SS, N, and P loads in different forms across world major rivers.

First, the free parameter of terrestrial soil erosion,  $C_1$ , was adapted from Pelletier (2012) to account for the input and model resolution difference. It is a scale parameter, as demonstrated in Eq. (3), that affects the overall amount of SS and was calibrated to reduce prediction errors of SS loads.

Second, the generalized algal mortality constant,  $k_m$ , was adapted from Dunne et al. (2005) to reduce the complexity of various algal mortality pathways. This parameter has been tuned to produce reasonable chlorophyll a concentrations in globally distributed lakes, acknowledging the limitation of the present lake biogeochemistry in LM3-FANSY (see Sect. 4.5 further discussion). We adapted an approach of Chapra (2008) to partition nutrient fluxes from the algae mortality to different pools of nutrients in different forms (i.e., particulate organic, dissolved organic, and inorganic), based on fixed fractions. Uncertainties in these fixed fractions due to the lack of theoretical and empirical evidence have been investigated in the sensitivity analysis (Sect. 4.4).

Third, we find that the rate coefficients for hydrolysis, nitrification, and denitrification are highly influential parameters for determining river nutrient loads in different forms, relative to those for decay processes that breakdown complex organic compounds into simpler organic and inorganic compounds. These highly influential parameters have been calibrated to reduce prediction errors of nutrient loads in different forms. We adapted an approach of Bowie et al (1985) to partition fluxes from complex organic nutrient decomposition to more simpler organic vs. inorganic nutrient pools based on a uniform fraction. Uncertainties in the fraction have been investigated in the sensitivity analysis (Sect. 4.4).

*Line 1 in Table 2: “Many reactions are reported at 20°C” is not appropriate for the use of the ideal temperature ( $T_{ref}$ ) of all reactions. Detailed citations of mechanism studies are needed. Also, the dynamic feedback, which has been advertised so many times, cannot be well examined with the same  $T_{ref}$  for every biogeochemical process related to every element form*

***driven by every organism group. Many mechanism studies have shown the obvious difference in Tref of nitrification, denitrification, mineralization, primary production, etc. Using the same Tref constraint for all biogeochemical processes related to all element forms driven by organism groups will yield a wrong estimation of the process rates, element and algae processing and the consequent nutrient concentrations.***

The choice of a reference temperature is generally a decision made by the investigator fitting the relationships that we draw from in constructing the model. These relationships can generally be recast to mathematically identical forms using different reference temperatures, but 20 °C, is a common and thus intuitive reference, e.g.: “many reactions are reported at 20 °C” as stated in a book “Surface Water-Quality Modeling” by Chapra (1997).

Reported temperature correction factor,  $\theta$ , values for different reactions vary widely (Bowie et al., 1985 and references there). The value of 1.066, which we used for all the reactions, is a typical value for phytoplankton growth (Chapra, 1997; Eppley, 1972), but is also within the reported ranges for the other reactions. The  $\theta$  values for organic nutrient decompositions that break down complex organic nutrients into simpler ones (e.g., PON to DON, POP and SedP to DOP) range from 1.02 to 1.08. The  $\theta$  values for organic nutrient decompositions to inorganic ones (e.g, DON, PON, and SedN to  $\text{NH}_4^+$ ; DOP, POP, and SedP to  $\text{PO}_4^{3-}$ ) range from 1.02 to 1.14. The  $\theta$  values range from 1.02 to 1.08 for nitrification and range from 1.02 to 1.09 for denitrification.

Given the wide reported ranges of the  $\theta$  values, it is difficult to assign a specific value for each reaction, reflecting the limitation of any given global models including LM3-FANSY. We now have assigned the  $\theta$  value of 1.066 for algal dynamics and of 1.08 (a typical value for sediment oxygen demand) for the other reactions.

We have also conducted 4 sensitivity simulations to show model responses to different  $\theta$  values: (1) 1.066 for algal dynamics and 1.024 (a typical value for oxygen reaeration) for the other reactions, (2) 1.066 for algal dynamics and 1.047 (a typical value for biological oxygen demand) for the other reactions, (3) 1.066 for algal dynamics and 1.066 for the other reactions, and (4) 1.066 for algal dynamics, 1.08 for sediment decompositions, and 1.047 for the other reactions.

We find insensitiveness of the model to these variations, which can be found in the sensitivity analysis (Sect. 4.4, Table 7).

***Line 6 in Table 2 and Line 192: “R” symbol is used before, as term reaction. It is confusing to use the same letter for two different terms. Change one of them.***

For settling/resuspension dynamics and/or biogeochemical reactions, we have modified  $R_i$  to  $S_{Bi}$ .

***Equation 8: Please refer to the original reference (Ludwig and Probst). This equation is very old and only based on 19 observations. Are there not more observations to improve this?***

We were able to remove this equation after adopting the advice of reviewer 2 to link SS and PON fluxes from land to rivers. For details, please see [our earlier response to the reviewer's comment on Figure 1: \(1\)](#).

***Line 216-217: the definitions of algae growth and mortality differ from other studies! This should be mentioned to inform readers, or at least tell them these processes are "generalized" or "net". Otherwise, it is too confusing.***

We have attempted to clarify this in [our earlier response to the reviewer's comment on Figure 1: \(2\)](#):

We hope that addresses the reviewer's concern. If we have missed the mark, please let us know which other studies have adopted different definitions and we will clarify further.

***Line 273: it is hard to understand how equation #23 calculates the combined the net flux of a generalized mortality (non-predatory mortality + grazing + settling + excretion) as mentioned in Line 216-217. Contradictorily, according to Line 271-277, this equation does not reflect so-called "generalized mortality", but grazing (death because of predation).***

As discussed in Dunne et al. (2005), phytoplankton mortality processes are generally divided according to the degree of their density dependence. Grazing, for example, is often modeled as density-dependent loss with a quadratic relationship. Such a relationship implicitly assumes that the biomass of grazers scales in proportion to the biomass of their prey without lag. Similar assumptions have been used for losses to marine viruses (e.g., Stock et al., 2014) and aggregation (e.g., Doney et al., 1996). In reality, however, phytoplankton and their grazers are not perfectly coupled. Zooplankton responses lag phytoplankton fluctuations. There are also loss mechanisms such as excretion, basal respiration, and cell death that may not be density dependent. Dunne et

al. (2005) found that a density-dependent but weaker  $4/3$  scaling rather than a quadratic scaling did a better job of simulating dynamic shifts in phytoplankton biomass (e.g., blooms) while also setting an upper biomass bound in a given nutrient and light environment.

We have attempted to clarify these points while leaning on previously published works for the details:

“where  $k_m$  is temperature-dependent algae mortality rate ( $\text{kgN}^{-1/3} \text{s}^{-1}$ ) that reflecting the combined impacts of zooplankton grazing and other phytoplankton loss terms (e.g., viral-induced losses, cell death). Exponents between  $4/3$  and 2 have been commonly applied in this relationship, with higher values corresponding to more tightly coupled top-down control (Dunne et al., 2005). We have adopted a value of  $4/3$  to enable high biomass in nutrient rich environments.”

Doney, S.C., Glover, D.M., and Najjar, R.J.: A new coupled, one-dimensional biological–physical model for the upper ocean: applications to the JGOFS Bermuda Atlantic Time-series Study (BATS) site. *Deep-Sea Research Part II* 43, 591–624, 1996.

Stock, C. A., Dunne, J. P., and John, J. G.: Global-scale carbon and energy flows through the marine planktonic food web: An analysis with a coupled physical–biological model, *Progress in Oceanography*, 120, 1-28, 2014.

***Line 291: “assuming C weight content .... 1500 kg m<sup>3</sup>”. Does this mean that you use a global number for bulk density and for the C content of the soil? I hope not...***

Sorry for the misunderstanding. Dynamically simulated soil N and, by extension, C contents vary spatially and temporally. The simulated effective soil depth, as defined by Gerber et al., 2010, was used to calculate spatially and temporally varying N concentrations per unit volume from N contents per unit area. We have now clarified this in the text.

***Line 297-296: I wonder why PON cannot be calculated from TKN-NH<sub>4</sub>-DON for this calibration purpose.***

As shown in Table SI3-5, there is only one river data available for the all three species TKN,  $\text{NH}_4^+$ , and DON.

*Line 299-302: I wonder whether or not fPON can or should be used for compensating hydraulic controls like dams and reservoirs. I'm afraid this is not very appropriate for a mechanistic model. This factor is constant (Table 2) and from the calibration of measurements. But it is unclear how this calibration was performed. Line 295 claims that fPON slows down the N movement, but that that not make any sense.... Also the "limited measurement-based PON estimates" makes it a bit a mystery.*

We were able to remove Eq. (28) including fPON after adopting the advice of reviewer 2 to mechanistically link SS and PON fluxes from land to rivers. For details, please see [our earlier response to the reviewer's comment on Figure 1: \(1\)](#).

*Line 328: see comment for Line 273. I don't find any rationality for using a part of the mortality calculation of predation-caused algae death as DON excretion to NH<sub>4</sub><sup>+</sup>. Also, I don't know how fDON used here is calibrated, which is very strange to be done because generally it is unknown how much DON excretion will account for in algal mortality due to predation.*

We apologize for the misunderstanding: there is no path of "algae death as DON excretion to NH<sub>4</sub><sup>+</sup>" in FANSY. We have clarified the model calibration by adding a new section (Sect. 2.2.5 Model calibration). Please see for the full model calibration description in [our earlier response to the reviewer's comment on Line 170](#), which includes the description of fDON calibration:

**We adapted an approach of Chapra (2008) to partition nutrient fluxes from the algae mortality to different pools of nutrients in different forms (i.e., particulate organic, dissolved organic, and inorganic), based on fixed fractions. Uncertainties in these fixed fractions due to the lack of theoretical and empirical evidence have been investigated in the sensitivity analysis (Sect. 4.4).**

*Equations 35 – 38: What is CPN?*

We corrected the typo cPN to rPN. It is the algae P to N ratio used to diagnose algae P from algae N.

