REDUCTION OF A COMPLEX BIOGEOCHEMICAL MODEL WITH NEURAL NETWORK AND CLUSTERING TECHNIQUES

Knut Bernhardt[1], Tobias A. Sperr and Kai W. Wirtz

Institute for Chemistry and Biology of the Marine Environment, Carl-von-Ossietzky-Str. 9-11, P.O. Box 2503, 26111 Oldenburg, Germany
[1]Contact: knut.bernhardt@icbm.de

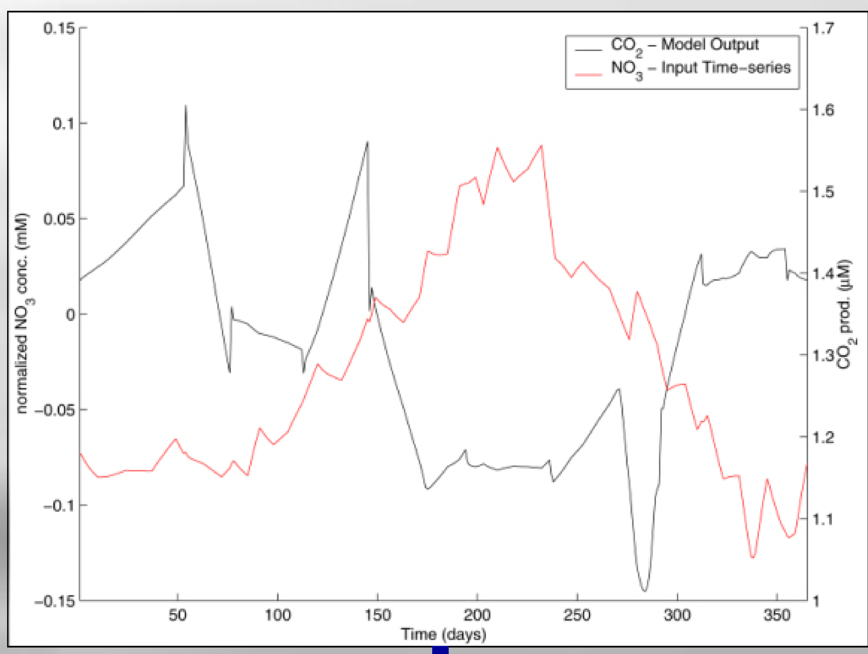

Figure 1: Time series of FNO3 (input) and TfCO2 (output)

The Integrated Sediment Model (ISM) proposed by Wirtz (2003) is a complex biogeochemical model incorporating microbial transformation of chemical species, local and non-local transport processes as well as temporally and spatially variable boundary conditions.
The model uses a set of time series (e.g. NO_3, NH_4) measured at the Spiekeroog backbarrier reef as input data and generates time-series/profiles of e.g. produced CO_2 (Figure 1). A special focus of this paper is laid on the biogeochemical cycle of nitrogen.
The aggregation/reduction of complex models is motivated by problems arising from the handling of these models, concerning e.g.

- Upscaling to global and regional models
- Dealing with parameter uncertainty for a better interpretation of model results
- Detection of effective modes to understand key variables of the system

The Self-Organizing Map (SOM), a neural network based vector-quantization technique (Kohonen 2001), in combination with clustering algorithms can serve as a means for aggregation of model variables of complex process-based models as shown below.

