PERPUSTAKAAN UMP
0000068591

An Evaluation of the Potential of Adaptive Neuro-Fuzzy Inference System in Hydrological Modelling and Prediction

by

Ngahzaifa Ab Ghani, BSc.

24/11/14 PERPUSTAKAAN
G UNIVERSITI MALAYSIA PAHANG
No. Perolehan 068591
Tarikh 30 NOV 2012
No. Panggilan GB 1207 .N43 2012 Thesis

Thesis submitted to The University of Nottingham for the degree of Doctor of Philosophy

October 2012

Abstract

The use of data-driven modelling (DDM) in hydrological forecasting has been in practice since decades ago. Nevertheless, despite the ease of use, DDM approach has also been criticised due its 'black box' nature where the physical insights of the modelled processes are far from reach. Whilst hydrologists are craving for the insight, the operational modellers are and will always prefer an easily applicable method regardless of whether the model is able to deliver knowledge. Hence, a method that could fulfil the need of both would be a perfect solution. ANFIS (Adaptive Neuro Fuzzy Inference System), for its advantages of having linguistic representation of models has been the interests of both groups and have been successfully tested on a number of international catchments. It is however still unclear as to what extent is ANFIS able to deliver the required knowledge; how capable is ANFIS in modelling sediment-discharge; and what are the advantages and disadvantages of using ANFIS as a modelling tool. This thesis explores ANFIS capability in order to address these issues. The methods involved include creating synthetic datasets that mimic the sediment-discharge relationships; experimenting with different ANFIS parameter settings; and observing and analysing the behaviour of models with the help of statistical and graphical evaluations. The results highlight that ANFIS is capable to model most of the tested relationships, but the produced model is very dependent on the parameters applied when training the model. Wrong choice of parameters may lead to the production of models with good metrics but low transferability, or even worse, not transferable at all. As a conclusion, ANFIS can be used in sediment-discharge modelling with certain restrictions on training parameters but this is mostly applicable to the simple and common rating curves. More studies are needed in order to explore the potential of ANFIS to model complex sediment-discharge processes.

Table of Contents

List of Figures

List of Tables

Glossary and Abbreviations

Chapter 1

ABM	Agent based modelling
ANFIS	Adaptive neuro-fuzzy inference system
DDM	Data driven modelling
GA	Genetic algorithm
NN	Neural network
SRC	Sediment rating curve

Chapter 2

mf, MF	Membership function
Gaussmf	Gaussian membership function
Gbellmf	Generalized bell membership function
Pimf	'Pi' membership function
Psigmf	Product of Sigmoid membership function
Trapmf	Trapezoidal membership function
Trimf	Triangular membership function

Chapter 3, 4

e	Epoch (training iteration)
C-Q	Sediment-discharge
NE	Synthetic dataset with normal distribution on y-axis and exponential distribution on x-axis
NG	Synthetic dataset with normal distribution on y-axis and Gumbel distribution on x-axis
NU	Synthetic dataset with normal distribution on y-axis and uniform distribution on x-axis
O-P	'observed versus predicted'
SSC	Suspended sediment concentration
Set I (dataset)	Synthetic dataset without heteroscedasticity
Set II (dataset)	Synthetic dataset with heteroscedasticity
SSL	Suspended sediment load

Chapter 5

NL1	Nonlinear1 (dataset of second degree polynomial relationship)
NL2	Nonlinear2 (dataset of higher degree polynomial relationship)
NL3	Nonlinear3 (dataset of higher degree polynomial relationship with higher sinuosity index than NL2)
SL	Dataset of simple linear
Q_a	Artificial discharge
C_a	Artificial sediment
LQ	low discharge value
MLQ	Medium low discharge value
MHQ	Medium high discharge value
HQ	High discharge value

Chapter 1 - Introduction

1.1 Data driven modelling - from conventional to machine learning

Models are the abstraction of reality. In hydrology, models are of central importance because of two main reasons i.e. to understand the hydrological processes; and for the purpose of prediction. The types of models can be broadly grouped into three main categories i.e empirical, conceptual, and physically based models.

Between the late 15th century and the late 18th century, the only way of learning the hydrological processes was by physical in-lab experiments (Babovic, 2005). Hydrologists had developed physical artificial models that mimic the real processes to study the response of each hydrological variable involved. Each response or process is then described or represented by process description or mathematical equations. In recent years, with the availability of advanced computational tools, a physically based model is also presented in a digital form such as with animations or 3D models where the numerical calculations are also done by computers.