***Line 367: Calculation for 1900 – 2010. But Beusen et al. (2015) has provided only data up to 2000. How did you calculate the last ten years?***

As shown in our earlier response to the reviewer’s major comment 5, we now have limited our analyses up to 2000.

***Line 368: Confusing: “reported point and nonpoint N and P inputs to rivers” Can you better describe what you used?***

The input description has been clarified in our earlier response to the reviewer’s comment on Line 119-123.

***Line 370: 1990 calibrated (constant) from partitioning factors for instream processes may be not suitable for the period 1982-2010 considering the possible temporal changes. Calibration process is not clear here.***

We have removed the statements. As shown in our earlier response to the reviewer’s major comment 5, we now have modified the years for the analyses to focus on 1990-2000. Please see our earlier response to the reviewer’s comment on Line 170 for the full description of model calibration (Sect. 2.2.5 Model calibration).

***Line 373-381: This atmospheric N deposition is calculated here, and can be different from Beusen et al. 2015. However, In Table 3 it is clearly written that atmospheric N deposition is from Beusen et al. 2015. So what are you using? Please clarify.***

As described in lines 380-381, the atmospheric N deposition is only applied to soils, not to rivers. The description of inputs to rivers has been clarified in our earlier response to the reviewer’s comment on Line 119-123.

***Table 3: So LM3-FANSY simulates soil N budget and soil budget and their delivery to rivers and lakes, but no P budget. This means that P is not coupled to the LM3. How consistent are***

*the assumptions of P of Beusen et al. (2015) with the assumptions of LM3? Do they both have agricultural land on the same spots? This is a major concern of me, because one of the results is the N:P ratio. If the assumptions do not match, then the N:P ratio does not mean anything. Perhaps the N:P ratio is a bridge to far for this model at the moment.....*

We agree and have addressed this concern in our earlier response to the reviewer's major comment 3.

*Table 3: The numbers in Table 3 are not the same as in in Vilmin et al. (2018). The fractions for wastewater are dependent on the treatment level. Here it is assumed that whole the world has primary treatment, which is not true. The fractions of agricultural surface runoff and aquaculture are different than Vilmin et al. (2018). How is this determined?*

We agree that the description of using the fractions from Vilmin et al. (2018) was not clear in the previous manuscript. Since only aggregated inputs were available from Buesen et al. (2015), we had to adopt an admittedly simplified assumption for the wastewater additions to each of the three species. We have extended the description, and also conducted sensitivity tests to show that uncertainties associated with the fractions had only a small impact on our results (Table SI10). The clarified descriptions have been added in Table 3 and line 404 as follows:

**Vilmin et al (2018) suggest fractions to divide TN and TP inputs to rivers from different sources into three N and P species respectively. For sewage, three groups of fractions are given for three types of sewage (i.e., untreated, primary treated, and secondary/tertiary treated). The sewage inputs from Beusen et al. (2015) that we used are aggregated for all three types of sewage. To divide Beusen et al. (2015)'s sewage TN and TP inputs into three N and P species respectively, we have taken a middle ground by using the fractions for primary treated sewage. Although we acknowledge that use of the fractions is unrealistic, we find that our results are relatively insensitive to a wide range of alternatives, as assessed by 4 parameter perturbation experiments (see Table 3 and the sensitivity analysis in Sect. 4.4). The first 2 uncertainty simulations used 1) the fractions for untreated sewage and 2) the fractions for secondary/tertiary treated sewage. The next 2 uncertainty simulations used assumed fractions based on an estimate that globally, over 80% of wastewater is discharged without "adequate treatment" (Environment and Natural Resources Department, 2022). Specifically, the fractions are driven by assuming that 1) 80%, 10%, and 10% of Beusen et al. (2015)'s sewage are untreated, primary treated, and secondary/tertiary treated respectively and 2) 40%, 40%, and 20% of Beusen et al. (2015)'s sewage are untreated, primary treated, and secondary/tertiary treated respectively. In all**

cases, the simulated river loads of each species changed by < 9%, and the simulated total loads did not change.

For TP fluxes from agricultural lands to rivers, two groups of fractions are given for two sources (i.e., surficial runoff and soil loss). Agricultural surface runoff TP inputs from Beusen et al. (2015) are aggregated for both sources. To divide Beusen et al. (2015)'s agricultural surface runoff TP inputs into three P species, we have used nearly equal fractions. Although we acknowledge that use of the fractions produces uncertainties, we used perturbation experiments to show that our results are relatively insensitive to a wide range of the fractions (see Table 3 and the sensitivity analysis in Sect. 4.4). The 6 uncertainty simulations used 1) the fractions for surficial runoff for all agricultural fluxes, 2) the fractions for soil loss for all agricultural fluxes, fractions driven by assuming that 3) 40% and 60% of Beusen et al. (2015)'s agricultural surface runoff TP inputs are from surficial runoff and soil loss respectively, 4) 60% and 40% of those are from surficial runoff and soil loss respectively, 5) 20% and 80% of those are from surficial runoff and soil loss respectively, and 6) 80% and 20% of those are from surficial runoff and soil loss respectively. In all cases, the simulated riverine loads of each species changed by < 13%, and the simulated total load changed by < 1%.

For aquaculture, two groups of fractions are given for particulate and dissolved sources. The fractions in Table 3 are driven by assuming that about 12% and 31% of aquaculture TN and TP inputs from Beusen et al. (2015) are particulate, based on Figure 3 of Bouwman et al (2013). We did not conduct sensitivity simulations for aquaculture, because aquaculture nutrient inputs are very small compared to the other sources (< 1% of total inputs for both N and P), and thus uncertainties associated with the fractions should be negligible for the global application herein.

The following reference is newly added.

Environment and Natural Resources Department, Wastewater as a resource, European Investment Bank, <https://www.eib.org/en/publications/wastewater-as-a-resource>, 2022

*Line 434-445: The use of observation data for this model is very strange. There are a lot of published global measurement data available for validation for the period 1982-2010, but this paper intentionally chooses the measurement data during the 1970s-1990s which Global NEWS used. This introduced high temporal uncertainty may be larger than the model uncertainty itself. It is not clear how the comparison is made. The timestep is 30 minutes. So how is the value that is used for calibration constructed? Discharge weighed? Mouth of the river or the location of the observation? Only the year 1990? Or average over yearly*



***concentrations of 1970 – 1990 or 1982 – 2010? This is very unclear. The time period difference is very strange and is also a concern. A lot has happened with the N and P loads/concentrations in the period 1982 – 2010.***

GEMS-GLORI is a widely cited database (Meybeck and Ragu, 2012) that has been used by many global watershed modeling studies (e.g., He et al., 2011; Beusen et al., 2015), in addition to Global NEWS. We have also used the database, because it provides perhaps the most comprehensive river chemistry. The GEMS-GLORI database was compiled in a consistent manner and provides a detailed data description.

As noted in our earlier response, we further evaluated our results against Global NEWS considering it as a “gold standard” and widely cited and pioneering community benchmark.

All the dynamics in FANSY described in Sect. 2.2 are updated every 30 minutes. This means that the model outputs (including water, algae, solid, and nutrients throughout the river network) are given at the 30 minute time step. The annual results for yields, loads, and concentrations are the averages of 30 minute results with applicable unit conversions. This high temporal resolution does not require discharge-weighted calculations.

We have clarified the years used for the analyses in our earlier response to the reviewer’s major comment 5.

Also, we have clarified the description of the comparisons of measurement-based and modeled estimates in our earlier response to the reviewer’s major comment 1.

***Line 444-445: “Cross-watershed .... and trends.” Here the purpose of the model is defined. I do not agree with this. If this is true, then you don’t need to spinup 11000 years, and do all the dynamic calculations. Just use the Global NEWS model. This model is far to complicated for this purpose and this purpose does not fit in the LM3 framework.***

We have extensively clarified the purpose of this model and this manuscript at the beginning of this response.

A land model spin-up is to create steady state conditions of terrestrial systems before forcing the model with contemporary conditions. It is a necessary procedure even for evaluating a single time period results of mostly unperturbed systems (e.g., Zhang et al., 2020).

***Line 450-452: It is unclear whether it is because of the calibrated parameter of instream solid dynamics that improves the model performance, because the results without calibrated parameter are not shown.***

As demonstrated in Eq. (3) and lines 167-172,  $C_1$  is a single global value scale parameter to account for the difference in input data and model resolution. Eq. (3) demonstrates that the model predictive capacity for the spatial variation has little to do with this scale parameter, but is more likely explained by spatially varying inputs (e.g., slope, rainfall).

***Table S11 – S19: Nice to have all the observations together, but I miss the results from FANSY in these tables. Then the reader can also see, which rivers are doing fine or not.***

The LM3-FANSY results corresponding to the observational estimates in Table S11-S19, along with all the other LM3-FANSY results, are reported in the supplementary excel file. Thus, the readers can see the LM3-FANSY performance relative to the observational estimates and perform additional analysis if they desire.

***Table 4: Again, “R” is used as a new abbreviation for another term in the same paper.***

The symbol “R” in Table 4 has been removed.

***Table 4 and 5: Not clear how this 15% is applied. Is the change applied only for 1990 and the calculation is only done for 1990? Constant over the year or not? Or total N or P and then some forms get more change than others? Please elaborate on this in the methods section.***

We have extended and clarified the description of sensitivity simulations in Sect. 3.2 as follows:

### **3.2 Sensitivity simulations**

**Model sensitivities to the inputs, components, and parameters are examined by analyzing the responses of river solid and nutrient loads to their changes for the period 1948-2000. Model sensitivities to the nutrient inputs to rivers are examined by increasing each input source by 15% and removing it. When the inputs are from LM3, the 15% increase was**

applied to each input variable (i.e.,  $\text{NO}_3^-$ ,  $\text{NH}_4^+$ , DON, and PON). When the inputs are externally prescribed, the 15% increase was applied to each input source listed in Table 3. In both cases, the increases were applied uniformly over the grid and time.

