ESA Climate Space Tipping Elements Project

RESET lakes

Uncertainty and Error Characterization

Deliverable 4.1 Version 1.1

13 May 2026

Authors

Caren Binding

Laura Boyall

Elisa Calamita

Daniel Odermatt

Luiz Oliveira

Ann Scheliga

Iestyn Woolway

Christiane Zarfl

eawag aquatic research

Prifysgol Bangor University

Environment and Climate Change Canada / Environnement et Changement climatique Canada

Eberhard Karls Universität Tübingen

Content

1 Purpose

In this document, we provide an initial description of uncertainty handling for the scientific analysis of lake tipping points. In Section 2, we perform an assessment the quantification and sources of uncertainty of the Earth Observation (EO) products we use in the project. This review focuses initially on data services available from the ESA Climate Change Initiative (CCI) for lakes and is thus the basis for providing feedback to the corresponding CCI team. In Section 3, we describe sources and quantification of uncertainty for the lake model we use, and how the lake model and EO-based analyses will complement each other. In Section 4, we provide an overview of how we will propagate errors during the tipping point analysis. In Section 5, we present the results of a validation study of chlorophyll-*a* in the Laurentian Great Lakes. In Section 6, we cover uncertainty and error handling in a multi-variate EO case study on Great Bear Lake. Version 2 of this document will be submitted as part of the Analysis and Validation Review in 12 months.

2 EO Product Uncertainties

As summarized in the Lake CCI User Requirement Document, and as highlighted in guidance from the Global Climate Observing System (GCOS), uncertainty characterization is an important factor in generating trustworthy, reproducible lake climate data products [GCOS, Pinardi et al. 2025]. To this end, the RESETlakes project will utilize the provided Lake CCI uncertainties and describe how EO uncertainty impacts our ability to detect lake tipping points. In this section, we provide an overview of uncertainties in lake surface water temperature (LSWT, Section 1.1), chlorophyll-a and turbidity (Section 1.2), and other Lake CCI variables (Section 1.3). While we have not yet used these other Lake CCI variables (lake ice cover, water level, and extent) in our analysis, we include descriptions of the product uncertainty quantification, as we anticipate future utilization of the products.

2.1 Lake surface water temperature (LSWT)

The Lake CCI LSWT product includes a quality flag and uncertainty as data variables. From the recommendation of the Lake CCI Development Team [Simis et al., 2025], we use data with only the recommended quality levels 4 and 5. The total uncertainty values are provided for all space and time in units of Kelvin. The overall sources of uncertainty for LSWT come from three general sources: propagated uncertainty, retrieval uncertainty, and sampling uncertainty. The typical total uncertainty for LSWT ranges from 0.4-0.7 K.

The Lake CCI LSWT product is derived from an Optimal Estimator (OE) algorithm. The OE algorithm estimates LSWT from an observation of brightness temperature (BT), the prior LSWT and water vapour states, and meteorological inputs [MacCallum and Merchant 2012]. The OE algorithm implements a forward model (for LSWT, a radiative transfer model) to determine partial derivatives or sensitivities between the observations and states. The OE algorithm also includes error covariance matrices that represent forward model, observational, and prior state errors. The algorithm then calculates the new estimate of LSWT based on the difference between the current observation and prior simulated state, incorporating the sensitivities and errors. See MacCallum and Merchant 2012 for further details.

Importantly, the OE algorithm quantifies the propagated uncertainty (from error in the radiative transfer model and uncertainty in the prior estimates) and retrieval uncertainty (from sensor noise and atmospheric conditions). These uncertainties are calculated per pixel.

Sampling uncertainty comes from the aggregation of single observed pixels into a gridded product. A pixel swath may not capture an entire grid cell, or cloud cover may obscure the observation. The sampling uncertainty is higher for grid cells with low fractional coverage and/or highly variable LSWT. This uncertainty component is calculated per grid cell.

2.2 Chlorophyll-*a* and turbidity

Sources of uncertainty in chlorophyll-a and turbidity products

Uncertainty in the CCI Lakes chlorophyll-*a* and turbidity products may arise from multiple sources, primarily falling into the categories of algorithm performance and inter-sensor continuity. Atmospheric correction (particularly over optically complex inland waters), adjacency effects, and optically shallow waters, contribute significantly to the variability in algorithm performance and resulting retrieval uncertainties. Product uncertainty is also driven by the wide range of optical water types and inherent optical properties that may or may not be well represented in the algorithm calibration datasets. Additionally, limitations in cloud masking, sun glint removal, and ice or mixed pixel contamination can introduce bias. With merged multi-mission datasets, differences in the spatial, spectral, and temporal characteristics of contributing sensor data may result in temporally variable retrieval uncertainties, thereby introducing inconsistencies in time-series data. It is paramount to identify and where possible, correct for, such data artifacts prior to any tipping point analysis.

Quantification of uncertainty in chlorophyll-a and turbidity products

Uncertainties for the lake water leaving reflectance (LWLR) variables are fully based on in-situ validation. This approach quantifies the end-to-end (overall) uncertainty instead of propagating through the processing chain. It represents only uncertainty at in-situ sampling locations, which are typically in pelagic, deep water and hence exclude error sources such as adjacency effects or unmasked benthic reflectance. The validation results for chlorophyll-a and turbidity are separated by sensor, algorithm, and optical water type. For the individual wavebands within the LWLR product, validation is separated by sensor and waveband. If there are insufficient in-situ observations (~50-100 observations across a range of lakes and reflectance amplitudes) for a sensor-waveband category, then the waveband uncertainty is marked as unknown.

The uncertainty for each waveband is provided per pixel as bias and root mean squared difference (RMSD). Chlorophyll-*a* and turbidity (products derived from the wavebands) have relative uncertainty per pixel, expressed as a percentage.

Following the End-to-End Uncertainty Budget [Simis et al. 2025], we use the lwlr_poor_consistency and lwlr_low_consistency quality flags to remove poor data. These flags are especially important for LWLR retrievals in locations where partial lake ice cover may contaminate the signal.

Further assessment of retrieval uncertainties was conducted leveraging extensive in situ water quality datasets in the Laurentian Great Lakes to quantify bias and percentage error across the entire period of the CCI Lakes products (See section 5).

2.3 Other Lake CCI variables

We have not yet focused on lake water level (LWL) or lake water extent (LWE), variables available within Lake CCI dataset. From the End-to-End ECV Uncertainty Budget [Simis et al. 2025], the recommendations for LWL are to use the quality flag which is based on the uncertainty (good quality has an uncertainty below 10 cm, medium quality below 30 cm). For LWE, uncertainty is provided in square kilometers, and the quality flag is based on the ratio between the uncertainty and the total lake area: good is below 5% of total area and medium is below 10% of total area.

For lake ice cover (LIC), uncertainty is from a classification error for ice (2.23%), water (0.83%), and clouds (3.07%). These values are only indicative of overall performance of the product, not spatially resolved uncertainty estimates.

