Supporting Information for "Improved understanding of eutrophication trends, indicators and problem areas using machine learning"

Deep S. Banerjee[1,2] and Jozef Skákala[1,2]

[1]Plymouth Marine Laboratory, PL1 3DH Plymouth, United Kingdom,
[2]National Centre for Earth Observation, PL1 3DH Plymouth, United Kingdom.

Correspondence: Deep S. Banerjee (dba@pml.ac.uk)

Contents of this file

1 Figures

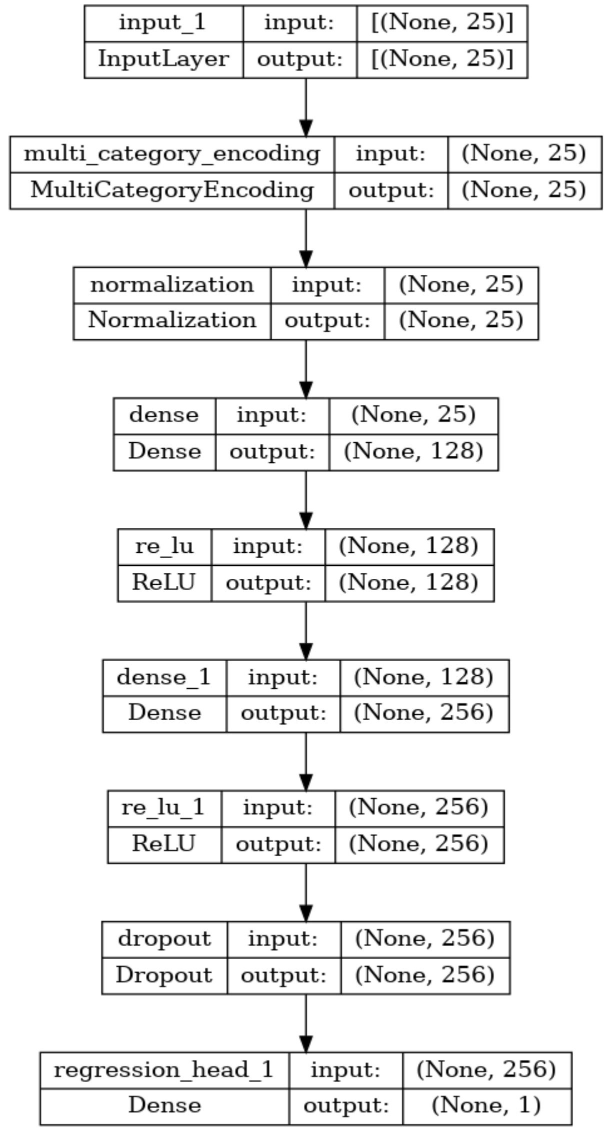


Figure S1. The architecture of the NN model used in this study.

References

Improved understanding of eutrophication trends, indicators and problem areas using machine learning Banerjee Deep S.1,2 and Skákala Jozef1,2 1Plymouth Marine Laboratory, PL1 3DH Plymouth, United Kingdom, 2National Centre for Earth Observation, PL1 3DH Plymouth, United Kingdom. Correspondence: Banerjee Deep S. (dba@pml.ac.uk) Abstract. Nitrate is an essential inorganic nutrient limiting phytoplankton growth in many marine environments. Eutrophica- tion, often caused by

