Ecological Modelling 291 (2014) 205–217

Contents lists available at ScienceDirect

Ecological Modelling

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

Improvement of complex and refractory ecological models: Riverine water quality modelling using evolutionary computation

MinHyeok Kim[a], Namyong Park[b], R.I. (Bob) McKay[b], Haisoo Shin[b], Yun-Geun Lee[b], Kwang-Seuk Jeong[c,d], Dong-Kyun Kim[e,*]

[a] Institute of Computer Technology, Seoul National University, Seoul 151-721, Republic of Korea
[b] Department of Computer Science & Engineering, Seoul National University, Seoul 151-721, Republic of Korea
[c] Department of Biological Sciences, Pusan National University, Busan 609-735, Republic of Korea
[d] Institute of Environmental Technology & Industry, Pusan National University, Busan 609-735, Republic of Korea
[e] Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario M1C 1A4, Canada

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 24 July 2014
Accepted 25 July 2014

Keywords:
Ecological model
River process
Parameterisation
Optimisation
Evolutionary algorithms
Model improvement

ABSTRACT

Complex environmental models have frequently suffered from large discrepancies between prediction and reality, inaccurate quantification of multivariate parameters, and difficulties in dealing with non-linearities. We introduce an interdisciplinary project combining an ecological river-process model and evolutionary optimisation of model parameters, resulting in tools for more effective water resource management. The aim is to more tightly integrate the expert's knowledge and the evolutionary system through an iterated cycle of knowledge refinement and evolutionary search. This requires new methods to specify the expert knowledge in ways that can be integrated into the search. We used an evolutionary algorithm to optimise the multivariate values of the model parameters while retaining their acceptability, verifying that their ranges and values were consistent with ecological knowledge and constraints. The best model had a significantly lower predictive error than the initial process model parameterised from literature estimates. Its error was also over 50% less than those of the purely empirical modelling methods of linear regression and neural network learning. We conclude that combining process knowledge with evolutionary learning can play an important role in ecological modelling.

1. Introduction

Ecological modelling problems frequently combine coarse datasets and weak domain theories (Shan et al., 2006). Ecosystem processes embody enormous complexity in both the structure and functional mechanisms of the system. Striving to model these complex and nonlinear systems, we inevitably introduce simplifications and approximations. Modelling efforts have often aimed to do so without compromising the predictive or explanatory power of the model.

In general, ecological modelling encompasses three types of mathematical or computational methods. The first is process (or mechanistic) modelling. A model is built up from known processes, and parameter values are determined based on the best available knowledge, with perhaps some judicious parameter adjustment at the end of the modelling process to obtain better fit. In this method, the modeller's knowledge and expertise are paramount, and available data is mainly used to validate the model (Brown and Barnwell, 1987; Pei and Ma, 2002). It can work well for simpler model structures that embody a relatively small number of model parameters. As the number of parameters increases, it becomes increasingly likely that some of the expertise-derived parameter values are sub-optimal, so that validating the model becomes more difficult.

The second is heuristic modelling such as machine learning. Some form of machine learning – for example neural network learning (Recknagel, 2001; Yao and Liu, 2001) or genetic programming (Whigham and Recknagel, 2001a; Peterson et al., 2002; Jeong et al., 2003; Kim et al., 2007b) – is used to generate a model from data. The data alone determines the model, which is highly data-driven. These approaches have an important limitation: the amount of data required to learn a model of a given complexity. Since fairly limited data is typically available in ecosystems modelling, this imposes stringent restrictions on the complexity of the models that can be learnt.

The third method is a variant of process modelling, in which the model itself is built from the best available knowledge, but

* Corresponding author. Tel.: +1 4162084878.
E-mail address: dkkim1004@gmail.com (D.-K. Kim).

http://dx.doi.org/10.1016/j.ecolmodel.2014.07.021
0304-3800/

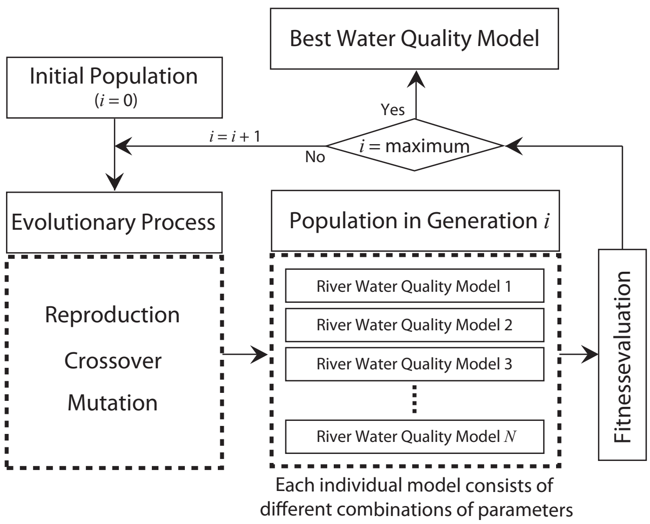

Fig. 1. Conceptual diagram of the use of evolutionary parameter optimisation in river water quality modelling.

the parameter values, instead of being estimated from knowledge, are optimised to give the best fit to the available data (Whigham and Recknagel, 2001b; Cho et al., 2004; Cao et al., 2008). While the second method may sometimes give greater predictive power, it cannot always guarantee greater explanatory power. This third method balances the influences between expertise about the model structure and data observation of the real ecosystem – the model is determined from expertise but parameter values are estimated from observed data.

