Supplement of Biogeosciences, 22, 5349–5385, 2025
https://doi.org/10.5194/bg-22-5349-2025-supplement

Biogeosciences Open Access EGU

Supplement of

Optimization of the World Ocean Model of Biogeochemistry and Trophic dynamics (WOMBAT) using surrogate machine learning methods

Pearse J. Buchanan et al.

Correspondence to: Pearse J. Buchanan (pearse.buchanan@csiro.au)

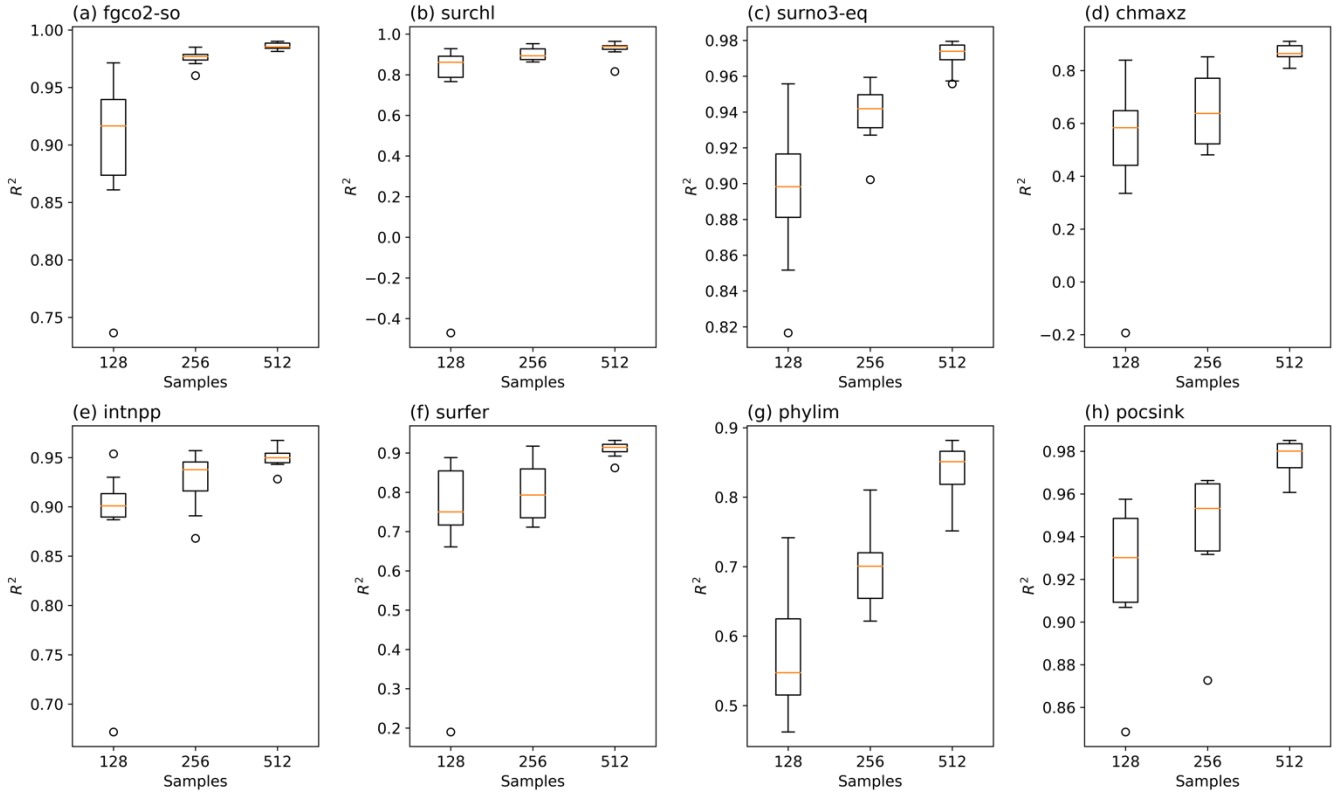


Figure S1. Sensitivity of the Gaussian Process Regression (GPR) model to true model output sample size during the sensitivity analysis. Performance of the GPR is measured as the R^2 when comparing predicted versus definite root mean square errors (RMSE). Variables are (a) downward flux of CO_2 in the Southern Ocean south of 30ºS, (b), surface chlorophyll concentrations, (c) surface nitrate concentrations in the equatorial zone between 20ºS and 20ºN, (d) the depth of the chlorophyll maximum, (e) depth-integrated net primary production, (f) surface concentrations of dissolved iron, (g) the primary limiting nutrient (iron or nitrogen) for phytoplankton growth, and (h) the sinking flux of particulate organic carbon.

