UNIVERSITY
OF TECHNOLOGY
SYDNEY

Improving Reliability of Hydrological Flow Estimation using Hydroinformatics Approach

By Md Atiqzzaman

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Associate Professor Jaya Kandasamy and
Professor Saravanamuthu Vigneswaran

University of Technology Sydney
Faculty of Engineering and Information Technology

June 2019

CERTIFICATE OF AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for any degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

•••••••••••••••••••

Md Atiquzzaman

June 2019

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere and deep gratitude to my honourable supervisors who have provided continuous and unlimited supports throughout this academic research. It was not possible for me to complete this thesis without their valuable suggestions, timely guidance and helps during the course of my study. My deepest admiration goes to my principal supervisor Dr. Jaya Kandasamy, Associate Professor of Faculty of Engineering and Information Technology (FEIT), University of Technology, Sydney (UTS), Australia, for many fruitful discussions to share his research experience relevant to my project work. Dr. Jaya directed me towards the correct path with great patience and understanding about my abilities and limitations. He has always engaged me in various research activities, and hence, has encouraged me to expand my ideas and thoughts. I also thank my co-supervisor Professor Saravanamuthu Vigneswaran at UTS for his support and guidance.

I appreciate the great help of Mr Gurmeet Singh in WaterNSW and would like to thank for supplying data that I used in this assessment. I would also like to acknowledge DHI Water and Environment Pty Ltd (Australia) for supplying the MIKE11-NAM license for this research.

My acknowledgement will never be completed without thanking my parents, my family and friends. My wife Hasina Akter Lovely has always inspired me to hold positive thinking towards research activities, and my sons Jaish Zaman and Jergees Zaman have also been considerate about my engagement in the research work. I would also like to thank my siblings for their never ending love and supports. Above all, I would like to express my true faith and gratitude to ALMIGHTY who gave me the opportunity to bring this research to an end through a life path with many ups and downs.

LIST OF PUBLICATIONS

Journal Publications:

- Atiquzzaman, M and Kandasamy, J. (2016). "Prediction of Inflows from Dam Catchment using Genetic Programming", International Journal of Hydrology Science and Technology, vol 6, No. 2, pp103-117, http://dx.doi.org/10.1504/IJHST.2016.075560.
- Atiquzzaman, M and Kandasamy, J. (2016). "Prediction of Hydrological Time-Series using Extreme Learning Machine", Journal of Hydroinformatics. 18.2, pp. 345-353, http://dx.doi.org/10.2166/hydro.2015.020.
- Atiquzzaman, M and Kandasamy, J. (2018). "Robustness of Extreme Learning Machine in the prediction of hydrological flow series.", Computers & Geosciences Journal, 120, pp. 105-114, http://dx.doi.org/10.1016/j.cageo.2018.08.003.
- Atiquzzaman, M and Kandasamy, J. (2019). "Kernel and node based extreme learning machines to predict hydrological time-series.", Paper prepared for the submission to a journal.

TABLE OF CONTENTS

NOMENCLATURE

$(Q_m)_t$ = “Modelled flow at time *t*”;

$(Q_o)_t$ = “Observed flow at time *t*”;

a_i = “Weight vector connecting the i^{th} hidden node and the input variables”;

b_i = “Bias of the i^{th} hidden node”;

t_j = “Target at time *j*”;

y_j = “Output at time *j*”;

$\Delta w_{ij}(s)$ = “Weight adjustment between node *j* in layer *s* and node *i* in layer (*s*-1)”;

$F_j(s)$ = “Output of the neuron *j* in layer *s*”;

H = “Hidden layer output matrix”;

$H^{'}$ = “Moore-Penrose generalized inverse of hidden layer output matrix”;

L = “Number of random hidden nodes;

Qgp = “Predicted flow by GP”;

Qnam = “Predicted flow by NAM”;

Q_t = “Flow at time *t*”;

R_t = “Rainfall at time *t*”;

$w_{ij}(s\text{-}1)$= “Weight in the link between neuron *j* in layer *s* and neuron *i* in layer (*s*-1)”;

$x_{\mathrm{i}}(s)$ = “Iinput of neuron *j* from previous layer’s neuron *I*”;

$x_i(s\text{-}1)$ = “Input from neuron *i* in layer *s*-1”;

$Y_j(s)$ = “Weighted sum for neuron *j* in layer *s*”;

β_i = “Weight connecting the hidden node and the output node”;

$g(x)$ = “Activation function (example, sigmoidal function)”;

$\delta_j(s)$	=	"Local or instantaneous gradient"; and
ε	=	"Error Value".

