Environmental Modelling & Software 61 (2014) 380–392

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Model ensemble for the simulation of plankton community dynamics of Lake Kinneret (Israel) induced from *in situ* predictor variables by evolutionary computation☆

Friedrich Recknagel [a],*, Ilia Ostrovsky [b], Hongqing Cao [a]

[a] *University of Adelaide, School of Earth and Environmental Sciences, Seaton 5023, Adelaide 5005, Australia*
[b] *Israel Oceanographic and Limnological Research, Kinneret Limnological Laboratory, Migdal 14950, Israel*

ARTICLE INFO

Article history:
Received 29 April 2013
Received in revised form
4 February 2014
Accepted 25 March 2014
Available online 5 May 2014

Keywords:
Model ensemble
Hybrid evolutionary algorithm HEA
Model operationality
Lake Kinneret
Plankton community dynamics
In situ predictor variables
Forecasting
Ecological thresholds
Sensitivity analysis

ABSTRACT

This study addresses the need for operational models in view of rapidly advancing *in situ* sensor technology that puts lakes into online surveillance mode. A model ensemble for simulating plankton community dynamics in Lake Kinneret (Israel) from 1988 to 1999 has been induced from electronically-measurable predictor variables (EMPV) such as water temperature, pH, turbidity, electrical conductivity and dissolved oxygen by the hybrid evolutionary algorithm HEA. It cascade wise predicts the total nitrogen to total phosphorus ratios TN/TP, concentrations of chlorophyta, baccilariophyta, cyanophyta and dinophyta, as well as densities of rotifera, cladocera and copepoda solely from EMPV. The best coefficients of determination (r^2) have been achieved with 0.6 by the dinophyta model, 0.45 by the rotifera model and 0.44 by the bacillariophyta model. The worst coefficients of determination (r^2) have been produced by the cladocera model with 0.24 and by the TN/TP model with 0.28. Despite the differences in the r^2 values and apart from the cladocera model, the remaining models matched reasonably well seasonal and interannual plankton dynamics observed over 11 years in Lake Kinneret.

The model ensemble developed by HEA also revealed ecological thresholds and relationships determining plankton community dynamics in Lake Kinneret solely based on *in situ* predictor variables.

1. Introduction

Modelling and simulation of plankton community dynamics in lakes is considered to be the domain of deductive process-based modelling. Ordinary differential equations (ODE) are traditionally used for the simulation of temporal dynamics of ‘key’ state variables (e.g. Park et al., 1974; Benndorf and Recknagel, 1982; Arhonditsis and Brett, 2005), and partial differential equations (PDE) are used for the simulation of spatio-temporal dynamics of ‘key’ state variables (e.g. Hamilton and Schladow, 1997; Mao et al., 2008; Gal et al., 2009). ODE are site-specifically calibrated and validated by continuously monitored water quality data and normally don’t require *ad hoc* monitoring but fail to give insights into spatial dynamics. By contrast, PDE need to be calibrated and validated for close-mashed vertical and horizontal grids that require extensive *ad hoc* monitoring for the benefit of simulating 2- or 3-dimensional lake behaviour. Scenario analysis, supported by both approaches informs hypotheses testing and medium- to long-term management decisions.

Both, the ODE and the PDE modelling approach have several short comings. Firstly, the current knowledge on processes that determine mass balances of nutrients, phytoplankton phyla and zooplankton functional groups is rather vague, and process equations are largely statistically fitted rather than deduced from causal knowledge. That in turn leads to rigid model structures that are *ad hoc* calibrated for specific lake sites. Secondly, the current knowledge on species specific processes that determine mass balances and food–web interactions of plankton genera or species is limited, and – in terms of parameter requirements – cause a level of complexity that can hardly be comprehended and managed. This applies in particular to PDE models where additional transport processes between grid cells need to be calibrated and validated by spatio-temporal monitoring data.

Overall properly validated process-based lake models serve as powerful tools for scenario analysis, but face severe constraints in terms of rigidity, complexity and operationality.

☆ Thematic Issue on Novel Approaches to Challenges in Aquatic Ecosystem Modelling.

* Corresponding author.

E-mail address: friedrich.recknagel@adelaide.edu.au (F. Recknagel).

http://dx.doi.org/10.1016/j.envsoft.2014.03.014

Novel computational concepts provide opportunities to overcome some constraints of process-based modelling. To overcome model rigidity, Recknagel et al. (2008a,b) applied object-oriented programming to build a library of alternative process models for algal growth and grazing, zooplankton growth and predation etc. which allows to identify best performing model structures for lakes and lake categories. It allows running and testing alternative process models in the framework of a process-based lake model. Resulting dynamic model structures served both, optimum model design and model generalization for lake categories as demonstrated for warm-monomictic hypertrophic lakes in South Africa, dimictic mesotrophic and dimictic eutrophic lakes in Germany (Recknagel et al., 2008a,b). Cao and Recknagel (2009) embedded evolutionary algorithms into categorized process-based models to evolve optimal process equations previously statistically fitted, and optimum temperature and nutrient functions as substitutes for constant parameters. Both approaches improved structural flexibility and generality of ODE models, but did not achieve operationality.

Model operationality becomes of growing importance since *in situ* sensor technology puts lakes into online surveillance mode and generates instantaneous data suitable for short-term forecasting by compatible lake models. Inductive modelling by evolutionary computation proves to be superior in developing operational models as demonstrated by 7-day-ahead forecasting models for the tropical cyanobacterium *Cylindrospermopsis* developed from four years of on-line monitored water quality data of the Wivenhoe Reservoir, Australia (Recknagel et al., 2014). Model operationality is of growing demand, as implied in the scope of international initiatives such as COST NETLAKE (https://www.dkit.ie/netlake/action-details) and GLEON (http://www.gleon.org/).

The present model ensemble for plankton community dynamics in Lake Kinneret (Israel) has been designed through inductive reasoning by evolutionary computation. It utilizes electronically-measurable predictor variables (EMPV) such as water temperature, pH, turbidity, electrical conductivity, dissolved oxygen and cascade-wise predicts: firstly, the total nitrogen to total phosphorus ratio TN/TP from EMPV; secondly, chlorophyta, baccillariophyta, cyanophyta and dinophyta from EMPV and predicted TN/TP; and thirdly, rotifera, cladocera and copepoda from EMPV, predicted TN/TP as well as from predicted chlorophyta, baccilariophyta, cyanophyta and dinophyta. The proposed approach opens the prospect to implement and run model ensembles for forecasting phyto- and zooplankton community dynamics in real-time mode.

2. Materials and methods

2.1. Data of Lake Kinneret

The meso-eutrophic monomictic Lake Kinneret (Sea of Galilee) is located in northern part of Israel and supplies 30–50% of Israel's freshwater. It covers an area of ~170 km^2 and its maximum depth ranges between 39 and 44 m due to seasonal changes in rainfall and water consumption. Lake Kinneret receives most of its water and allochthonous nutrients from the northern Jordan River.

The lake has been monitored regularly since 1969 on weekly and biweekly bases. In this paper we illustrate how the dynamics of major plankton components between 1988 and 1999 can be associated with electronically-measurable ecosystem variables (Table 1). Data on biological, chemical and physical parameters of Lake Kinneret were collected by the Kinneret Limnological Laboratory at a central station; data on chemical and hydrological parameters of the Jordan River were monitored by the Mekorot Water Company. In Lake Kinneret most of variables measured at several standard depths were averaged over the upper 10-m layer where most of the algal community develops. Monitoring of water temperature and estimation of epilimnion depth is described by Rimmer et al. (2011). Chemical parameters in Lake Kinneret and Jordan River were measured using standard procedure (APHA, 1998; Nishri, 2011). Phytoplankton and zooplankton abundances were estimated for the upper oxygenated layer and described by Zohary (2004) and Gophen (1972), respectively. Phytoplankton was categorized based on the following phyla: cyanophyta, dinophyta, chlorophyta, and, bacillariophyta (diatoms). Zooplankton was categorized using systematic affiliation (copepoda, cladocera and rotifera). Chemical analyses were performed by the Watershed Unit, Mekorot Water Company.

Table 1
Limnological variables measured at a central site of Lake Kinneret from 1988 to 1999.

	Water quality variables	Units	Mean ± SD
Electronically Measurable Predictor Variables (EMPV)	WT (mean surface water temperature)	°C	22.3 ± 5.1
	TURB (turbidity)	NTU	2.29 ± 0.98
	pH		8.6 ± 0.3
	DO (dissolved oxygen)	mg/L	9.25 ± 2.14
	COND (electrical conductivity)	μS/cm	445.5 ± 43.3
Output Variables	TN_TP (TN: TP ratio)		30.1 ± 12.2
	Chlorophyta	g/m^2	11.6 ± 9.5
	Cyanophyta	g/m^2	4.2 ± 9.3
	Dinophyta	g/m^2	57.9 ± 86.2
	Bacillariophyta	g/m^2	12.1 ± 32.5
	Copepoda	lnd/L	112.1 ± 84.8
	Cladocera	lnd/L	46.6 ± 35
	Rotifera	lnd/L	66.9 ± 109.6

2.2. Design of the model ensemble

Fig. 1 illustrates the design of the model ensemble that cascade-wise predicts TN/TP from electronically-measurable predictor variables (EMPV), then chlorophyta, baccilariophyta, cyanophyta and dinophyta from EMPV and predicted TN/TP, and finally rotifera, cladocera and copepoda from EMPV, predicted TN/TP and predicted chlorophyta, baccilariophyta, cyanophyta and dinophyta.

