

Response to Reviewer

We thank the reviewer for their positive assessment and constructive comments, which have helped us further improve the manuscript. We have addressed each point as detailed below. In the following, the reviewer's comments are reproduced in italics, followed by our response and the corresponding changes in the manuscript. Line numbers refer to the revised manuscript.

L89: Synechococcus, Trichodesmium, and Prochlorococcus must be italicized. Please check this throughout the entire manuscript.

We thank the reviewer for pointing this out. We have carefully checked the italicisation of all genus and species names throughout the entire manuscript and corrected the inconsistencies.

L90: "Kodama and Taniuchi" should be "Kodama et al."

Corrected.

Changes in manuscript: The in-text citation has been changed to "Kodama et al., 2022" (L89).

L155 & L157: Chlorophyll a: Pigment names (e.g., chlorophyll) should start with lowercase letters as they are not proper nouns. Also, the "a" must be italicized.

We agree and have corrected this throughout the manuscript.

Changes in manuscript: All pigment names now start with lowercase letters, and the "a" (and "b", "c") suffixes are italicised, both in the Table 1 caption (L155–157) and throughout the main text (e.g., L53, L155–157, L236–237, L240).

L156: Use the full names for first mention: 19'-hexanoyloxyfucoxanthin (19'-HF) and 19'-Butanoyloxyfucoxanthin (19'-BF).

We have updated the Table 1 caption and the first mention in the main text accordingly.

Changes in manuscript: Full names with abbreviations now provided at first mention in Table 1 caption (L156) and in Section 2.3 (L238–241).

L240: As this is the first mention of pigment names in the main text, please define all abbreviations here (similar to Table 1).

We have expanded Section 2.3 to define all pigment abbreviations at their first mention in the main text.

Changes in manuscript: Section 2.3 (L238–241) now reads: "peridinin (Peri), 19'-butanoyloxyfucoxanthin (19'-BF), fucoxanthin (Fuco), 19'-hexanoyloxyfucoxanthin (19'-HF), alloxanthin (Allo), zeaxanthin (Zea), divinyl-chlorophyll-*a* (DV-Chl*a*), divinyl-chlorophyll-*b* (DV-Chl*b*), and chlorophyll *b* (Chl*b*), with total chlorophyll *b* (TChl*b*) defined as the sum of DV-Chl*b* and Chl*b*."

L254: Please specify which software was used for the statistical analyses.

We have added a detailed description of the software and packages used for the statistical analyses.

Changes in manuscript: Section 2.4 (L269–272) now specifies: "All analyses were carried out in R (v4.3.1; R Core Team) and Python (v3.14). Data processing, quality control, and the correspondence-analysis clustering were performed in R using the tidyverse, vegan, and FactoMineR packages, with figures produced via ggplot2, patchwork, and ggpubr. The machine-learning classification was implemented in Python with scikit-learn (v1.8.0) and imbalanced-learn (v0.14.1), supported by pandas, NumPy, Matplotlib, and seaborn."

L299: Section 2.5.2 is missing or incorrectly referenced.

We thank the reviewer for catching this. The cross-reference was a remnant from an earlier section numbering. The manuscript structure now goes from Section 2.4 to subsections 2.4.1, 2.4.2 and 2.4.3, and we have verified all internal cross-references.

Changes in manuscript: All internal cross-references to sections have been verified and corrected (notably L300 referring to Section 2.4.2).

L324: The calculation for "impurity" is unclear. Please provide a brief description or formula.

We have added an explicit formula (Gini index, equation 3) and a brief description of how the impurity decrease is calculated and attributed to each descriptor.

Changes in manuscript: Section 2.4.3 (L344–357) now reads: "Here, impurity is quantified by the Gini index (equation 3), where p_k is the proportion of samples of class k at node t . At each split (i.e., branch separation), the decrease in impurity (weighted by the fraction of samples reaching the node) is attributed to the descriptor used. The importance of each descriptor is then obtained by averaging these decreases over all trees in the ensemble, so that larger values indicate more influential descriptors." Equation 3 (the Gini index formula) has been added at L349.

L416 & L542: Use the abbreviations for fucoxanthin and alloxanthin consistently.

We have homogenised the use of pigment abbreviations throughout the manuscript: "Fuco" and "Allo" are now used systematically after their first definition in Section 2.3.

L500: Are the y-axis labels in Figure 4 correct? As currently shown, the precision scores appear extremely low. Please verify. Please check Figure 6 as well.

We thank the reviewer for catching this error. In the previous version, the y-axis of Figure 4 ranged from 0 to 1 while still being labelled "Precision (%)" and "Recall (%)", which was inconsistent and made the values appear to be on the wrong scale. The figures have now been corrected so that the y-axis ranges from 0 to 100 and the axis labels correctly read "Precision score (%)" and "Recall score (%)". We have also verified Figure 6, where the y-axis is now correctly labelled "Recall score (%)" on a 0–100 scale.

Changes in manuscript: Figures 4 and 6 have been regenerated with corrected y-axes (now 0–100, in percent), consistent with the values reported in Appendix Table A1.

cp should be c_p .

Corrected. All instances of "cp" and "bbp" now appear with proper subscripts (c_p , b_{bp}) throughout the manuscript, figures, and tables.

Changes in manuscript: Subscript formatting applied consistently throughout the main text, figure captions, axis labels (Figs. 4–6), and Tables 2 and A1.

Here, the authors used a paired t-test, but they should adjust the p-value using the Bonferroni method or use ANOVA.

We thank the reviewer for raising this point. After verification, the statistical test actually used in this study is a Kruskal–Wallis test, not a paired t-test as was incorrectly stated in the previous version of the manuscript. The Kruskal–Wallis test is a non-parametric test appropriate for comparing more than two independent groups (here, the six sensor configurations within each cluster, and the three cluster-number configurations within each sensor configuration), and it provides a single global test per group, so no multiple-comparisons correction is required at this level. We have corrected the description in the figure captions accordingly.

Changes in manuscript: The figure captions of Figs. 4 (L527–528) and 6 (L651–653) now correctly describe the test used: "differences among the six sensor configurations were assessed with a Kruskal–Wallis test, with significance indicated by stars (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$)."

L530 (Figure 5): There are rendering issues in my environment (see the red-circled parts in the attached). Please fix these. Additionally, since these are mean values, please add error bars (standard deviations).

We have regenerated Figure 5 to fix the rendering issues affecting the x-axis labels and the y-axis tick at zero. Error bars representing \pm one standard deviation across the 20 cross-validation runs have been added, and we have updated the figure caption accordingly. The methodology for computing these error bars has also been added to Section 2.4.3.

Changes in manuscript: Figure 5 has been regenerated with clean x-axis labels and added error bars. The caption (L547–550) now reads: "Error bars indicate standard deviation of the mean decrease impurity across the 20 cross-validations." The corresponding explanation has been added to Section 2.4.3 (L351–357).

Before the next submission, please check the entire manuscript. The contents are good but there are several errors in the text. I pointed out that I noticed, but this is the author's responsibility, not a reviewer's.

We thank the reviewer for this fair remark. We have carefully proofread the entire manuscript to correct typographical errors, harmonise spelling and abbreviation usage, verify all internal cross-references, and ensure that every in-text citation has a matching entry in the reference list (and vice versa).

We thank the reviewer once again for their thoughtful and constructive comments, which have substantially improved the clarity and quality of the manuscript.