One of the distinct features of LM3-FANSY is the capability of modeling interactions of algae and nutrient dynamics with solid dynamics. Light shading by SS and algae themselves modulates the strength of algal productivity and, in turn, river solids and nutrients. In LM3-FANSY, the light extinction coefficient is dynamically simulated as a function of ISS, POM, and CHL (Eq. 23), instead of using a prescribed parameter. To evaluate how critical the dynamic light extinction component is for model predictive capacity, the component was replaced with a prescribed parameter value ( $k_e = 0.15$  or  $0.45$ ) and the river load responses to more active algal populations are examined.

Model sensitivities to the parameters are examined by decreasing each of all calibrated parameters listed in Table 2 by half and increasing it by twice. For the parameters which have much smaller or much larger impacts compared to the other parameters, additional sensitivity tests have been performed to show responses of river loads to the parameter changes in their broad observed ranges.

*Tables 4-5: hard to interpret. To me, this sensitivity does not make any sense. First of all because the output parameter is the Pearson correlation coefficient. I can not understand this, how this helps me to understand the model mechanics. Better to do this on output parameters of the model. Then also an explanation could be added on why this difference is big or not. Secondly, I think that a one factor analysis is very limited for this kind of models.*

As shown in our response to the reviewer's opening comments, the sensitivity analyses are based on both 1) the Pearson correlation coefficient (Table 4 in the original manuscript), which tests the impact of a parameter, input, or component change on the spatial agreement between the simulated and observed relative river loads, and 2) the amount of river solid and nutrient loads (Table 5 in the original manuscript), which tests the effect of a change on the magnitude of the river loads. In the sensitivity analysis, we show the load differences between the baseline and sensitivity simulations and the model responses to different inputs/components (e.g. lines 640-643, 651-658).

In the modified manuscript, as done in a previous watershed modeling study (Harrison et al., 2010), we have further extended analysis by adding the Nash–Sutcliffe model efficiency coefficient which reflects “Model Efficiency”. This addition is to evaluate the effects of each input source or model component removal on model predictive capacity. According to Harrison et al. (2010),

“Components with little impact on the model had little impact on Nash-Sutcliffe efficiency, whereas removal of critical model components had a large impact on Nash-Sutcliffe efficiency.”

As shown in our response to the reviewer’s earlier comment, we have substantially expanded sensitivity/uncertainty analyses by changing all calibrated parameters and by changing the fractions of dividing the externally prescribed TN and TP inputs into different N and P species, in addition to the previous model component and input manipulations. This has increased the total number of sensitivity/uncertainty simulations from 22 to 87, requiring 4611 years of total sensitivity/uncertainty simulations. Discussion of the insights gained by these expanded simulations has also been extended considerably.

We do not feel that Monte Carlo experiments changing all parameters at once are appropriate at this initial development stage because 1) it is challenging, if not impossible, to robustly attribute a response to a parameter or process when one is changing everything at once and 2) many of the simulations arising from such experiments prove implausible when compared with observations. We agree that such approaches are proper when constructing an ensemble of similarly plausible model permutations, but only after proper filtering for model fidelity and once the more basic single factor sensitivities are understood.

We feel that the significant expansion of single factor sensitivities/uncertainties and discussion made in response to the reviewer’s comments and those of reviewer 2 have strengthened this model description paper considerably. The extended sensitivity analysis is as follows:

#### **4.4. Model sensitivity with changes in parameter settings and nutrient inputs**

**For solid dynamics, the free scale parameter of terrestrial soil erosion ( $C_1$ ) plays a significant role in determining the overall amount of river SS loads, with its decrease (increase) by half (twice) leading to a 50% decrease (99% increase) in global river SS loads (Table 7). The decrease (increase) in  $C_1$  also reduces (enhances) the erosion associated fluxes from terrestrial litter and soils and, in turn, river PON loads. The decrease (increase) in  $C_1$  also reduces (enhances) sorption of  $PO_4^{3-}$  to solid particles, as reflected by DIP vs. PIP load changes. These results associated with terrestrial soil erosion are, however, found to be insensitive to one order of magnitude changes in POM-to-PON ratios in eroded terrestrial soils and/or freshwaters (Sect. 2.2.1).**

**For nutrient dynamics, the responses of river loads to removal of each nutrient input source or to an increase of it by 15% suggests that terrestrial runoff and erosion are the dominant drivers of river N loads, followed by wastewater (Table 7). For river P loads, terrestrial runoff and erosion are also the dominant drivers, but unlike for N, the second largest driver is weathering, followed by wastewater. Terrestrial runoff and erosion also**

play a critical role in explaining model efficiency and spatial distribution of river nutrient loads (Table 8). Removal of this source reduces NSE and r values substantially. Wastewater plays a modest role in explaining model efficiency and spatial distribution of river P loads. Removal of aquaculture or atmospheric deposition has almost no impacts on the amount, model efficiency, and spatial variation of river loads, suggesting that inaccuracies in these inputs have minor impacts on regional and global model estimates, relative to the inaccuracies associated with the other model inputs. However, the importance of each source is likely to vary, depending on the dominant control on river nutrient loads in a specific region.

The parameter sensitivity tests show the relatively insensitive responses of river nutrient loads to the changes in the rate coefficients of decay processes that break down complex organic compounds into simpler organic or inorganic compounds. The changes in the rate coefficients for hydrolysis, nitrification, and denitrification, however, have relatively large impacts on river nutrient loads. This is especially for the freshwater N cycle, which includes an additional loss pathway to the atmosphere via denitrification unlike the freshwater P cycle. The role of freshwater denitrification on global river N loads, however, has not been explicitly investigated by previous global watershed modeling studies. Our sensitivity results imply a prominent role of freshwater denitrification in determining the amount of N loss to the atmosphere vs. N exports to the coastal ocean at both global (Table 7) and regional (Fig. 8) scales.

Algae dynamics play a significant role in determining the relative composition of inorganic vs. organic nutrients in freshwaters. Decreasing the algal mortality rate constant by half enhances algal uptake, decreasing DIN (the sum of  $\text{NO}_3^-$  and  $\text{NH}_4^+$ ) and IP (the sum of  $\text{PO}_4^{3-}$  and PIP) by -11% and -33% respectively, while it increases ON (the sum of DON and PON) and OP (the sum of DOP and POP) by 24% and 33% respectively. Similarly, increasing the maximum photosynthesis rate or chlorophyll a-specific initial slope of the photosynthesis-light curve by twice enhances algal uptake, decreasing DIN by -6 or -7% and IP by -21% or -25% respectively, while it increases ON by 18% or 21% and OP by 23% and 28% respectively. The opposite holds for the parameter changes that reduce algal uptake. An analysis of the model sensitivity simulations, wherein the dynamic contributors to light extinction (i.e., ISS, POM, and CHL in Eq. (23)) were removed, further suggests that proper light limitation of algal growth is also important for skillful estimates of freshwater inorganic vs. organic nutrients. Removing the dynamic light shading component leads to a ~26 % and ~35% overestimation of ON and OP loads and an underestimation of DIN and IP loads by ~10 % and ~32% respectively. Inorganic nutrient levels are suppressed by invigorated algal populations and more nutrients end up in organic forms. Algal controls thus offer an effective means of calibrating the mix of inorganic and organic constituents. We note uncertainty associated with the fractions that partition nutrient

fluxes from algae mortality to different forms, which appears to have a modest effect on organic nutrient loads (Table 7).

Finally, additional uncertainty tests have shown the relatively insensitive responses of river loads to a broad range in 1) the fractions that divide externally prescribed TN and TP inputs into different N and P forms (Table SI10), 2) the fractions that partition fluxes from complex organic nutrient decomposition to simpler organic vs. inorganic nutrients, and 3) the temperature correction factor,  $\theta$ , values that account for the temperature effect on freshwater biogeochemical reactions.

*Table 6: Needs to mention the estimates from IMAGE-GNM are for 2000 instead of the 1990s as in other studies.*

We have added years for the estimates listed in Table 6 (which is Table 5 in the modified manuscript).

*Table 6: LM3-FANSY results needed to be added in this table for straightforward comparison.*

We have added the LM3-FANSY results in Table 6 (which is Table 5 in the modified manuscript).

*Line 485-486: This is not true. global river chemical measurement data are not limited, but they are simply not used in this paper.*

The statement in lines 485-486 means that measurement-based estimates of particulate nutrient compounds are limited. The relative lack of particulate nutrient measurements is also noted in the description of the GEMS-GLORI database (Meybeck and Ragu, 1997). To date, we have not seen any global watershed model evaluated against observational PON or POP across globally distributed rivers. We have clarified the statement as follows:

**Measurement-based estimates of particulate nutrient compounds to evaluate our model results at 1 degree resolution are limited.**

***Line 495-498: Explanations are needed for the estimate ranges like 36.4-41.3. Not clear whether this is because of temporal change or uncertainty or something different.***

The definition of the ranges (which are due to interannual variability) was previously defined in the caption of Table 5 wherein the estimates were reported. As shown in our earlier response to the reviewer's major comment 5, we now have modified the years for the analyses.

