Response to Reviewer 2

This paper examines a VIIRS high resolution band for determination of water turbidity, based on Secchi depth for a number of lakes in China, showing patterns. This is an interesting result for these lakes, and shows potentially useful results with this satellite for turbidity in this region. The method suggests broader applicability of the VIIRS or the relationship to Secchi depth.

There are editorial and statistical issues. The paper needs a close read. Figures 3 and 4 are mislabeled. Figure captions do not explain what is in the figures.

Response: Thank you for the positive feedback and constructive suggestions on this work. As the reviewer suggested, we checked the numbering and revised the descriptions associated with Figs. 3 and 4. We also expanded the relevant figure captions and revised the statistical interpretation, as detailed in the point-by-point responses below.

1. Figure "1" (which isn't Figure 1) line 162, STD is meaningless for a log distributed data set. Report the IQR or another appropriate metric of dispersion.

Response: Thank you for this suggestion. We have replaced STD with IQR in Fig. 3 and defined IQR as Q3 − Q1 in the figure caption. The mean and median are retained to describe the asymmetric distributions of the in situ observations (line 161-163). *"Each panel reports the sample size, mean, median, and interquartile range (IQR = Q3 − Q1), where Q1 and Q3 are the 25th and 75th percentiles, respectively."*

2. Line 90 Wei et al, Wei et al, Wei et al, Wei et al. ????

Response: We have corrected repeated and malformed citations and consolidated the duplicate references.

3. 3.5 Better statistics. statistical significance is not meaningful, and is not scientific or ecological significance (Wasserstein et al., 2019, American Statistician)

Response: We agree that statistical significance should not be equated with scientific or ecological importance. We have expanded the statistical analysis and revised the interpretation to assess the direction and magnitude of long-term SDD changes in relation to typical lake clarity and temporal variability, rather than relying solely on p-value thresholds.

Specifically, Section 3.5 now clarifies the use of the seasonal Mann–Kendall test to account for seasonality and the corresponding seasonal Sen slope to quantify trend magnitude. We additionally report decadal relative changes normalized by each lake's median SDD and the trend-to-variability ratio. For the TVR calculation, seasonal adjustment was performed by subtracting each lake's calendar-month median from the corresponding monthly observations.

These analyses quantify both the observed changes and their heterogeneity among lakes. We have expanded Section 6.2 to discuss these findings and removed the statistical-significance arrows from Fig. 12. Together, these revisions provide quantitative context beyond statistical significance. Relative change and TVR are interpreted as descriptive measures of the scale of change, rather than universal thresholds of ecological importance; they do not, by themselves, establish ecological consequences (Sections 3.5 and 6.2) (line 309-323) (line 507-514). *"Temporal trends in SDD were assessed using the seasonal Mann–Kendall test implemented in the seasonal test function of pyMannKendall (Hirsch et al., 1982; Hussain and Mahmud, 2019). To account for the pronounced annual cycle in lake water clarity, monthly observations were grouped by calendar month, and comparisons were made between observations from the same month in different years. The within-month test statistics were then combined to assess the overall monotonic trend. This procedure accounts for seasonality through stratified comparisons, without requiring a prior subtraction of the seasonal cycle. Kendall's tau (τ), ranging from −1 to 1, was used to describe the direction and strength of the rank-based temporal association. Positive and negative values indicate increasing and decreasing tendencies, respectively, while larger absolute values indicate more consistent directional changes among the compared observations. Trend magnitude was quantified using the corresponding seasonal Sen slope, calculated as the median of pairwise slopes between observations from the same calendar month across years, and expressed in m decade-1. Thus, τ describes the consistency of the temporal association, whereas the seasonal Sen slope quantifies the rate of change." "To complement the pooled-pixel mean trajectories in Fig. 13, Fig. 14 presents the distribution of relative SDD change rates calculated separately for each lake. Among the 767 lakes, 533 showed positive and 234 showed negative rates, with a median of 14.9 % decade-1 (IQR: −5.2 to 37.1 % decade-1). The regional medians were the highest in*

the TPL (17.9 % decade-1) and YGPL (16.0 % decade-1), although 3 of the 13 YGPL lakes showed negative rates. The EPL had the lowest median (7.4 % decade-1). The IQR intervals in the TPL, NPML, IMXL, and EPL included zero, showing that positive regional medians did not represent a uniform increase across lakes."

