Journal of Hydrology 531 (2015) 863–876

Contents lists available at ScienceDirect

Journal of Hydrology

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

A fuzzy rule based metamodel for monthly catchment nitrate fate simulations

S. van der Heijden *, U. Haberlandt

Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz Universität Hannover, Hanover, Germany

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form 14 August 2015
Accepted 12 October 2015
Available online 23 October 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Fuzzy rules
Metamodel
Nitrate
Catchment scale
SWAT

SUMMARY

The high complexity of nitrate dynamics and corresponding deterministic models make it very appealing to employ easy, fast, and parsimonious modelling alternatives for decision support. This study presents a fuzzy rule based metamodel consisting of eight fuzzy modules, which is able to simulate nitrate fluxes in large watersheds from their diffuse sources via surface runoff, interflow, and base flow to the catchment outlet. The fuzzy rules are trained on a database established with a calibrated SWAT model for an investigation area of 1000 km^2. The metamodel performs well on this training area and on two out of three validation areas in different landscapes, with a Nash–Sutcliffe coefficient of around 0.5–0.7 for the monthly nitrate calculations. The fuzzy model proves to be fast, requires only few readily available input data, and the rule based model structure facilitates a common-sense interpretation of the model, which deems the presented approach suitable for the development of decision support tools.

1. Introduction

Nitrate emissions from diffuse sources are one of the main causes of degraded water quality in many rivers in Europe. Therefore nitrate is an important topic in European and German legislature and many policies have been established at different levels in order to reduce nitrate input into water bodies (Sutton et al., 2011). Due to the complex nature of nitrate dynamics, the planning and evaluation of different conservation measures and the identification of their optimal distribution in space and time can only be carried out with the help of numerical models. Thus, a multitude of nitrate models exists at different scales and for different application purposes. The model types range from field scale applications (see e.g. Hansen et al., 1990; Franko et al., 1995), reactive transport models in groundwater (e.g. Prommer and Post, 2010) or surface waters (e.g. Chapra et al., 2008), to catchment scale models (see Section 2.2) and large scale balance models (e.g. Venohr et al., 2009; Gebel et al., 2010). Decision support for nitrate management generally requires large areas to be considered and the evaluation and optimisation of different measures and their combinations. However, data availability is often low and non-experts concerning modelling may be involved in the decision process. Thus several criteria can be identified which can enhance the suitability of a model for decision support. These include a simple and easy to understand model concept, parsimony in data requirements and computing times, a time step which allows capturing relevant processes and variations in the dynamics, the ability for different scenario analysis, and for regionalisation. To our knowledge, there is no model available which concurrently fulfils all of the above criteria. Process-based models are usually too complex and need large amounts of data which prohibit their application on large scales and/or in optimisation procedures, where a large number of model runs is required. More aggregated balance models on the other hand cannot deliver sufficient resolutions in space and time to evaluate individual protection measures.

The aim of this study is therefore to demonstrate a new model approach, which has the potential to fulfil all mentioned criteria. As a simple model approach, fuzzy rule based modelling (FRBM) was selected. Fuzzy rules, like all machine learning techniques, are usually trained on data. However, measured water quality data and many other influencing variables are rarely available in sufficient spatial and/or temporal resolution to allow for sound rule training. One possible solution to this problem is the metamodelling approach. Here, a well-calibrated process-based model is employed to generate an input–output database regarding the target variable, on which the fuzzy rules can be trained. Thus the

* Corresponding author.
E-mail address: vdheijden@iww.uni-hannover.de (S. van der Heijden).

http://dx.doi.org/10.1016/j.jhydrol.2015.10.039
0022-1694/

rule system becomes an emulator of the original model with regard to the target variable in question, in our case nitrate load in water bodies. Once the metamodel is trained it should be applicable independently from the process-based model.

Metamodelling thus is the reproduction of highly complex process-based models via a simple emulator (Bouzaher et al., 1993; Razavi et al., 2012a). The idea of metamodelling is that not all details of the original model are necessary to describe the system behaviour with regard to a specific goal. Thus, by concentrating only on relevant processes, a simplification in model structure and a reduction in necessary input variables is possible. Once trained, the metamodel no longer requires model parameters to be derived or estimated. All this promises a potentially more robust model with better regionalisation ability when compared to the original model. If only easily obtainable input variables are chosen, it is also possible that an application for unobserved catchments is facilitated. Additionally, when considering optimisation studies with possibly hundreds or even thousands of model runs, the original goal of metamodels, specifically the reduction in computational time, remains an important benefit.

All methods used for metamodelling are data-driven function approximation techniques. These methods have in common that they approximate the response surface of the original model by fitting simplified functions to previously calculated design sites. Therefore a metamodel is also called a response surface surrogate within the literature (Razavi et al., 2012b). Metamodels can be distinguished between exact emulators and inexact emulators. Exact emulators reproduce values at design sites without error by interpolating between sites. Inexact emulators approximate the whole response surface and thus introduce a varying bias at different design sites. Exact emulators are best applied when working with noise-free data, e.g. when the focus is on the best possible emulation of the original model (e.g. Kleijnen, 2005; Mullur and Messac, 2006). If the purpose of the metamodel however is to apply it to real world problems with measurement data containing noise, it is advisable to use inexact emulators as they are less sensitive to noise (e.g. Khu and Werner, 2003; Ratto and Pagano, 2010). Once training on data produced by the original model is complete, the purpose of the metamodel presented in this study is to simulate nitrate transport using real data, hence with fuzzy rule based modelling an inexact emulator has been chosen for this study.