***Line 507: This model estimates a global export of 6.5-7.8 TgP. But if I correctly understand this article, you used the same riverine input as Beusen et al (2015) who estimated a global export of 4 TgP. So there is problem here. The global input of P is around 9 TgP (stated in Beusen et al. (2015)). So the P loss in the river is around 30% whereas the P loss in Beusen et al is above 50%. This is a huge difference and should be explained here. Also the N loss should be checked!***

We note that the overall freshwater N and P budgets have been described in our earlier response to the reviewer's major comment 2 as follows:

**Globally, TN inputs to rivers in LM3-FANSY are 85 (85-91) TgN yr<sup>-1</sup>, of which about 59 (56-59)% are lost to the atmosphere and the other 41 (40-44)% are exported to the coastal ocean.**

**Globally, TP inputs to rivers in LM3-FANSY are 8 (8-9) TgN yr<sup>-1</sup>, of which about 9 (6-10)% are stored within freshwaters and 91 (90-94)% are exported to the coastal ocean.**

In line 507 with the P budget, we now have added the following statements:

**IMAGE-GNM estimates that ~56% (5 of 9 Tg yr<sup>-1</sup>) of global TP inputs to rivers are stored within freshwaters (Beusen et al., 2016). This is a large difference from our estimate of 9 (6-10)% storage, but the difference is around a very uncertain number as it has not been directly measured. Our estimate of global river TP loads to the coastal ocean (7, 7-8 Tg yr<sup>-1</sup>) falls within the range of other estimates (9 Tg yr<sup>-1</sup> from Global NEWS, Mayorga et al., 2010 and 4 Tg yr<sup>-1</sup> from IMAGE-GNM, Beusen et al., 2016, Table 5). The relatively low freshwater P retention in LM3-FANSY may arise, in part, from the lack of dams and reservoirs (see Sect. 4.5 for further discussion). The overall consistency of our SS and P estimates with the observed cross-watershed constraints (Fig. 2, 4), however, suggests that the bias introduced by the simplification may not be large.**

***Line 521: should be “Arctic”.***

The word has been removed with the reduction of the N:P analysis following the reviewer’s earlier comments.

***Line 551: grammar error: “are the among the ...”***

We have modified “are the among”  
to  
“are among”.

***Figures 6-8: The maps are for 1990 but the comparisons of concentrations, loads and yields are performed for 1982-2010. Why is there a difference in time between the two result analyses?***

As shown in our earlier response to the reviewer’s major comment 5, we now have modified the years for the analyses.

## ***References***

- Akbarzadeh, Z., Maavara, T., Slowinski, S., Van Cappellen, P., 2019. Effects of damming on river nitrogen fluxes: a global analysis. Global Biogeochem. Cycles 33, 1339-1357.***
- Beusen, A.H.W., Bouwman, A.F., Van Beek, L.P.H., Mogollón, J.M., Middelburg, J.J., 2016. Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. Biogeosciences 13, 2441-2451.***
- Beusen, A.H.W., Van Beek, L.P.H., Bouwman, A.F., Mogollón, J.M., Middelburg, J.J., 2015. 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. Geosci. Model Dev. 8, 4045–4067.***



*Bian, Z., Pan, S., Wang, Z., Yao, Y., Xu, R., Shi, H., Kalin, L., Anderson, C., Justic, D., Lohrenz, S., Tian, H., 2022. A Century-Long Trajectory of Phosphorus Loading and Export From Mississippi River Basin to the Gulf of Mexico: Contributions of Multiple Environmental Changes. Global Biogeochem. Cycles 36, e2022GB007347.*

*Tian, H., Xu, R., Pan, S., Yao, Y., Bian, Z., Cai, W.-J., Hopkinson, C.S., Justic, D., Lohrenz, S., Lu, C., Ren, W., Yang, J., 2020. Long-term trajectory of nitrogen loading and delivery from Mississippi River Basin to the Gulf of Mexico. Global Biogeochem. Cycles 34, e2019GB006475.*

*Vilmin, L., Mogollón, J. M., Beusen, A. H. W., and Bouwman, A. F.: Forms and subannual variability of nitrogen and phosphorus loading to global river networks over the 20th century, Glob. Planet. Change, 163, 67–85, 2018.*

*Vilmin, L., Bouwman, A.F., Beusen, A.H.W., van Hoek, W.J., Mogollón, J.M., 2022. Past anthropogenic activities offset dissolved inorganic phosphorus retention in the Mississippi River basin. Biogeochemistry 161, 157-169.*

*Vilmin, L., Mogollón, J.M., Beusen, A.H.W., van Hoek, W.J., Liu, X., Middelburg, J.J., Bouwman, A.F., 2020. Modeling process-based biogeochemical dynamics in surface fresh waters of large watersheds with the IMAGE-DGNM framework. Journal of Advances in Modeling Earth Systems 12, 1-19.*

*Wade, A.J., Durand, P., Beaujouan, V., Wessel, W.W., Raat, K.J., Whitehead, P.G., Butterfield, D., Rankinen, K., Lepisto, A., 2002. A nitrogen model for European catchments: INCA, new model structure and equations. Hydrol. Earth Syst. Sci. 6, 559-582.*

*Whitehead, P.G., Wilson, E.J., Butterfield, D., 1998a. A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I - Model structure and process equations. Sci. Total Environ. 210-211, 547-558.*

*Whitehead, P.G., Wilson, E.J., Butterfield, D., Seed, K., 1998b. A semi-distributed integrated flow and nitrogen model for multiple source assessment in catchments (INCA): Part II - Application to large river basins in south Wales and eastern England. Sci. Total Environ. 210-211, 559-583.*

*Yao, Y., Tian, H., Shi, H., Pan, S., Xu, R., Pan, N., Canadell, J.G., 2020. Increased global nitrous oxide emissions from streams and rivers in the Anthropocene. Nat. Clim. Chang. 10, 138-142.*

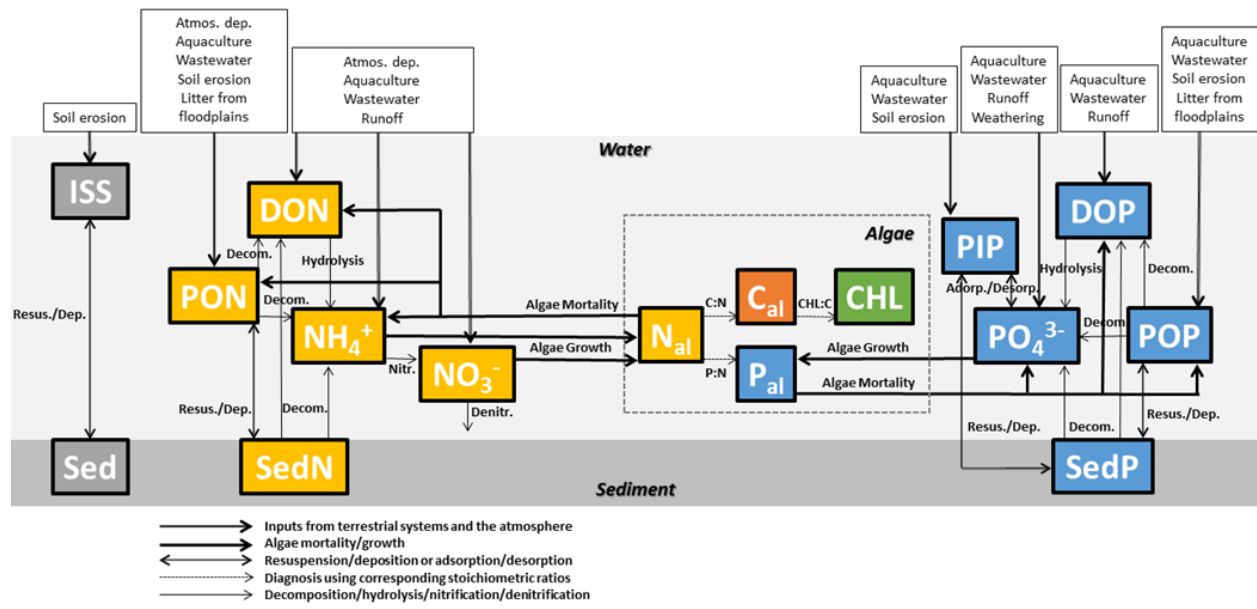
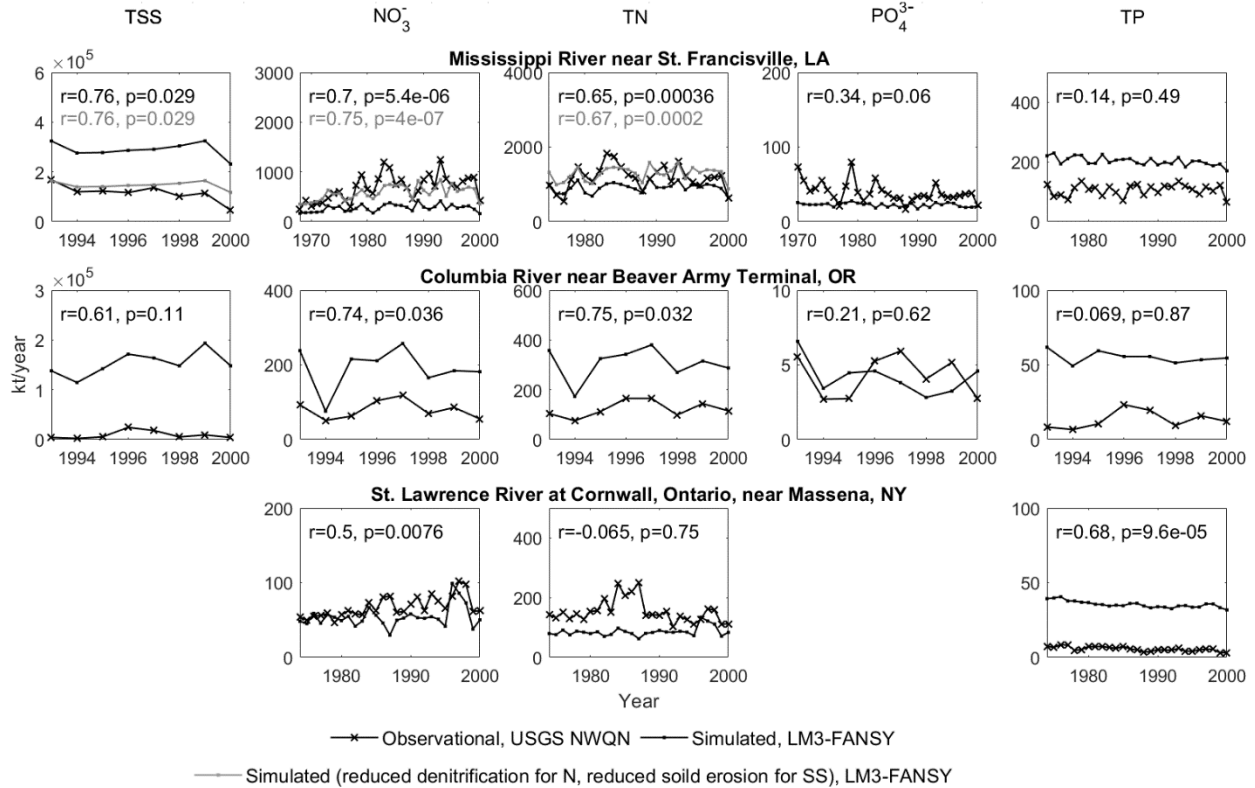
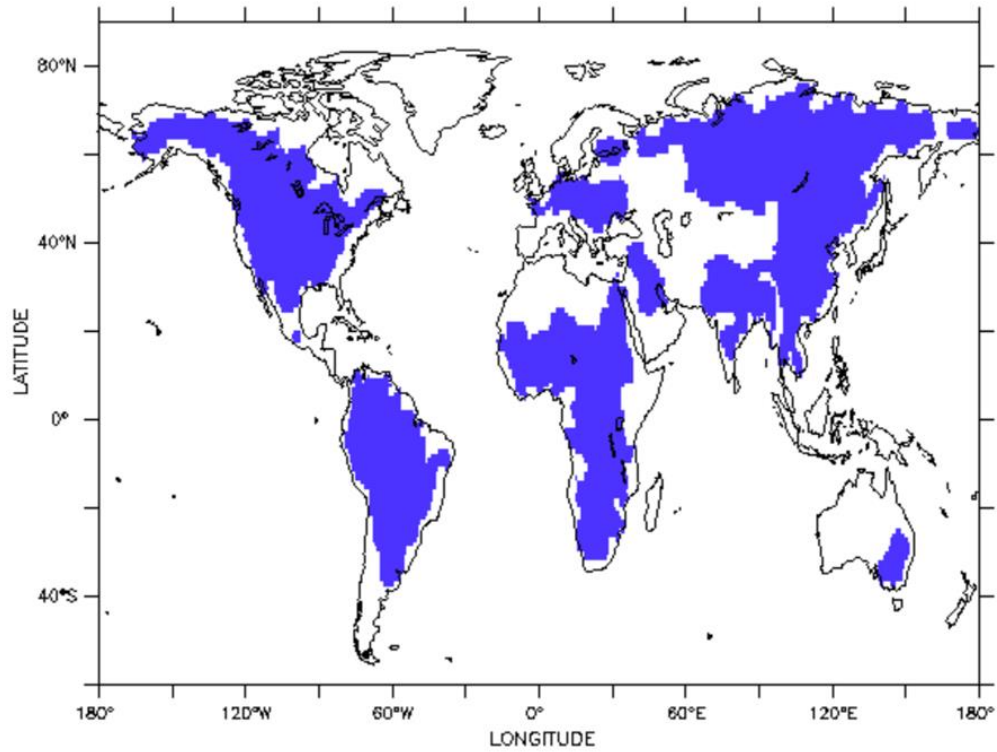