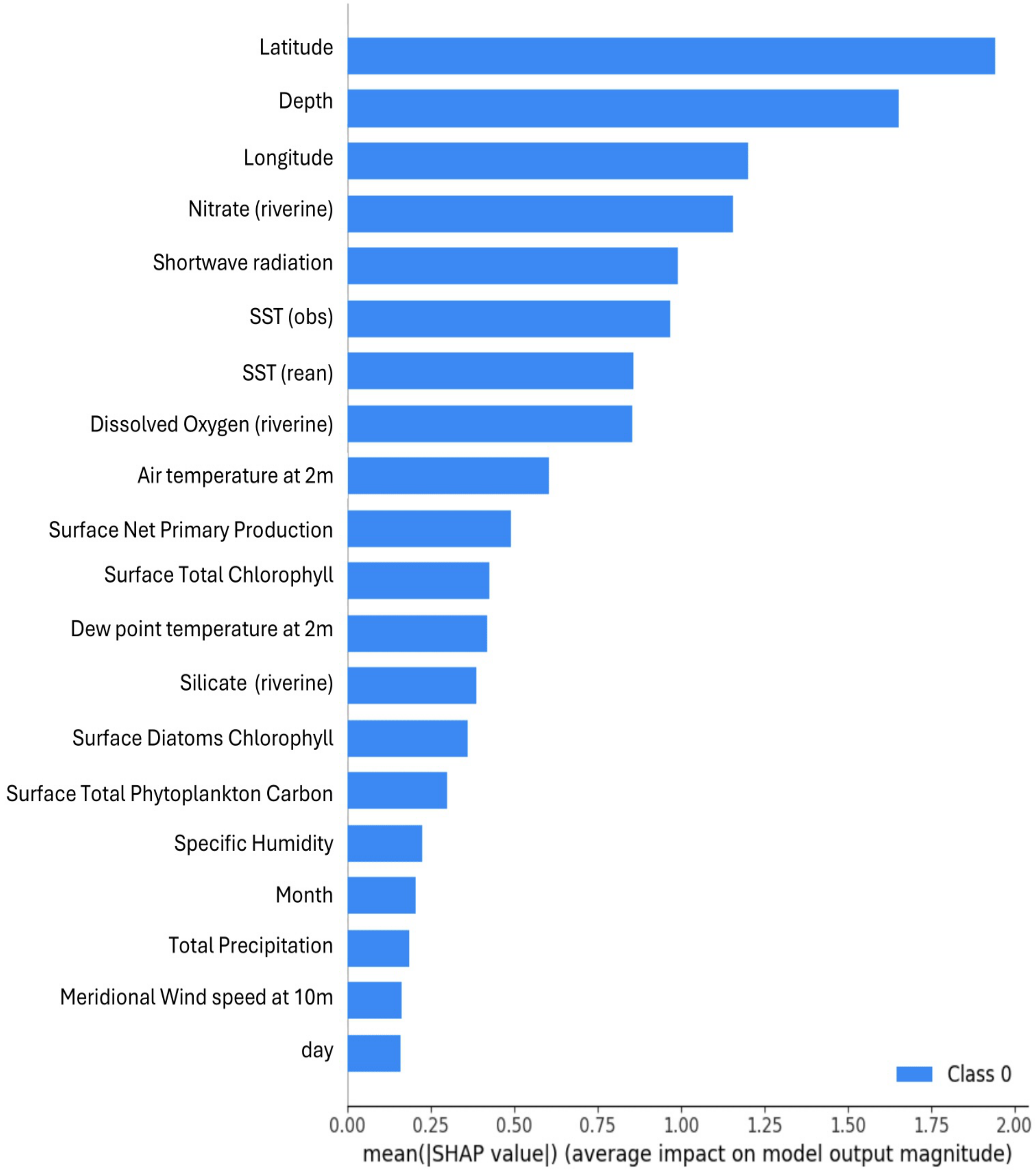
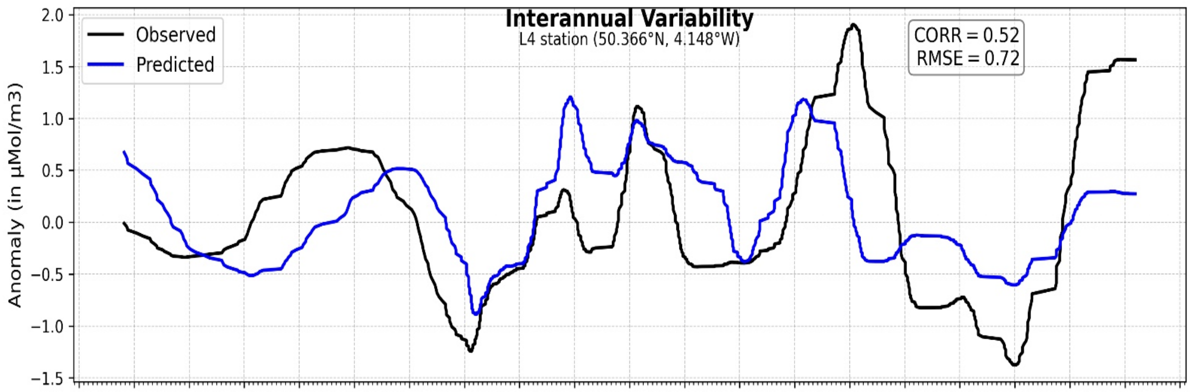


Figure S2. Evaluation of input feature importance using SHAP analysis. The features are ordered according to their importance, with 20 most important features shown.