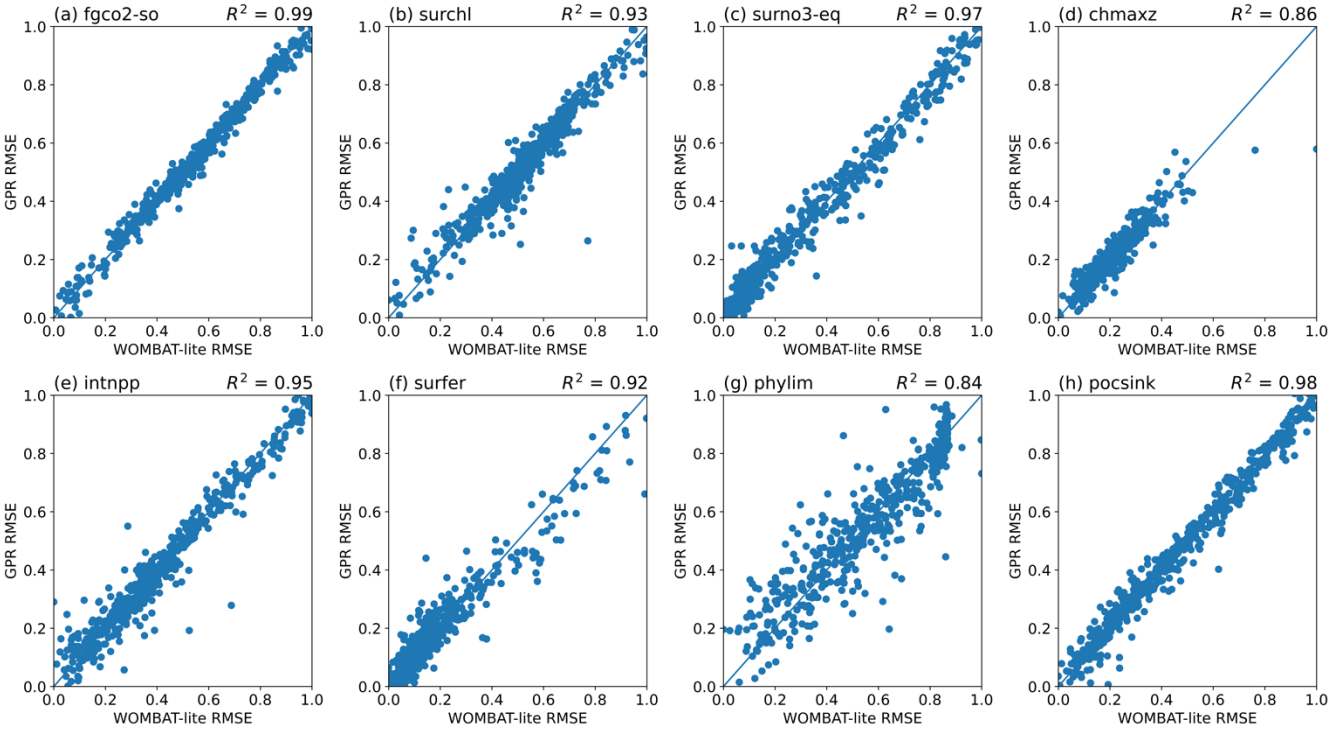


Figure S2. Accuracy of the Gaussian Process Regression (GPR) model to true model output (512 samples) during the sensivity analysis. Performance of the GPR is measured as the R^2 when comparing predicted versus definite root mean square errors (RMSE) and is shown above each panel for each target variable. Variables are (a) downward flux of CO_2 in the Southern Ocean south of 30ºS, (b), surface chlorophyll concentrations, (c) surface nitrate concentrations in the equatorial zone between 20ºS and 20ºN, (d) the depth of the chlorophyll maximum, (e) depth-integrated net primary production, (f) surface concentrations of dissolved iron, (g) the primary limiting nutrient (iron or nitrogen) for phytoplankton growth, and (h) the sinking flux of particulate organic carbon.