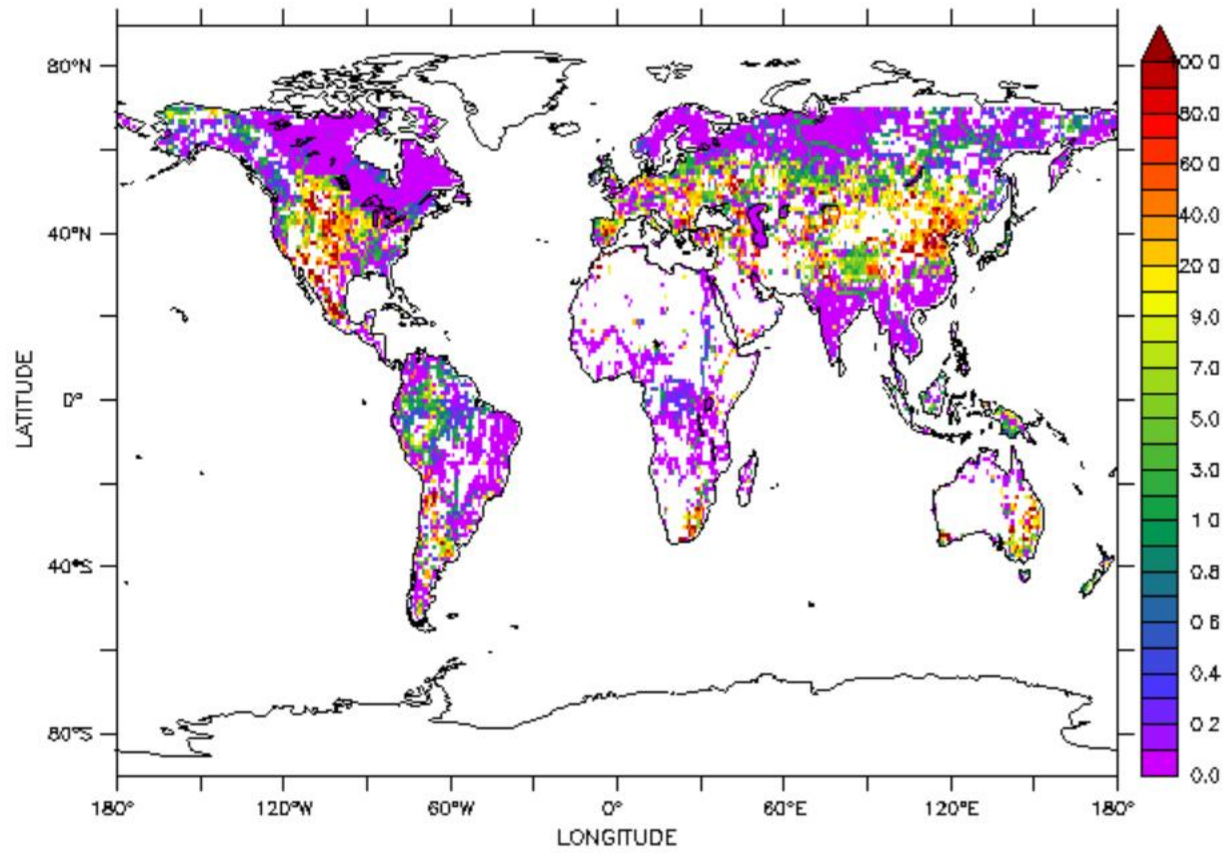
Figure 1: FANCY structure with arrows depicting fluxes of constituents of algae, nutrients, and solids in rivers and lakes. The constituents are listed in Table 1.



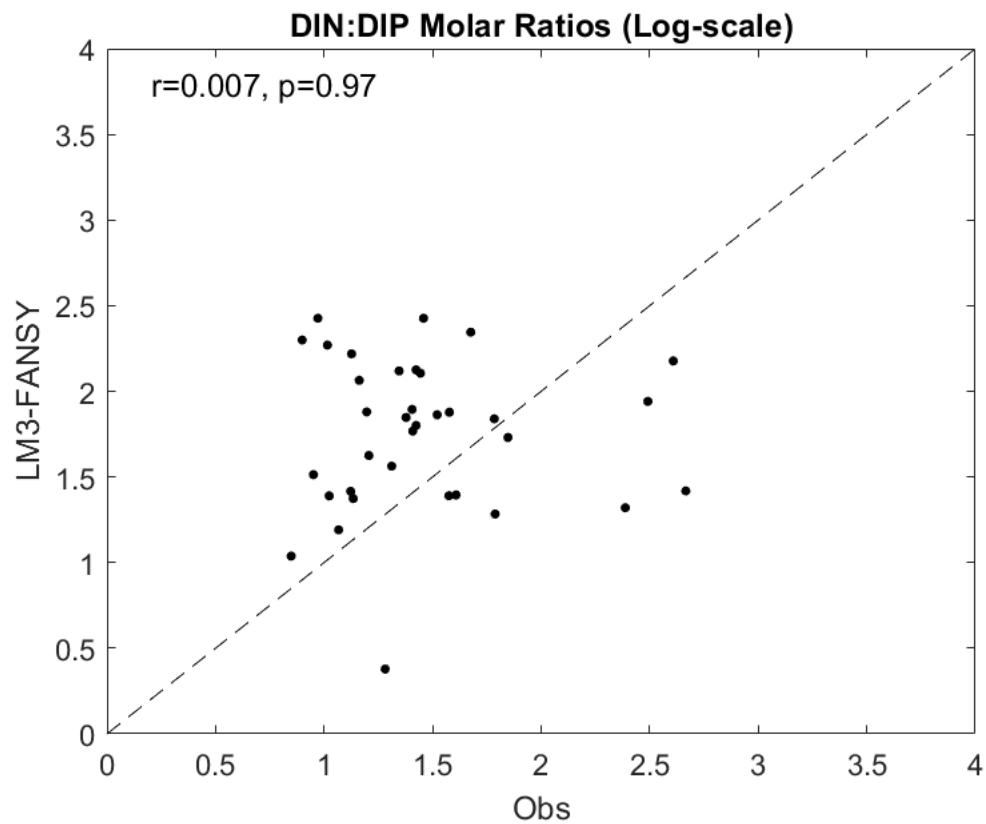
**Figure 8: Pearson correlation coefficients ( $r$ ) and  $p$  values ( $p$ ) between the measurement-based vs. simulated annual loads from large U. S. rivers for the period ~1968-2000. The grey color represents load responses to the changes in the freshwater denitrification rate coefficient from  $0.15 \text{ day}^{-1}$  to  $0.05 \text{ day}^{-1}$  for N dynamics and the free parameter of terrestrial soil erosion by half for solid dynamics (Table 7).**