SOM - TRAINING

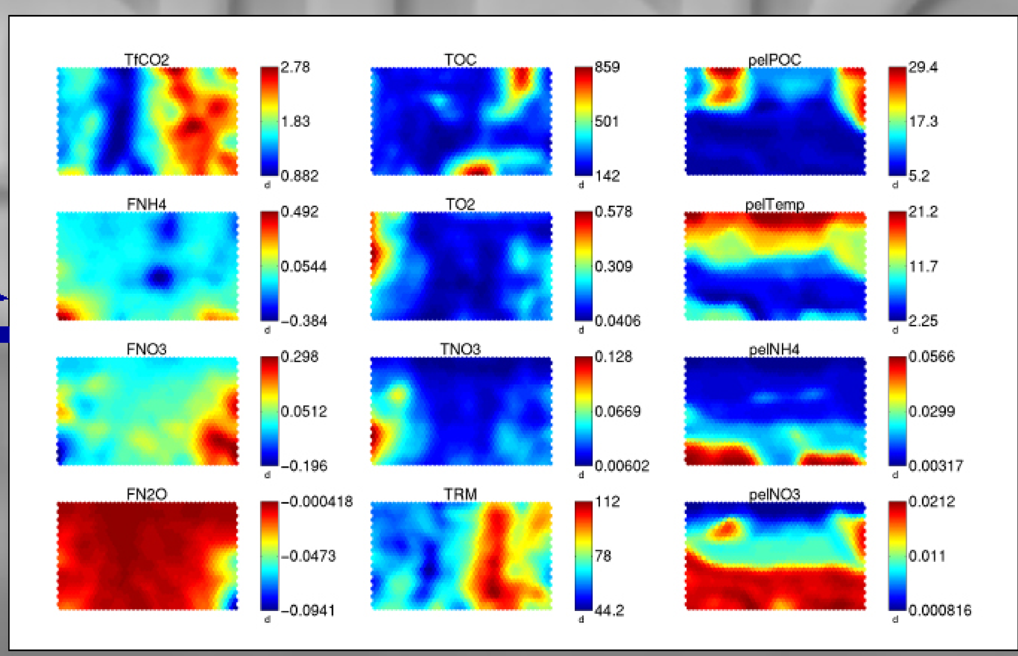

Figure 2: Component planes of the trained SOM for 16 model parameters.

A SOM provides a mapping of a multidimensional dataset onto a set of topologically ordered prototype vectors arranged in a rectangular grid with hexagonal neighbourhood relationship.
We obtained the dataset by calculating 1000 x 365 simulation days with different parameter settings. 12 model variables were selected to capture the microbial activity and the flux of N-components.

Components planes for each reference vector of the trained map are shown in Figure 2. Beside a global ordering e.g. into a northern summer and a southern winter part (see variable pelTemp) local structures emerge as, for example, the south-eastern corner of the map showing high negative values of N_20 flux (FN2O).

CLUSTERING

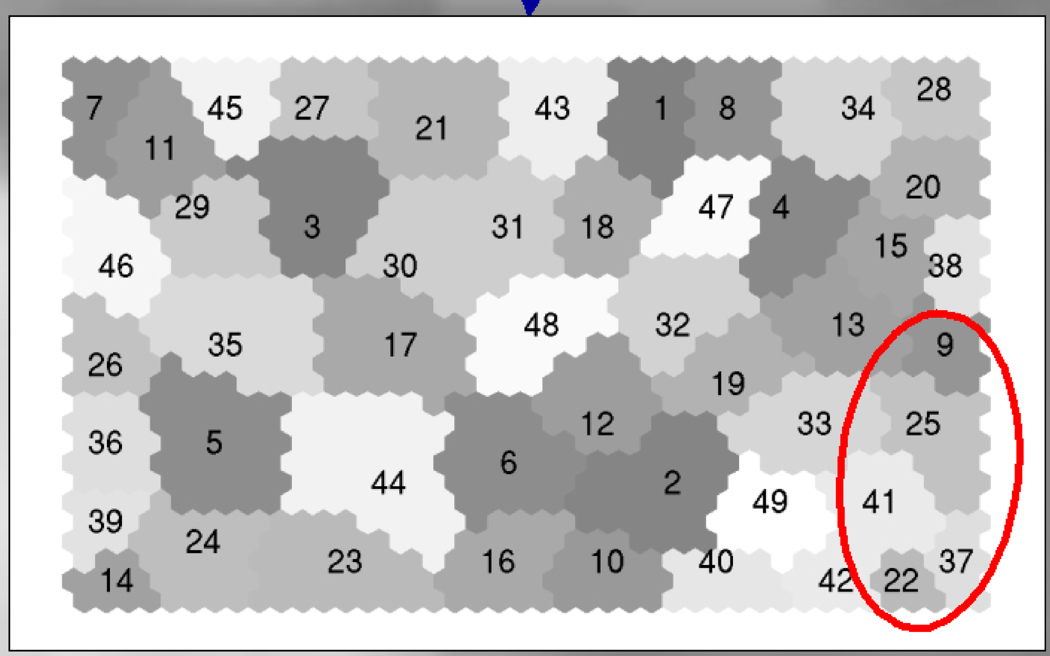

Figure 3: Clustered version of the SOM shown in Figure 2. Clusters with high NO_2 emission are marked in red.