The underlying models are induced by the hybrid evolutionary algorithm HEA (Cao et al. 2013) that has been previously successfully applied for predictive modelling of cyanobacteria growth for a variety of lakes and rivers worldwide (e.g. Chan et al., 2007; Kim et al., 2007; Recknagel et al. (2008a,b)) as well as for the identification of ecological thresholds and relationships determining cyanobacteria growth (Recknagel et al., 2013, 2014). It combines genetic programming (GP) for structure optimizing and differential evolution (DE) for parameter optimization of evolved models (see Fig. 2). Alternative inductive modelling techniques either fail to perform fairly accurate short-term forecasting such as decision trees or fail to explicitly disclose models and thresholds such as artificial neural networks (Recknagel et al., 2002).

We use GP according to Koza (1992) to evolve the rule model structure. Since GP typically operates on parse trees rather than traditionally used bit strings, it suits well to evolve equations or IF-THEN-ELSE-rules for multivariate relationships. GP uses logic functions (FL = {AND, OR}), comparison functions (FC = { >, <, ≥, ≤ }) and arithmetic functions (FA = { +, −, *,/, exp, ln}), to represent IF-THEN-ELSE-rules as vector of multiple trees. Tree1denotes the IF condition with the function set: Ftree1 = FL ∪ FC ∪ FA, tree2 and tree3 respectively denote the THEN and ELSE branches with the function set Ftree2/tree3 = FA.

Figs. 3–5 illustrate exemplarily one crossover step by GP for the optimisation of the IF-trees of two IF-THEN-ELSE models for predicting TN/TP ratios. Figs. 3 and 4 show the tree representations by GP of model 1: IF (COND>393) OR (56.8-WT) ≥ 42.1 THEN TN/TP = 23/(WT-25.8)+(WT-37.6)/DO+43.6 ELSE TN/TP = 28.7, and model 2: IF (COND>382) AND (TURB<2.9) THEN TN_TP= (WT+(WT/exp(TURB)))*2 ELSE TN_TP = 191.6/WT+20.3. After crossover of the IF-trees of both models by GP at the crossover points highlighted in Figs. 3a and 4a, GP creates a new model 1 (Fig. 5a): IF (TURB<2.9) OR (COND<393) THEN TN_TP = 23/(WT-25.8)+(WT-37.6)/DO+43.6 ELSE TN_TP = 28.7, and a new model 2 (Fig. 5b): (b) IF (COND>382) AND (56.8-WT) ≥ 42.1 THEN TN_TP= (WT+(WT/exp(TURB)))*2 ELSE TN_TP = 191.6/WT+20.3.

Differential evolution (DE) has been proposed by Storn and Price (1997) as effective evolutionary algorithm for global optimization that extracts the differential information (i.e., information on distance and direction towards global optimum) from the current population of solutions and guides to further search. It doesn't require separate probability distribution that makes the scheme completely self-organizing. DE has been implemented in HEA for multi-objective optimization of model parameters as described by Cao et al. (in press).

The boot-strap scheme has been implemented in HEA that randomly selects r_{max} data-subsets for training (75%) and testing (25%) for each of which t_{max} generations of models are evolved. After r_{max} boot-strap runs it determines the overall "fittest model" of all generations evolved by principles of genetic programming and differential evolution. This research evolved $r_{max} = 100$ generations of models in each experimental run. 100 generations proved to be sufficient for minimising the RMSE and approximating global optima of the different modelling experiments conducted by a Corvus Supercomputer (SGI, Altix XE1300).

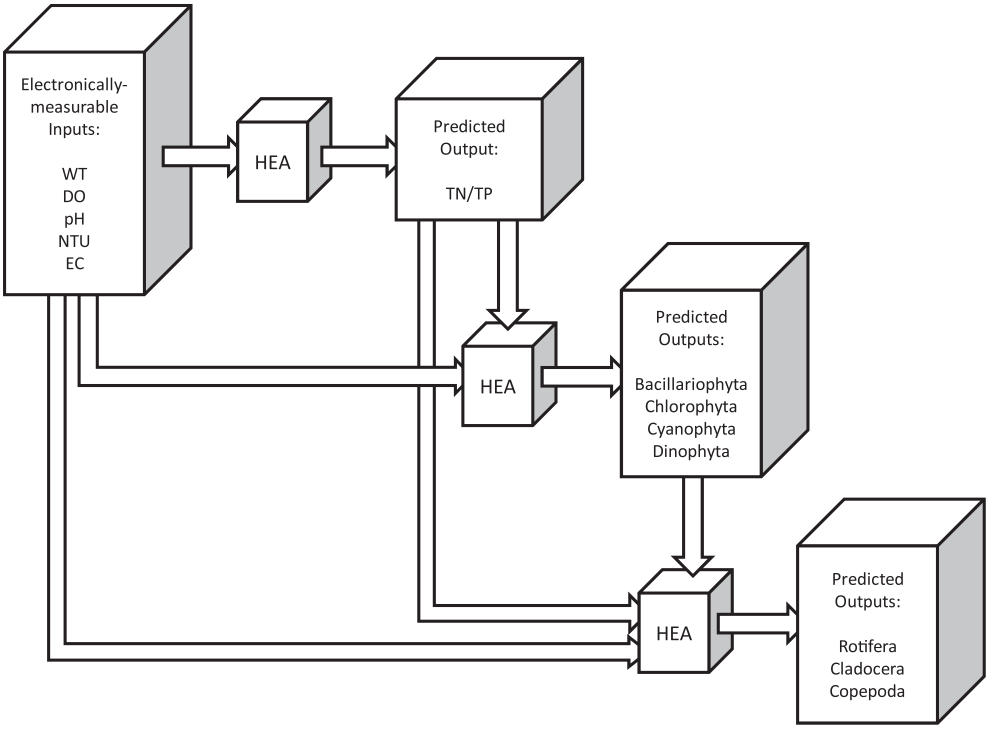

Fig. 1. Model ensemble for the prediction of TN/TP, the phytoplankton phyla bacillariophyta, chlorophyta, cyanophyta, dinophyta, and the zooplankton groups cladocera, copepoda, rotifera by HEA solely based on electronically-measurable predictor variables (EMPV).

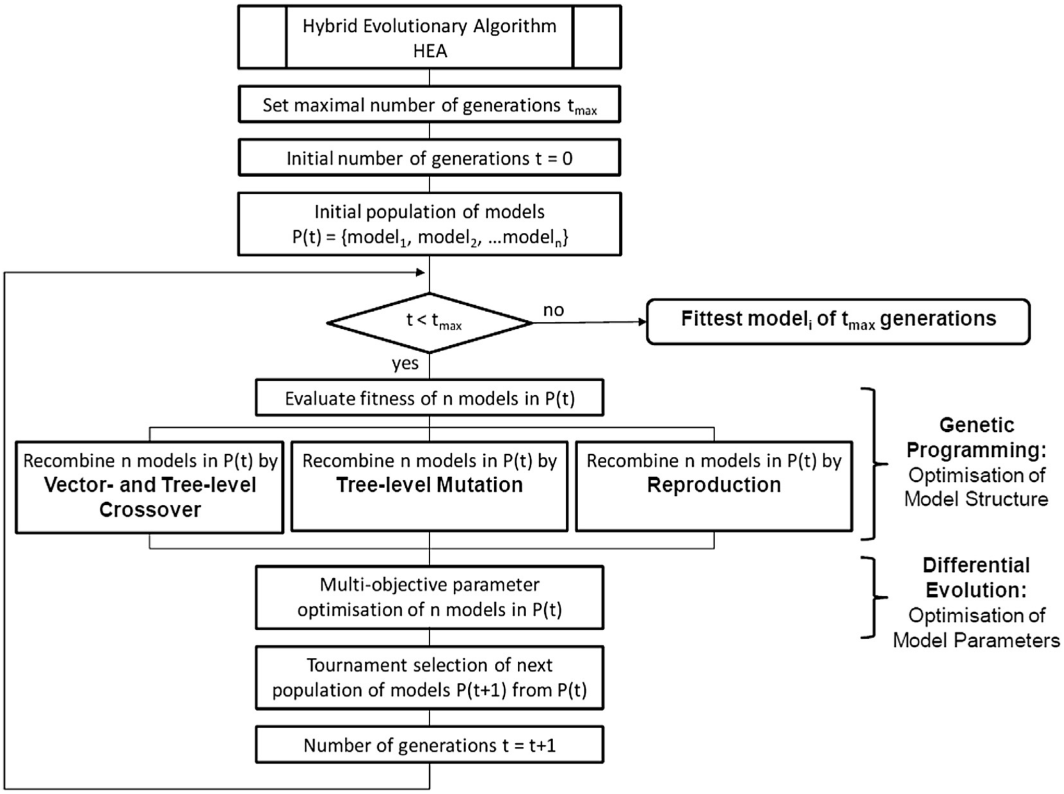

Fig. 2. Design and functioning of the hybrid evolutionary algorithm HEA (Recknagel et al., 2013).