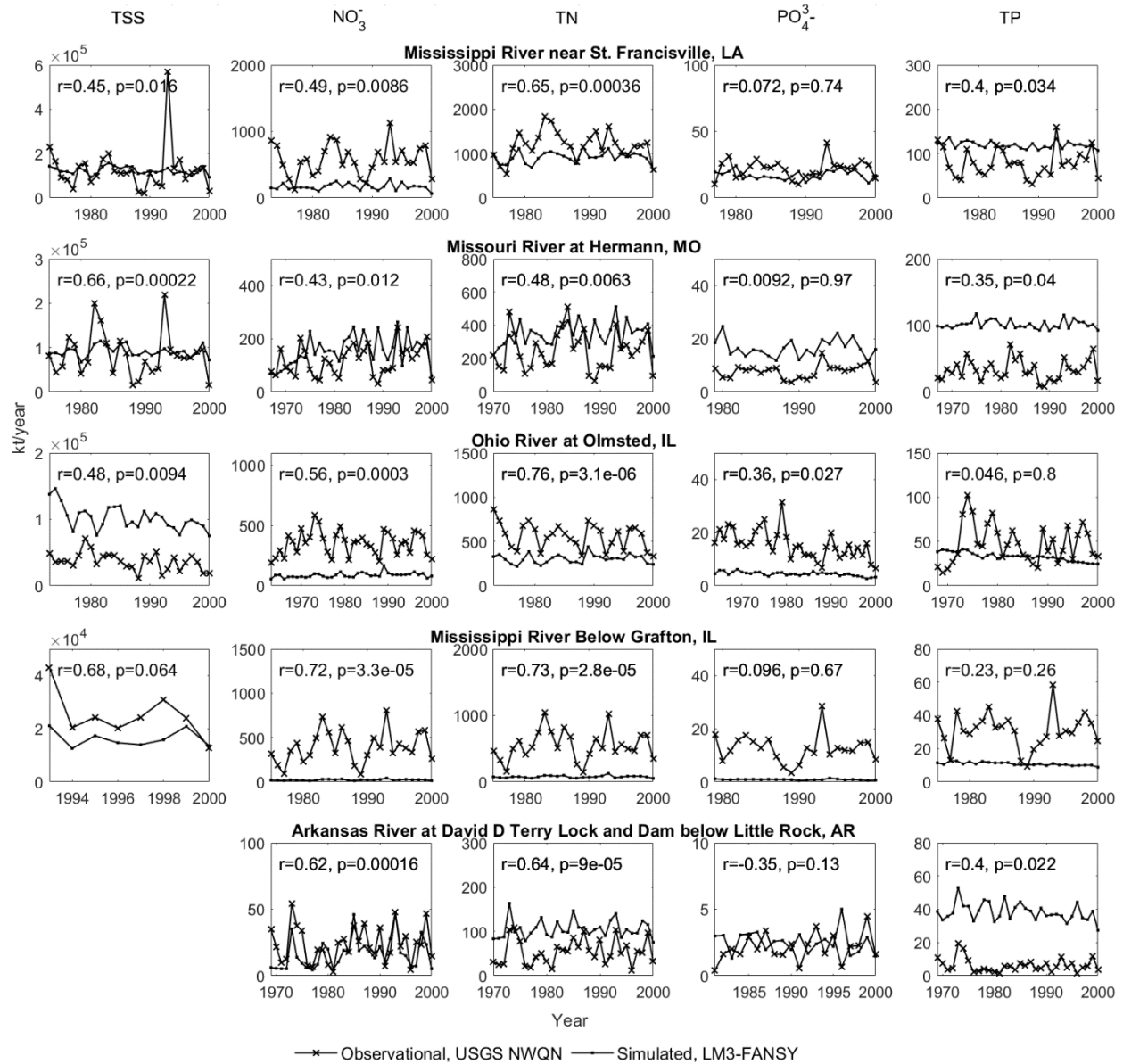
**Figure SII: The 69 rivers considered for analyses are distributed broadly covering various climate and land use. The basin area of the 69 rivers covers 55% of global land area (excluding the Antarctic).**



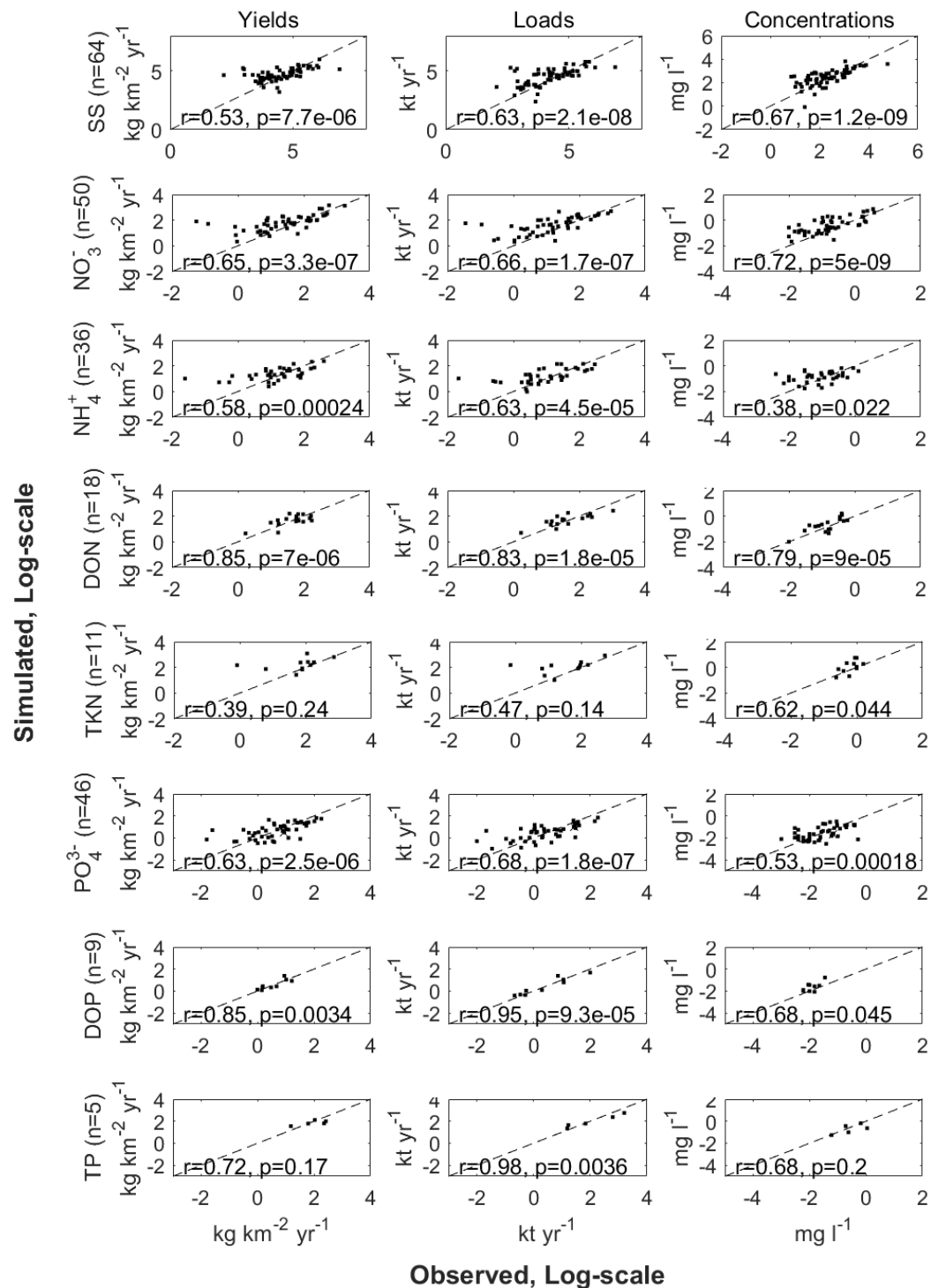
**Figure SI3: Chlorophyll a concentrations ( $\text{mg m}^{-3}$ ) in lakes for the year 1990.**



**Figure SI4: Pearson correlation coefficient (r) and p value (p) between the measurement-based vs. simulated DIN:DIP molar ratios across 35 rivers for the year 1990.**



**Figure SI5: Pearson correlation coefficients (r) and p values (p) between the measurement-based vs. simulated annual loads across 5 stations in the Mississippi River Basin for the periods ~1963-2000.**



**Figure SI6: Pearson correlation coefficients (r) and p values (p) between the measurement-based vs. simulated yields, loads, and concentrations across the world major rivers for the year 1990. Here the actual (instead of natural) water discharges of Meybeck and Ragu (2012) were used when calculating loads and yields from their multi-year average concentrations.**



Input source	N			P			Solid	
	NH <sub>4</sub> <sup>+</sup>	NO <sub>3</sub> <sup>-</sup>	ON	PO <sub>4</sub>	PIP	OP	POM	ISS
Subsurface runoff								
Litter from floodplains	LM3-FANCY	LM3-FANCY	LM3-FANCY	0	0	100	LM3-FANCY	LM3-FANCY
Natural surface runoff				0	25	75		
Agricultural surface runoff				0.34	0.33	0.33		
Weathering				100	0	0		
Atmospheric deposition	35	35	30					
Aquaculture	0.26	0.62	0.12	0.69	0.06	0.25		
Wastewater	0.90	0.00	0.10	0.80	0.10	0.10		
Agricultural surface runoff uncertainty tests (see Sect. 3 Model forcing and simulations for details), assuming that the agricultural surface runoff from Beusen et al (2015) contains the followings:								
Surficial runoff 100%				60	0	40		
Soil loss 100%				0	75	25		
Surficial runoff 40% soil loss 60%				24	45	31		
Surficial runoff 60% soil loss 40%				36	30	34		
Surficial runoff 20% soil loss 80%				12	60	28		
Surficial runoff 80% soil loss 20%				48	15	37		
Wastewater uncertainty tests (see Sect. 3 Model forcing and simulations for details), assuming that the wastewater from Beusen et al (2015) contains the followings:								
Untreated	65	0	35	15	30	55		
Secondary/tertiary treated 100%	60	30	10	80	10	10		
Untreated 80% Primary treated 10% Secondary/tertiary treated 10%	67	3	30	28	26	46		
Untreated 40% Primary treated 40% Secondary/tertiary treated 20%	74	6	20	54	18	28		

