NETLAKE toolbox for the analysis of high-frequency data from lakes

Factsheet #11

Inferential modelling of time series by evolutionary computation

Friedrich Recknagel and Ilia Ostrovsky

Objective

The hybrid evolutionary algorithm (HEA) has been designed: 1) to represent and forecast multivariate relationships between environmental conditions and population densities by inferential (IF-THEN-ELSE) models, and 2) to quantify 'tipping points' for population outbreaks by IF-conditions (Figure 1). During the course of hundreds of iterations, HEA discovers the 'best-fitting' model after optimising model structures by genetic programming and model parameters by differential evolution towards the lowest RMSE and highest R^2 (Cao et al. 2013).

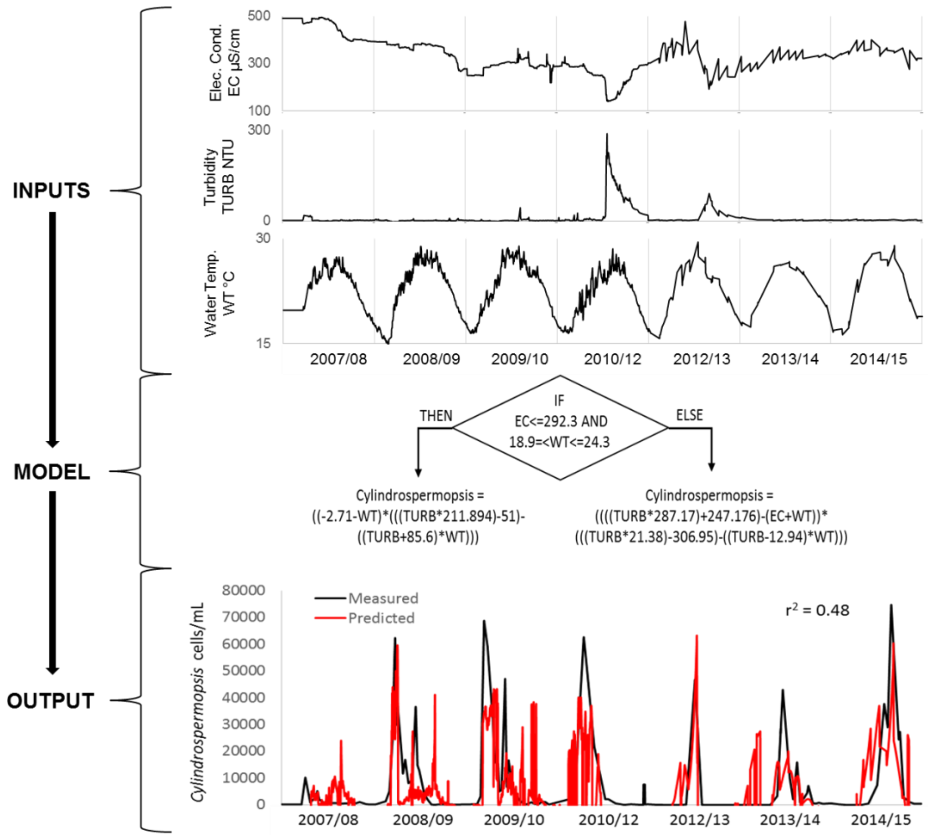

Figure 1. 20-day-ahead forecasting of *Cylindrospermopsis raciborskii* in Lake Wivenhoe (Australia) by means of inferential modelling based on HEA. The IF-condition suggests that fast population growth of *C. raciborskii* in Lake Wivenhoe may occur within the temperature range of 18.9 to 24.3 °C and at conductivity levels lower than 292 µS/cm.

The forecasting accuracy of inferential models by HEA suits early warning of population outbreaks. Ensembles of inferential models allow scenario analysis of how shifts in physical-chemical boundaries impact on aquatic communities. Meta-analysis of 'tipping points' and ecological relationships across lakes with the same stratification regime and trophic state allows the generalisation of knowledge inherent in complex ecological data.

Specific application

Quantifying ecological tipping points and relationships has been demonstrated successfully by case studies for Lakes Müggelsee (Germany), Kinneret (Israel), Taihu (China) and Lajes (Brazil) (Recknagel et al. 2016; Recknagel et al. 2015; Recknagel et al. 2014; Recknagel et al. 2013). **Short-term forecasting and early warning** of cyanobacteria blooms as well as meta-analysis of tipping points have been demonstrated successfully by case studies for Lakes Wivenhoe, Somerset and Samsonvale (Australia) (Recknagel et al. 2014). **Spatially-explicit short-term forecasting** of cyanobacteria blooms has been demonstrated successfully by case studies for Lakes Lajes (Brazil), Taihu (China) and Wivenhoe (Australia) (Recknagel et al. 2015; Zhang et al. 2015; Cao et al. 2016).

Background

The tool is available as user-friendly software written in C++. To use the tool requires basic programming skills. To execute evolutionary computations by HEA can be very time-consuming. It is therefore recommended to run HEA on supercomputers in cloud mode.