Despite having the advantage of 'real' control of both temporal and spatial variables and the improvement by computational advancement, this approach is still not always preferred by modellers. This is because physically based models are often complex thus require more computing time and effort (Yu, 2002). Moreover, physically based model has also been criticized due to scaling issues where the scale of measurement for many variables is usually incompatible with their use in hydrological models (Grayson et al., 1992) This type of models are more feasible for small catchment studies where the variables are small and well under control (Gan et al., 1991).

In its simplest form, a conceptual model is presented in the form of flowcharts in order to describe the modelled processes. Whilst this basic conceptual model is the simplest type of model and can be easily understood, it is one of the hardest to develop (Lane, 2003). The flowcharts are very useful to get some sort of general ideas and knowledge about the relationships involved in the particularly modelled processes. Because it is not transferable, this type of model cannot be used for prediction.

In hydrological modelling and prediction, the use of conceptual model is not limited to the basic processes description only. In fact, it is far more complex i.e. with the integration of complex mathematical description to describe each process or phase in the conceptual model and of course requires computational tool. In regards to the process description, the variables involved are usually not directly measurable and must be calibrated from the observed data (Beven and Binley, 1992).

Empirical models are obtained by establishing relationships between the observed parameters involved. Most commonly, the relationships are achieved by statistical methods. One of the popular empirical approach in hydrological modelling is the Sediment Rating Curve (Campbell and Bauder, 1940) that is used to predict suspended sediment from discharge value.

Any model can be based on either inductive or deductive modelling approach. Deductive approach is theory or knowledge driven and starts with using knowledge to develop hypothesis in order to control or restrict the production of model. Conceptual and physically-based models are based on this approach.

Contrary to deductive, inductive approach is data-driven and works by learning the data to achieve the model. It starts with data observations, establishing pattern/relationships on the observed data, developing hypothesis, and lastly producing the model (theory). All data-driven models including empirical models are based on this approach.

The fact that data-driven model (DDM) is easy to produce and does not require critical expertise, has contributed to its popularity. Moreover, since decades ago, DDM has improved from the conventional statistical methods to more sophisticated artificial intelligence (AI) (McCarthy, 1956) method, which combines more advanced statistics and requires computer processing power.

AI method is mostly inspired by natural processes. Artificial Neural Network (NN) (McCulloch and Pitts, 1943) was developed on the concept of human brain function, Genetic Algorithms (GA) on the behaviour of chromosomes in genetic science (Goldberg and Holland, 1988), and Agent Based Modelling (ABM) on the body cells behaviour where they function as the agents in human body (Bonabeau, 2002). NN, GA and ABM are just a few of many other available AI methods.

In hydrological modelling, NN is one of the AI methods that have been the attention of hydrologists because of its learning capability towards non-linear relationships. Although NN has been very popular in hydrological modelling with increasing success, it has nevertheless been criticised due to its 'black box' behaviour: it is not understandable in terms of physical parameters (Johannet et al., 2007).

1.2 The motivation of research

In 1993, Jang introduced Adaptive Network Based Fuzzy Inference System (ANFIS) (Jang, 1993) which has been considered more conceptually advanced than ANN. The main advantage of ANFIS over ANN is that it is developed by two parts: ANN to learn the data and model it; and 'fuzzy components' to describe the model. Being one step advanced than

ANN in terms of transparency, ANFIS is indeed a hope for hydrologists in getting the solution over the 'black-box' issues.

Interestingly enough, in most publications of ANFIS application on water resources modelling , majority of the studies concluded that ANFIS has performed very well, or even better, superior to other modelling methods compared in most cases. Moreover, ANFIS has been believed to be a tool that is capable of providing further insight into the process being modelled (Cobaner et al., 2009, Kisi, 2009, Kisi et al., 2009, Kisi et al., 2008, Sayed and Razavi, 2000, Sayed et al., 2003). The growing interests towards ANFIS in hydrological modelling especially suspended sediment and the promising results of testing has been one of the motivations of this research.

One of the areas that would benefit from ANFIS application is sediment forecasting. Sediment forecasting is important for many purposes such as for reservoir designs, water pollution controls i.e. for continuous monitoring of water quality in rivers and reservoirs, and flood-risk assessment. By in-field observation, sediment measurement is obtained using standardised samplers and sampling methods. Although this traditional method provides accurate and reliable measurement, the data is usually temporally sparse and very expensive to collect (Gray and Glysson, 2004).