**Table 3: Description of solid and nutrient inputs from terrestrial systems and the atmosphere to rivers, which were either simulated by LM3-FANCY or provided by Beusen et al (2015). The numbers are the fractions of dividing TN and TP inputs from Beusen et al (2015) into different N and P species.**

		SS	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>	DON	TKN	PO <sub>4</sub> <sup>3-</sup>	DOP	TP
r		.76 (.71, .76)	.79 (.75, .81)	.66 (.65, .74)	.85 (.85, .93)	.70 (.65, .74)	.70 (.66, .74)	.93 (.93, .95)	.98 (.95, .98)
	NSE	.55 (.48, .55)	.42 (.33, .49)	.37 (.34, .49)	.66 (.62, .77)	.00 (.00, .22)	.48 (.42, .53)	.85 (.85, .88)	.82 (.81, .86)
Prediction error	Min	-97 (-98, -97)	-86 (-92, -79)	-76 (-90, -76)	-75 (-76, -68)	-56 (-60, -24)	-98 (-99, -86)	-51 (-55, -47)	-64 (-66, -61)
	25 <sup>th</sup>	-35 (-53, -35)	1 (-12, 8)	-48 (-49, -34)	-43 (-48, -7)	31 (-8, 31)	-59 (-62, -51)	-39 (-50, -38)	-62 (-65, -60)
	Med	39 (18, 50)	146 (78, 203)	18 (-1, 45)	35 (2, 75)	60 (33, 68)	-22 (-29, -5)	-3 (-17, 18)	-10 (-16, 0)
	75 <sup>th</sup>	260 (227, 339)	475 (396, 693)	540 (300, 541)	243 (140, 293)	287 (259, 396)	171 (137, 234)	155 (114, 155)	78 (26, 78)
	Max	3219 (2324, 4238)	1796 (1346, 3256)	2946 (1099, 4158)	610 (358, 733)	1040 (911, 1374)	1956 (1956, 2955)	365 (274, 449)	182 (103, 208)
	IQR	296 (273, 384)	474 (405, 705)	589 (349, 589)	286 (176, 317)	256 (248, 383)	230 (198, 293)	194 (156, 194)	141 (88, 141)

**Table 4: Pearson correlation coefficients (r) and Nash–Sutcliffe model efficiency coefficient (NSE) between the measurement-based vs. simulated loads across world major rivers for the year 1990 (range for the years 1990-2000 in parenthesis). The prediction error is computed as the difference between the simulated and measurement-based estimates of loads expressed as a percentage of the measurement-based load.**

	SS	TN	DIN	DON	PON	TP	DIP	DOP	PP
LM3-FANSY	10 (10, 11)	35 (34, 38)	14 (14, 16)	13 (13, 14)	7 (7, 8)	7 (7, 8)	2	1	5
Global NEWS 1 (year 1995) Global NEWS 2 (year 2000)	19 (11, 27) Beusen et al (2005)	44.9 NEWS 2, Mayorga et al (2010)	18.9 NEWS 2, Mayorga et al (2010)	10 NEWS- DON, Harrison et al. (2005)	13.5 NEWS 2, Mayorga et al (2010)	9.04 NEWS 2, Mayorga et al (2010)	1.45 NEWS- DIP-HD, Harrison et al (2010)	0.6 NEWS- DOP, Harrison et al (2005)	6.56 NEWS 2, Mayorga et al (2010)
QRT; (years 1960-1995)	12.6 Syvitski et al (2005)								
IMAGE-GNM (year 2000)		36.5 Beusen et al (2016)				4 Beusen et al (2016)			
Boyer et al. (2006) (mid-1990s)		48							
Galloway et al. (2004) (early-1990s)		47.8							
Green et al., 2004 (mid-1990s)		40	14.5						
Smith et al., 2003 (1990s)			18.9				2.3		

**Table 5: LM3-FANSY and published estimates of global river loads to the coastal ocean in  $\text{Pg yr}^{-1}$  for SS and  $\text{Tg yr}^{-1}$  for nutrients. LM3-FANSY results are for the year 1990 (ranges for the years 1990-2000).**

	Data period, to the year 2000 from the year shown in the 1 <sup>st</sup> line r, p value shown in the 2 <sup>nd</sup> line Prediction error shown in the 3 <sup>rd</sup> line				
8 NWQN stations	NO <sub>3</sub> <sup>-</sup>	TN	PO <sub>4</sub> <sup>3-</sup>	TP	SS
Mississippi River near St. Francisville, LA	1968 .70, <.01 -60	1975 .65, <.01 -23	1970 .34, .06 -43	1974 .14, .49 91	1993 .76, .03 150
Mississippi River at Thebes, IL	1973 .49, <.01 -71	1973 .60, <.01 -53	1977 .07, .74 -21	1973 .40, .03 51	1973 .45, .02 -4
Missouri River at Hermann, MO	1967 .43, .01 34	1970 .48, <.01 41	1979 .01, .97 116	1967 .35, .04 208	1975 .66, <.01 8
Ohio River at Olmsted, IL	1963 .56, <.01 -75	1973 .76, <.01 -46	1964 .36, .03 -71	1968 .05, .80 -31	1973 .48, .01 178
Mississippi River Below Grafton, IL	1975 .72, <.01 -95	1975 .73, <.01 -86	1979 .10, .67 -93	1975 .23, .26 -65	1993 .68, .06 -35
Arkansas River at David D Terry Lock and Dam below Little Rock, AR	1969 .62, <.01 -29	1970 .64, <.01 92	1981 -0.35, .13 20	1969 .4, .02 515	No Data
Columbia River near Beaver Army Terminal, OR	1993 .74, .04 140	1993 .75, .03 150	1993 .21, .62 -2	1993 .07, .87 318	1993 .61, .11 1588
St. Lawrence River at Cornwall, Ontario, near Massena, NY	1974 .50, <.01 -20	1974 -.07, .75 -44	No data	1974 .68, <.01 530	No data

**Table 6: Pearson correlation coefficients (r) and p values (p) between the measurement-based vs. simulated annual loads across 8 stations in large U. S. rivers for the periods ~1963-2000. The prediction error is computed as the difference between the simulated and measurement-based load averages over the period expressed as a percentage of the measurement-based load average over the period.**

		SS	TN	DIN	DON	PON	TP	DIP	PIP	DOP	POP
Runoff and erosion	Removal	-1	-90	-84	-99	-84	-74	-57	-61	-91	-94
	+15%	0	13	12	15	13	11	11	7	13	14
Wastewater	Removal	0	-10	-14	-1	-19	-10	-16	-7	-14	-7
	+15%	0	2	2	0	3	2	3	1	2	1
Aquaculture	Removal	0	-1	-1	0	-1	0	-1	0	-1	0
	+15%	0	0	0	0	0	0	0	0	0	0
Atmospheric deposition	Removal	0	-1	-1	0	-2	0	0	0	0	0
	+15%	0	0	0	0	0	0	0	0	0	0
Weathering	Removal	0	-2	1	-1	-11	-15	-32	-13	-11	-4
	+15%	0	0	0	0	2	2	5	2	2	1
$\Gamma_{DN,Ero}$	$\times 0.1$ (1.39)	0	0	0	0	0	0	0	0	0	0
	$\times 10$ (139)	-4	0	0	0	0	0	1	-2	0	0
$\Gamma_{DN,Ero}, \Gamma_{DN}$	$\times 0.1$ (1.39)	0	0	0	0	2	0	-2	-1	2	1
	$\times 10$ (139)	-4	-1	0	0	-5	0	6	-1	-5	-2
d	Half (0.005, small silt)	0	0	0	0	0	0	0	0	0	0
	Twice (0.02, large silt)	0	0	0	0	0	0	0	0	0	0
	(0.084, large phytoplankton)	-1	0	0	0	-2	0	0	0	0	-1
$\theta_{Al}, \theta, \theta_{Sed}$	(1.066, 1.024, 1.024)	0	2	4	1	-2	0	0	1	-1	0
	(1.066, 1.047, 1.047)	0	1	2	0	0	0	0	0	0	0
	(1.066, 1.066, 1.066)	0	0	1	0	0	0	0	0	0	0
	(1.066, 1.047, 1.08)	0	1	2	0	0	0	0	0	0	0
$P_{max}^C$	$\times 0.5$	0	-10	4	-2	-48	0	28	15	-43	-19
	$\times 2$	0	9	-6	3	45	0	-23	-19	43	18
$\alpha^{CHL}$	$\times 0.5$	0	-9	4	-2	-42	0	24	13	-38	-17
	$\times 2$	1	11	-7	4	53	0	-27	-23	52	21
$k_{NO_{23}}$	$\times 0.5$	0	1	0	0	4	0	-2	-1	3	1
	$\times 2$	0	-1	0	0	-4	0	2	1	-4	-1
$k_{NH_4}$	$\times 0.5$	0	1	0	0	3	0	-2	-1	2	1
	$\times 2$	0	-1	0	0	-3	0	2	1	-3	-1
$k_{PO_4}$	$\times 0.5$	0	0	0	0	1	0	-1	-1	1	1
	$\times 2$	0	0	0	0	-2	0	1	1	-2	-1
$k_{NO_{23}}, k_{NH_4}, k_{PO_4}$	$\times 0.1$ (~lower bound)	0	3	-2	1	16	0	-7	-6	13	6
	$\times 4$ (~upper bound)	0	-4	2	-1	-22	0	15	7	-23	-9
$k_{ew}$	$\times 0.5$	0	0	0	0	0	0	0	0	0	0
	$\times 2$	0	0	0	0	0	0	0	0	0	0
$k_e$	(0.15)	1	13	-10	5	66	0	-34	-29	66	26
	(0.45)	1	12	-8	5	61	0	-31	-27	61	24
$k_{SedN,d}$	$\times 0.5$	0	0	0	0	0	0	0	0	-1	0
	$\times 2$	0	0	0	0	0	0	0	0	0	0
$k_{SedP,d}$	$\times 0.5$	0	0	0	0	0	0	0	0	0	0
	$\times 2$	0	0	0	0	0	0	0	0	0	0
$k_{PON,d}$	$\times 0.5$	0	1	0	0	3	0	0	1	-1	-1
	$\times 2$	0	-1	0	0	-6	0	-1	-1	1	1
$k_{POP,d}$	$\times 0.5$	0	0	0	0	-1	0	-1	-1	-5	3
	$\times 2$	0	0	0	0	2	0	2	1	8	-5
$k_{DON,d}$	$\times 0.5$	0	7	-8	29	0	0	1	0	-1	0
	$\times 2$	0	-5	9	-22	0	0	0	0	0	0
$k_{DOP,d}$	$\times 0.5$	0	-1	0	0	-4	0	-5	-2	22	-2
	$\times 2$	0	1	0	0	5	0	4	2	-22	2
$k_{nitr}$	$\times 0.5$	0	5	9	0	4	0	-2	-1	4	2
	$\times 2$	0	-3	-5	0	-4	0	2	1	-4	-1
$k_{denitr}$	(0.05)	0	25	54	1	10	0	-5	-4	9	4
	(0.075)	0	15	31	0	7	0	-4	-3	6	3