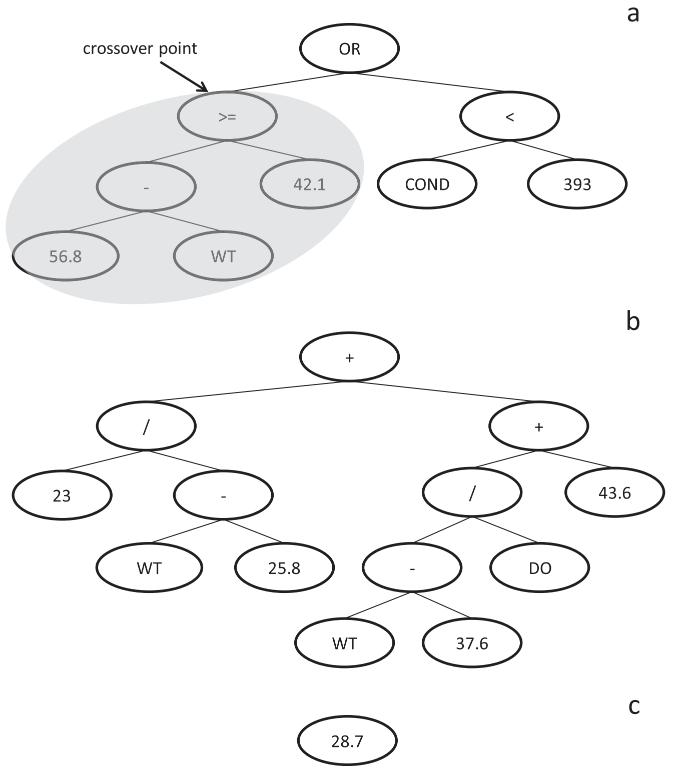

Fig. 3. IF-, THEN- and ELSE-tree representations of model 1 by genetic programming before crossover: a) IF (56.8-WT) ≥ 42.1 OR (COND<393), b) THEN TN/TP = 23/(WT-25.8)+((WT-37.6)/DO+43.6), c) ELSE TN/TP = 28.7.

The fitness of each model is evaluated by the root mean squared error (RMSE) between the measured training data and the predicted data by the model defined as:

$$\text{Fitness} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(\hat{y}_i - y_i\right)^2}.$$

The models' performances are measured by means of coefficients of determination (r^2). However in accordance with Bennett et al. (2013), the visual comparison between measured and calculated data proved to be the most relevant approach for the validation of models related to this highly complex data.

The software HEA automatically carries out sensitivity analyses for the input variables of each discovered model. It calculates output trajectories separately for each input range (mean ± SD) by keeping remaining input variables constant at mean values. Resulting sensitivity curves visualise the output trajectories in percentage terms (0–100%) within their range of each input.

HEA induces models from data patterns. It therefore requires cleansed and formatted cross-sectional or time-series data that are representative for the system to be modelled both in terms of number of observations and of relevance for the modelling purpose. Daily data interpolation is required in order to match dissimilar monitoring frequencies between physical, chemical and biological data, and to allow short-term forecasting for days ahead. Ecosystem evolution requires that models become regularly revised by updated data.

3. Results

Fig. 6 illustrates the best-performing model (Fig. 6a) and daily forecasting results of the TN/TP ratio (Fig. 6b) in Lake Kinneret for 11 consecutive years from 1988 to 1999. The model selected COND, WT, DO and TURB as predictor variables. It achieved a low coefficient of determination (r^2) of 0.28 and the predicted trajectories fail to approximate seasonal dynamics of TN/TP. However the trajectories clearly distinguish between years with high ratios such as 1992 and 1993 and years with low ratios such as 1989, 1997 and 1998.

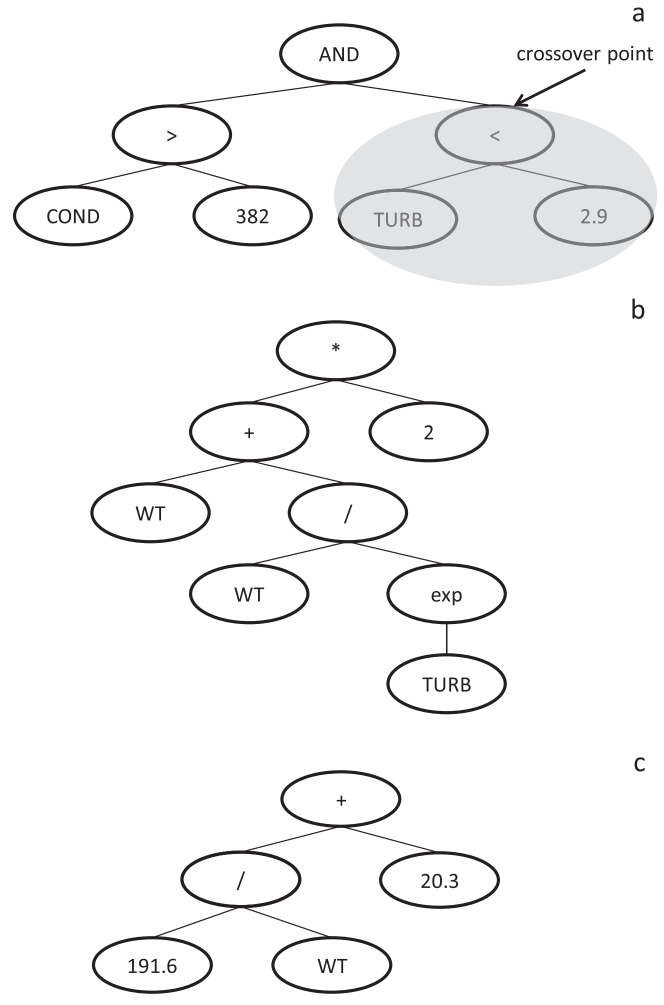

Fig. 4. IF-, THEN- and ELSE-tree representations by genetic programming of the model 2 before crossover:a) IF (COND>382) AND (TURB<2.9) b) THEN TN_TP = (WT + WT/exp(TURB))*2 c) ELSE TN_TP = 191.6/WT+20.3.

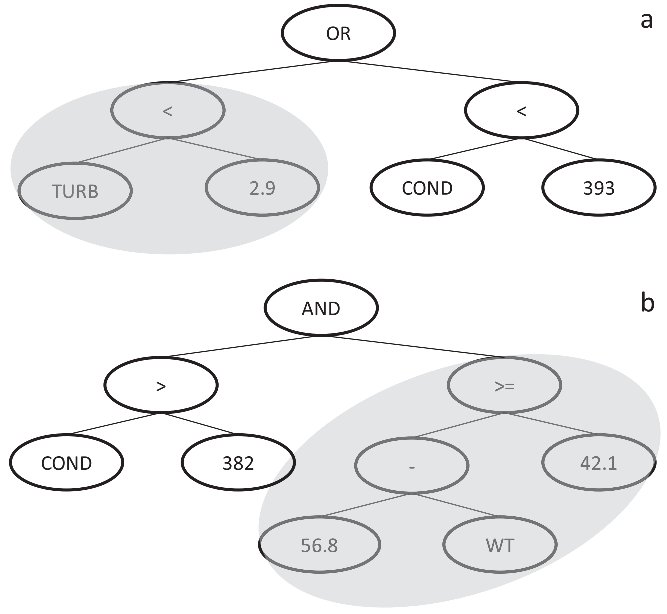

Fig. 5. Two IF-tree offspring representations by genetic programming after crossover between model 1 IF-tree and model 2 IF-tree: a) IF (TURB<2.9) OR (COND<393) (b) IF (COND>382) AND (56.8-WT) ≥ 42.1.