3 Model uncertainties

Ensemble modelling approaches are commonly used to better characterise uncertainty in environmental simulations. Ensemble frameworks are widely used across climate science, including General Circulation Model (GCM) experiments, regional climate modelling, and more recently have been applied to exploring process-based lake models. Rather than relying on a single model realisation, ensembles create a suite of simulations that contain variants of or perturbations to a realisation, such as: different model structures, parameterisations, assumptions, or input forcings. This approach provides a range of plausible outcomes and allows uncertainties associated with variation or perturbation to be quantified [Trolle et al., 2014; Woolway et al., 2021; Moore et al., 2021].

A common feature of ensemble simulations is a wide ensemble spread in model outputs, reflecting influence of model structure, forcing data, and/or parameterisation schemes on the simulations. While this spread can lead to reduced coherence among individual model simulations, it is also a key indicator of uncertainty within the modelling framework. Consequently, ensemble statistics such as the multi-model mean or median are often used to represent the central tendency of model projections and have been shown to provide more robust estimates of system behaviour than individual models alone. However, the full ensemble distribution remains important to consider, as the spread captures the range of possible system responses and can provide insight into extreme or low-probability outcomes.

Within the RESETlakes analysis, the ensemble contains the simulations from the lake model (ALBM) with input forcings produced by different individual global climate models (e.g., IPSL-CM6A-LR, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2-0, and UKESM1-0-LL). As a result, the ensemble spread primarily reflects uncertainty associated with the driving climate inputs rather than uncertainty in lake model formulation.

3.1 Sources of uncertainty in model simulations

Differences among the simulations can arise because the global climate models used as an input to the lake model vary in their representation of atmospheric circulation, temperature, precipitation, and radiation fields, which in turn influence the lake processes and characteristics being modelled (e.g. lake ice cover or thermal properties). Consequently, even when using the same lake model, the resulting lake simulations can exhibit substantial variability depending on the climate model forcing. Analysing the ensemble therefore allows the sensitivity of lake responses to different climate trajectories to be assessed.

3.2 Quantification of uncertainty in model simulations

To properly quantify the ensemble spread and assess the uncertainties within the ensemble, we will not only be looking at the multi-model mean and median but also conducting some simple descriptive statistics including calculating the interquartile ranges, standard deviation percentage contribution of variability for each simulation. Additionally, we will conduct more detailed statistical tests including the analysis of variance (ANOVA) for each of the lake variables and main climate drivers (e.g. temperature and precipitation).

3.3 Validation of model simulations

The availability of the high-resolution lake cci data provides a unique opportunity to validate the ALBM simulations with specific lakes (e.g. explore the temporal evolution of a particular lake such as Great Bear), or spatially (e.g. explore the spatial trends of global lakes) for multiple lake variables. For this we will be quantitatively validating the full simulation uncertainty against the lake cci data and producing a

series of plots, such as target diagrams and Taylor diagrams to explore the errors between true and predicted (RMSE), the standard deviations and the correlations between true and predicted.

4 Error propagation

Given the uncertainties in the EO products and model, in this section, we describe how we move from pixel-by-pixel uncertainties to uncertainty in tipping point detection. The main considerations are spatial aggregation, temporal aggregation, and uncertainty quantification in the regime shift detection.

The first step in error propagation is to convert pixel-wise uncertainties into spatiotemporally aggregated uncertainties. We use the standard deviation (σ) of the pixels and express this as the normalized standard deviation ($\sigma/\sqrt{n_{px}}$, where n_{px} is the number of observations), as in the Great Bear case study in Section 6.

Once we have lake-aggregated uncertainties, we will determine the uncertainty in the tipping point detection with a cross-validation and bootstrapping. To tune parameters for multi-variate breakpoint detection (see RESETlakes Methodology), we will use blocked cross-validation [Bergemir and Benítez 2012]. When using cross-validation with changepoint detection, we can improve validation performance by using absolute error loss [Pein and Shah 2025]. Finally, we will use a bootstrapping procedure to produce a confidence interval or standard deviation to describe the uncertainty in the regime shift [Antoch et al. 1995, Gilarranz et al. 2022].

Independent validation of tipping point detection remains an area of development. For individual lakes, we can perform in-situ validation (such as Section 5). But for more comprehensive spatial coverage, we will compare uncertainties in regime shifts detected by EO and by model reanalyses. Additionally, while multivariate EO analyses (such as Section 6) do not provide a fully independent validation assessment of tipping point uncertainty, we do expect a correlation or cascade between lake variables during a regime shift.

5 Validation study: Laurentian Great Lakes

We compare performance over the Laurentian Great Lakes for the chlorophyll-*a* products from Lake CCI and Ocean Colour CCI (OC CCI). We analyze retrieval uncertainties for the two products against in-situ observations to assess the overall suitability for trend and regime shift analysis.

The OC CCI chlorophyll-*a* product showed overall good agreement with the in-situ data but underestimated high chlorophyll-*a* and overestimated low chlorophyll-*a* values. The Lake CCI dataset showed good agreement for high chlorophyll ranges, underestimated low chlorophyll-*a* values, and displayed changes in performance between sensors.

5.1 In-situ dataset

The Environment and Climate Change Canada (ECCC) project collaborator has access to in-situ chlorophyll-*a* data from a regular monitoring program of the Laurentian Great Lakes run by the US Environmental Protection Agency (EPA) and ECCC Water Quality Monitoring and Surveillance (WQMS) program. The two programs provide regular in situ observations across all five lakes, going back to 1998. We use in-situ samples taken at depths of less than 5 meters. Figure 1 shows a map of the lakes and sampling locations.

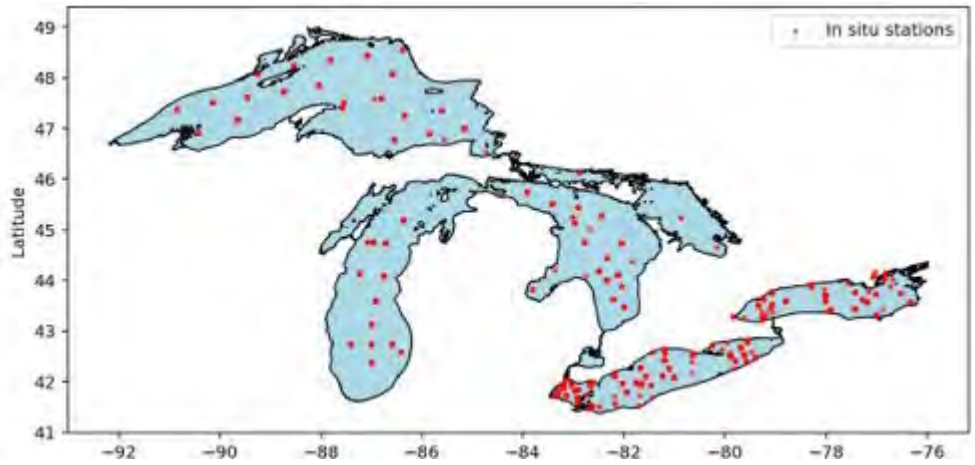

Figure 1: Map of Laurentian Great Lakes with in-situ sampling stations. From West to East, the lake names are: Superior, Michigan, Huron, Erie, and Ontario.