Fuzzy rules have been applied successfully in hydrology and water resources management. For example, Haberlandt et al. (2002) developed a fuzzy model to simulate monthly nitrate leaching into groundwater. Jacquin and Shamseldin (2006) demonstrated a fuzzy model for rainfall–runoff process. Casper et al. (2007) used only two input variables to calculate runoff from a small catchment with expert-defined fuzzy rules. Haberlandt and van der Heijden (2007) showed that a combination of measurement and simulation data from many different sources can be used to successfully train fuzzy rules for annual nitrate leaching. Shrestha et al. (2007) also combined observed and simulated data for rule training to simulate daily nitrate concentrations in a small river. Lohani et al. (2011) compared FRBM to neural networks and linear transfer functions for daily rainfall–runoff simulations, with the fuzzy model proving superior for the selected catchment.

More details on FRBM are given in the next Section 2.1, followed by a short description of the SWAT model, which has been chosen as the process-based model for this study (Section 2.2). The study areas and available data are presented in Section 3 while Section 4 describes the precise steps in the creation process of our metamodel. Section 5 of the paper presents the results, which together with the whole model concept are discussed in Section 6.

2. Methods and models

2.1. Fuzzy rule based modelling

Fuzzy rule based modelling is based on Fuzzy Logic, which was introduced by Zadeh (1965). Since it is not possible to explain the method in detail here, the reader is referred to text books like Bárdossy and Duckstein (1995) for more information. The crucial point in Fuzzy Logic is that when defining a set of elements, it is not only possible for each element to belong completely to the set (membership = 1) or not at all (membership = 0), it is also possible to have partial membership with any real number assigned between 0 and 1. This corresponds well to how we handle the linguistic expression of real world phenomena. When speaking of heavy rainfall or high temperature there is no sharp numeric threshold value in rainfall intensity or degrees Celsius, which divides heavy from light rainfall or low from high temperature, rather it is a smooth transition. With this concept in mind, it is possible to define continuous fuzzy sets across the whole range of a variable of interest. Those defined fuzzy sets overlap and thus allow a seamless transition between linguistic steps, like for example very low, low, medium, high, and extremely high temperature (see Fig. 1). The fuzzy sets shown in the example are also called fuzzy numbers, since they fulfil the two conditions of normality (maximum membership value $\mu_A(z) = 1$) and convexity ($\mu_A(c) \geqslant \min(\mu_A(a), \mu_A(b))$ for all real numbers a, b, c with $a < c < b$). If all describing variables, as well as the target variable are classified in such a way, it is possible to define fuzzy if-then-rules of the pattern:

$$\text{IF } (x_1 \text{ is } A_{i,1}) \text{ AND } (x_2 \text{ is } A_{i,2}) \text{ AND } \ldots \text{ AND } (x_K \text{ is } A_{i,K}) \text{ THEN } (y \text{ is } B_i) \tag{1}$$

where x_j is the value of input variable j ($j = 1,\ldots,K$), y is the value of the dependent variable, $A_{i,j}$ is the fuzzy number for variable j and rule i ($i = 1,\ldots,M$), and B_i is the fuzzy response for rule i. Such rules have a direct verbal translation, for example "IF (temperature is high) AND (rainfall is low) THEN (discharge is low)" (see Fig. 2).

Many such rules can be found to describe different situations of the system in question. Together they constitute the fuzzy rule system, which, in a perfect case, is able to describe the complete system behaviour across the domains of the input and target variables. To an individual case, a fuzzy rule is said to be applicable if the membership value $\mu_{A_{i,j}}(x_j) > 0$ for all input variables j. Due to the overlapping of the fuzzy numbers, it is possible that several rules are applicable at the same time for one specific case. This allows the model to produce stepless output and is therefore required. A measure of applicability to a certain case is the degree

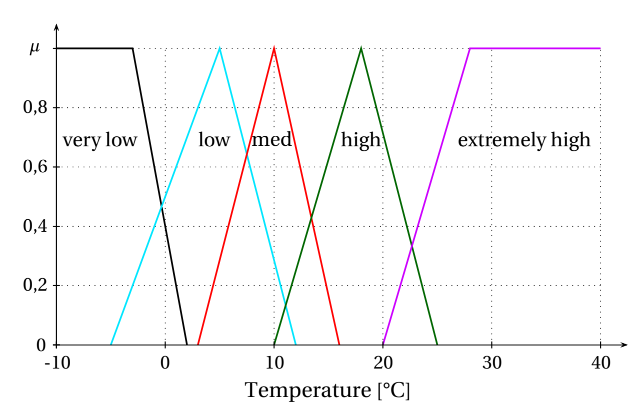

Fig. 1. An example of fuzzy number allocation to the domain of a variable. μ: membership value.