Although these three methods are very different, they share one important characteristic: limited interaction between the expertise of the ecosystems modeller and that of the computer system developer. An increased use of informed expertise could play an important role in determining both predictability and interpretability of the model. In this context, some modelling studies using Bayesian approaches have generated results allocating different confidences to predicted values based on prior distributions specified for model parameters (Gelman, 2006; Wellen et al., 2014). However their emphasis is on finding relevant parametric bounds within assessed uncertainties, thus differing from our aim of better prediction.

The overarching goal of our research is to obtain an ecologically sound model through cooperation between an expert and a computationally robust system. The knowledge of the ecosystems modeller can influence the behaviour of the optimisation/learning system in intricate ways, and conversely, it is possible to increase the level of self-tuning of the system. This deeper interaction can generate more detailed, and hopefully more accurate, models. In this paper, we investigate how this can benefit modelling methods of the third type noted above.

This paper focuses on the combination of process-based modelling and data-driven parameter estimation. It presents model improvement by use of evolutionary algorithms in parameter adjustment of the process model (i.e. the third type of modelling described above, Fig. 1). We specifically address a water quality modelling program implemented in the lower Nakdong River, Korea. Our study aims to generate better models to predict plankton dynamics in a river ecosystem using evolutionary methods. We strive to improve an existing process-based model through adaptive implementation via evolutionary algorithms. The model we develop is based on generic limnological knowledge of a freshwater ecosystem. The methodological techniques are general, and can be readily extended to other problems. We emphasise that the underlying philosophy is adaptable to a much wider range of ecological modelling problems.

2. Background of research

2.1. Eutrophication and water resource management in fresh waters

Large open freshwater ecosystems contain numerous internal components, but are also affected by unpredictable external forces (Moss, 1998), both natural (e.g. weather variations) and anthropogenic (land use changes, dams and barrages etc.). A river ecosystem is generally seen as a very ambitious domain for modelling. As a consequence of eutrophication of freshwater ecosystems, algal blooms have become ubiquitous in favourable conditions. Rivers around the world are subject to increasing development, and subsequent algal proliferations have become a major concern in many countries. To resolve these problems, establishing guidelines and assisting decision-making through modelling is one of the most promising options for water resources management (Chapra, 1997).

Effective ecosystems management requires robust and reliable predictive models of ecological phenomena. It is almost impossible to manage a river ecosystem effectively without understanding the potential effects of management decisions (Calow and Petts, 1992). In the case of algal blooms, we need to model the effects both of regulatory management – e.g. decisions on water discharge from dams, or controls on nutrient export – and of shifts in Nature – e.g. changes in precipitation levels and timing as a result of climate change. While currently available algal bloom models have often helped to elucidate the ecosystem properties and dynamics, they may be unsatisfactory in terms of structural complexity. The complexity, even of the known system processes, is far beyond what we can hope to model. Thus models can only approximate what we think are the most important influences; but this leaves us hostage to fortune: if we are even slightly wrong in what we choose to model, our models will not perform accurately, as indeed we often find.

Ecological models provide the capability to explain and predict ecosystem dynamics, ranging from specific components to system structure. Both quantitative and qualitative properties of the data sets are crucial in determining the performance and robustness of the ecological models. In building the process structure of the ecological model, the most straightforward method is to base it directly on expert knowledge. However a knowledge based approach does not guarantee an effective model. The data may be too coarse in quality and/or quantity to be used without treatment, and the knowledge used to generate the process may be inaccurate – or more commonly, may not use the most suitable abstractions (Shan et al., 2006).

2.2. Necessity of water quality prediction and evolution of the predictive models in a complex ecosystem

Freshwater ecosystems fall naturally into two groups, lentic and lotic, based on the flow rate of the water body. Lakes, the extreme form of lentic systems, are one of the most popular domains for water quality process modelling. Notably, biogeochemical models have played a key role in lake ecosystem research, and have been used to elucidate ecological patterns from the perspective of system dynamics (Mieleitner and Reichert, 2006). The classical models of lake eutrophication started from empirical models (e.g. statistical regression analysis) and have developed into the mass balance approach (Chapra, 1997; Mooij et al., 2010). To date, a wide variety of lake ecosystem models have been proposed and developed.