Over the full OC CCI timeframe (1998-2024), there are 2551 matchups between the satellite product and in-situ samples. For the full Lake CCI v3.0 timeframe (2002-2023), there are 1116 matchups. The distribution of in-situ sampling by year and lake is shown in Figure 2 for OC CCI and in Figure 3 for Lake CCI. No in-situ samples were taken in 2020 due to pandemic-related fieldwork restrictions.

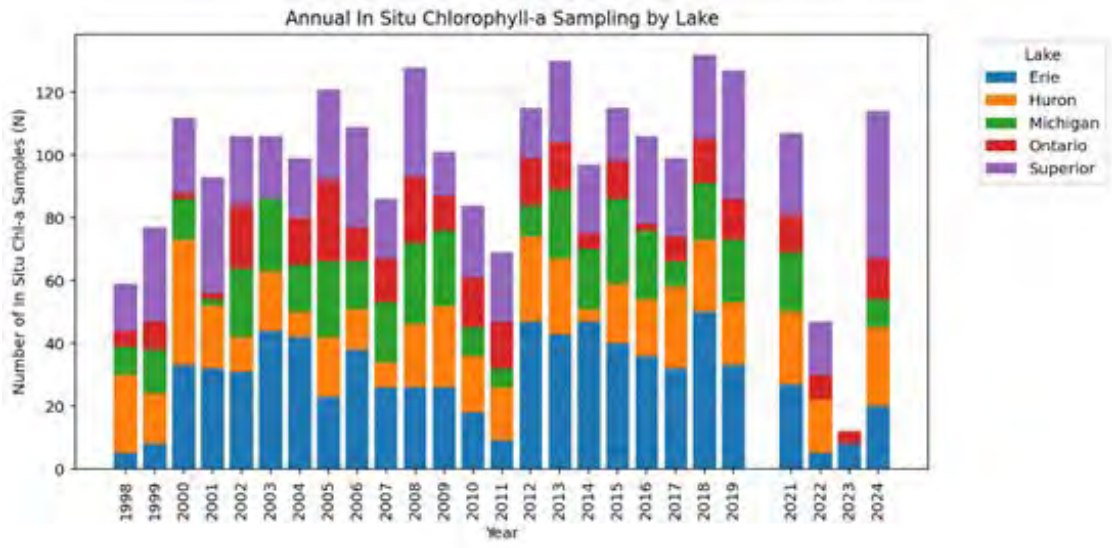

Figure 2: Number of in-situ samples by year and by lake for OC CCI.

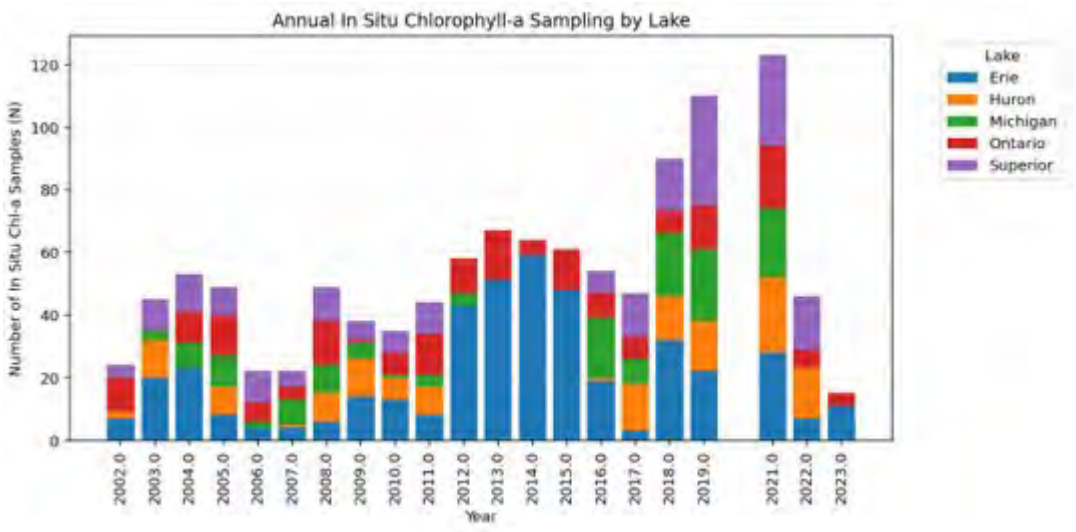

Figure 3: Number of in-situ samples by year and by lake for Lake CCI.

5.2 Performance by sensor period

The OC CCI product uses several overlapping satellite sensors to estimate chlorophyll-*a*. In contrast, the Lake CCI product uses sensors sequentially, as shown in Figure 4. For Lake CCI, we break the data record into 3 periods P1, P2, and P3 to correspond with the MERIS, MODIS, and Sentinel-3 OLCI periods, respectively. For OC CCI, we split the data record every year a new sensor is incorporated, resulting in P1 (SeaWiFS only), P2 (SeaWiFS, MERIS and MODIS), P3 (MODIS and VIIRS), and P4 (OLCI and VIIRS).

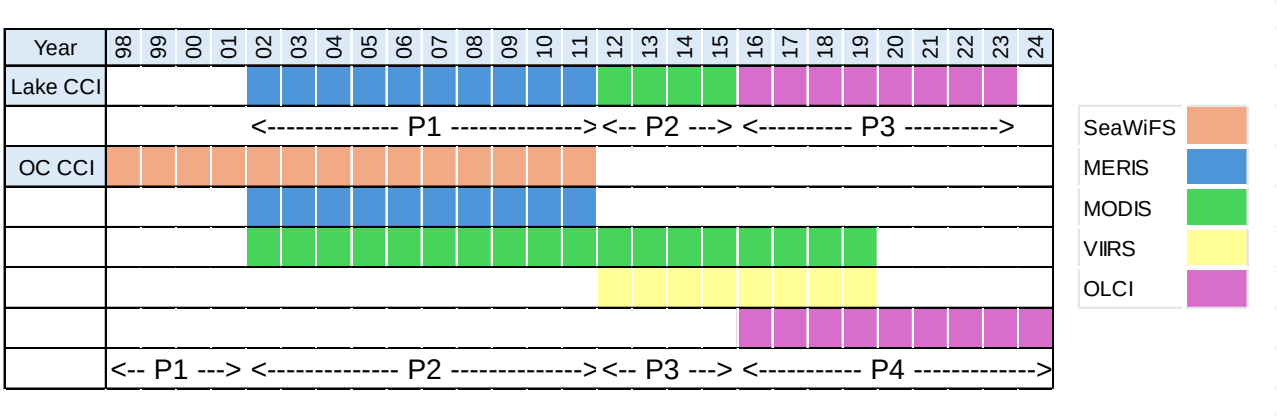

Figure 4: Timeline of satellite sensor coverage for Lake CCI and OC CCI.