Type of data and requirements

Ecological time series are recorded in .xls spreadsheets where rows contain input- and output parameters of interest (e.g. physical, chemical and biological data) for consecutive equidistant time steps. Since the HEA software learns from patterns, modelling of seasonal and inter-annual dynamics requires at least 3 years of data, but it generalises best with decades of data containing a wealth of patterns. If data are missing or have been measured at non-equidistant time steps, interpolation of data to the smallest measured time step is required (HEA licence includes a software tool for flexible linear data interpolation of time series). Whilst 'day' is the recommended time step for 'several-day-ahead' predictive modelling, there is no restriction to the choice of the smallest time step. Data for spatially-explicit modelling of same ecological attribute measured simultaneously at multiple sites has the same requirements as for modelling single-site data (HEA licence includes detailed manual and data examples for single- and multi-site modelling experiments).

The .xls spreadsheets need to be completed by specifying HEA control parameters such as numbers of inputs, outputs, generations, boot-strap loops etc. before being saved as Text (Tab delimited) files. To run HEA, the HEA *exe*-file together with the Text file need to be submitted to a supercomputer.

Basic procedures

1. Prepare equidistant input and output data as well as HEA control parameters in .xls files before saving them as Text (Tab delimited) files.
2. Submit HEA *exe*-file together with Text file to supercomputer.
3. Review the modelling protocol documenting 10 'best fitting' models by: IF-THEN-ELSE rules, graphical validation, root mean squared error (RMSE), R^2, ranking inputs by sensitivity, input sensitivity functions.

Pitfalls and tips

- Since HEA ranks inputs by sensitivity after each run, noise from the least sensitive inputs can be removed for consecutive runs that may improve model validity.
- To avoid bias by relying on a single model, averages and Min-Max envelopes of an ensemble of 3 to 5 best-fitting models can be utilised for validation.
- Since HEA infers IF-THEN-ELSE rules for the underlying research question, the IF conditions reveal quantitative thresholds that explain causes for high and low output magnitudes.

Further reading

Key References:

Cao, H., Recknagel, F., Orr, P. 2014. Parameter optimisation algorithms for evolving rule models applied to freshwater ecosystem. *IEEE Transactions on Evolutionary Computation* 18: 793-806.

Cao, H., Recknagel, F., Bartkow, M. 2016. Spatially-explicit forecasting of cyanobacteria assemblages in freshwater lakes by multi-objective hybrid evolutionary algorithms. *Ecological Modelling*, 342, 97-112.

Recknagel, F., Adrian, R., Köhler, J., Cao, H. 2016. Threshold quantification and short-term forecasting of *Anabaena, Aphanizomenon* and *Microcystis* in the polymictic eutrophic Lake Müggelsee (Germany) by inferential modelling using the hybrid evolutionary algorithm HEA. *Hydrobiologia* 778: 61-74.

Other useful references:

Recknagel, F., Branco, C.W., Cao, H., Huszar, V.L., Sousa-Filho, I.F. 2015. Modelling and forecasting the heterogeneous distribution of picocyanobacteria in the tropical Lajes Reservoir (Brazil) by evolutionary computation. *Hydrobiologia* 749: 53-67.

Recknagel, F., Orr, P., Cao, H. 2014. Inductive reasoning and forecasting of population dynamics of *Cylindrospermopsis raciborskii* in three sub-tropical reservoirs by evolutionary computation. *Harmful Algae* 31: 26–34.

Recknagel, F., Ostrovsky, I., Cao, H. 2014. Model ensemble for the simulation of plankton community dynamics of Lake Kinneret (Israel) induced from in situ predictor variables by evolutionary computation. *Environmental Modelling & Software* 61: 380-392.

Recknagel, F., Ostrovsky, I., Cao, H., Chen, Q. 2014. Hybrid evolutionary computation quantifies environmental thresholds for recurrent outbreaks of population density. *Ecological Informatics* 24: 85–89.

Recknagel, F., Ostrovsky, I., Cao, H., Zohary, T., Zhang, X. 2013. Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets. *Ecological Modelling* 255: 70-86.

Zhang, X., Recknagel, F., Chen, Q., Cao, H., Li, R. 2015. Spatially-explicit modelling and forecasting of cyanobacteria growth in Lake Taihu by evolutionary computation. *Ecological Modelling* 306: 216-225.

Code

HEA has been coded in C++ language and is not yet freely available. The authors offer short courses on inferential and process-based modelling, and welcome collaboration on data processing and modelling (for more details please contact friedrich.recknagel@adelaide.edu.au).

Contact details

Friedrich Recknagel. University of Adelaide, School of Biological Sciences, Adelaide, Australia. friedrich.recknagel@adelaide.edu.au

Ilia Ostrovsky. Israel Oceanographic & Limnological Research, Kinneret Limnological Laboratory, P.O.B. 447, Migdal 14950, Israel. ostrovsky@ocean.org.il

Suggested citation

Recknagel, F. and Ostrovsky, I. 2016. Inferential modelling of time series by evolutionary computation. In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from lakes* (Factsheet 11). Technical report. NETLAKE COST Action ES1201. pp. 57-60. http://eprints.dkit.ie/id/eprint/542.

Acknowledgement

This factsheet is based upon work from the NETLAKE COST Action ES1201, supported by COST (European Cooperation in Science and Technology).