

SUPPLEMENTARY MATERIAL: A Comparative Study of Biological Production in Eastern Boundary Upwelling Systems Using an Artificial Neural Network

Z. Lachkar¹ and N. Gruber¹

¹Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics,, ETH Zurich,
Universitätstrasse 16, CH-8092 Zurich, Switzerland

Correspondence to: Z. Lachkar (zouhair.lachkar@env.ethz.ch)

Abstract

We used a feature selection method to reduce the number of potential drivers and retain only the most important NPP predictors. This was performed using a SOM-based algorithm relying on the SOM topology preserving property. This supplementary material provides the detail of our feature selection algorithm. Two subsets of drivers emerge as the most relevant for explaining NPP variability: those include the upwelling index and nitrate concentration at 50 m depth, in addition to three common drivers: the eddy kinetic energy, the mixed layer depth and the continental shelf width.

1 Feature selection method

In this study we use SOM to perform the task of feature selection (see [Laha \(2005\)](#), [Laine and Simil \(2004\)](#) and [Benabdeslem and Lebbah \(2007\)](#) for other applications). The ability of SOM to assist in feature selection is a result of SOM's topology preservation property, permitting to derive a quantitative measure of the relative importance of predictors to the target data to be explained. However, SOMs topology preservation ability is affected severely if the intrinsic dimension of the data (which is usually lower than the absolute dimension) is much higher than the dimension (here 2) of the SOM lattice [Ritter and Schulten \(1988\)](#). Because of this dimensionality mismatch the SOM folds and twists to achieve the mapping, which leads to higher violation of topology preservation. The intrinsic dimension of the data is artificially increased when the least important predictors are included. Conversely, a fully dependent variable improves the neighborhood relationship, resulting in a reduction of the topological violation. Therefore, by measuring for each subset of variables the corresponding topological violation as well as the improvement in topology preservation due to including the dependant variable, NPP, we were able to rank the different combinations of drivers according to their respective degree of relevance. More specifically, the set of drivers which ensures the lowest topological violation and the highest NPP-induced improvement of the topology preservation should provide the best compact representation of data with only the most relevant features. Our algorithm does an exhaustive search to find the best subset of variables providing the most compact mapping according to our topological violation criterion. The topology violation is measured as the proportion of all data vectors, for which first and second best-matching units (BMU) (i.e. the two most closest neurons) are not adjacent vectors on the map. This is traditionally referred to as the topographic error [Kohonen \(2000\)](#). The algorithm browses subsets with more than 3 drivers, as the topological violation criterion can not be used to discriminate between subsets with a number of variables approaching the map dimension (2). [Figure 1](#) shows the results of this ranking. According to this criterion, two subsets of drivers appear to be the most relevant for explaining NPP variability: those include the upwelling index and nitrate concentration at 50 m depth, in addition to three common drivers: the eddy kinetic energy, the mixed layer depth

and the continental shelf width.

2

References

- Benabdeslem, K. and Lebbah, M.: Feature Selection for Self-Organizing Map, in: Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on, pp. 45–50, doi:{10.1109/ITI.2007.4283742}, 2007.
- Kohonen, T.: Self-Organizing Maps, Springer, 3rd edn., 2000.
- Laha, A.: Detecting Topology Preserving Feature Subset with SOM, pp. 40–48, <http://www.springerlink.com/content/5rwt58mka001pw06>, 2005.
- Laine, S. and Simil, T.: Using SOM-Based Data Binning to Support Supervised Variable Selection, pp. 172–180, <http://www.springerlink.com/content/j8bpmfrggek86fex>, 2004.
- Ritter, H. and Schulten, K.: Kohonen’s self-organizing maps: exploring their computational capabilities, in: Neural Networks, 1988., IEEE International Conference on, pp. 109–116, doi:{10.1109/ICNN.1988.23838}, 1988.

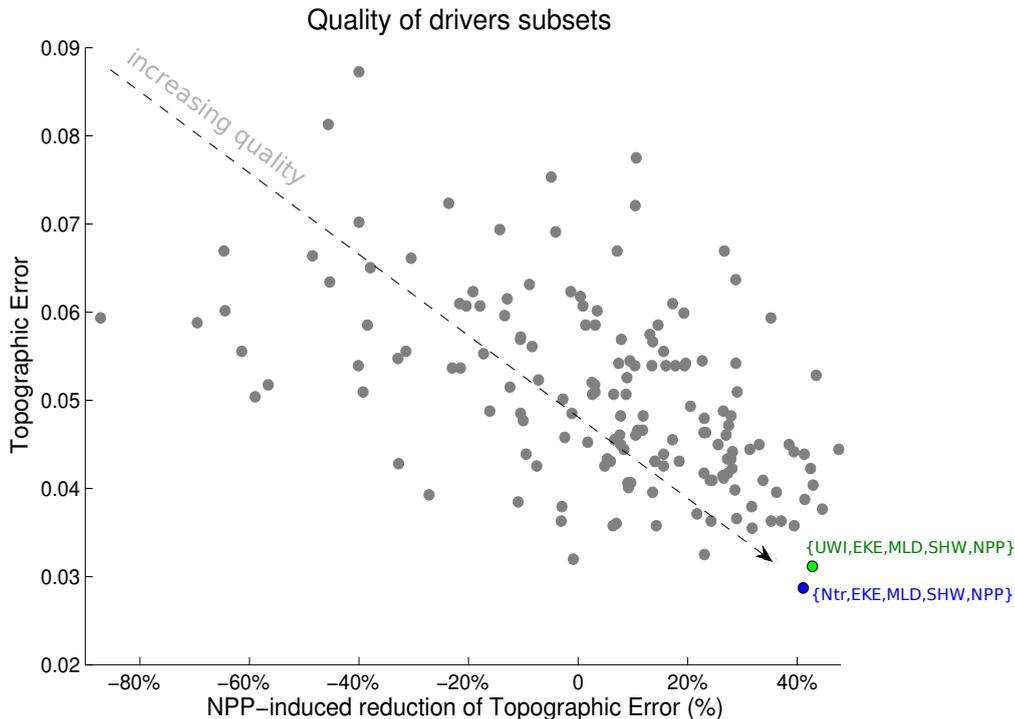


Fig. 1: Quality of various sets of NPP drivers based on the SOM topology preservation criterion. The quality of drivers is lowest in the top left corner (high topographic error (TE) and low NPP-induced reduction of TE) and highest in the bottom right corner (low TE and high NPP-induced reduction of TE). The blue (nitrate availability at 50m, eddy kinetic energy, mixed layer depth, shelf width) and green (upwelling index, eddy kinetic energy, mixed layer depth, shelf width) dots correspond to the best subsets of drivers.