Figure 5 shows the performance of Lake CCI against the in-situ samples split by sensor period. Notably, for Lake CCI v3.0, Lake Superior, Michigan, and Huron no longer contain data during the MODIS sensor period (P2) from 2012 to 2016 (Figure 3) because the selection criteria to determine which lakes are suitable for the MODIS sensor changed. By excluding the oligotrophic Great Lakes, the data in the MODIS period for Lake CCI, do not contain many low chlorophyll-*a* values, as shown in the middle panel of Figure 5. When comparing the performance across sensor periods for Lake CCI, we see a moderate difference in mean absolute percentage error (MAPE) during the MODIS period, and an extreme difference in bias (Figure 6).

The shifts in bias and MAPE may be detrimental to regime shift analysis, as these abrupt statistical changes imitate physical tipping points in the lake system. Normalization techniques may alleviate the inter-sensor shifts, but further consideration would be needed on how empirical post-processing of the chlorophyll-*a* signal would impact the representation of relative uncertainty.

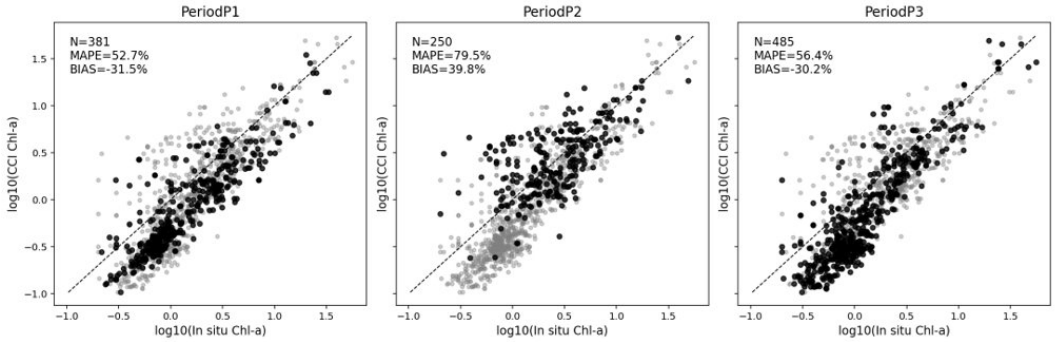

Figure 5: Performance of each sensor period (black points) against the overall performance (grey points) for Lake CCI. From left to right, the columns are P1 (MERIS), P2 (MODIS), P3 (OLCI).

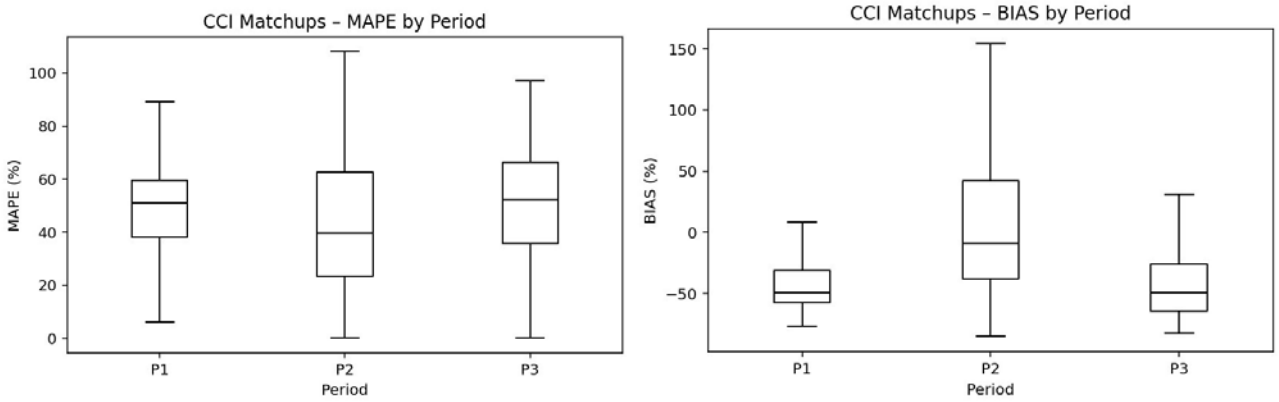

Figure 6: Mean absolute percentage error (MAPE, left panel) and bias (right panel) for each sensor period in Lake CCI.

For OC CCI, the intersensory period is more stable, as shown in the performance plots by sensor (Figure 7) and the MAPE and bias (Figure 8). The P1 period from 1998-2002 with only the SeaWiFS sensor performs more poorly than the other periods, which would be an issue for tipping point detection. However, this period is at the beginning of the time frame, so simply cropping the data is an easier and more complete solution than attempting any post-processing and would have no impact on the uncertainty quantification.

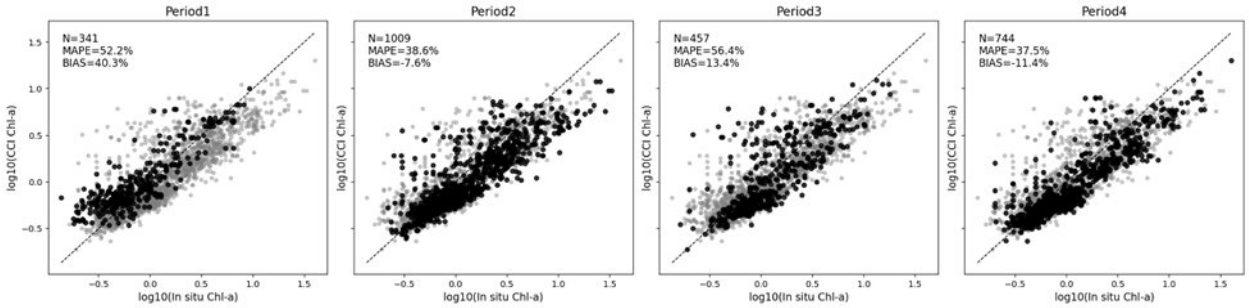

Figure 7: Performance of each sensor period (black points) against the overall performance (grey points) for OC CCI. From left to right, the columns are P1, P2, P3, and P4.