4. Use Seegers et al., 2018 for metrics, e.g., multiplicative error, and multiplicative bias. These are particularly necessary given the large range of data. MAPE is ok instead of multiplicative error, as long as it is small (e.g., < 50%) and is not used for algorithm comparison (MAPE biases towards the algorithms that underestimate). However, as Figure 7d has a MAPE > 100%, which is physically meaningless (implies negative values), use multiplicative error and bias throughout.

Response: Thank you for recommending Seegers et al. (2018). We report multiplicative MAE and Bias and have removed MAPE from the *performance assessment. Both metrics are dimensionless factors calculated from positive values in logarithmic space. Their definitions are given in Sect. 3.5, Eqs. (2) and (3)* (line 279-288)*:*

$$\mathrm{MAE} = 10^{\frac{1}{N}\sum_{i=1}^{N}\left|\log_{10}\left(\frac{y_i}{x_i}\right)\right|} \quad (2)$$

$$\mathrm{Bias} = 10^{\frac{1}{N}\sum_{i=1}^{N}\log_{10}\left(\frac{y_i}{x_i}\right)}, \quad (3)$$

where N is the number of valid paired observations, xi is the measured value, and yi is the estimated value. Both metrics have an ideal value of 1. MAE describes the magnitude of the multiplicative deviation, while Bias describes its direction: values below 1 indicate underestimation and values above 1 indicate overestimation.

5. Lines 235-236 and Figure 5. P is not an error metric. P has no value, given the 906 samples, p<0.001 is achieved once Rsquared > 0.012, which is not "significant" by any standard.

Response: We agree that a small *p* value does not establish the strength or practical importance of a relationship and is not a measure of prediction error. We have removed the *p* values from Fig. 5 and the corresponding text in Section 3.4.1. Figure 5 describes associations between in situ SDD and Chl-a or SPM, rather than retrieval validation. We retain R^2 as a descriptive

measure of goodness of fit and show 95% confidence bands for the fitted mean relationships and 95% prediction bands for individual observations. These bands are explained in the caption.

6. Figure 6 and after. Rsquared scales the RMSE to the range of data, which is ok if you are building a model. Otherwise, R is useful for saying there is a general relationship (e.g., Figure 5, Figure 7a), but not to validate a model. Report the multiplicative error, or better yet, explain what the pink zone is. It's probably an error metric, and may be a suitable one, like one that encompasses X% of the data.

Response: Figure 6 reports multiplicative MAE and Bias following Seegers et al. (2018). The figure caption now identifies the shaded area as the 95% prediction band of the fitted relationship in logarithmic space and distinguishes the fitted line from the 1:1 line (line 373-376). *"The black dashed line denotes the 1:1 relationship, and the red dashed line is a descriptive fit. The shaded area in plot a shows the 95% prediction band of the fitted relationship in logarithmic space."*

7. For validation, an excellent model can have a low Rsquared, if the range is small. (And it is trivial to demonstrate a model with Rsquared =0.99 that is completely wrong). Because Rsquared scales to range, Rsquares can often not be compared. There is no reason to show a regression line for validation, unless it is explained what value it provides. Yes, people do it frequently, but that doesn't make it valid or useful. A model that performs well except at the lowest (or highest) values may have a regression line that significantly deviates from 1, while a model that is severely biased in the middle of the range can have a slope of 1.

Response: We agree that R^2 and fitted slopes do not establish agreement between estimated and measured values and may obscure range-specific biases. We retain fitted lines only as descriptive aids. Sect. 3.5 now states (line 288-294): *"In the validation plots (Figs. 6a and 7b–d), the fitted regression lines provide descriptive summaries of the overall relationships between estimated and measured values. Their departures from the 1:1 line help visualize broad tendencies toward overestimation or underestimation but may obscure biases within specific value ranges. Therefore, neither the fitted slopes nor R^2 are used to establish predictive accuracy or rank algorithm performance; validation relies on multiplicative MAE and Bias,*

together with the distribution of individual observations relative to the 1:1 line."

