nature portfolio

Peer Review File

Substantial increase of organic carbon storage in Chinese lakes

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript, titled “Substantial increase of organic carbon storage in Chinese lakes”, investigated spatiotemporal changes in OC concentrations and OC storge in Chinese lakes using extensive field data and satellite monitoring techniques. This is an important paper that provides a new perspective for estimating organic carbon storage in lakes. It is an interesting topic to the researchers in the related areas but the paper needs improvement before acceptance for publication. My detailed suggestions and questions are as follows:
1. Main Text, line 31, “lakes receive 2.9–5.1 Tg C/yr from terrestrial ecosystems” change “lakes” to “inland waters”;

line31-32, “bury 0.06–0.25 Mt C/yr into the sediment”, change “lakes” to “lakes and reservoirs” or “inland waters”;

line32-33,“emit 0.06 – 0.84 Tg C/yr of CO2 to the atmosphere”,
lakes and reservoirs emit 0.06 – 0.84 Tg C/yr of CO2 to the atmosphere

2. Main Text, line 47-48, the sentence “The total water volume of many lakes (79.4%) has also increased since 2003” expressed not clearly, what does 79.4% refer to ?

3. Main Text, line 82-83, “suggested by a positive correlation between the logarithmic in-situ DOC and chlorophyll-a (Chl-a) concentration (r = 0.23, N = 3,624)”, what is statistical p value ?

4. Main Text, line 103-104, “concentration, indicated by the linear relationship between the logarithmic in-situ TSM and POC concentration (N = 2,026, R2 = 0.14)”, what is statistical p value ?

5. Main Text, references, line 359, change “Mendonca” to “Mendonça”.

Reviewer #2 (Remarks to the Author):

Lakes play an essential role in regulating the global carbon cycle and climate change. Although many studies have provided insights into the carbon cycle of lakes, the spatiotemporal variations of lake OC storage on a large regional scale have not been reported. Previous studies have commonly assumed a fixed value for OC storage in lakes. As a matter of fact, OC storage in lakes showed obvious spatiotemporal variations along with water eutrophication, water level rise, and human activities. It is of great significance to obtain OC storage spatiotemporally in lakes. On the one hand, OC storage reflects how much carbon have been stored in lakes. On the other hand, changes in OC storage also have important influences on lake CO2 emissions and carbon burial in sediments, two essential aspects of the global carbon cycle. Therefore, the dynamic acquisition of lake OC storage, as studied in this paper, plays an irreplaceable role in accurately quantifying the

role of lakes in the global carbon cycle.

Based on measurements at 4,201 stations from 348 lakes across China, this study quantified the spatiotemporal variability of DOC and POC storage for 24,366 Chinese lakes during 1984-2023 using advanced satellite monitoring techniques. The authors report that DOC and POC storage in Chinese lakes increased by 44.6% and 33.5%, respectively, which hasn't been noticed before and is of great importance for understanding the carbon cycle in lakes. Using remote sensing techniques, the authors innovatively conducted a remote sensing study of OC storage in a large region of Chinese lakes by integrating basin characteristics and satellite reflectance, and revealed the spatiotemporal variation characteristics of OC storage in all lakes in China for the first time. I believe the study has an important promoting role and reference value for the dynamic monitoring of global lake carbon cycle by remote sensing and deep learning techniques, and can compensate for the insufficient spatiotemporal coverage of global lake monitoring OC data.

Overall, the extensive field data are very valuable, the methods used are innovative and reasonable, and the findings are very important. The satellite-derived datasets of DOC and POC concentrations are also valuable for understanding carbon emission and carbon burial in Chinese lakes. In addition, I think the manuscript is well organized and written. Some revisions and clarifications are required before publication in the NC journal. I have the following suggestions for the authors to improve the manuscript.

1. "OC storage is expected to vary … water volume changes". According to the above definition, OC concentration and water volume variations will change the OC storage in the lake. In addition, the OC concentration in the lake also has obvious temporal variation characteristics. The authors here say that OC storage varies with water volume, which is incomplete.
2. "Across the studied lakes …". How many lakes did the authors study in total? Do these lakes include all lakes in China? If not, how did the authors identify the lakes they studied? Although the authors may have included relevant information in the online methods, it is still necessary to explain it here so that readers can understand the relevant content.
3. "the southeast NPML and EPL zones" and "the northwest TPL zone". In my opinion, the TPL zone is located in southwest China, while the NPML zone is located in northeast China. How do the authors divide the different lake zones? Why do the authors divide them this way? What is the reference standard? Please add any relevant information.
4. "…demonstrated clear increasing tends…". I don't quite understand what the authors mean by this. Do the authors mean increasing trends in different lakes? Please clarify.
5. "80.6% of lake area showed increasing POC concentration". How is the lake area ratio calculated here? Did the authors calculate the changes in POC concentrations of all pixels in Chinese lakes? From the whole paper, I think the authors only counted the average POC concentrations of lakes in China, but did not perform statistics at the remote sensing pixel level.
6. "Lake DOC was primarily concentrated in the TPL zone… increased substantially". The statement is not clear. Do the authors mean that lake DOC is mainly stored in the TPL zone? Is the increase in lake DOC mainly in the TPL zone? Or do Chinese lakes show a significant increase in DOC storage?
7. Does the lake proportion in Figure 2 represent the number of lakes or the area of lakes? How are "changes of DOC storage" obtained? Does the "changes of DOC storage" mean the linear fitted slope between DOC storage and year during 1986-2023? Also, I think the figure captions need to be labelled separately.
8. "… were identified as the main drivers …". How did the authors identify the main drivers of