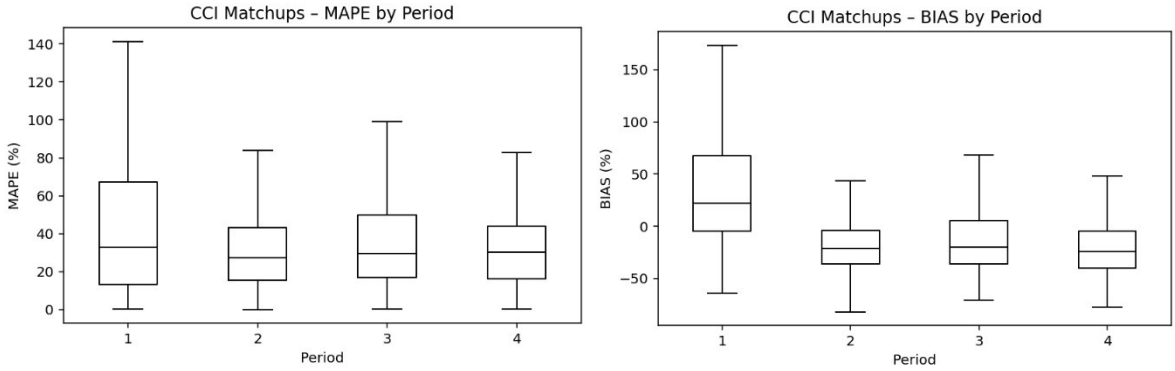

Figure 8: Mean absolute percentage error (MAPE, left panel) and bias (right panel) for each sensor period in OC CCI.

5.3 Performance by lake

Performance between lakes varied. Generally, Lakes Erie and Ontario are more turbid, with higher chlorophyll-*a* concentration ranges, while Lakes Superior, Huron, and Michigan are clearer and more oligotrophic. The performance plots for each lake with each dataset are shown in Appendix A.

6 Case study: Great Bear Lake

In this section, we describe the uncertainties associated with the temporal and spatial averaging procedures applied to the lake surface water temperature (LSWT) and the chlorophyll-*a* (chl-*a*) products. Uncertainty was quantified using the standard deviation (σ) and its contribution to the averaged value, expressed as the normalized standard deviation ($\sigma/\sqrt{n_{px}}$, where n_{px} is the number of observations).

The lake surface water temperature (LSWT) input dataset corresponds to the gap-filled product developed within the ESA LakeCREST project. As such, spatial uncertainties were not reassessed in this study and should be referred to the original product documentation. We used the chlorophyll-*a* (chl-*a*) dataset from Lake CCI (v2.1); therefore, both spatial and temporal uncertainties are assessed.

6.1 Anomalous versus non-anomalous years

June was selected as the representative month, as it exhibited significant differences between anomalous and non-anomalous years (cf. RESETlakes Methodology v1.0). Although the LSWT time series was truncated in the main analysis to match the chl-*a* temporal coverage, the uncertainty assessment presented here considers the full available LSWT time series.

Figure 9 shows the spatially and temporally averaged LSWT, along with its associated standard deviation, for the month of June across the time series. Anomalous years characterized by mixing events (red) generally exhibit higher standard deviations than non-anomalous years. This increased variability is hypothesized to result from accelerated ice-off and lake warming, leading to greater day-to-day temperature fluctuations. In contrast, non-anomalous years likely experience more gradual and stable warming, resulting in a more homogeneous temporal distribution.

Figure 10 presents the normalized standard deviation ($\sigma/\sqrt{n_{px}}$) associated with the temporal averaging procedure. This metric quantifies the contribution of variability to the averaged LSWT values used in the statistical analyses. Overall, the contribution of the standard deviation is low, ranging from 0 to 0.2 °C throughout the entire time series, indicating that temporal averaging introduces limited uncertainty into the derived LSWT metrics.

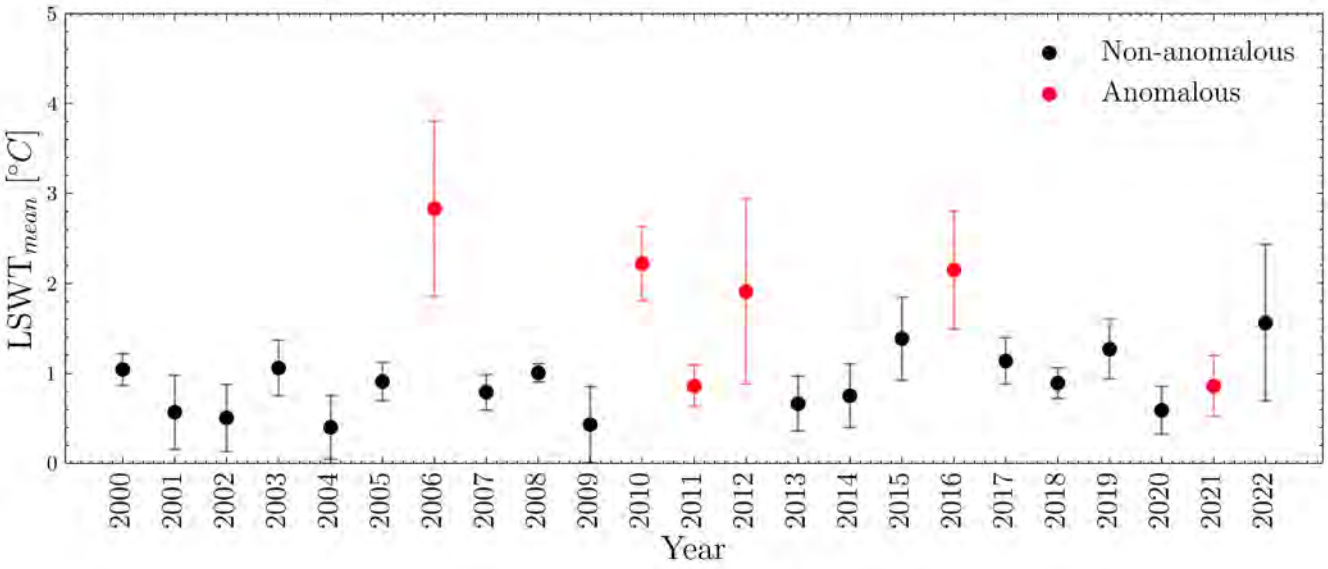

Figure 9: Spatially and temporally averaged lake surface water temperature (LSWT) and its temporal standard deviation for the month of June across the time series. Mixing anomalous years are highlighted in red, while non-anomalous years are shown in black. The y-axis limits differ from those in the Methodology document to enhance the visibility of the standard deviation.