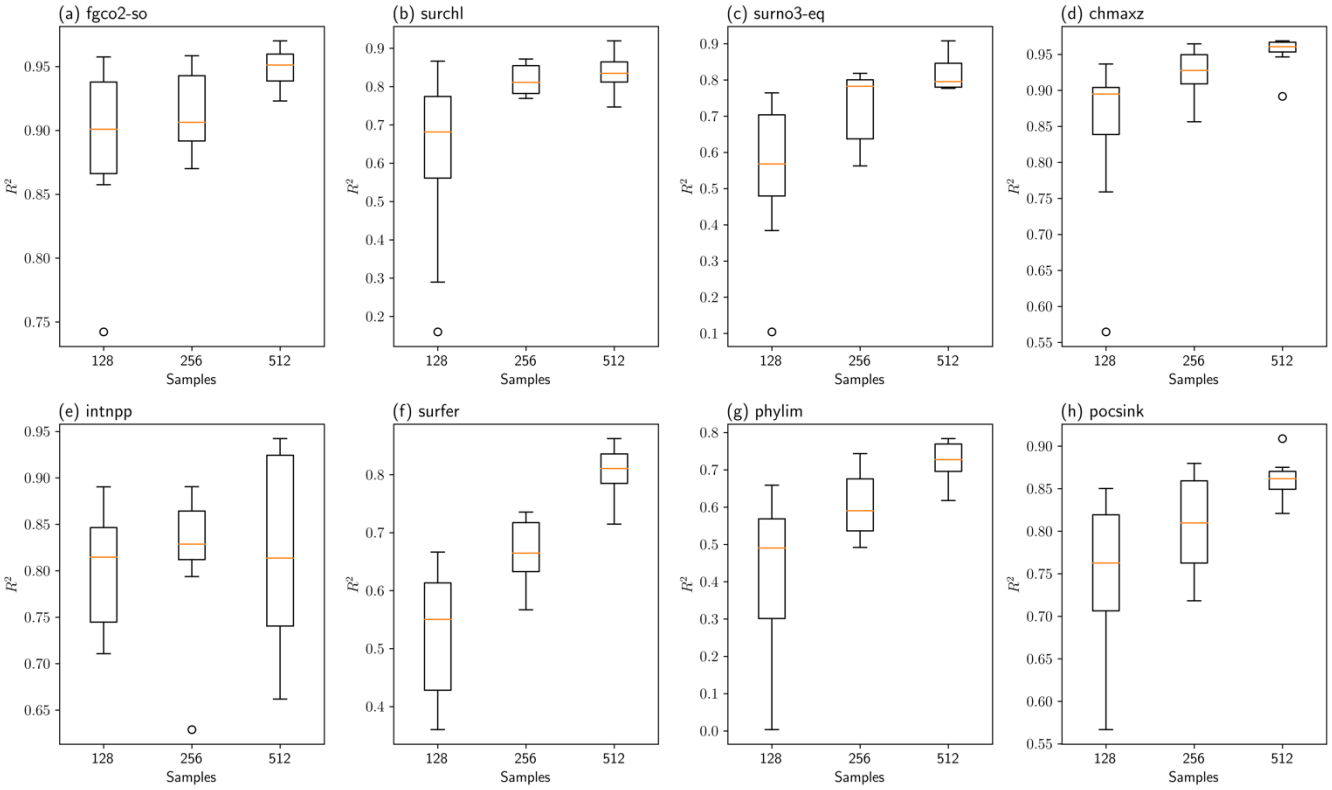


Figure S3. Sensitivity of the Gaussian Process Regression (GPR) model to true model output sample size during the optimization procedure. Performance of the GPR is measured as the R^2 when comparing the cost function of each variable. Variables are (a) downward flux of CO_2 in the Southern Ocean south of 30ºS, (b), surface chlorophyll concentrations, (c) surface nitrate concentrations in the equatorial zone between 20ºS and 20ºN, (d) the depth of the chlorophyll maximum, (e) depth-integrated net primary production, (f) surface concentrations of dissolved iron, (g) the primary limiting nutrient (iron or nitrogen) for phytoplankton growth, and (h) the sinking flux of particulate organic carbon.