changes in DOC and POC concentrations in different lakes? Are the primary drivers those that contribute more than 50%? Or does it refer to the factor that has the largest contribution among the factors studied? The authors need to explain both in the main text and in the supplementary information, as the reader will often not necessarily read the supplementary information.
9. “glacial meltwater and rainwater … DOC concentrations … 1.0 mg/L”. The DOC concentration in rainwater also has obvious spatiotemporal variation characteristics. Does 1.0 mg/L refer to the average DOC concentration in lakes in the IMXL and TPL zones? Is the DOC concentration in glacial meltwater also 1.0 mg/L? The study area of Gao et al. (2020) is Central Asia. Is the DOC concentration in the glacial meltwater of Central Asia the same as in the IMXL and TPL zones? The author needs to provide more information or a clear introduction.
10. “… through pore water or photoproduction”. Pore water is usually derived from sediments. Does photoproduction also refer to the oxidative decomposition of OC in sediments to produce DOC? Eutrophic lakes are usually very turbid and have very low turbidity, so light has little or no access to the bottom of the turbid water. If so, how can light oxidize OC in sediments and break it down into DOC? Please explain further.
11. “OCAR was exponentially related to lake-based mean POC concentration”. How was the POC concentration from the satellite remote sensing inversion matched to the OCAR data? Both POC concentration and OCAR vary with time, and different time-matching schemes may give different results. In addition, did the authors get annual OCAR data, like POC concentration?
12. “…Europe exhibited OCAR values…”. As the previous comment, the authors need to give a specific time for the OCAR values.
13. “… measure the lake carbon cycle… using satellite data”. The "lake carbon cycle" is a very vague definition that includes various aspects of carbon composition, decomposition, and transformation. What exactly do the authors mean by "lake carbon cycle"? As far as I know, although satellite data can be used to estimate CO2 emissions and carbon storage studied in this paper, there is no report on the use of satellite data to estimate carbon decomposition and transformation.
14. “monthly averaged basin property … and wind speed”. Where do these data come from? What is the temporal coverage of these data? This sentence is incomplete.
15. “… applied to retrieve annual mean DOC concentrations …” and “… was applied to retrieve annual mean POC concentrations …”. The models constructed in this paper should first be applied to Landsat data to calculate the daily DOC and POC concentrations, and then calculate the average values according to the daily remote sensing results. The authors need to clearly explain the relevant processes. Also, how did the authors remove clouds and other non-water pixels? Did the authors use the methods constructed in this study, or did they follow the published methods?
16. For a given factor, the nonlinear Random Forest analysis indicated its relative contribution using the Gini importance. The authors must explain why the Gini Importance index was chosen. Compared to the Permutation Importance and Boruta indexes, can the Gini Importance index better reflect the importance of the results of the Random Forest analysis?

Authors' responses to reviewers' comments

Substantial increase of organic carbon storage in Chinese lakes

Dong Liu, Kun Shi, Peng Chen, Nuoxiao Yan, Lishan Ran, Tiit Kutser, Andrew N. Tyler, Evangelos Spyrakos, R. Iestyn Woolway, Yunlin Zhang, Hongtao Duan

Submission ID: NCOMMS-24-39360

Reviewer #1 (Remarks to the Author):

This manuscript, titled "Substantial increase of organic carbon storage in Chinese lakes", investigated spatiotemporal changes in OC concentrations and OC storge in Chinese lakes using extensive field data and satellite monitoring techniques. This is an important paper that provides a new perspective for estimating organic carbon storage in lakes. It is an interesting topic to the researchers in the related areas but the paper needs improvement before acceptance for publication. My detailed suggestions and questions are as follows:

Response: We thank the reviewer very much for these positive comments. In this revision, we have addressed all the comments raised by the reviewer. To facilitate your assessment of the changes, we have highlighted our changes in red throughout the revised manuscript. According to the reviewers' comments, we have made the following major modifications:

(1) We have added statistical *p* values for the model between the logarithmic DOC and Chl-a and the model between the logarithmic TSM and POC. Please refer to the following Responses #1.3 and #1.4.

(2) We have corrected the inappropriate statements about the lake carbon cycle and the incorrect spelling of the author's name in the references. Please refer to the following Responses #1.1 and #15.

(3) We have rephrased the statement about the figure showing the percent of global lakes with increasing water volumes since 2003. Please refer to the following Response #1.2. With your help, we believe the manuscript has greatly improved, especially for model assessment and language expression. We hope that these changes have addressed your main concerns. Your comments are invaluable in helping us improve our work.

1. Main Text, line 31, “lakes receive 2.9–5.1 Tg C/yr from terrestrial ecosystems” change “lakes” to “inland waters”;

line31-32, “bury 0.06–0.25 Mt C/yr into the sediment”, change “lakes” to “lakes and reservoirs” or “inland waters”;

line32-33,“emit 0.06 – 0.84 Tg C/yr of CO2 to the atmosphere”， lakes and reservoirs emit 0.06 – 0.84 Tg C/yr of CO2 to the atmosphere

Response #1.1: We apologize for the inappropriate statements in the writing of this manuscript and we thank the reviewer for pointing them out. According to the reviewer's corrections, we have changed “lakes” to “inland waters” for the statement about carbon from terrestrial ecosystems, changed “lakes” to “lakes and reservoirs” for the statement about carbon burial into the sediment, and changed “lakes” to “lakes and reservoirs” for the statement about carbon emission to the atmosphere. Please refer to lines 30-33 in the revised manuscript:

*Globally, **inland waters** receive 2.9 – 5.1 Tg C yr^{-1} from terrestrial ecosystems[1,3]; **lakes and reservoirs** bury 0.06 – 0.25 Mt C yr^{-1} into the sediment[4]; and **lakes and reservoirs** emit 0.06 – 0.84 Tg C yr^{-1} of CO_2 to the atmosphere[5], and export 0.8 – 1.1 Tg C yr^{-1} to the oceans[2].*

2. Main Text, line 47-48, the sentence “The total water volume of many lakes (79.4%) has also increased since 2003” expressed not clearly, what does 79.4% refer to ?