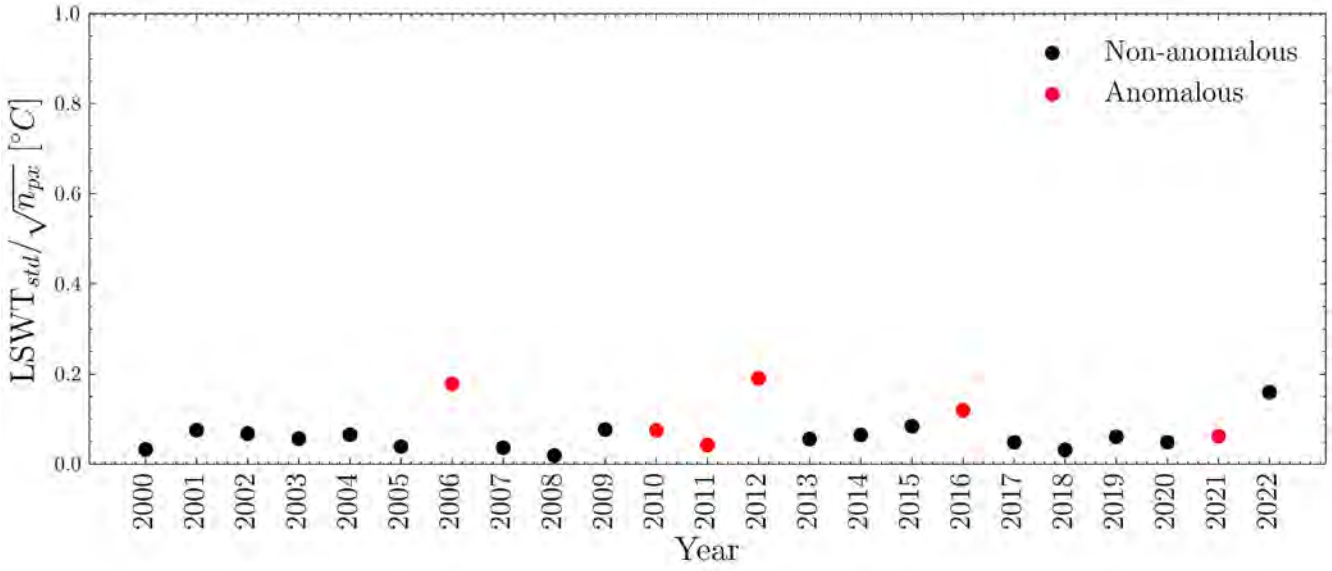

Figure 10: Normalized temporal standard deviation of lake surface water temperature (LSWT) for the month of June across the time series. Mixing anomalous years are highlighted in red, while non-anomalous years are shown in black.

Figure 11 shows the spatially and temporally averaged chl-*a* concentration along with its associated temporal standard deviation. Anomalous years exhibit lower temporal variability compared to non-anomalous years. The normalized temporal standard deviation remains consistently low across the time series, ranging from 0 to 0.2 mg/m^3 (Figure 12), indicating a limited contribution of temporal variability to the averaged chl-*a* values.

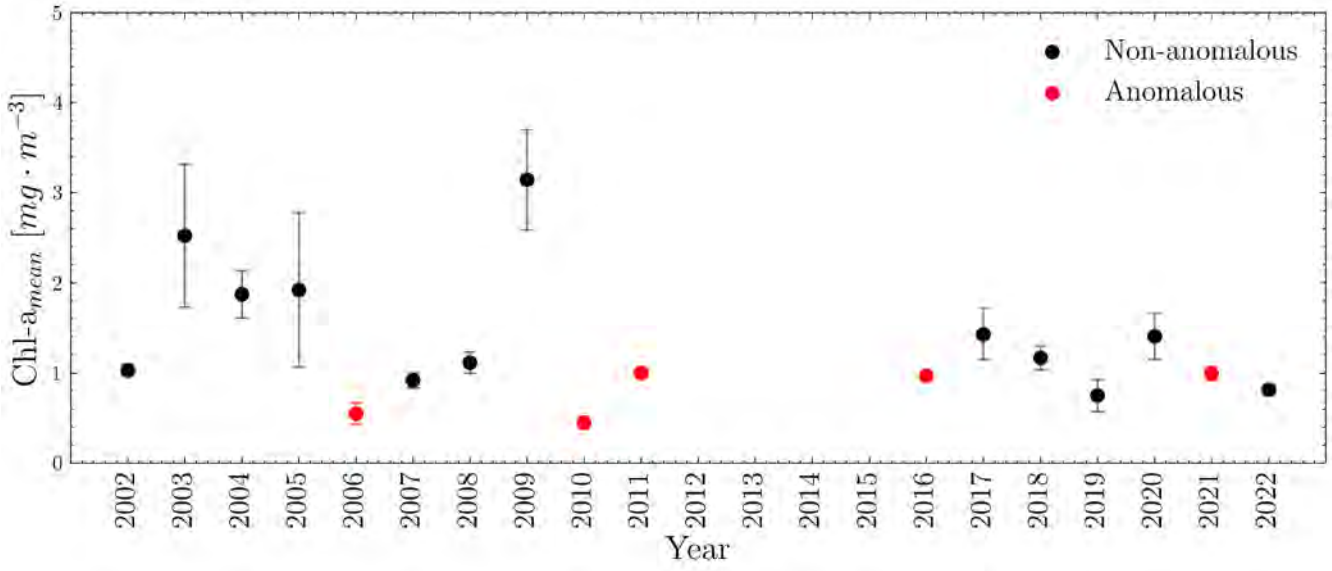

Figure 11: Spatially and temporally averaged chlorophyll-a (chl-a) concentration and its temporal standard deviation for the month of June across the time series. Mixing anomalous years are highlighted in red, while non-anomalous years are shown in black. The y-axis limits differ from those in the Methodology document to enhance the visibility of the standard deviation.

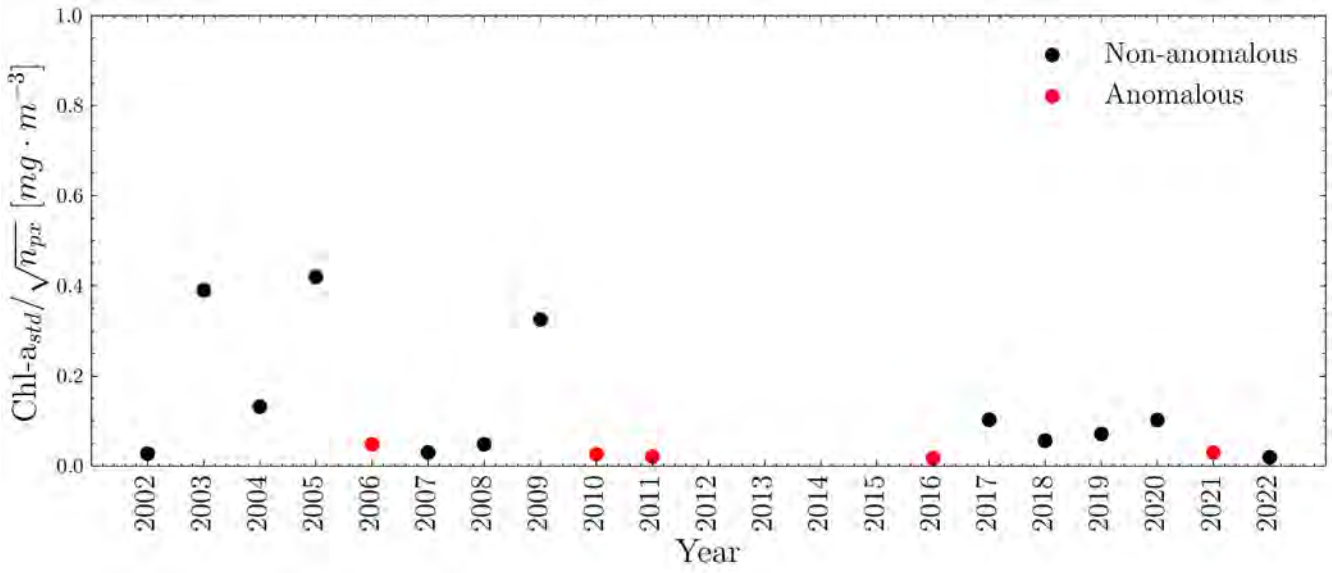

Figure 12: Normalized temporal standard deviation of chlorophyll-a (chl-a) concentration for the month of June across the time series. Mixing anomalous years are highlighted in red, while non-anomalous years are shown in black.