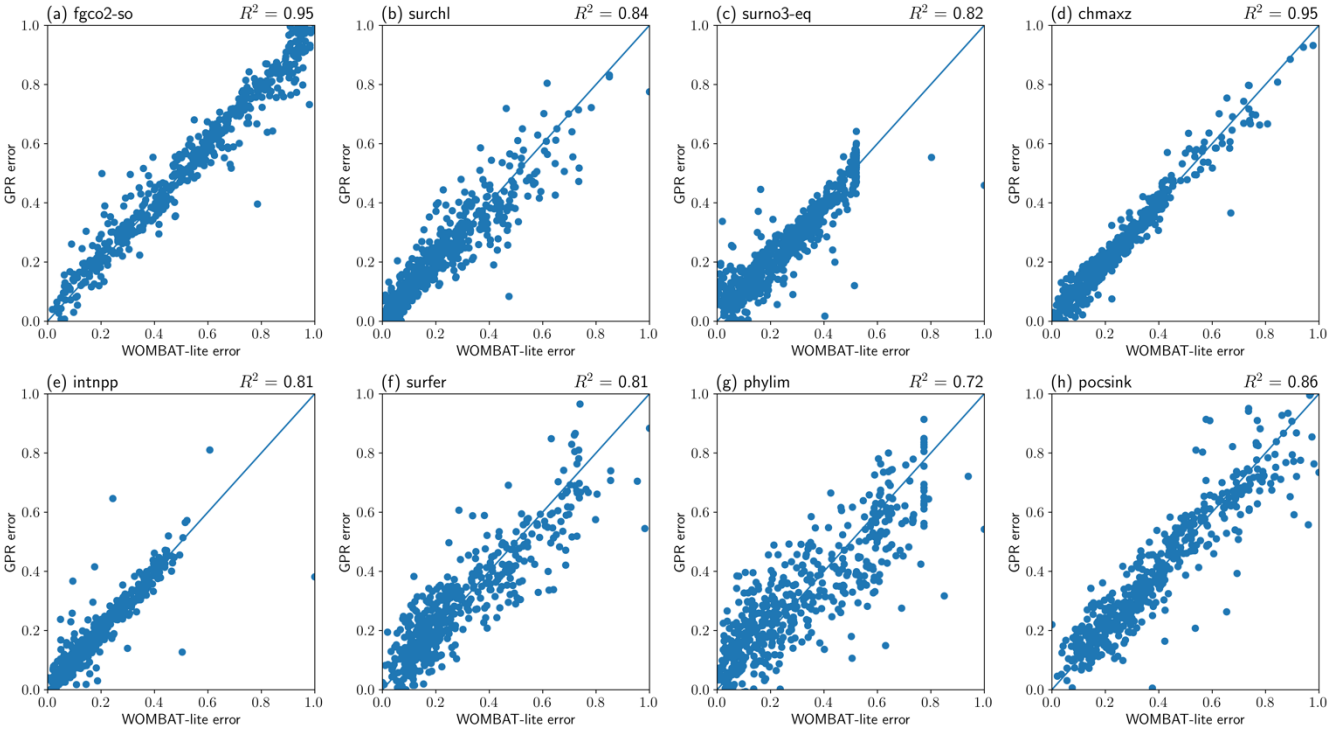