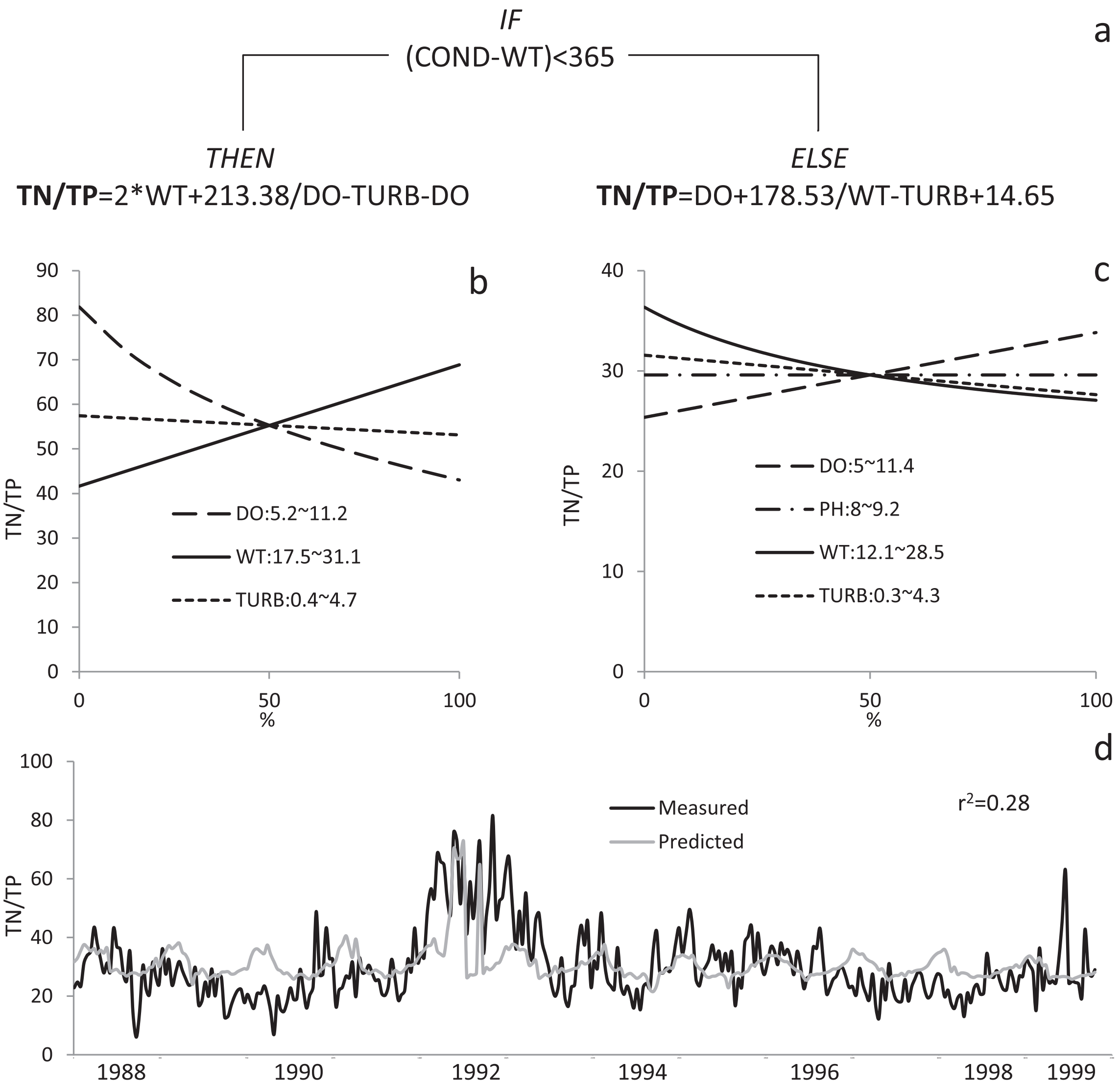

Fig. 6. TN/TP model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

The threshold condition of the model in Fig. 7a predicts chlorophyta biomass greater than 10 g/m^2 by DO and TN/TP conditions, and biomass lower than 12 g/m^2 by WT and TURB conditions. As the sensitivity analyses (Fig. 7b,c) suggests, chlorophyta concentrations are negatively correlated with DO, TURB and TN/TP > 20, and positively correlated with WT. Despite a relatively low $r^2 = 0.34$, the model predicts reasonably well the timing and magnitudes of peak concentrations of chlorophyta (Fig. 7d).

The dinophyta model (Fig. 8a) predicts high concentrations for WT ≤ 24.5 °C depending on COND and TURB conditions and low concentrations for WT > 24.5 °C depending on WT conditions. The model predicts for the majority of the ten years the timing and magnitudes of peak concentrations of chlorophyta well that is also reflected by a rather high $r^2 = 0.6$ (Fig. 8d).

The threshold condition of the cyanophyta model (Fig. 9a) predicts high concentrations up to 80 g/m^2 by TURB, COND and TN/TP, and lower concentrations up to 30 g/m^2 by pH, DO, TURB and WT conditions. As the sensitivity analyses (Fig. 9b,c) suggest, in this model cyanophyta concentrations are negatively correlated with COND and positively correlated with DO $\sim$ 8.5 mg/l and the remaining predictor variables. The model achieved an $r^2 = 0.42$ and predicts reasonably well the major peak concentrations in 1994 and

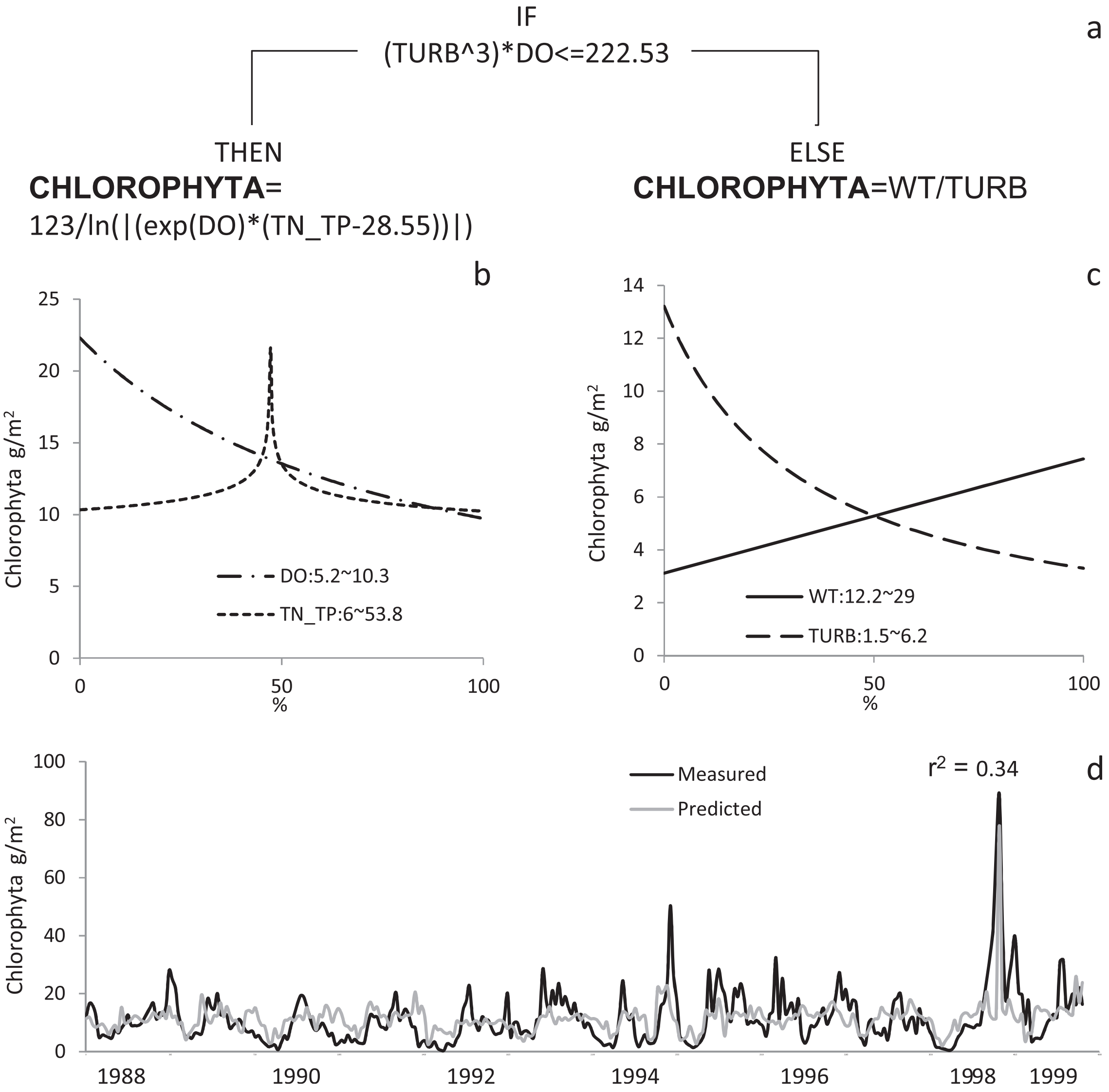

Fig. 7. Chlorophyta model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

1995 but underestimates the magnitudes of medium–high concentrations between 1997 and 1999 (Fig. 9d).