6.2 Time series analysis

Figure 13 presents the time series of the standard deviation associated with the spatial averaging procedure. Higher standard deviation values are observed at the beginning and end of the time series, likely reflecting the influence of ice cover and its interaction with lake bathymetry.

The contribution of spatial variability to the averaged chl-*a* concentration is shown in Figure 14, with values ranging from 0 to 5 mg/m^3. The lake is typically fully frozen from November to May [Kang et al., 2012; Carmack et al., 2024]. From Figure 14, the highest spatial variability is in early June and late October, consistent with lake ice-off and lake ice-on windows, respectively. Beyond these ice-on and ice-off transitions, approximately from late June to early October, the spatial variability contribution is substantially reduced, generally ranging between 0 and 1 mg/m^3.

7 Conclusion

Overall, the goal of the uncertainty and error characterization is to ensure that regime shift estimation incorporates the underlying EO uncertainty and model ensemble spread. The Lake CCI data products contain uncertainty and quality flags that enable us to determine how EO uncertainty propagates to and impacts our ability to detect lake tipping points. The model ensemble provides us with an estimate of the sensitivity of regime shifts and early warning indicators to climate forcings.

In the next stage of the project, we will develop comparisons between EO-based vs modelled early warning indicators and regime shifts. With these comparisons, we can also gain insight into detectability levels for EO-based regime shift detection.

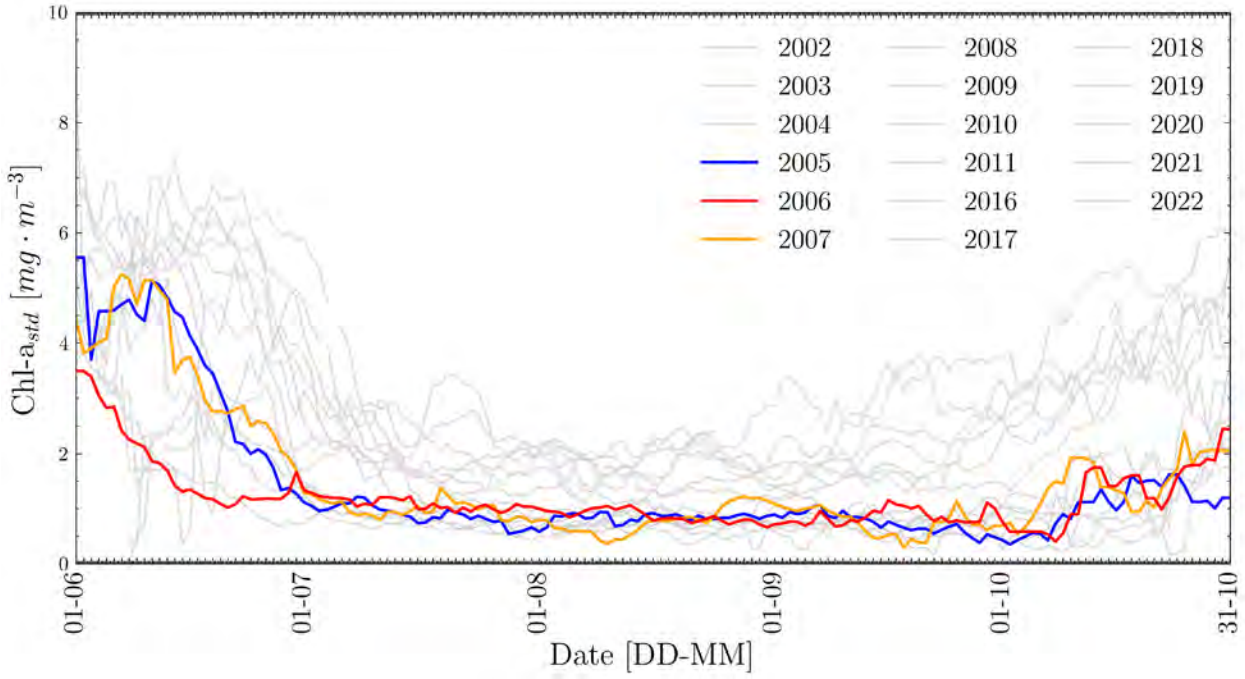

Figure 13: Time series of the spatial standard deviation of chlorophyll-a (chl-a) concentration for the entire lake, computed using an 8-day moving window. The red line represents an anomalous year, while the blue and orange lines denote the preceding and subsequent years, respectively.