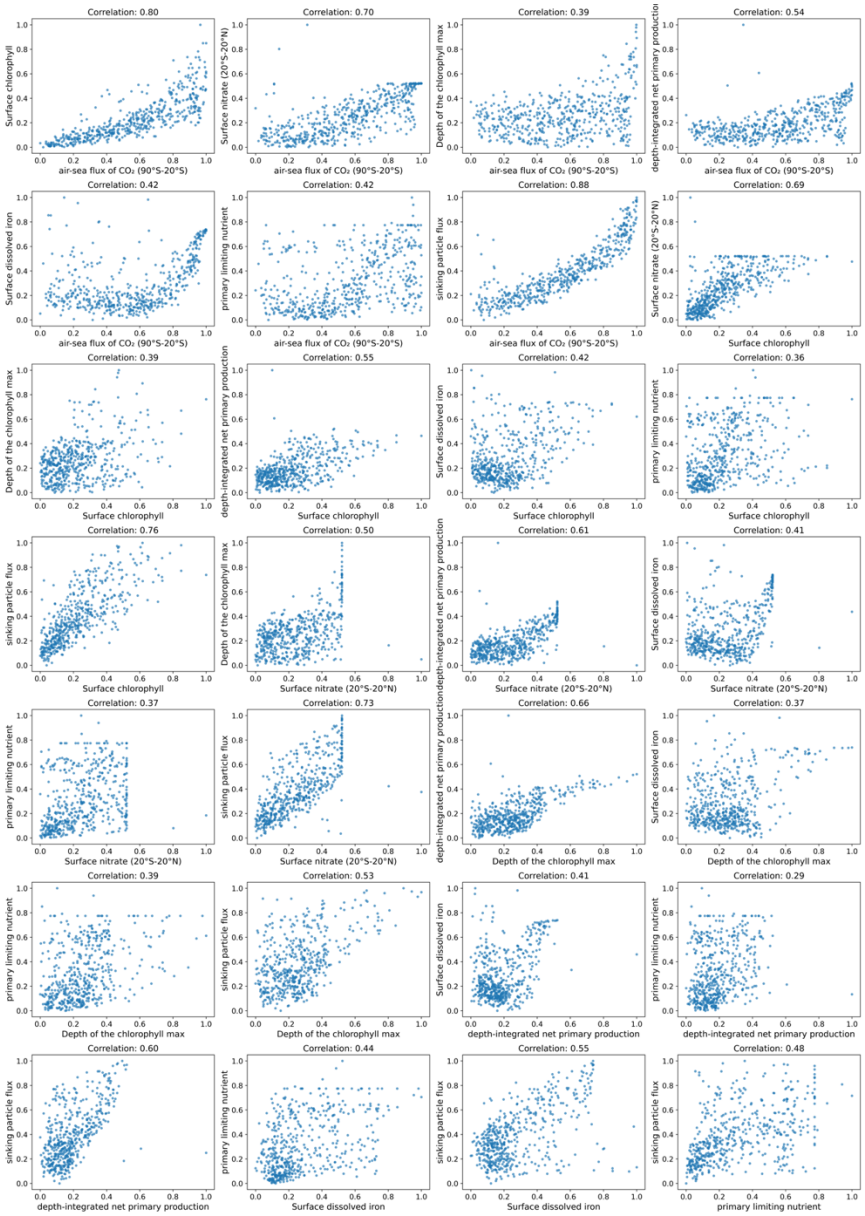