The bacillariophyta model (Fig. 10a) predicts high concentrations for pH $\leq$ 8.3 by DO, PH, WT and TURB conditions, and low concentrations up to 50 g/m^2 for PH > 8.3 by TURB conditions. As the sensitivity analyses (Fig. 10b,c) are revealing, in this model bacillariophyta concentrations are negatively correlated with WT > 17 °C and TURB, and slightly positively correlated with a DO and PH. The model achieved an $r^2 = 0.44$ and predicts well the timing and magnitudes of peak concentrations of bacillariophyta in 1988, 1996 and 1997 but underestimates peak concentrations in 1998 (Fig. 10d).

Fig. 11 documents the rotifer model that predicts high individual numbers depending on DO, dinophyta and cyanophyta, and lower individual numbers by DO and dinophyta. As the sensitivity analyses (Fig. 11b,c) reveal, the model correlates rotifers negatively with cyanophyta, neutrally with DO, and both, positively and negatively with dinophyta. The model achieved a coefficient of determination (r^2) of 0.45 and predicted well the peak density from 1994 to 1997 and in 1999, but mismatched peak numbers in 1988 and 1993.

The model in Fig. 12 predicts high or low individual numbers of cladocera depending on threshold condition cyanophyta of about 30 g/m^2. If the cyanophyta concentration is above this threshold

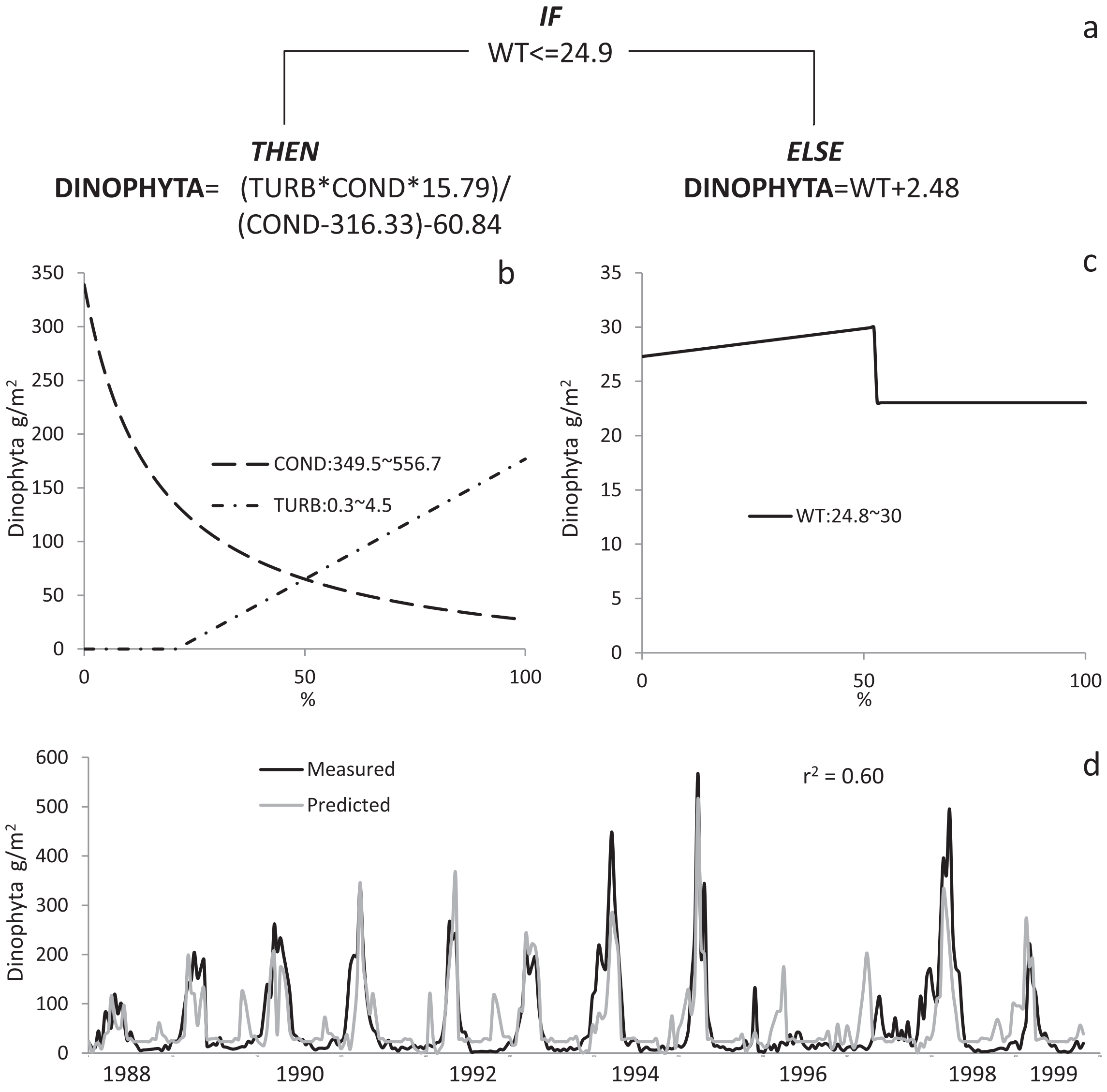

Fig. 8. Dinophyta model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

then predicted cladocera density is below 120 Ind/L, if it is below the threshold then predicted cladocera density is greater than 120 Ind/L. The input sensitivities in Fig. 12b,c indicate that cladocera is negatively correlated with chlorophyta concentrations and neutrally correlated with DO. Achieving a coefficient of determination (r^2) of 0.24 only, is reflected by poor prediction results between 1988 and 1994 failing to match observed dynamics of cladocera abundance. Results for the following years 1995–1999 are still poor whereby at least timing and magnitudes of the highest individual numbers observed in 1997 and 1998 are well predicted.

The model in Fig. 13 predicts high or low copepod density depending on the alternative threshold conditions cyanophyta of 3.6 g/m^2 or chlorophyta of 6 g/m^2. If biomasses of the two phyla don't exceed these threshold conditions, then predicted copepoda density is lower than 160 Ind/L, otherwise are higher than 160 Ind/L. The input sensitivities in Fig. 13b,c suggest that copepoda is negatively correlated with DO and PH, and positively correlated with COND and bacillariophyta biomass. The model achieves an $r^2 = 0.41$ and matches well observed dynamics of copepoda density between 1994 and 1999 with less satisfying results for the years before.