A clustering of the trained SOM using the k-means algorithm (see e.g. Baraldi and Blonda 1999) helps to identify typical states of the system.

A number of 42 clusters for our SOM with 28 x 42 nodes was found to be an optimal choice and the set of clusters obtained is depicted in Figure 3.
Clusters 9, 22, 25, 37 and 41(marked in red) coincide with regions of high N_20 emission (see Figure 2) and were analysed in the following steps.

TRANSITION DYNAMICS

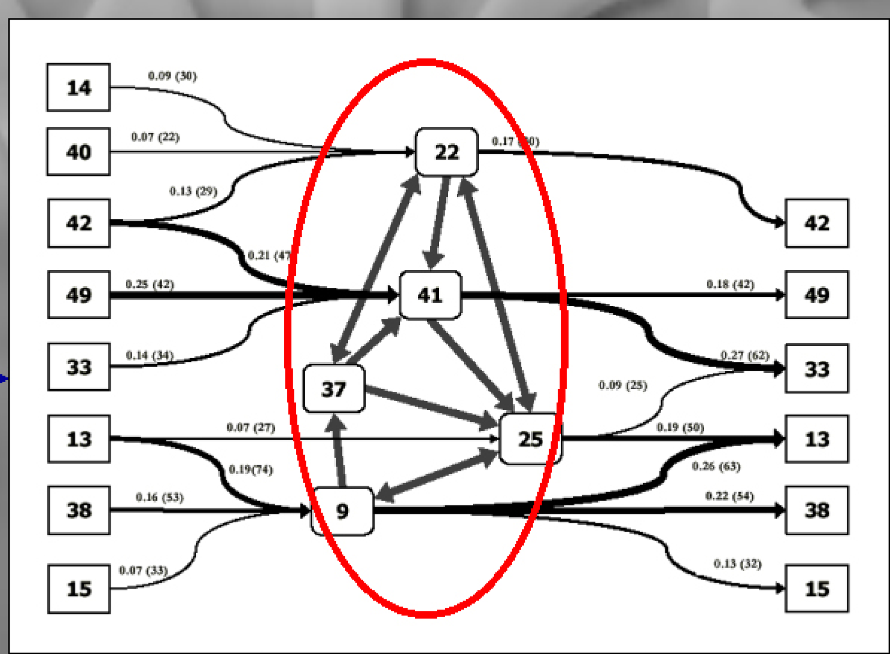

Figure 4: Transition graph for clusters showing high NO2 efflux (red). Input and output-clusters are also shown.

Temporal dynamics can be re-inserted into the SOM by mapping each day of the input data onto the nearest cluster of the network and by determining empirical transition-probabilities between the clusters.
An example for the clusters characterized by high N_2O emissions (marked in red) is shown in Figure 4 together with clusters representing input- and output-states for this part of the system.
A physical meaning of the clusters can be obtained by inversely mapping each cluster to approximate values of standard model variables.
After doing so, the transition matrix obtained in this study could, for example, be implemented as a lookup-model usable in global change assessments and pelagic ecosystem models.

Future work:

- Analysis and improvement of drawbacks of the SOM-algorithm outlined by Baraldi and Blonda (1999) concerning e.g. the topology-preservation of high-dimensional input data
- Improvement of the clustering method
- Reformulation of the presented transformation approach in terms of measurable variables

References:

Baraldi, A. & Blonda, P. (1999). A Survey of Fuzzy Clustering Algorithms for Pattern Recognition - Part II. IEEE Trans. Syst., Man, Cybern., **29**(6), 778-785.
Kohonen, T. (2001). Self-organizing Maps (3. ed.). Springer, Berlin, New York.
Vesanto, J. & Alhoniemi, E. (2000). Clustering of the Self-Organizing Map. IEEE Trans. Neural Networks, **11**(3), 586-600.
Wirtz, K. W. (2003). Control of biogeochemical cycling by mobility and metabolic strategies of microbes in the sediments: an integrated model study. FEMS Microbiology Ecology, accepted.