8. 295-298. Kendall's tau is not defined. That is an uncertainty metric. P is not. There is no explanation of what is done. Given the strong seasonality, was deseasonalized version used? (figure 13, and related).

Response: We expanded Sect. 3.5 to define Kendall's tau and explain how seasonality is handled (line 309-323): *"Temporal trends in SDD were assessed using the seasonal Mann–Kendall test implemented in the seasonal test function of pyMannKendall (Hirsch et al., 1982; Hussain and Mahmud, 2019). To account for the pronounced annual cycle in lake water clarity, monthly observations were grouped by calendar month, and comparisons were made between observations from the same month in different years. The within-month test statistics were then combined to assess the overall monotonic trend. This procedure accounts for seasonality through stratified comparisons, without requiring a prior subtraction of the seasonal cycle. Kendall's tau (τ), ranging from −1 to 1, was used to describe the direction and strength of the rank-based temporal association. Positive and negative values indicate increasing and decreasing tendencies, respectively, while larger absolute values indicate more consistent directional changes among the compared observations. Trend magnitude was quantified using the corresponding seasonal Sen slope, calculated as the median of pairwise slopes between observations from the same calendar month across years, and expressed in m decade-1. Thus, τ describes the consistency of the temporal association, whereas the seasonal Sen slope quantifies the rate of change."*

9. Figures 5 and 6, the pink zone is not explained.

Response: The captions now distinguish the shaded regions. In Fig. 5, the darker and lighter bands are the 95% confidence and prediction bands, respectively. In Fig. 6a, the shaded region is the 95% prediction band. These are regression-based intervals, not pixel-specific uncertainty estimates for the distributed product(line250-252). *"The darker and lighter shaded bands show the 95% confidence bands for the fitted mean relationships and the 95% prediction bands for individual observations, respectively."*

10. Line 265. Reference figure 7 here (The relationship of SDD to *Rrs*(638)).

Response: We have cited Figure 7 as you suggested.

11. As far as R2= 0.87, p<0.001. Report the error/uncertainty around the regression. Again p is meaningless. As a note, no metric should be reported that is not explained. The purported probability of a null hypothesis is not relevant to your hypothesis.

Response: We have removed the *p*-value statement from the model-development description and clarified the interpretation of the reported statistics in Sect. 3.5. In Fig. 7a, R^2 is retained only to describe the proportion of variation in log-transformed SDD explained by the fitted regression; it is not used as evidence of predictive accuracy. The 95% prediction band represents uncertainty for a new individual log-transformed SDD observation at a given $R_{rs}(638)$, accounting for both uncertainty in the fitted mean relationship and residual variability under the regression assumptions. Its interpretation is now explained in Section 3.5. Independent retrieval performance is evaluated separately using multiplicative MAE and Bias in Fig. 7b, with both metrics defined in Section 3.5 (line 279-288). *"Retrieval performance was evaluated using multiplicative mean absolute error (MAE) and multiplicative bias (Bias), following Seegers et al. (2018). These metrics are calculated from positive measured and estimated values in logarithmic space and are expressed as dimensionless factors. Their definitions are given below:*

$$\mathrm{MAE} = 10^{\frac{1}{N}\sum_{i=1}^{N}\left|\log_{10}\left(\frac{y_i}{x_i}\right)\right|}, \quad (2)$$

$$\mathrm{Bias} = 10^{\frac{1}{N}\sum_{i=1}^{N}\log_{10}\left(\frac{y_i}{x_i}\right)}, \quad (3)$$

where N is the number of valid paired observations, xi is the measured value, and yi is the estimated value. Both metrics have an ideal value of 1. MAE describes the magnitude of the multiplicative deviation, while Bias describes its direction: values below 1 indicate underestimation and values above 1 indicate overestimation.."

12. Figures 7 c and d. Use multiplicative error. A percentile error of > 100% is meaningless, this implies negative SDD. Explain the pink region.

Response: Panels (b)–(d) of Fig. 7 now report multiplicative MAE and Bias. The caption explains that (line 406-408) *the shaded areas show the 95% prediction bands of the fitted*

relationships in logarithmic space. MAE and Bias are dimensionless multiplicative metrics; fitted lines and R^2 are not used to rank predictive accuracy..

