Illuminating interactions in microbial lake ecology using General Learning Chain Ensembles

Theodor Sperlea[1], Daniela Beißer[2], Jens Boenigk[2], Dominik Heider[1]

[1]Department of Bioinformatics, Faculty of Mathematics and Computer Science, Philipps-Universität Marburg, Germany; [2]Biodiversity Department, Center for Water and Environmental Research, University of Duisburg-Essen, Germany

In the recent years, legislation such as the EU's Water Framework Directive is putting a focus on the issue of anthropogenic stressors heavily affecting aquatic biomes (Hering et al., 2010). Impacting both biotic as well as abiotic factors of ecosystems, these anthropogenic stressors lead to a global decline of ecosystem quality and increase the probability of extinctions. In order to slow down, halt, or even revert this process, it is necessary to develop methods to assess the status of an ecosystem as well as to identify the requirements that different organisms pose on their surroundings.
In the context of microbial lake ecology, it is straightforward to log and count the microorganisms in a water sample based on their genomic sequences (Tan et al., 2015; Grossmann et al., 2016). However, the connections between the number of these microorganisms and the status of their surroundings remain mostly unknown.
In the current study, we introduce General Learning Chain Ensembles (GLCEs), that are able to efficiently model multi-target datasets containing inhomogenous data types. We further show that GLCEs are able to model the complex interactions between biotic and abiotic factors in microbial ecology. From these models, we were able to gather information on which parameters are important for the well-being of certain species as well as on microscopic bioindicators for water quality.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., Heiskanen, A.-S., Johnson, R. K., Moe, J., and Pont, D. (2010). The european water frameworkdirective at the age of 10: A critical review of the achievements with recommendations for the future. Science of The Total Environment, 408(19), 4007–4019.

Grossmann, L., Jensen, M., Heider, D., Jost, S., Glücksman, E., Hartikainen, H., Mahamdallie, S. S., Gardner, M., Hoffmann, D., Bass, D., and Boenigk, J. (2016). Protistan community analysis: key findings of a large-scale molecular sampling. The ISME Journal , 10(9), 2269–2279.

Tan, B., Ng, C., Nshimyimana, J. P., Loh, L. L., Gin, K. Y.-H., and Thompson, J. R. (2015). Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges and future opportunities. Frontiers in Microbiology, 6.