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Application of hydroinformatics tools in water resources has been very common in water industry due to the rapid advancement of digital computer. Over the last few decades, there are several tools have been developed and applied with success. The most commonly used Artificial Intelligence (AI) based hydroinformatics tools in hydrology are Genetic Programming (GP), Artificial Neural Network (ANN), Fuzzy Logic (FL), Standard Chaos Technique, Inverse Approach, Support Vector Machine (SVM) and Evolutionary Computation (Genetic Algorithm (GA), Shuffled Complex Evolution (SCE), Particle Swarm Optimization (PSO), Ant Colony Optimization Algorithm (ACOA)) based AI techniques including SVM (EC-SVM). These tools including Genetic Programming (GP) have been proven to be efficient in prediction of flows from event based rainfalls series.

The driving factor behind the application of hydroinformatics tools was to ease the complex numerical modelling process. In principal, both conceptual and physically based distributed models require a large number of parameters such as catchment characteristics, losses, flow paths, meteorological and flow data. The values of some of these parameters are evaluated through calibration. The calibration process of complex models may be cumbersome and requires considerable effort and experience particularly when the number of the calibration parameters is large. Even though the model is calibrated, the application of the parameters is catchment specific. Model parameters from one catchment may not be representative for the other catchment. In this case, hydroinformatics tools like GP and/or ANN can be used where no parameters associated with catchment and soil characteristic are necessary. GP has been successfully applied for calibration of numerous event based rainfall and runoff models. However, application of GP for the prediction of long term time series is

limited.

The application of GP for long term runoff prediction from a dam catchment is demonstrated. The model is developed and calibrated for a dam catchment located in New South Wales, Australia. The calibration shows excellent agreement between the observed and simulated flows recorded over thirty years and the results are better than traditional Sacramento model and ANN. GP is also linked to MIKE11-NAM to build a hybrid model. The concept of this hybrid model is to fill the data gaps and generate long term (100 years) predictions. The calibrated GP model is then applied for the assessment of two future rainfall scenarios where future hundred year flows are predicted using rainfall input generated from different assumed climatic conditions. The analysis results provide some basis for making future water management plans including water supply from alternative sources. While the application was successful and produced better results, it was found that GP suffered from computational overhead in the learning process from input data. To improve the prediction accuracy, relatively new AI technique, called Extreme Learning Machine (ELM) is proposed.

ELM is applied to partly overcome the slow learning problems of GP and ANN and to predict the hydrological time-series very quickly. ELM, which is also called single-hidden layer feed-forward neural networks (SLFNs), is able to well generalize the performance for extremely complex problems. ELM randomly chooses a single hidden layer and analytically determines the weights to predict the output. The ELM method was applied to predict hydrological flow series for the Tryggevælde Catchment, Denmark and for the Mississippi River at Vicksburg, USA. The results confirmed that ELM's performance was similar or better in terms of Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (NRMSE) compared to ANN and other previously published techniques, namely

Evolutionary Computation based Support Vector Machine (EC-SVM), Standard Chaotic Approach and Inverse Approach. In this analysis, the sensitivity of ELM's input parameters on the prediction accuracy were not investigated. The influence of input parameters was then analysed to further improve the model results.

The robustness of ELM's performances based on number of lagged input variables, the number of hidden nodes in ELM, higher lead days prediction and extrapolation capability using four goodness-of-fit measures is demonstrated. The results show that (1) ELM yields reasonable results with all combinations of lagged input variables (flows) for 1-day lead prediction. The minimum errors were obtained when 4-day lagged flows were applied as input variables; (2) ELM produced satisfactory results very rapidly for any number of hidden nodes ranging from ten to six thousand in the hidden layer. The time required to train ELM varies from less than a second to two minutes as only single iteration is required. A larger number of hidden nodes generally gives slightly better results; (3) ELM generated reasonable results for higher number of lead days (second and third) predictions; (4) ELM was able to extrapolate when the highest magnitude of input variables were excluded from training dataset; (5) ELM was shown to be computationally much faster and capable of producing better results compared with GP and EC-SVM for prediction of flow series from the same catchment. This demonstrates ELM potential for forecasting real-time hydrological time-series. This analysis was based on node based ELM (NELM) method. The performance of ELM is further improved by introducing Kernel function (KELM) in the learning process in the subsequent analysis.

In addition to node based ELM, Kernel based ELM (KELM) is also applied. The performance of KELM was also compared against hidden node based ELM (NELM). The predictive capabilities of both NELM and KELM were investigated using data from three different catchments located in three different climatic regions (Tryggevælde catchment, Denmark, Mississippi River at Vicksburg, USA and Duckmaloi Weir catchment, Australia). The results were compared with those obtained with Genetic Programming (GP) and evolutionary computation based Support Vector Machine (EC-SVM), the later obtained from literature. The results show that KELM predictions were better than NELM, GP and EC-SVM. KELM ran faster than any other model.

ELM's fast learning capability from a training dataset for the prediction of hydrological flows means that it would be more suitable for on-line and real-time applications where quick processing time is important or vital. The study demonstrates ELM's ability for rapid prediction and has potential application in real-time forecasting and in water resources planning and management.