13. Figure 8, what is black? Are the SDD statistics by pixel or by lake? Is this the mean of all pixels? STD is meaningless. The distribution is clearly not Gaussian, so STD is not only incorrect, but an STD of 3.06 with a mean of 1.17, means that negative SDD occur. The IQR would be a good choice.

Response: Grey represents land or invalid pixels, whereas black is part of the upper end of the SDD color scale. The statistics are calculated from the valid pixels in each displayed scene, not from an equal-weighted set of lake means. As you suggested, STD has been replaced with IQR. We have explained it in the caption (line 425-428): *"The lower-right inset in each panel shows the mean, median, and IQR calculated from valid SDD pixels in the displayed scene. Grey represents land or invalid pixels, while black represents the upper end of the displayed SDD color scale."*

14. Line 405. Change to "highlights substantive differences associated with lake morphology".

Response: Corrected.

15. Sen's slope is appropriate for the trends. However, this should be deseasonalized. Was it? No information in the methods. Figure 12. Remove the "statistical significance" arrows. What is of interest is how much the trend compares to the variability, or to the typical lake turbidity. There are lakes with "significant" trends that have a rate of < 0.1 m per decade (given there is only one decade of data). I'm skeptical that these are ecologically significant changes (the only significance that matters). What is the ratio of the change to the mean condition of the lake over the time period? A ratio of < 10-15% over the decade is not particularly significant. This gets close to Cohen's d differences. The other question is how the trend compares to the variability, Kendalls' tau threshold as a marker, e.g. tau > 0.3, or better the "trend to variability ratio". No arrow required. This requires deseasonalized.

Response: Thank you for these constructive suggestions.

(1) We used the seasonal Mann–Kendall test and its corresponding seasonal Sen slope, thereby

accounting for the seasonality in trend estimation.

(2) We have removed the statistical-significance arrows from Fig. 12, which now shows signed seasonal Sen slopes expressed in m decade^{-1}.

(3) Figure A2 presents decadal relative change normalized by each lake's median SDD, used as a robust measure of its typical condition. Figure A3 presents the trend-to-variability ratio, defined as the absolute decadal seasonal Sen slope divided by the IQR of the seasonally adjusted lake-level monthly series. For this comparison, seasonal adjustment was performed by subtracting the corresponding calendar-month median.

We use relative change and TVR as descriptive measures, rather than exclusion criteria or universal thresholds of ecological importance. Small changes remain in the reported distributions and are interpreted relative to typical clarity and temporal variability. These measures do not by themselves establish ecological consequences. Sect. 6.2 describes the patterns shown in Figs. A2 and A3 (line 481-496): *"Relative SDD changes exhibited pronounced spatial heterogeneity across the five lake regions (Fig. A2). Lakes with SDD increases exceeding 10% decade-1 were widely distributed across the TPL and IMXL, including many with increases exceeding 25% decade-1. Both positive and negative changes occurred in the EPL and NPML. In particular, lakes with declines exceeding 25% decade-1 were interspersed with lakes showing increasing clarity along the middle and lower Yangtze River. These patterns indicate that regional trends mask substantial differences among individual lakes. The spatial distribution of the TVR provided a complementary perspective (Fig. A3). Lakes with TVR > 1 were concentrated in the northern and eastern TPL and occurred in parts of the NPML, indicating relatively large decadal changes compared with their seasonally adjusted variability. Several large lakes in the EPL fell within the intermediate TVR classes of 0.33–1.00. Comparison with results in Fig. A2 shows that large relative increases in SDD did not consistently correspond to high TVR values; relatively high TVR values occurred in lakes with either increasing or decreasing clarity. Together, the two maps reveal substantial differences among lakes in the direction of change, their magnitude relative to typical clarity and temporal variability."*

16. Abstract and Lines 443-445, Figure 13. First, is this mean trend based on the trend on each lake, vs all the pixels? Next a bulk trend is not informative. The mean is driven by a specific region, as noted. The trends should be presented as a distribution, what is the median in each region, and the IQR. Histograms might be useful. An issue, for example with YGPL, is that the summer is not getting better, but the winter, in bulk appears to have higher SDD, clearer water.