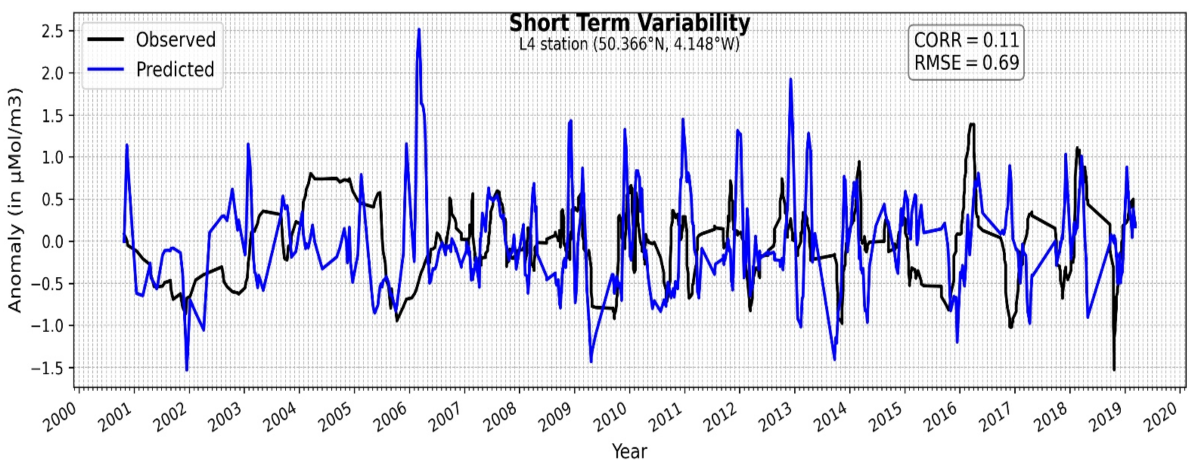


Figure S3. The comparison of interannual series (upper plot) and sub-monthly series (bottom plot) for surface nitrate at L4 station, between the NN-prediction and the L4 observations. The time-series were obtained by (i) subtracting seasonal climatology (smoothed on daily time-scale) from nitrate time-series and then (ii) either low-pass filtering the data (inter-annual time-series), or high-pass filtering the data (sub-monthly time-series), both on a 100 day scale.

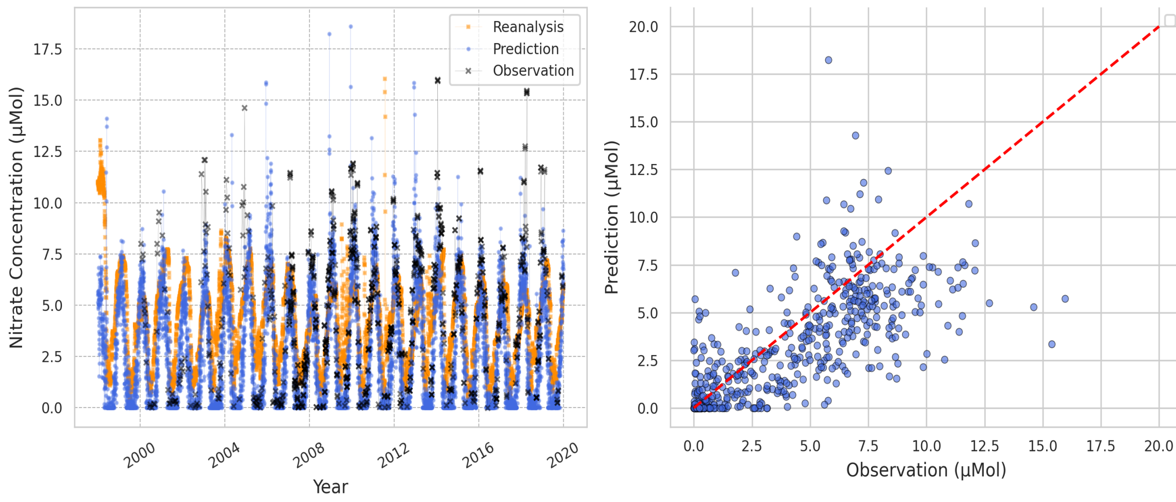


Figure S4. The comparison of the surface nitrate NN model prediction with the Copernicus reanalysis and the observations at the L4 station. The NN model prediction is from the NN-generated 1998-2020 gridded nitrate product. Both NN-model prediction and reanalysis are structured on the same spatial grid and were taken from the nearest model grid point to L4.