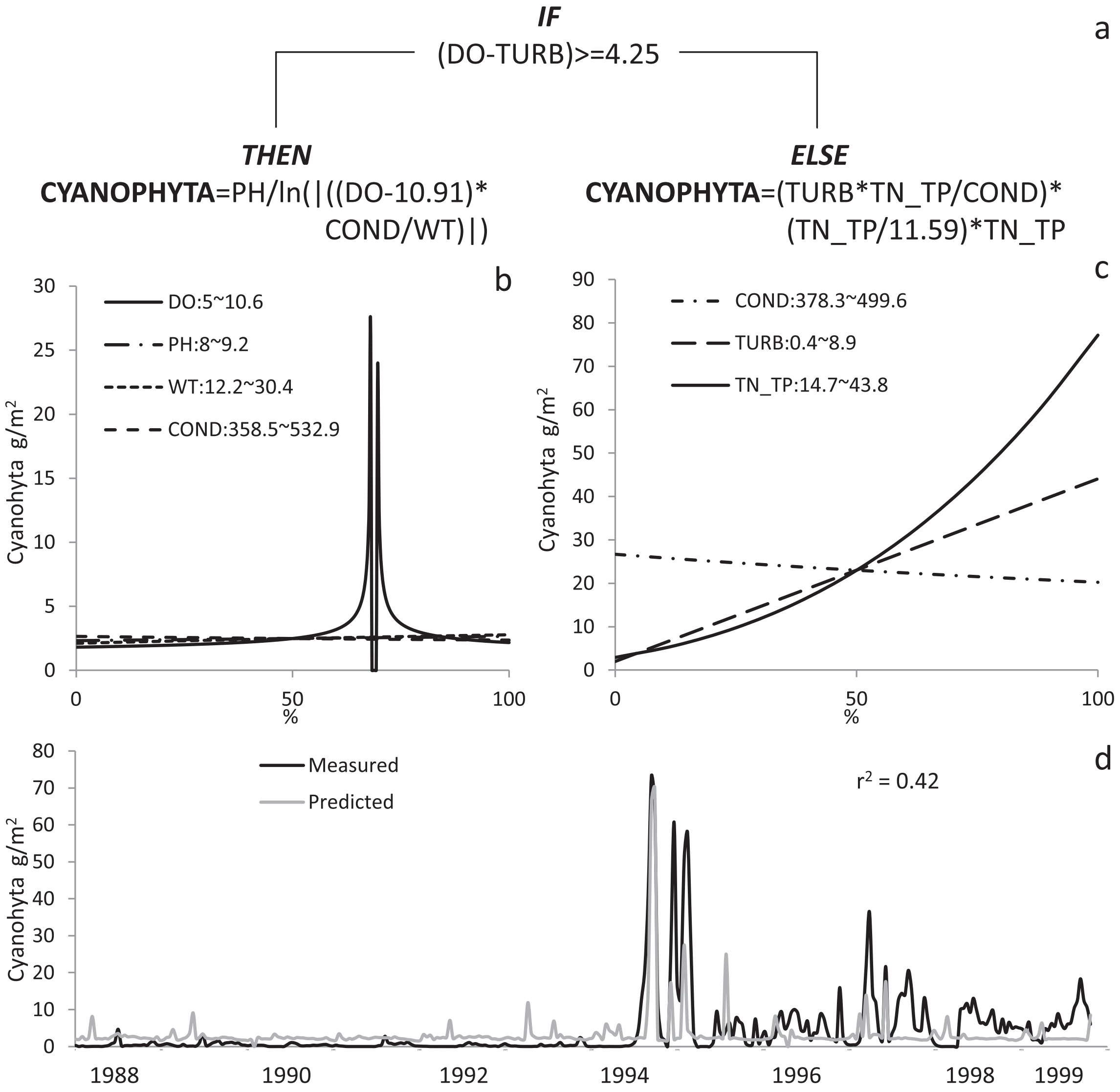

Fig. 9. Cyanophyta model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

4. Discussion

The model ensemble evolved from electronically-measurable predictor variables (EMPV) by the hybrid evolutionary algorithm HEA has demonstrated its capacity to simulate seasonal and interannual plankton community dynamics in Lake Kinneret over 10 years. The model for TN/TP had a relatively low coefficient of determination (r^2) of 0.28 but still matched major features of seasonal and interannual dynamics. It is interesting that the model suggests that TN/TP is strongly correlated with DO and WT pointing at seasonal differences in external nutrient loadings, which can be associated with microbial nutrient recycling and nutrient uptake by phytoplankton.

The models for the four phytoplankton phyla showed reasonably good validity in terms of forecasting timing and magnitudes of peak biomasses of the phyla over the ten-year period. The best result has been achieved for dinophyta with an $r^2 = 0.6$ that is based on a model limiting high growth events by the threshold condition WT ≤ 25 °C. Recknagel et al. (2013) discovered a slightly higher water temperature of 25.5 °C as threshold above which the

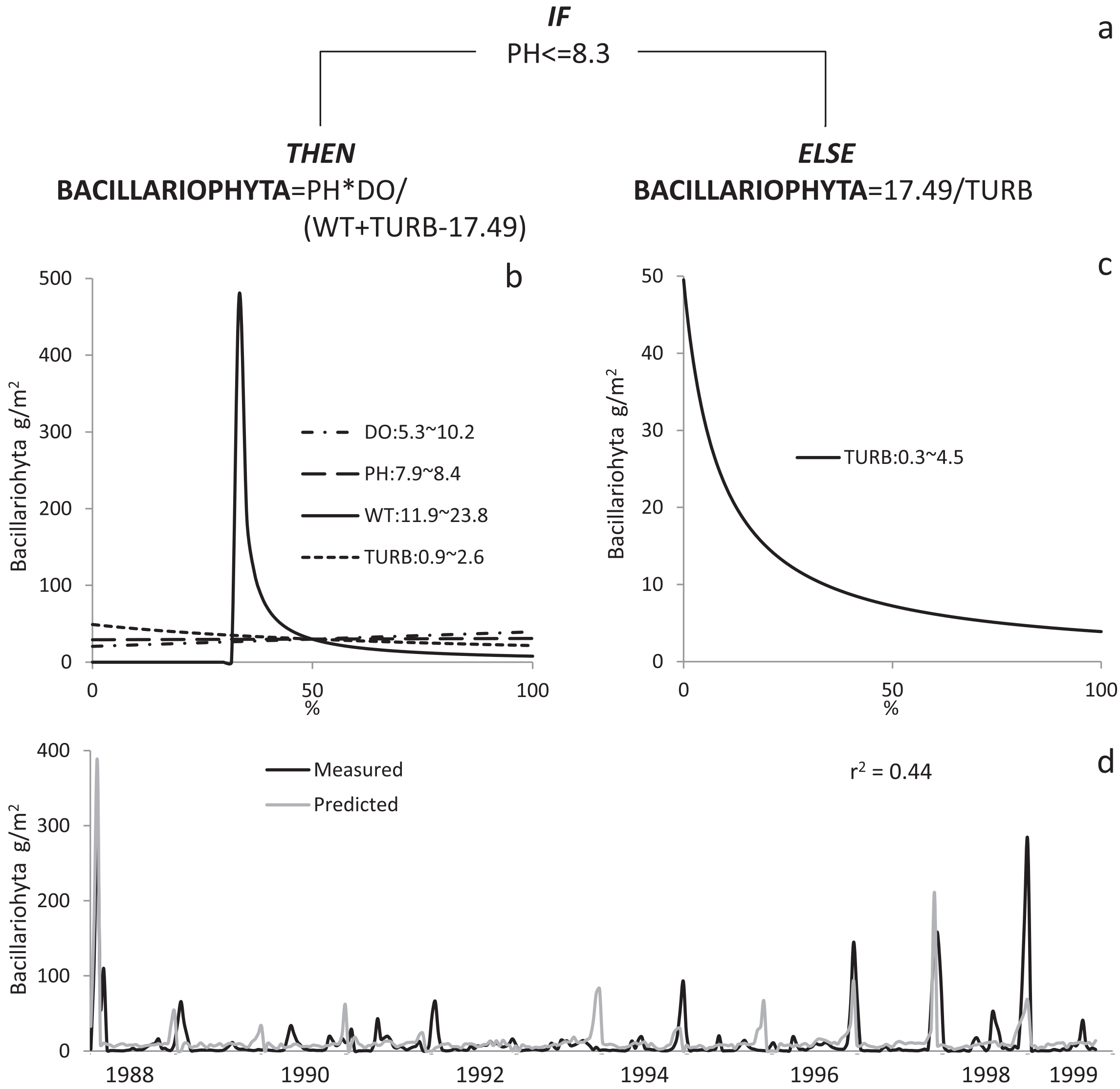

Fig. 10. Bacillariophyta model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

development of dinophyta biomass is inhibited, and explained it by findings of Berman-Frank et al. (1994), that unrestrained metabolism is susceptible to warming because at an increased temperature the supply of CO_2 for photosynthesis in water to the cell surface is reduced. Therefore, warmer waters increase the rate at which resources are needed for both routine metabolism and for achieving maximum growth rate (Atkinson et al., 2003).

The chlorophyta model suggested strong negative correlations with both, DO and TURB. The finding for TURB may relate to inhibitive underwater light condition by turbid waters during rain events or shading effects during algal mass developments. The sensitivity analysis also reveals a positive correlation between chlorophyta and TN/TP ratios up to 20, but inhibitory effects of higher ratios. This finding may hint to periods of nitrogen limitation as observed by Berman et al. (1995). The sensitivity analysis displayed a strong inhibitory correlation of bacillariophyta at $WT > 17$ °C, i.e. at time when the lake becomes thermally stratified. Recknagel et al. (2013) linked similar findings to bacillariophyta's dependence on lake turnover in winter and spring for minimizing biomass losses by cell sedimentation.