Response #1.2: We thank the reviewer for pointing out this unclear expression. The figure “79.4%” refers to the percent of studied global lakes, which have experienced an increase in

water volume since 2003. According to the reviewer's comment, we have rephrased the sentence to make the expression clearer. Please refer to lines 47-49 in the revised manuscript:

Water volumes in 79.4% of global lakes has also increased since 2003, and one of the five hotspots of such change is China's Tibetan Plateau[21-23].

3. Main Text, line 82-83, "suggested by a positive correlation between the logarithmic in-situ DOC and chlorophyll-a (Chl-a) concentration (r = 0.23, N = 3,624)", what is statistical p value ?

Response #1.3: We thank the reviewer for this comment. The statistical *p* value for the positive correlation between the logarithmic *in-situ* DOC and Chl-a concentration was 4.33×10^{-45} ($p < 0.0001$). According to the reviewer's comment, we have added the statistical *p* value to the statement. Please refer to lines 86-89 in the revised manuscript:

… as suggested by a positive correlation between the logarithmic in-situ DOC and chlorophyll-a (Chl-a) concentrations (r = 0.23, N = 3,624, ***p < 0.0001****).*

4. Main Text, line 103-104, "concentration, indicated by the linear relationship between the logarithmic in-situ TSM and POC concentration (N = 2,026, R2 = 0.14)", what is statistical p value ?

Response #1.4: Thank you to the reviewer for this comment. The statistical *p* value for the positive correlation between the logarithmic *in-situ* DOC and Chl-a concentration was 2.36×10^{-70} ($p < 0.0001$). According to the reviewer's comment, we had added the statistical *p* value to the statement. Please refer to lines 110-113 in the revised manuscript:

… indicated by the linear relationship between the logarithmic in-situ TSM and POC concentrations

*(N = 2,026, R^2 = 0.14, **p < 0.0001**).*

5. Main Text, references, line 359, change “Mendonca” to “Mendonça”.

Response #1.5: We apologize for the error in writing this reference and thank the reviewer for pointing it out. According to the reviewer’s correction, we have changed “Mendonca” to “Mendonça” in this revision. Please refer to lines 402-403 in the revised manuscript:

4. ***Mendonça**, R. et al. Organic carbon burial in global lakes and reservoirs. Nat. Commun. **8**, 1694 (2017).*

Reviewer #2 (Remarks to the Author):

Lakes play an essential role in regulating the global carbon cycle and climate change. Although many studies have provided insights into the carbon cycle of lakes, the spatiotemporal variations of lake OC storage on a large regional scale have not been reported. Previous studies have commonly assumed a fixed value for OC storage in lakes. As a matter of fact, OC storage in lakes showed obvious spatiotemporal variations along with water eutrophication, water level rise, and human activities. It is of great significance to obtain OC storage spatiotemporally in lakes. On the one hand, OC storage reflects how much carbon have been stored in lakes. On the other hand, changes in OC storage also have important influences on lake CO2 emissions and carbon burial in sediments, two essential aspects of the global carbon cycle. Therefore, the dynamic acquisition of lake OC storage, as studied in this paper, plays an irreplaceable role in accurately quantifying the role of lakes in the global carbon cycle.

Based on measurements at 4,201 stations from 348 lakes across China, this study quantified the spatiotemporal variability of DOC and POC storage for 24,366 Chinese lakes during 1984-2023 using advanced satellite monitoring techniques. The authors report that DOC and POC storage in Chinese lakes increased by 44.6% and 33.5%, respectively, which hasn't been noticed before and is of great importance for understanding the carbon cycle in lakes. Using remote sensing techniques, the authors innovatively conducted a remote sensing study of OC storage in a large region of Chinese lakes by integrating basin characteristics and satellite reflectance, and revealed the spatiotemporal variation characteristics of OC storage in all lakes in China for the first time. I believe the study has an important promoting role and

reference value for the dynamic monitoring of global lake carbon cycle by remote sensing and deep learning techniques, and can compensate for the insufficient spatiotemporal coverage of global lake monitoring OC data.

Overall, the extensive field data are very valuable, the methods used are innovative and reasonable, and the findings are very important. The satellite-derived datasets of DOC and POC concentrations are also valuable for understanding carbon emission and carbon burial in Chinese lakes. In addition, I think the manuscript is well organized and written. Some revisions and clarifications are required before publication in the NC journal. I have the following suggestions for the authors to improve the manuscript.

Response: We appreciate the reviewer for highly recognizing the research value of this study, confirming the innovation of OC storage remote sensing in a large region of Chinese lakes by integrating basin characteristics and satellite reflectance, and approving the reliability of the results. We also thank the reviewer for providing detailed comments and suggestions, which are of great help to us to improve the manuscript. To facilitate your assessment of the changes, we have highlighted our changes in red throughout the revised manuscript. According to the reviewers' comments and suggestions, we have made the following major changes to the manuscript:

(1) We have made some minor changes to Fig. S1, Fig. 2, and Fig. 3. Please refer to the following Responses #2.3, #2.7, and #2.8, respectively.

(2) We have added some statements, which are essential for understanding the manuscript. The new additions include: further clarification on lake identification (Response #2.2); the dividing criteria for the five lake zones (Response #2.3); the meaning of

the main drivers (Response #2.8); the time-matching schemes for OCAR and POC data (Responses #2.11 and #2.12); the temporal coverage of basin property data (Response #2.14); the removal processes of clouds and other non-water pixels (Response #2.15); and the reason for choosing the Gini importance index (Response #2.16).

(3) We have modified some incorrected and/or unclear statements. Modifications have been made to the statements on: the impact factors on OC storage changes (Response #2.1); the increasing trends of POC concentrations in different lakes (Response #2.4); the calculation method for the area ratio of lakes with increasing POC concentrations (Response #2.5); DOC storage in the TPL zone (Response #2.6); the assumption of DOC concentration in glacial meltwater and rainwater (Response #2.9); photoproduction of OC in suspended sediment (Response #2.10); and the vague phrase of "lake carbon cycle" (Response #2.13).