nitrogen deposition, is a reoccurring problem in coastal regions, including the North-West European Shelf (NWES). Despite of their importance, nitrate observations on the NWES are difficult to obtain and thus sparse both in time and space. We demonstrate that machine learning (ML) can generate, from sparse observations, a skilled, gap-free, bi-decadal (1998-2020) surface nitrate data-set. We demonstrate that the effective resolution (scales on which the data-set is skilled) is5 slightly coarser than the 7 km and daily resolution of the product, but still completely sufficient to analyse nitrate dynamics on a monthly scale. With such a data-set we can address questions that would be otherwise hard to answer: (i) We show that nitrate-limited regions on the NWES, potentially vulnerable to eutrophication, extend beyond the eutrophication-problem areas already identified by the monitoring bodies (i.e. OSPAR). The newly identified regions include southern Irish coastline and parts of Irish Sea, indicating that these areas could become problematic under sub-optimal policy, or management changes. (ii)10 We demonstrate that bi-decadal 1998-2020 trends in coastal nitrate, responding to long-term policy-driven reduction in riverine discharge, are mostly modest with a notable exception of the Bay of Biscay. (iii) We show that winter nitrate plays relatively minor direct role in the phytoplankton bloom intensity the following spring, which can have some implications for using winter inorganic nitrogen as eutrophication indicator (as often included by OSPAR). 1 Introduction15 Nitrogen is one of the most important components of organic matter, needed in relatively large concentrations, as demonstrated by the Redfield ratios (Tett et al., 1985). Despite of its large abundance (the Earth's atmosphere comprises 78as N2), it is non-trivial to obtain nitrogen in forms useful for plants. As a consequence of this, nitrogen is often the most limiting nutrient for plant, or algae growth, including