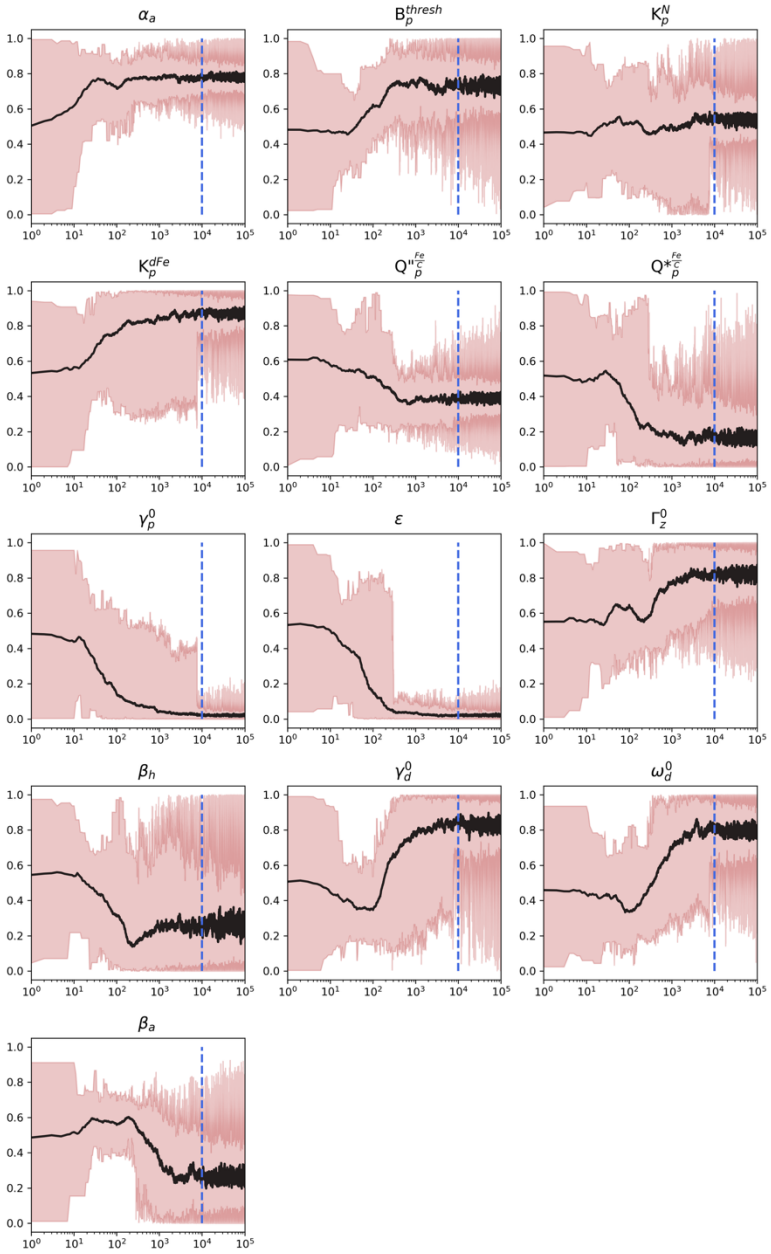