We would like to thank the reviewer for his/her valuable comments and suggestions, which are very beneficial for the improvement of our manuscript. We sincerely hope that you will be satisfied with the revised manuscript.

1. "OC storage is expected to vary … water volume changes". According to the above definition, OC concentration and water volume variations will change the OC storage in the lake. In addition, the OC concentration in the lake also has obvious temporal variation characteristics. The authors here say that OC storage varies with water volume, which is incomplete.

Response #2.1: Yes, the reviewer is correct. Apart from water volumes, OC concentrations in lakes also have temporal variation characteristics. Therefore, the statement that OC

storage varies with water volume is incomplete. In this revision, we have changed the statement to emphasize that lake OC storage also varies with changes in OC concentration. Please refer to lines 52-53 in the revised manuscript:

Lake OC storage is expected to vary greatly, ***as the changes in OC concentration and/or water volume****.*

2. "Across the studied lakes …". How many lakes did the authors study in total? Do these lakes include all lakes in China? If not, how did the authors identify the lakes they studied? Although the authors may have included relevant information in the online methods, it is still necessary to explain it here so that readers can understand the relevant content.

Response #2.2: We thank the reviewer for pointing out this unclear statement. We agree with the reviewer that it is necessary to explain how many lakes were studied here so that readers can understand the relevant content. In this study, we studied 24,366 lakes in total. Based on the HydroLAKES dataset (Supplementary Table S1), we identified the studied lakes using an area threshold of > 0.01 km^2. The studied lakes are those that could be observed using Landsat satellite data with a spatial resolution of 30 m. In this revision, we have added this information. Please refer to lines 76-79 in the revised manuscript:

Across the studied lakes ***with an area > 0.01 km^2 from the HydroLAKES dataset i.e., that could be observed using Landsat satellite data (N = 24,366)*** *….*

3. "the southeast NPML and EPL zones" and "the northwest TPL zone". In my opinion, the TPL zone is located in southwest China, while the NPML zone is located in northeast China.

How do the authors divide the different lake zones? Why do the authors divide them this way? What is the reference standard? Please add any relevant information.

Response #2.3: We thank the reviewer for bringing this to our attention. We forgot to illustrate the dividing criteria of the five lake zones across China in the original manuscript. Because climatic conditions and human activities in China are clearly different on both sides of the Hu-line, we further divided the five lake zones into the northwest and southeast parts according to the Hu-line (Supplementary Fig. S3). The northwest part includes the IMXL and TPL zones, and southeast part includes the YGPL, NPML, and EPL zones. In this revision, we had added the information about the dividing criteria and delineated the Hu-line in Supplementary Fig. S3. Please refer to lines 62-65 in the revised manuscript:

Across China, climatic conditions and human activities are clearly different on both sides of the Hu-line (Supplementary Fig. S3). Compared to the northwest IXML and TPL zones of the Hu-line, the three southeast lake zones had higher precipitation and more human activities[25].