Response: Thank you for these constructive comments.

(1) The trends in Fig. 13 were estimated from mean SDD time series obtained by pooling valid pixels within each region or across all regions. Sect. 6.2 states (line 492-493): *"Long-term seasonal trajectories of pixel-pooled mean SDD from February 2012 to February 2025 showed an increasing aggregate trend across Chinese lakes (Fig. 13)."*

(2) We agree that aggregate trends can conceal differences among lakes. Figure 14 therefore presents lake-level distributions, with one value per lake (Sect. 6.2) (line 507-514): *"To complement the pooled-pixel mean trajectories in Fig. 13, Fig. 14 presents the distribution of relative SDD change rates calculated separately for each lake. Among the 767 lakes, 533 showed positive and 234 showed negative rates, with a median of 14.9 % decade-1 (IQR: −5.2 to 37.1 % decade-1). The regional medians were the highest in the TPL (17.9 % decade-1) and YGPL (16.0 % decade-1), although 3 of the 13 YGPL lakes showed negative rates. The EPL had the lowest median (7.4 % decade-1). The IQR intervals in the TPL, NPML, IMXL, and EPL included zero, showing that positive regional medians did not represent a uniform increase across lakes."*

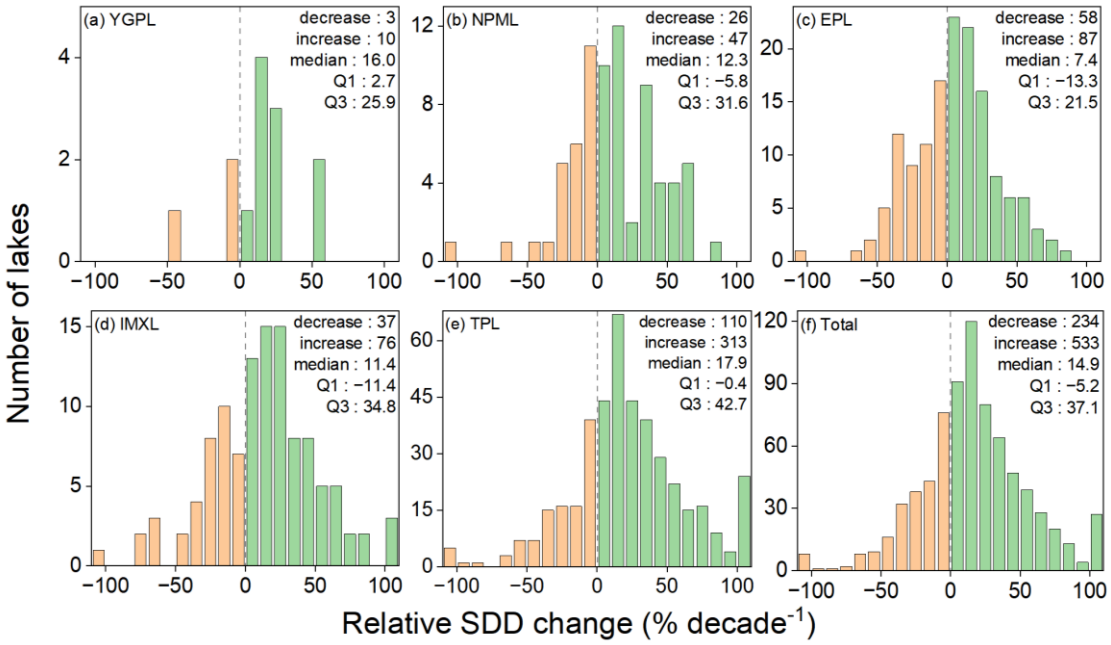


Fig. 14. Distributions of lake-level relative SDD change rates (% decade-1) across Chinese lakes from February 2012 to February 2025 for (a) YGPL, (b) NPML, (c) EPL, (d) IMXL, (e) TPL, and (f) all lakes. No valid SDD observations were available from January to April or from November to December for lakes in the NPML, IMXL, and TPL.