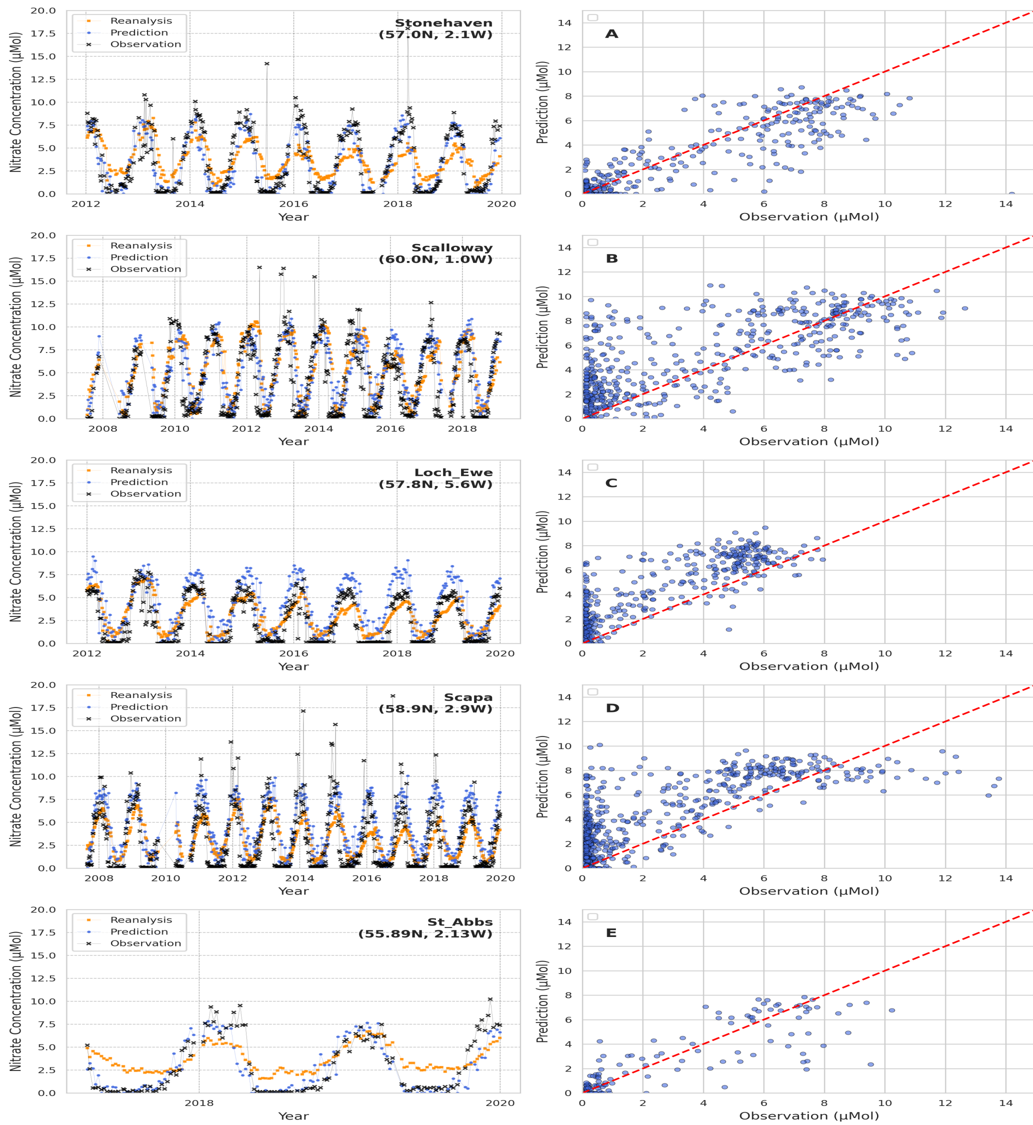


Figure S5. The comparison of the surface nitrate NN model prediction with the Copernicus reanalysis and the observations at five Scottish coastal stations. The NN model prediction is from the NN-generated 1998-2020 gridded nitrate product. Both NN-model prediction and reanalysis are structured on the same spatial grid and were taken from the nearest model grid points to the different stations.

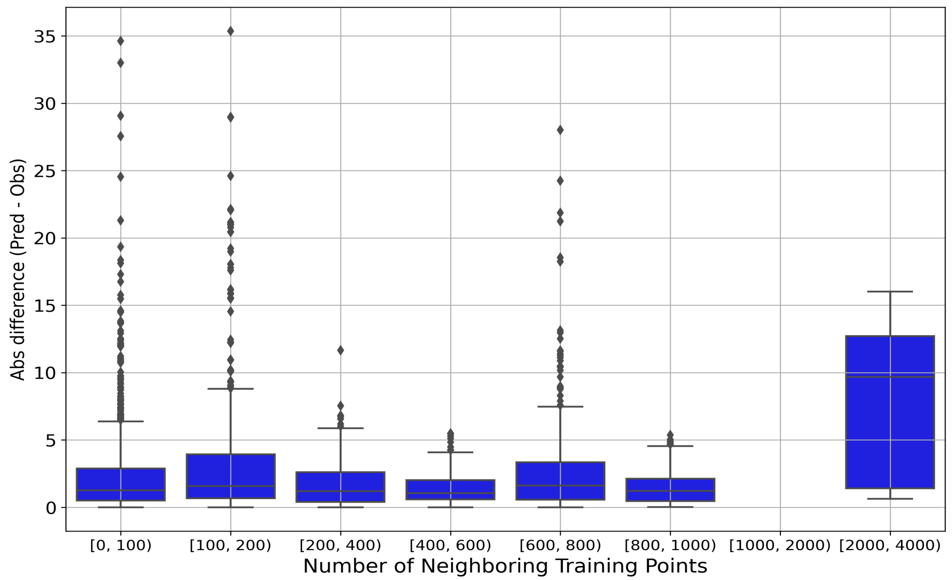


Figure S6. The RMSE skill score (μmol/m^3) of the NN-model calculated as a distance from the training data. The score is calculated as follows: (i) for each ICES test data-point we determine the number of training data that fall into the 50 km neighborhood of the test data point. The spatial separation of that test data-point from the training data can be then defined as inversely proportional to the number of training data that fall into the 50 km neighborhood of that test data-point. (ii) The test data are then binned into categories based on their spatial separation from the training data. The x-axis shows the bins, with their defining feature: the [X,Y] bin includes test data that have between X and Y training data in their 50 km spatial neighborhood. The y-axis then shows how well the NN model predicts the test data from the specific bin, i.e. it shows the distribution of absolute values of difference between the NN-model prediction and the ICES test data. The black horizontal lines show median values, the blue bars around those lines show two quantiles around the median, the error-bars show the interval with 95% of the values and the remaining extreme values are shown as separate dots.

the coastal marine environment (Ryther and Dunstan, 1971; Board et al., 2000). Nitrogen fixation, converting atmospheric nitrogen to forms useful for life, happens through various biotic and abiotic20 pathways, resulting in ammonium, nitrite and nitrate (Noxon, 1976; Hill et al., 1980; Postgate, 1998; Beman et al., 2008; Voss et al., 2013). Nitrate in the ocean is the primary nutrient for phytoplankton, with phytoplankton uptake enabling nitrogen flows into higher trophic levels and various detrital and dissolved forms of organic matter. In a nitrogen-limited environment, excess nitrate concentrations, primarily originating from agricultural runoff and industrial wastewater discharge, can stimulate harmful 1