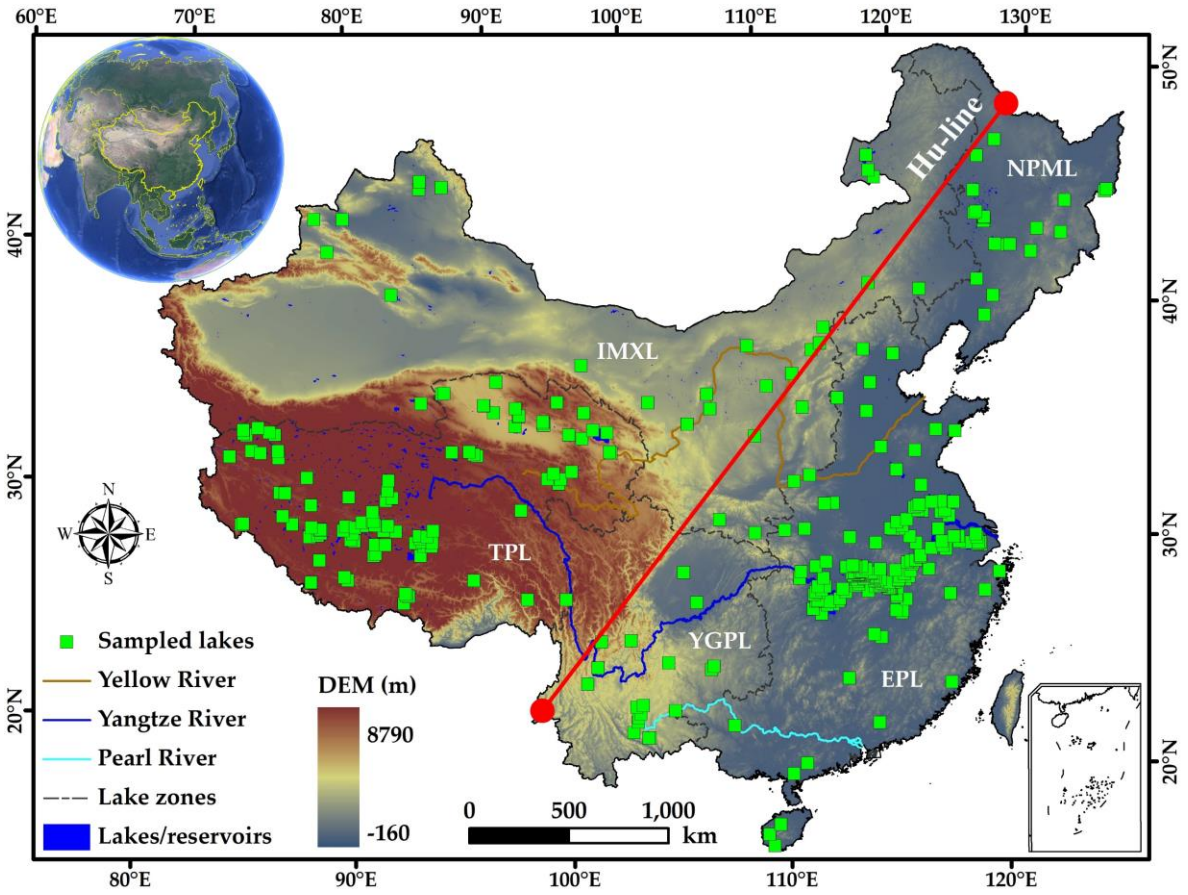

Fig. S3. The spatial distribution of the 348 sampled lakes during 2004-2023 (Dataset I in Supplementary Table S1). The inserted global map was obtained from Google Earth.

4. "…demonstrated clear increasing tends…". I don't quite understand what the authors mean by this. Do the authors mean increasing trends in different lakes? Please clarify.

Response #2.4: We apologize for this unclear statement. Yes, the sentence means that DOC concentrations demonstrated clear increasing trends in different lakes across China during 1984-2023. In this revision, we have clarified the statement. Moreover, we have also made similar changes to the statement on POC concentrations. Please refer to lines 76-79 and lines 103-105 in the revised manuscript:

Lines 76-79: *… and showed general increasing trends **in different lakes** during 1984-2023*.

Lines 103-105: *… and demonstrated clear increasing trends **in different lakes** during 1984-2023 across China.*

5. “80.6% of lake area showed increasing POC concentration”. How is the lake area ratio calculated here? Did the authors calculate the changes in POC concentrations of all pixels in Chinese lakes? From the whole paper, I think the authors only counted the average POC concentrations of lakes in China, but did not perform statistics at the remote sensing pixel level.

Response #2.5: We thank the reviewer for this thoughtful suggestion. Yes, we only counted the average lake POC concentrations and did not perform statistics at the remote sensing pixel level. That is, we did not calculate the changes in POC concentrations of all pixels. The lake area ratio here was calculated by dividing the total area of lakes with increasing POC concentrations by the total area of all studied lakes. In this revision, we have rephrased the sentence to clarify it. Please refer to lines 114-116 in the revised manuscript:

Therefore, as population density generally grew in China (Supplementary Fig. S7h), ***16,007 lakes whose area in total accounted for 80.6%*** *showed increasing POC* ***concentrations*** *during 1984-2023 (Fig. 1d).*

6. “Lake DOC was primarily concentrated in the TPL zone… increased substantially”. The statement is not clear. Do the authors mean that lake DOC is mainly stored in the TPL zone? Is the increase in lake DOC mainly in the TPL zone? Or do Chinese lakes show a significant increase in DOC storage?

Response #2.6: We apologize for this unclear statement. Yes, we meant that lake DOC was mainly stored in the TPL zone and the increase in DOC storage mainly occurred in the TPL

zone. Chinese lakes collectively contained 39.43 Tg C of DOC in 2015, and 83.2% of this total DOC storage was stored in the TPL zone. Furthermore, of the 1,125 lakes with time-series of water volume, 63.7% experienced an increase in DOC storage, and 92.0% of the total increase occurred in the TPL zone. In this revision, we have rephrased the sentence to clarify them. Please refer to lines 123-124 in the revised manuscript:

Lake DOC was ***mainly stored*** *in the TPL zone, and* ***the increase in lake DOC storage during 1984-2023 also mainly occurred in the TPL zone****.*

7. Does the lake proportion in Figure 2 represent the number of lakes or the area of lakes? How are "changes of DOC storage" obtained? Does the "changes of DOC storage" mean the linear fitted slope between DOC storage and year during 1986-2023? Also, I think the figure captions need to be labelled separately.

Response #2.7: We appreciate this comment by the reviewer and have modified Figure 2 accordingly. In Fig. 2, "the lake proportion" represents the number of lakes, and the "changes of DOC storage" represents the linearly fitted slope between DOC storage and year during 1986-2023. In this revision, we have modified Fig. 2 and its caption to clarify this information. Moreover, we have also separately labelled the figure caption. Please refer to Fig. 2 in the revised manuscript:

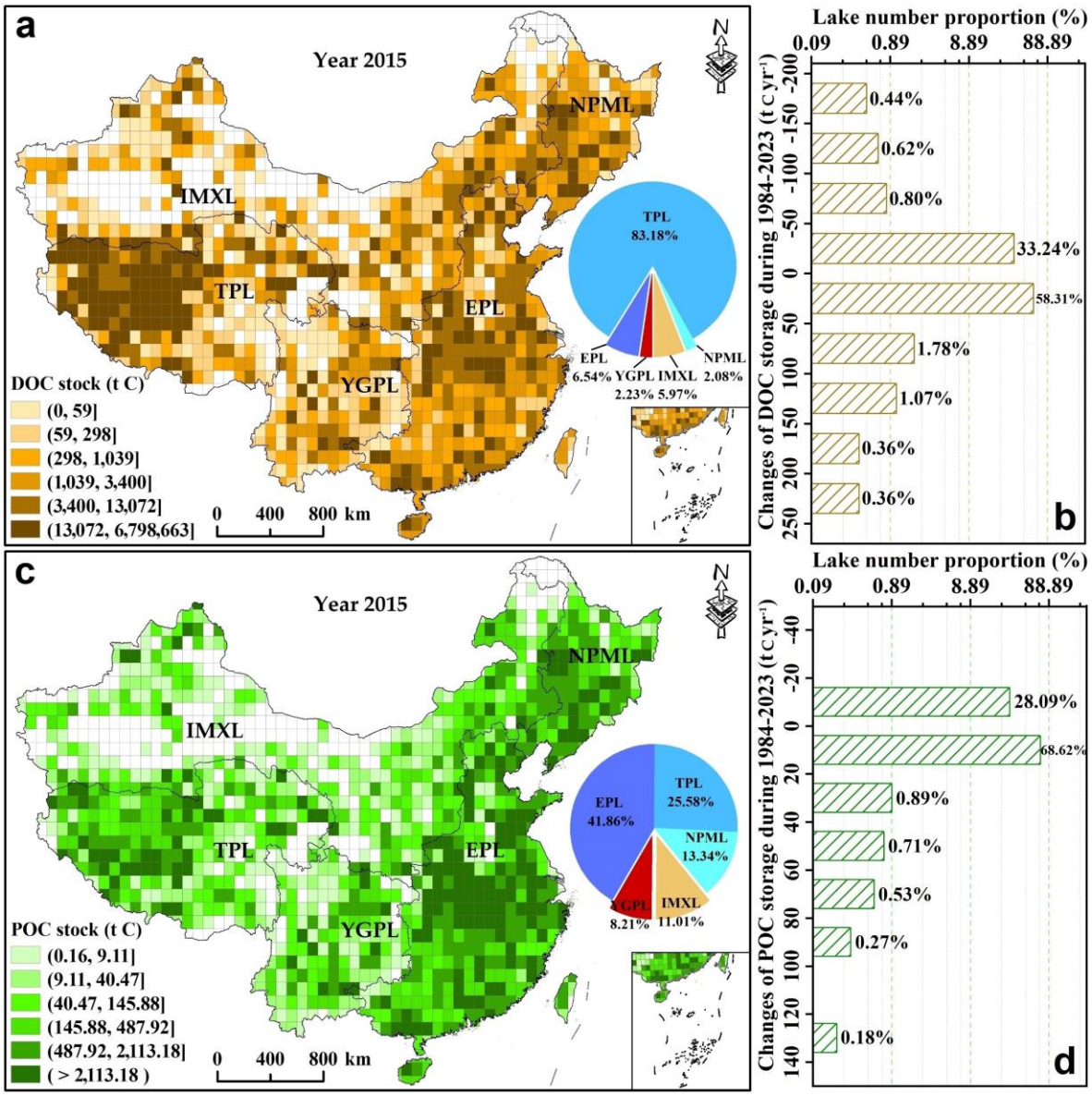

Fig. 2. Spatiotemporal variations in OC storage. **a** 1.0° grid-based DOC stock for 24,366 lakes in 2015; **b** lake-based changes for 1,125 lakes during 1984-2023; **c** 1.0° grid-based POC stock for 24,366 lakes in 2015; and **d** lake-based changes for 1,125 lakes during 1984-2023. **The changes of DOC storage mean the linearly fitted slope between DOC storage and year during 1984-2023. Source data are provided as a Supplementary Data file**.

8. "… were identified as the main drivers …". How did the authors identify the main drivers of changes in DOC and POC concentrations in different lakes? Are the primary drivers those that contribute more than 50%? Or does it refer to the factor that has the largest contribution among the factors studied? The authors need to explain both in the main text and in the supplementary information, as the reader will often not necessarily read the supplementary information.

Response #2.8: Thank you for bringing this to our attention. We apologize for that we did not explain the meaning of the main driver in the original manuscript. The primary driver refers to the factor that had the largest contribution among the impact factors examined. According to the reviewer's comments, we have added this information in lines 176-177, the caption of Fig. 3, and the Supplementary Note S4:

Lines 176-177: *The main driving factors indicate those which have the largest contributions*.

Supplementary Note S4: *According to the relative contributions of different factors, we determined the main driving factors which have the largest contributions for different lakes*.

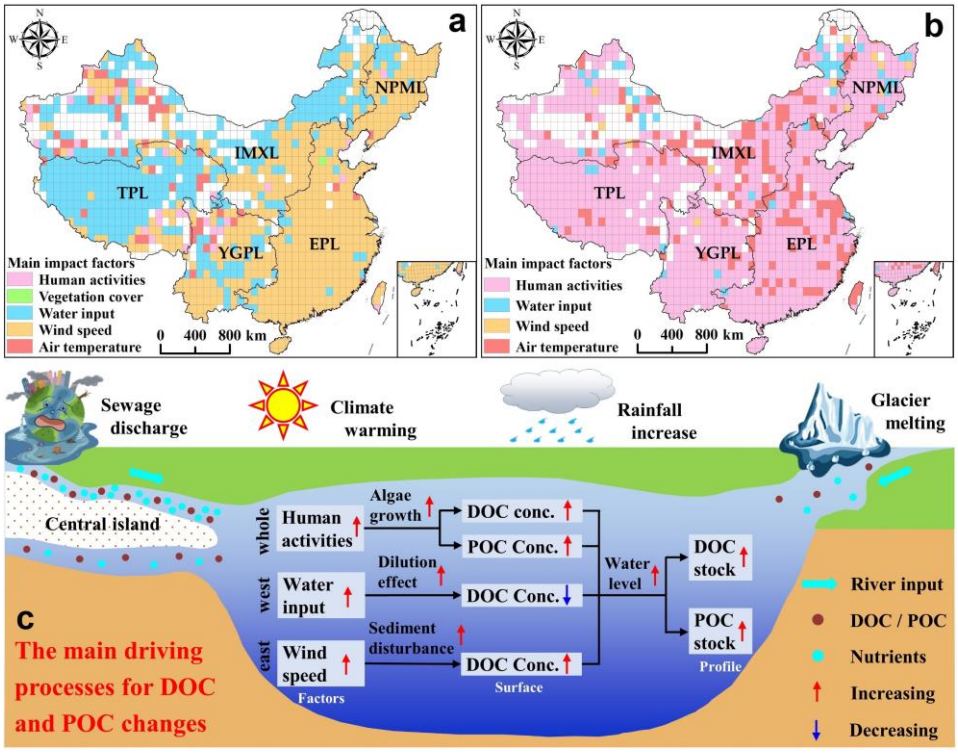

Fig. 3. Influences of different factors on DOC and POC variations. The main driving factors for the annual changes of **a** DOC and **b** POC concentrations during 1984-2023. **The main driving factors indicate those which have the largest contributions**. The figures show the main factors with the largest mean contributions for lakes in the 1.0° grids. **c** The driving processes of the main factors for the changes in DOC and POC. DOC Conc.: DOC concentration. Source data are provided as a Supplementary Data file.