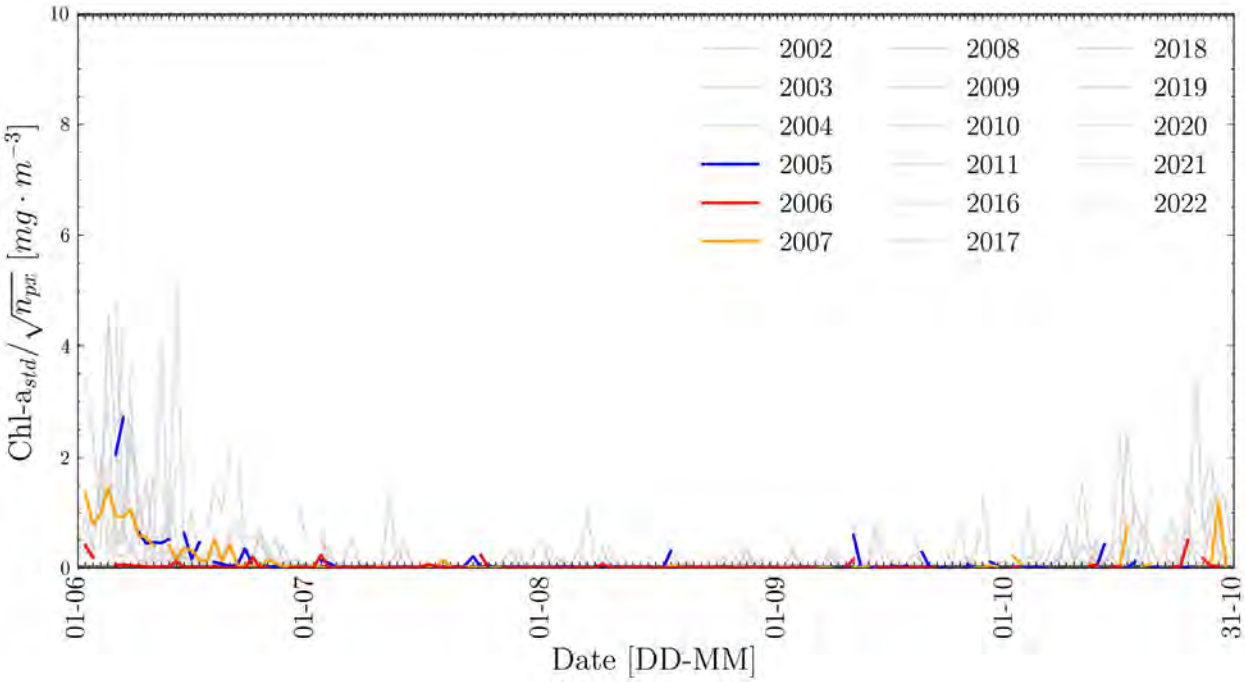

Figure 14: Normalized spatial standard deviation of chlorophyll-a (chl-a) concentration for the entire lake, computed using an 8-day moving window. The red line represents an anomalous year, while the blue and orange lines denote the preceding and subsequent years, respectively.

8 References

Antoch, J., Hušková, M., & Veraverbeke, N. (1995). Change-point problem and bootstrap. Journal of Nonparametric Statistics, 5(2), 123–144. https://doi.org/10.1080/10485259508832639

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, Data Mining for Software Trustworthiness, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Gilarranz, L. J., Narwani, A., Odermatt, D., Siber, R., & Dakos, V. (2022). Regime shifts, trends, and variability of lake productivity at a global scale. *Proceedings of the National Academy of Sciences*, *119*(35), e2116413119. https://doi.org/10.1073/pnas.2116413119

Global Climate Observing System programme (GCOS), (GCOS 244), 2022a. The GCOS 2022 Implementation Plan. https://library.wmo.int/index.php?lvl=notice_display&id=22134.

MacCallum, S. N., & Merchant, C. J. (2012). Surface water temperature observations of large lakes by optimal estimation. *Canadian Journal of Remote Sensing*, *38*(1), 25–45. https://doi.org/10.5589/m12-010.

Moore, T.N., Mesman, J.P., Ladwig, R., Feldbauer, J., Olsson, F., Pilla, R.M., Shatwell, T., Venkiteswaran, J.J., Delany, A.D., Dugan, H., Rose, K.C., and Read, J.S. 2021. LakeEnsemblR: An R package that facilitates ensemble modelling of lakes. Environmental Modelling & Software. 143. 105101. https://doi.org/10.1016/j.envsoft.2021.105101

Pein, F., & Shah, R. D. (2025). Cross-validation for change-point regression: Pitfalls and solutions. *Bernoulli*, *31*(1). https://doi.org/10.3150/24-BEJ1732

Pinardi M., et al. 2024. Lakes Climate Change Initiative User Requirement Docurment (URD), version 3.0.1. European Space Agency.

Simis, S., Carrea, L., Calmettes, B., Crétaux, J-F., Duguay, C., C. Fatras, Liu, X., Mangili, A., Yésou, H. 2025. Lakes Climate Change Initiative End-to-End ECV Uncertainty Budget (E3UB), version 3.0.0. European Space Agency. DOI: 10.5281/zenodo.17288518

Trolle, D., Elliott, J.A., Mooij, W.M., Janse, J.H., Bolding, K., Hamilton, D.P., and Jeppesen, E. 2014. Advancing projections of phytoplankton responses to climate change through ensemble modelling. Environmental Modelling & Software. 61. 371-379. https://doi.org/10.1016/j.envsoft.2014.01.032

9 Appendix A: Correlation of chlorophyll-*a* products for the Laurentian Great Lakes

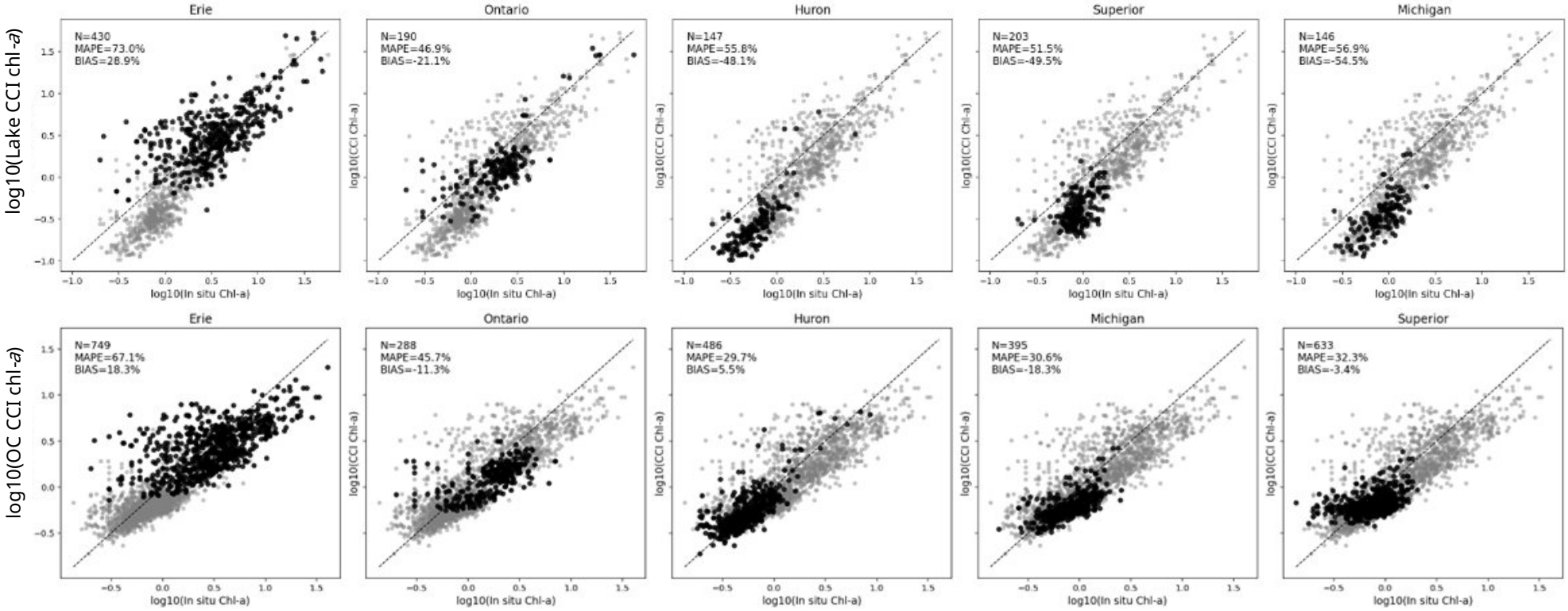

Figure 15: Performance of each lake (black points) against the overall performance (gray points) for Lake CCI (top row) and OC CCI (bottom row) against in-situ measurements. From left to right, the columns are Lake Erie, Ontario, Huron, Superior, and Michigan.