The need for sediment forecasting is crucial. The failure to reach sediment information may lead to serious consequences. For example, a continuous observation of sediment can help in monitoring the stability of a dam thus can possibly avoid incidence such as damn collapse that is due to underlying sediment failure. The forecasting of sediment is also very useful to prevent flood as a flood event can be caused by accumulated sediment blocking the river course.

From the good impression on the performance of ANFIS as a modelling tool for hydrological problems (based on the publications cited before), perhaps, the use of ANFIS for sediment forecasting can help in managing water resources and preventing environmental problems caused by sediment such as flooding and water pollution.

1.3 Research aim and objectives

This research aims to provide enhanced knowledge that can be developed into theories and guides that will improve the efficiency of ANFIS application in hydrological modelling especially in suspended sediment-discharge relationships in operational and research level. The objectives of this research are:

Objective 1

To identify problems and issues in ANFIS modelling of suspended sediment-discharge by reviewing and critically discuss the progress and gaps in existing research.

Objective 2

To create a series of synthetic datasets in order to provide a controlled environment for further testing and investigations of the problems identified from Objective 1.

Objective 3

To explore the capability, pros and cons of ANFIS models with different modelling parameters by experimenting with the synthetic datasets.

Objective 4

To propose a set of guides in the form of rules and theories for ANFIS modelling of suspended sediment-discharge.

Objective 5

To test and evaluate the applicability of the proposed guides and theories by applying the guides and theories to (two) river case studies.

1.4 Thesis Overview

Chapter 2: ANFIS mechanism

Chapter 2 provides the explanation of ANFIS working mechanism and its main components. The explanation covers how an ANFIS model is trained and developed. To ease the understanding of the role of each component, a numerical example of the step-by step calculation is also included.

Chapter 3: Literature Review

Literature Review in Chapter 3 is developed by two main parts. The first part presents the review on suspended sediment processes and the relationships of suspended sediment-discharge that have been commonly found in hydrologic research. The second part critically reviews ANFIS application in suspended sediment-discharge modelling and its progress in hydrological research community. The gaps in previous research are defined in the form of research questions.

Chapter 4: Methodology

Chapter 4 presents an overall methodology of the research where it is formed based on the understanding and research gaps defined in Chapter 3. From the literature review, it was identified that in existing studies, the approach of suspended sediment-discharge modelling has been on trial and error basis. Model evaluations were mostly dependent to statistics alone, with hardly deep understanding on either hydrological nature of the relationships or ANFIS components itself. Hence a series of synthetic datasets with different complexity based on the understanding of suspended sediment-discharge relationships were created for ANFIS experiments. The experiments involved a number of different parameter settings of type of membership functions, number of membership functions and training epochs.

The methods of experiments, analysis and validation of results are also explained briefly in this chapter.

Chapter 5: Results and Analysis of synthetic datasets experiments

This chapter presents the results and analysis of models trained from the synthetic datasets. The results are presented in the form of model plots and statistics (tabulated and graphs). In the first part of analysis, ANFIS models of common suspended sediment–discharge relationships were compared and thoroughly analysed in terms of its capability to model the datasets with regards to issues including sensitivity to outliers, stability with the changes of training parameters, and local over-fitting. In the second part, the capability of ANFIS to model more complex suspended sediment-discharge processes was examined. This chapter is concluded with the summaries of the two parts of analysis that are formed into guides and theories in modelling sediment-discharge. These guides and theories will be tested in Chapter 6.

Chapter 6: Case Studies – Real world validation

In Chapter 6, the guides and theories from Chapter 5 are re-applied to the modelling of datasets of two catchments to validate their applicability in real world. The catchments involved are Rio Valenciano (Puerto Rico) and Low Moor (River Tees, Northern England). The chapter is concluded with the finalised theories and guides for ANFIS modelling of suspended sediment-discharge.

Chapter 7: Discussion and Conclusion

In this chapter, the outcome of the research is discussed in a broad perspective. The discussion is organised into five sections. First section reviews the findings of this research and how this research fills the gaps in existing studies defined in Chapter 3. In the second section, the impacts of this research to the previous existing studies and the society are discussed. The third, fourth and fifth sections provide the limitations of the research, the conclusion and a set of recommendations for future research.