9. "glacial meltwater and rainwater … DOC concentrations … 1.0 mg/L". The DOC concentration in rainwater also has obvious spatiotemporal variation characteristics. Does 1.0 mg/L refer to the average DOC concentration in lakes in the IMXL and TPL zones? Is the

DOC concentration in glacial meltwater also 1.0 mg/L? The study area of Gao et al. (2020) is Central Asia. Is the DOC concentration in the glacial meltwater of Central Asia the same as in the IMXL and TPL zones? The author needs to provide more information or a clear introduction.

Response #2.9: We thank the reviewer for pointing out this incorrect assumption. In the original manuscript, we assumed a DOC concentration of about 1.0 mg L^{-1} for glacial meltwater and rainwater in the IMXL and TPL zones. However, DOC concentrations in both glacial meltwater and rainwater have obvious spatiotemporal variation characteristics. According to Gao et al. (2020), the DOC concentration of Muz Taw Glacier has a variation range of 0.22 – 7.71 mg L^{-1}, with an average value of 1.12 ± 1.66 mg L^{-1}. In addition, DOC concentration in the glacial meltwater of Central Asia is not the same as that in the IMXL and TPL zones. According to Safieddine and Heald (2017), the DOC concentration in rainwater varies from 0.01 mg L^{-1} to 10 mg L^{-1} globally. Therefore, the assumption of DOC concentration (about 1.0 mg L^{-1}) for glacial meltwater and rainwater in the IMXL and TPL zones was not correct. In this revision, we have removed the incorrect assumption and rephrased the related sentence. Please refer to lines 202-204 in the revised manuscript:

In contrast, glacial meltwater and rainwater ***commonly have low DOC concentrations (around 1.0 mg L^{-1} or lower)***[32,33] *and could dilute lake DOC in saline lakes.*

10. "… through pore water or photoproduction". Pore water is usually derived from sediments. Does photoproduction also refer to the oxidative decomposition of OC in sediments to produce DOC? Eutrophic lakes are usually very turbid and have very low

turbidity, so light has little or no access to the bottom of the turbid water. If so, how can light oxidize OC in sediments and break it down into DOC? Please explain further.

Response #2.10: We thank the reviewer for pointing out this misleading statement. As the reviewer said, eutrophic lakes in the southeast lake zones are usually turbid, so there is little or no light reaching the bottom of the turbid water. Therefore, light is less likely to oxidize the OC in the sediment and break it down into DOC. However, eutrophic lakes are usually characterized by shallow depths and high sediment resuspension, which favor DOC release from the photoproduction of OC in suspended sediment. Furthermore, high OC content in the sediment also favors DOC release from the pore water of the sediment. Based on these understandings, we have modified the misleading statement. Please refer to lines 213-215 in the revised manuscript:

*Second, high algal proliferation in these zones elevates the OC content of the sediment[19,37], which further leads to high DOC release **from the** pore water **of sediment** or from **the** photoproduction **of OC in suspended sediment**.*

11. "OCAR was exponentially related to lake-based mean POC concentration". How was the POC concentration from the satellite remote sensing inversion matched to the OCAR data? Both POC concentration and OCAR vary with time, and different time-matching schemes may give different results. In addition, did the authors get annual OCAR data, like POC concentration?

Response #2.11: Yes, we agree with the reviewer. Both POC concentration and OCAR vary with time, different time-matching schemes may give different results. At present, it is less

likely to get annual OCAR data, as with POC concentration. In this study, we only obtained OCAR data after the 1950s for 115 Chinese through meta-analysis. To match the lake-based OCAR data in time, we used the satellite-derived climatological mean POC concentrations of the relative lakes with OCAR data. In this revision, we have given the time spans for both the OCAR data and the POC concentrations. Please refer to lines 239-241 in the revised manuscript:

For different lakes, OCAR ***after the 1950s*** *was exponentially related to* ***the climatological mean POC concentration during 1984-2023****, with* $R^2 = 0.35$ *and* $p < 0.01$ *(Fig. 4b).*

12. "…Europe exhibited OCAR values…". As the previous comment, the authors need to give a specific time for the OCAR values.

Response #2.12: We thank the reviewer for this suggestion. The OCAR value was cited here with reference to Anderson et al. (2014). The time span for the OCAR value was "over the past century". In this revision, we have added this information. Please refer to lines 271-272 in the revised manuscript:

For example, eutrophic lowland lakes in Europe exhibited OCAR values in excess of 50 g C m^{-2} yr^{-1} ***over the past century***[42]*.*

13. "… measure the lake carbon cycle… using satellite data". The "lake carbon cycle" is a very vague definition that includes various aspects of carbon composition, decomposition, and transformation. What exactly do the authors mean by "lake carbon cycle"? As far as I know, although satellite data can be used to estimate CO2 emissions and carbon storage

studied in this paper, there is no report on the use of satellite data to estimate carbon decomposition and transformation.