	(0.1)	0	8	17	0	4	0	-2	-1	4	2
	(0.2)	0	-5	-10	0	-3	0	2	1	-3	-1
	(0.25)	0	-8	-17	0	-4	0	3	2	-5	-2
	(0.3)	0	-11	-23	0	-6	0	4	2	-6	-2
C <sub>1</sub>	Half	-50	-3	-2	1	-14	0	22	-26	7	3
	Twice	99	8	2	-1	38	0	-22	23	-5	-2
k <sub>m</sub>	Half	1	15	-11	4	61	0	-35	-31	60	24
	Twice	0	-10	4	-3	-48	0	28	15	-44	-19
f <sub>m,DON</sub> , f <sub>m,PON</sub>	(0.15, 0.3)	0	1	0	-1	7	0	1	1	-16	3
	(0.6, 0.3)	0	-2	0	1	-10	0	-2	-1	23	-4
	(0.3, 0.15)	0	-1	1	1	-12	0	-1	-1	17	-5
	(0.3, 0.6)	0	1	-1	-1	9	0	1	1	-14	3
f <sub>PON,DON</sub>	(0.4)	0	0	0	0	0	0	0	0	0	0
	(1.0)	0	0	0	0	0	0	0	0	0	0
f <sub>POP,DOP</sub>	(0.4)	0	0	0	0	1	0	1	0	-4	0
	(1.0)	0	0	0	0	0	0	0	0	2	0
f <sub>SedN,DON</sub>	(0.4)	0	0	0	0	0	0	0	0	0	0
	(1.0)	0	0	0	0	0	0	0	0	0	0
f <sub>SedP,DOP</sub>	(0.4)	0	0	0	0	0	0	1	0	-2	0
	(1.0)	0	0	0	0	0	0	0	0	1	0
f <sub>PON,DON</sub> , f <sub>POP,DOP</sub> , f <sub>SedN,DON</sub> , f <sub>SedP,DOP</sub>	(0.4)	0	0	0	0	1	0	1	0	-6	0
	(1.0)	0	0	0	0	0	0	-1	0	3	0
f <sub>SedP,POP</sub>	Half	0	0	0	0	0	0	0	0	0	0
	Twice	0	0	0	0	0	0	0	0	0	0

**Table 7: Model sensitivities to the inputs, components, and parameters examined based on the percentage (%) difference in global river loads between the sensitivity and baseline simulations for the year 1990. The changed parameter values other than a decrease by half or an increase by twice are given in parenthesis.**

	NSE for the year 1990 (range for the years 1990-2000) in the 1 <sup>st</sup> and 2 <sup>nd</sup> lines r for the year 1990 (range for the years 1990-2000) in the 3 <sup>rd</sup> and 4 <sup>th</sup> lines							
Treatment	SS	NO <sub>3</sub> <sup>-</sup>	NH <sub>4</sub> <sup>+</sup>	DON	TKN	PO <sub>4</sub> <sup>3-</sup>	DOP	TP
Baseline	.55 (.48,.55)	.42 (.33,.49)	.37 (.34,.49)	.66 (.62,.77)	.00 (.00,.22)	.48 (.42,.53)	.85 (.85,.88)	.82 (.81,.86)
	.76 (.71,.76)	.79 (.75,.81)	.66 (.65,.74)	.85 (.85,.93)	.70 (.65,.74)	.70 (.66,.74)	.93 (.93,.95)	.98 (.95,.98)
No nutrient runoff and erosion fluxes	.54 (.47,.55)	-1.91 (-2.01, -1.90)	-3.54 (-3.78, -3.54)	-15.41 (-16.30, -15.41)	-5.40 (-5.55, -5.04)	-1.38 (-1.68, -1.00)	-.52.35 (-.52.35, -25.79)	0.00 (-.10,0.02)
	.76 (.71,.76)	.62 (.61,.64)	.43 (.40,.44)	.06 (-.01,.07)	.24 (.18,.24)	.55 (.47,.62)	.86 (.70,.89)	.91 (.91,.94)
No wastewater	.55 (.48,.56)	.44 (.34,.50)	.32 (.32,.46)	.69 (.66,.81)	.21 (.21,.42)	.43 (.38,.49)	.86 (.86,.89)	.82 (.79,.85)
	.76 (.72,.76)	.79 (.75,.81)	.60 (.60,.69)	.85 (.85,.93)	.72 (.71,.78)	.67 (.63,.71)	.94 (.94,.95)	.96 (.93,.97)
No aquaculture	.55 (.48,.55)	.43 (.33,.49)	.37 (.35,.50)	.66 (.62,.77)	.00 (.00,.22)	.48 (.42,.53)	.85 (.85,.88)	.82 (.81,.86)
	.76 (.71,.76)	.79 (.75,.81)	.66 (.65,.74)	.85 (.85,.93)	.70 (.65,.74)	.70 (.66,.74)	.93 (.93,.95)	.98 (.95,.98)
No atmospheric deposition	.55 (.48,.55)	.43 (.33,.49)	.37 (.34,.49)	.66 (.62,.77)	.00 (.00,.21)	.48 (.42,.53)	.85 (.85,.88)	.82 (.81,.86)
	.76 (.71,.76)	.79 (.75,.81)	.66 (.65,.73)	.85 (.85,.93)	.68 (.63,.72)	.70 (.66,.74)	.93 (.93,.95)	.98 (.95,.98)
No weathering	.54 (.47,.55)	.42 (.33,.48)	.37 (.34,.50)	.66 (.62,.77)	.03 (.03,.24)	.28 (.28,.43)	.69 (.68,.73)	.73 (.73,.78)
	.76 (.71,.76)	.79 (.76,.81)	.67 (.65,.74)	.84 (.84,.93)	.70 (.65,.74)	.66 (.64,.72)	.85 (.85,.88)	.94 (.93,.97)
No dynamic light shading component of algae dynamics Ke=0.15	.54 (.48,.55)	.44 (.32,.51)	.39 (.36,.51)	.61 (.57,.71)	-.15 (-.19,.04)	.20 (.19,.26)	.85 (.81,.85)	.82 (.81,.86)
	.76 (.71,.76)	.77 (.71,.80)	.66 (.64,.73)	.84 (.84,.91)	.71 (.63,.74)	.58 (.56,.61)	.93 (.92,.94)	.98 (.95,.98)
Ke=0.45	.54 (.48,.55)	.44 (.32,.50)	.39 (.37,.50)	.61 (.58,.72)	-.14 (-.18,.05)	.27 (.26,.33)	.82 (.82,.85)	.82 (.81,.86)
	.76 (.71,.76)	.77 (.71,.80)	.67 (.65,.73)	.84 (.84,.92)	.71 (.64,.74)	.61 (.59,.63)	.92 (.92,.94)	.98 (.95,.98)

**Table 8: Model sensitivities to the inputs and components examined based on the Pearson correlation coefficients (r) and Nash–Sutcliffe model efficiency coefficient (NSE) between the measurement-based vs. simulated loads across world major rivers for the year 1990 (ranges for the years 1990-2000 in parenthesis).**

	SS	TN	DIN	DON	PON	TP	DIP	PIP	DOP	POP
Agricultural surface runoff (surficial runoff)	0	1	0	0	3	0	-4	-8	9	8
Agricultural surface runoff (soil loss)	0	-1	0	0	-7	0	7	11	-13	-10
Agricultural surface runoff (40% surficial runoff, 60% soil loss)	0	0	0	0	-2	0	1	3	-3	-3
Agricultural surface runoff (60% surficial runoff, 40% soil loss)	0	0	0	0	0	0	-1	-1	1	1
Agricultural surface runoff (20% surficial runoff, 80% soil loss)	0	-1	0	0	-4	0	4	6	-8	-6
Agricultural surface runoff (80% surficial runoff, 20% soil loss)	0	0	0	0	2	0	-3	-5	5	5
Wastewater (untreated)	0	0	-3	1	7	0	-9	-1	6	6
Wastewater (secondary/tertiary treated)	0	0	-1	0	0	0	0	0	0	0
Wastewater (80% untreated, 10% primary treated, 10% secondary/tertiary treated)	0	0	-3	1	5	0	-7	-1	4	5
Wastewater (40% untreated, 40% primary treated, 20% secondary/tertiary treated)	0	0	-1	0	3	0	-3	-1	2	2

**Table SI10: Model sensitivities to the fractions of dividing TN and TP inputs from Beusen et al (2015) into different N and P species (Table 3), examined based on the percentage (%) difference in global river loads between the sensitivity and baseline simulations for the year 1990.**