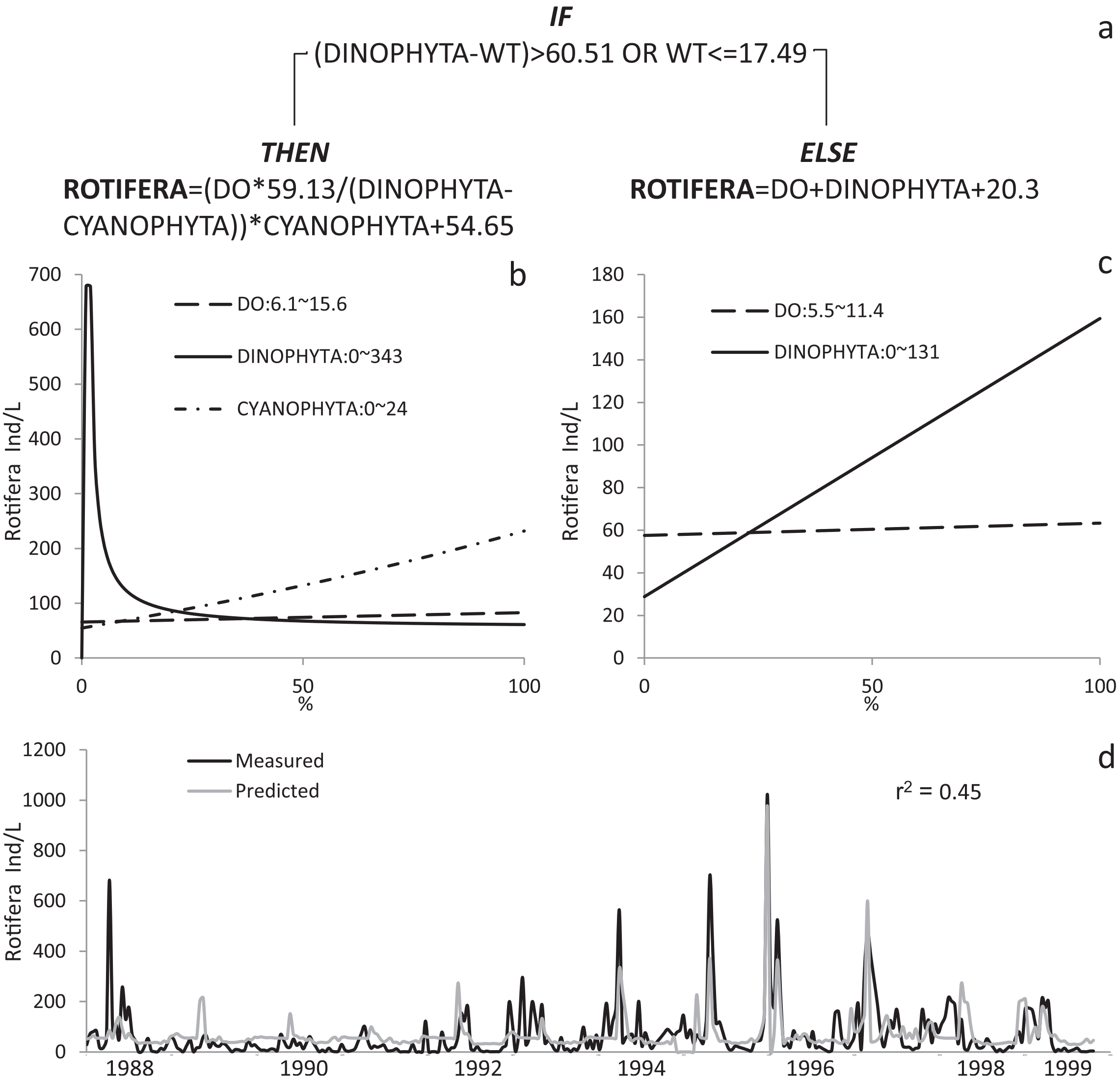

Fig. 11. Rotifera model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

The models for three zooplankton groups predicted well seasonal and interannual dynamics of rotifers and copepoda but failed to achieve valid prediction results for cladocera. A positive correlation of rotifers was observed for dinophyta concentrations up to ~130 g/m^2 that turned inhibitory at concentrations above 130 g/m^2. Strong negative correlations were found between cladocera and chlorophyta possibly reflecting high grazing losses of chlorophyta at times of high cladocera densities.

5. Conclusions

Lake modelling as tool for synthesizing complex limnological data, elucidating ecological relationships and informing sustainable management has a long history starting in the 1960s with the first inductive lake model by Vollenweider (1976), and the first deductive lake ecosystem model by Parker (1968). Since then hundreds of lake models have been developed by gradually sophisticating inductive and deductive methods that, however, are still just used for 'sand-box game' like experiments on ever-advancing computer platforms.

The present approach addresses the demand for operational models in view of the rapidly advancing *in situ* sensor technology that puts lakes into online surveillance mode. It demonstrates that evolutionary computation has the capacity to develop model ensembles for lake ecosystems that are solely based on *in situ* predictor variables. The model ensembles can be implemented and

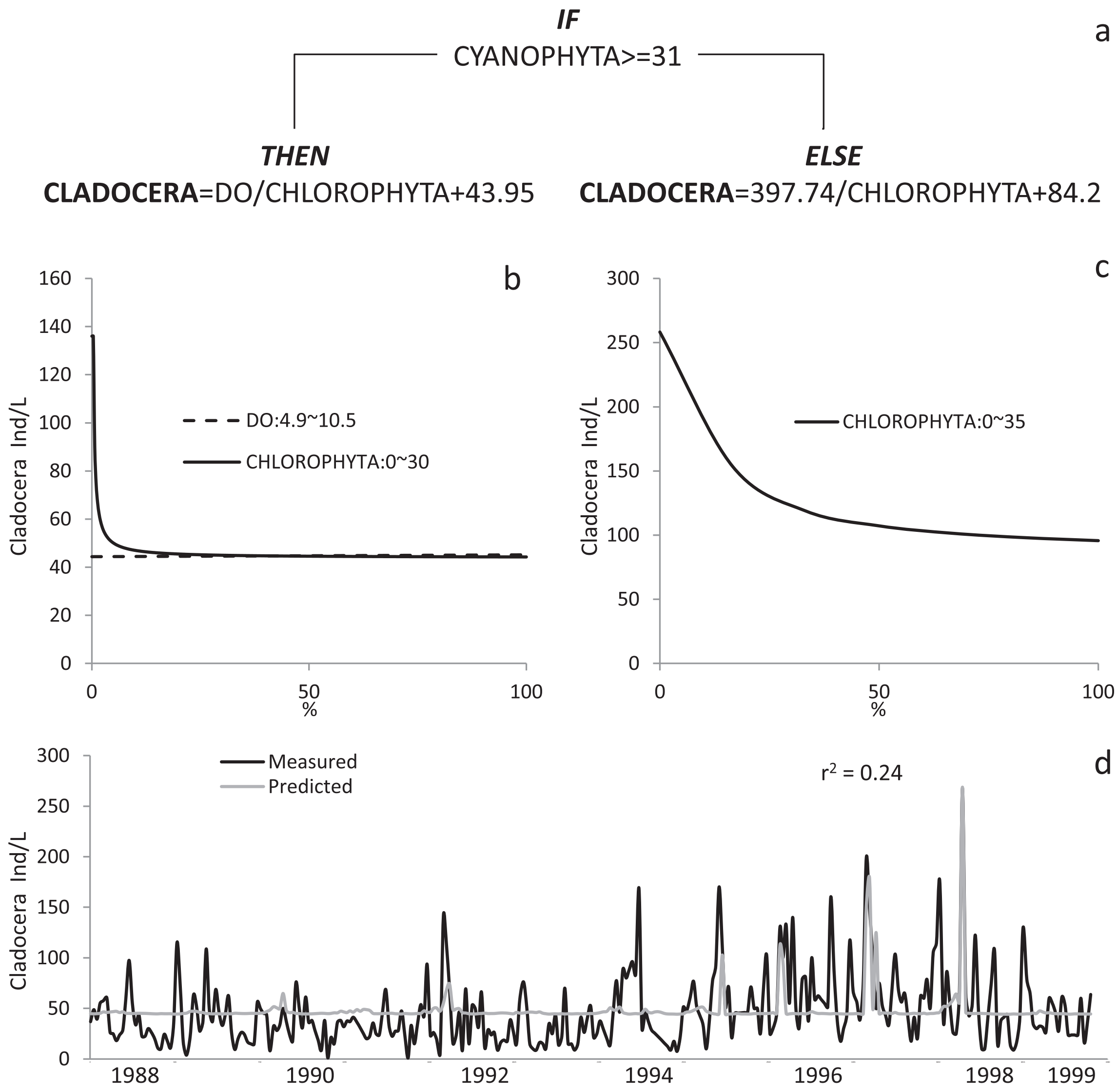

Fig. 12. Cladocera model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

operated in real-time mode serving as operational tools for lake management.

Future research will test opportunities to improve the validity of model ensembles by including NO_3—N, PO_4—P and SiO_2 models, and to develop species-specific phyto- and zooplankton models based on electronically-measurable predictor variables whereby inter-specific competition and mutualism will be considered in the model design.

Even though this research has demonstrated that model ensembles induced by HEA allow the prediction of interannual and seasonal dynamics of complex plankton communities by means of just five electronically-measurable water quality parameters, the models' accuracy does not yet suit their use as operational tools for early warning. To further improve the accuracy of induced models, the hybrid evolutionary algorithm HEA is currently upgraded by the integration of the *Strength Pareto*

Evolutionary Algorithm SPEA (Zitzler et al. 2003) that will perform multi-objective optimisation of both, the fitness function (RMSE) and the coefficient of determination (r^2).