Response #2.13: Yes, we agree with the reviewer that the phrase "lake carbon cycle" has a very vague definition here. Although satellite data can be used to estimate many parameters related to the lake carbon cycle, this study only reported information on carbon storage, CO_2 emissions, and carbon burial in the sediment. Moreover, at present, it is difficult to monitor carbon decomposition and transformation using satellite data. Therefore, according to the research scope of this study, we have replaced the vague phrase of "lake carbon cycle" with "carbon storage in the water column, CO_2 emission to the atmosphere, and carbon burial to the sediment". Please refer to lines 277-279 in the revised manuscript:

Therefore, urgent efforts are needed to monitor ***carbon storage in the water column, CO_2 emission to the atmosphere, and carbon burial in the sediment*** *using global revisit satellite data.*

14. "monthly averaged basin property … and wind speed". Where do these data come from? What is the temporal coverage of these data? This sentence is incomplete.

Response #2.14: We apologize for this incomplete sentence. The monthly averaged basin property data were obtained from the European Centre for Medium-Range Weather Forecasts. These data have a temporal coverage of 1984-2023. In this revision, we have supplemented this information. Please refer to lines 297-302 in the revised manuscript:

… moreover, ***from the European Centre for Medium-Range Weather Forecasts (ECMWF), we obtained*** *monthly averaged basin property data during 1984-2023,* ***including*** *population density,*

digital elevation model (DEM), evaporation, leaf area index of high vegetation (LAI_HighVeg), leaf area index of low vegetation (LAI_LowVeg), runoff, air temperature at 2 m height (Temp2m), precipitation, and wind speed.

15. "… applied to retrieve annual mean DOC concentrations …" and "… was applied to retrieve annual mean POC concentrations …". The models constructed in this paper should first be applied to Landsat data to calculate the daily DOC and POC concentrations, and then calculate the average values according to the daily remote sensing results. The authors need to clearly explain the relevant processes. Also, how did the authors remove clouds and other non-water pixels? Did the authors use the methods constructed in this study, or did they follow the published methods?

Response #2.15: We appreciate the reviewer's comments and have modified the manuscript accordingly. As the reviewer mentioned, the models constructed in this study were first applied to Landsat data to derive daily DOC and POC concentrations, and then we calculated the annual mean values according to the daily remote sensing results. Moreover, the clouds and other non-water pixels of the Landsat data were removed using the pixel quality attributes provided by GEE. In this revision, we have explained the relevant processes for the remote sensing inversion of DOC and POC concentrations. Please refer to lines 320-327 and lines 346-352 in the revised manuscript:

Lines 320-327: *The parameterized models were then applied to derive daily DOC concentrations from Landsat data, which had been processed to remove clouds and other non-water pixels using the pixel quality attributes provided by GEE. The developed models satisfactorily captured DOC*

concentrations from time-series Landsat data, with high values for saline/eutrophic lakes and low values for estuarine waters (Supplementary Fig. S10). Furthermore, annual mean DOC concentrations for 24,366 lakes during 1984-2023 were calculated using the satellite-derived daily DOC concentrations.

Lines 346-352: *When they were applied to time-series Landsat data, the developed models obtained reasonable POC concentrations, with high values for eutrophic/turbid lakes and estuarine waters (Supplementary Fig. S10). Then, similar to the DOC retrieval, the parameterized models were applied to obtain daily POC concentrations and annual mean POC concentrations for 24,366 lakes during 1984-2023, calculate POC storage for 24,366 lakes in 2015, and calculate annual POC storage for 1,125 lakes during 1984-2023.*

16. For a given factor, the nonlinear Random Forest analysis indicated its relative contribution using the Gini importance. The authors must explain why the Gini Importance index was chosen. Compared to the Permutation Importance and Boruta indexes, can the Gini Importance index better reflect the importance of the results of the Random Forest analysis?

Response #2.16: We thank the reviewer for this comment about the nonlinear Random Forest analysis in Supplementary Note S4. In the Random Forest analysis, the Gini, Permutation, and Boruta importance indexes are usually used to evaluate the importance of different features. Each of these three indexes have advantages and disadvantages. According to Rodríguez-Pérez and Bajorath (2021), the Gini importance index was chosen for this study because it is computationally efficient, suitable for large-scale analysis, and

easy to understand and explain. In this revision, we have supplemented the following statements to explain why we chose the Gini importance index in the Random Forest analysis. Please refer to Supplementary Note S4:

In the Random Forest analysis, the Gini, Permutation, and Boruta importance indexes are usually used to evaluate the importance of different features. Each of these three indexes has its advantages and disadvantages. According to Rodríguez-Pérez and Bajorath (2021)[10], we chose the Gini importance index because it is computationally efficient, suitable for large-scale analysis, and easy to understand and explain. For a given factor, the non-linear Random Forest analysis indicated its relative contribution using the Gini importance, which was equivalent to the mean decrease in Gini impurity calculated as the normalized sum of the impurity decrease values for all nodes[10]. The contribution of each factor had a value of 0 – 100%, and the sum of the contributions of the eight factors was 100%.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

I have carefully read the revision of the paper. The manuscript has improved greatly after the first revision round. The authors corrected the paper according to my comments and cleared up all my questions. I think that the paper can be accepted for publication. I do not have further questions or comments.

Authors' responses to reviewer's comments

Substantial increase of organic carbon storage in Chinese lakes

Dong Liu, Kun Shi, Peng Chen, Nuoxiao Yan, Lishan Ran, Tiit Kutser, Andrew N. Tyler, Evangelos Spyrakos, R. Iestyn Woolway, Yunlin Zhang, Hongtao Duan

Submission ID: NCOMMS-24-39360A

Reviewer #2 (Remarks to the Author):

I have carefully read the revision of the paper. The manuscript has improved greatly after the first revision round. The authors corrected the paper according to my comments and cleared up all my questions. I think that the paper can be accepted for publication. I do not have further questions or comments.

Response: We thank the reviewer for acknowledging our changes. The reviewer's comments and suggestions from the first round of review have help us to improve our manuscript considerably.