Figure S4. Accuracy of the Gaussian Process Regression (GPR) model to true model output (512 samples) during the optimization procedure. Performance of the GPR is measured as the cost function (Eq. 84) for each target variable. Variables are (a) downward flux of CO_2 in the Southern Ocean south of 30ºS, (b), surface chlorophyll concentrations, (c) surface nitrate concentrations in the equatorial zone between 20ºS and 20ºN, (d) the depth of the chlorophyll maximum, (e) depth-integrated net primary production, (f) surface concentrations of dissolved iron, (g) the primary limiting nutrient (iron or nitrogen) for phytoplankton growth, and (h) the sinking flux of particulate organic carbon.



30 **Figure S5. Pair-wise correlations between cost functions of the 8 target fields**. If positive and significant, the gains in the skill of the model to reproduce one target field will positively affect the models skill in reproducing the other.



35 **Figure S6. Mean (black line) and range (pink shading) of sampling by all 50 walkers during the Markov Chain Monte Carlo (MCMC) optimisation of the 13 parameters**. The blue dashed line demarcates the transition from the burn-in phase to the actual sampling used to construct the posterior. The x-axis is on a log10-scale, and the y axis is the same for all parameters because we normalised their ranges to 0-1 based on Max-Min scaling. A total of 100,000 steps is taken by the 50 workers.

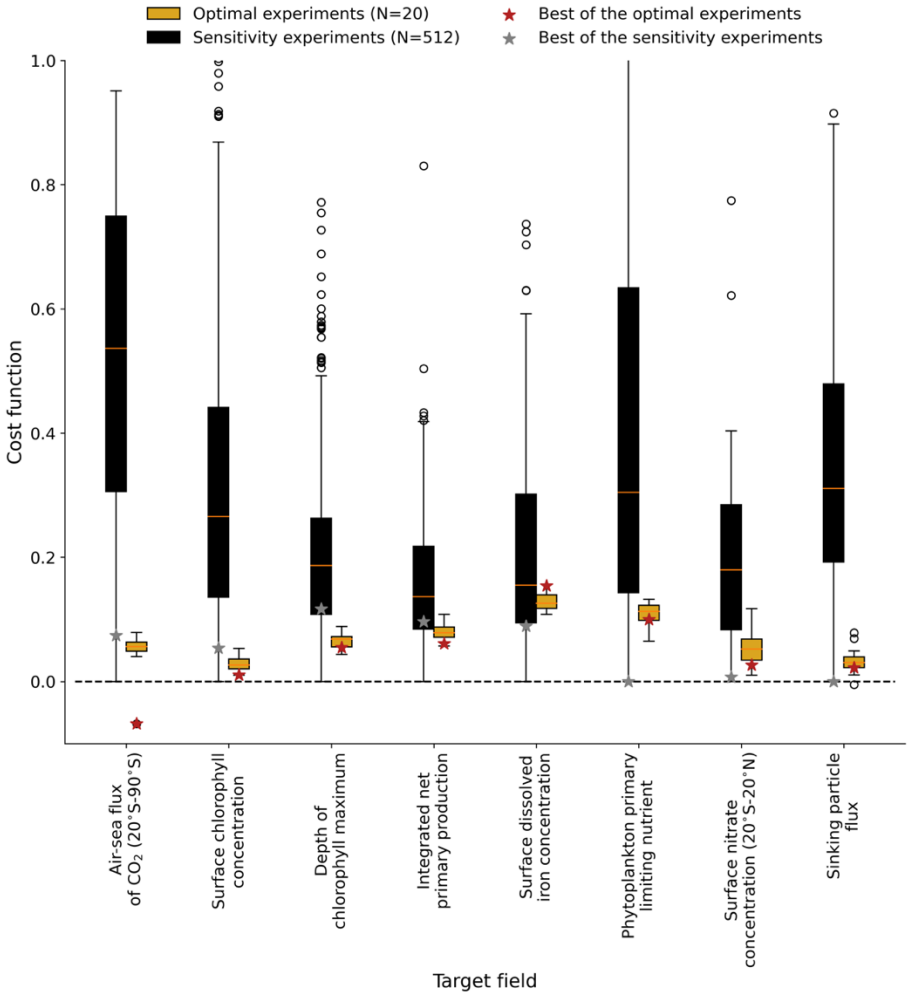


Figure S7. Performance of sensitivity and optimal experiments. Performance measured by the cost function of all 512 sensitivity experiments (black) against all 20 optimal experiments (gold) for each of the 8 target variables. The best performing of the 512 experiments (grey star) is compared with the best performing of the optimal experiments (red star). Best here is defined according to the global cost function. Negative values in the cost function are possible for the "optimal" experiments because their RMSE is normalised to the range of RMSE in the 512 sensitivity experiments, where zero is the minimum RMSE and one is the maximum RMSE.