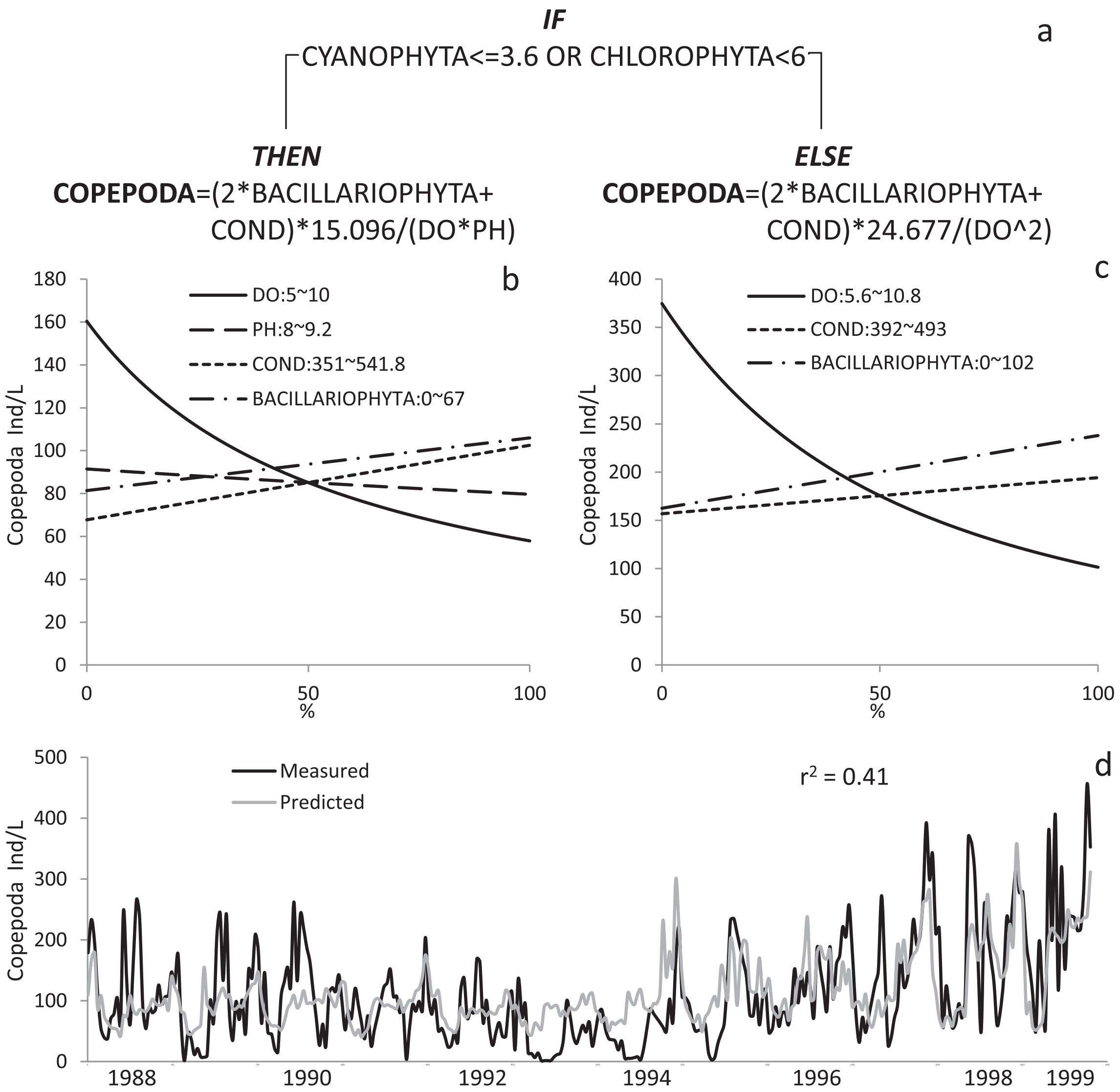

Fig. 13. Copepoda model: a) Rule structure; b) THEN-branch input sensitivity; c) ELSE-branch input sensitivity; d) Model validation.

Acknowledgements

This research has been funded by the Australian Research Council (ARC-L0990453). The authors thank anonymous reviewers for valuable comments that have significantly improved the manuscript.

References

APHA, 1998. Standard Methods for the Examination of Water and Wastewater. American Public Health Association, Washington, DC.

Arhonditsis, G.B., Brett, M.T., 2005. Eutrophication model for Lake Washington (USA) Part I. Model description and sensitivity analysis. Ecol. Model. 187, 140–178.

Atkinson, D., Ciotti, B.J., Montagnes, D.J.S., 2003. Protists decrease in size linearly with temperature: ca. 2.5% °C. Proc. Royal Soc. Lond. 270, 2605–2611.

Berman, T., Stone, L., Yakobi, Y.Z., Schlichter, M., Nishri, A., Pollingher, U., 1995. Primary production and phytoplankton in Lake Kinneret: a long term record (1972–1993. Limnol. Oceanogr. 40, 1064–1076.

Berman-Frank, I., Zohary, T., Erez, J., Dubinsky, Z., 1994. CO_2 availability, carbon anhydrase, and the annual dinoflagellate bloom in Lake Kinneret. Limnol. Oceanogr. 39 (8), 1822–1834.

Benndorf, J., Recknagel, F., 1982. Problems of application of the ecological model SALMO to lakes and reservoirs having various trophic states. Ecol. Model. 17, 129–145.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, G.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, St, Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20.

Cao, H., Recknagel, F., Orr, P., 2013. Enhanced functionality of the redesigned hybrid evolutionary algorithm HEA demonstrated by predictive modelling of algal growth in the Wivenhoe Reservoir, Queensland (Australia). Ecol. Model. 252, 32–43.

Cao H., Recknagel F. and Orr P., Parameter optimization algorithms for evolving rule models applied to freshwater ecosystems, IEEE Trans. Evol. Comput. (in press)

Cao, H., Recknagel, F., 2009. Hybridisation of process-based ecosystem models with evolutionary algorithms: multi-objective optimisation of process and parameter representations of the lake simulation library SALMO-OO. In: Jorgensen, S.E., Chon, T.-S., Recknagel, F. (Eds.), 2009. Handbook of Ecological Modelling and Informatics. WIT Press Southampton, Boston, pp. 169–185.

Chan, W.S., Recknagel, F., Cao, H., Park, H.-D., 2007. Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. Water Res. 41, 2247–2255.

COST NETLAKE. https://www.dkit.ie/netlake/action-details.

Gal, G., Hipsey, M.R., Parparov, A., Wagner, U., Makler, V., Zohary, T., 2009. Implementation of ecological modelling as an effective management and investigation tool. Ecol. Model. 220, 1697–1718.

GLEON. http://www.gleon.org/.

Gophen, M., 1972. Zooplankton distribution in Lake Kinneret (Israel). Isr. J. Zool. 21, 17–27.

Hamilton, D.P., Schladow, S.G., 1997. Prediction of water quality in lakes and reservoirs. Part 1 – model description. Ecol. Model. 96 (1–3), 91–110.

Kim, D.-K., Cao, H., Jeong, K.-S., Recknagel, F., Joo, G.-J., 2007. Predictive function and rules for population dynamics of Microcystis aeruginosa in the regulated Nakdong River (South Korea), discovered by evolutionary algorithms. Ecol. Model. 203, 147–156.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA.

Mao, J., Chen, Q., Chen, Y., 2008. Three-dimensional eutrophication model and application to Taihu Lake, China. J. Environ. Sci. 20, 278–284.

Nishri, A., 2011. Long-term impacts of draining a watershed wetland on a downstream lake, Lake Kinneret, Israel. Air, Soil. Water Res. 4, 57–70.

Park, R.A., et al., 1974. A generalised model for simulating lake ecosystems. Simulation, 33–50.

Parker, R.A., 1968. Simulation of an aquatic ecosystem. Biometrics 24 (4), 803–821.

Recknagel, F., Bobbin, J., Wilson, H., Whigham, P., 2002. Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. J. Hydroinform. 4 (2), 125–133.

Recknagel, F., Cao, H., van Ginkel, C., van der Molen, D., Park, H., Takamura, N., 2008a. Adaptive agents for forecasting seasonal outbreaks of blue-green algal populations in lakes categorised by circulation type and trophic state. Verh. Intern. Ver. Limnol. 30 (2), 191–197.

Recknagel, F., Cetin, L., Zhang, B., 2008b. Process-based simulation library SALMO-OO for lake ecosystems. Part 1: object-oriented implementation and validation. Ecol. Inform. 3, 170–180.

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. Ecol. Model. 225, 70–86.

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.

Rimmer, A., Gal, G., Opher, T., Lechinsky, Y., Yacobi, Y.Z., 2011. Mechanisms of long-term variations of the thermal structure in a warm lake. Limnol. Oceanogr. 56, 974–988.

Storn, R., Price, K., 1997. Differential evolution––a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11, 341–359.

Vollenweider, R., 1976. Advances in defining critical loading levels for phosphorus in lake eutrophication. Mem. Ist. Ital. Idrobiol. 33, 53–83.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V., 2003. Performance assessment of multiobjective optimizers: an analysis and review. IEEE Trans. Evol. Comput 7 (3), 117–132.

Zohary, T., 2004. Changes to the phytoplankton assemblage of Lake Kinneret after decades of predictable repetitive pattern. Freshw. Biol. 49, 1355–1371.