STREAMFLOW PREDICTION IN UNGAUGED RIVER BASIN USING GENE EXPRESSION PROGRAMMING

SALAUDEEN ABDULRAZAQ

A project report submitted in partial fulfillment of the

requirements for the award of the degree of

Master of Engineering (Hydraulics and Hydrology)

Faculty of Civil Engineering

Universiti Teknologi Malaysia

JANUARY, 2016

To my beloved parents, my wife, siblings and my entire family

ACKNOWLEDGEMENTS

All thanks and praise is due to Allah the Lord of the world, the source of all knowledge, master of the Day of Judgment who among His favors created me, nourished me, spares my life up to this moment and above all bestowed on me the mental ability to withstand all academic challenges. May His peace and blessings be upon our prophet Muhammad (SAW).

With immense and sincere gratitude, I wish to commend the efforts of my project supervisor Ass. Prof. Dr. Shamsuddin Shahid for his commitments and scholarly advice to make this work a reality, despite his tight schedules. My immediate colossal gratitude goes to Dr. Tarmizi Ismail for his immense contribution. I am greatly indebted to my parents for their care and unending supports up to this level. My immediate warm and sincere gratitude goes to my brothers Prof. O.B.E Salau, Dr. A.A. Salau, Justice A.O Salahuddeen, and their families for their both financial and moral supports. Also worth of commendations is my chairman Alh. Aliyu Badaki, my sister Haj. Bilkis Badaki for their moral supports. I must also in a candid way register my sincere appreciations to some of my brothers and friends Mal. Ibraheem Salahuddeen, Mal. Habeeb Salmon, Bar. Hassan Grema Gaji, Bar. Muhammad Alkali. My profound appreciation goes to Engr. Al-Amin Bello and Mal. Suleiman I. Musa for their immense contributions.

Lastly, I must not forget to register my colossal appreciation to my wife, Fatima Umar for her care, patience and moral support. My lovely daughters Ruqayyat and Saodat Salahuddeen are equally acknowledged. I also acknowledge my sincere gratitude to Hawwa Shittu for her kindness. To all, I say may Allah reward you abundantly.

ABSTRACT

Hydrologic studies are facilitated by abundant and continuous records of streamflows and indirect peak discharge measurements. This serves as the basis for design of hydraulic structures, water resources planning and management, hydropower operation, hydrological disaster risk management as well as in assessing the effects of environmental changes. Precipitation data, temperature, humidity, wind speed are some of the pertinent meteorological data required for appropriate studies. The backbone of hydrologic data for this type of study is continuous records of streamflow gauges. However, where streams are ungauged, recourse has to be made to rainfall – runoff processes competent to simulate the flow scenarios in the catchments of interest. Other pertinent data required include geomorphologic and soil characteristics of the catchments as well as the land use and land cover. The recurrence flooding episodes and the need to have better insight to flow variability in the states of Kelantan and Terengganu (some parts of the east coastal region of Peninsular Malaysia) has been a pointer to the need for the development of models that can serve as tools for flow simulations in any catchment within the study area. The main objective of this study is therefore to predict river discharge in ungauged river basins in the study area. For this purpose, a set of multivariate equations are developed; using Genetic Expression Programing (GEP) model available in soft computing software GeneXProTools 4.0 using 4 – 7 explanatory variables. These are: Rainfall, area, perimeter, main stream length, slope, drainage density and curve number. Thus; available streamflow data along with other catchment characteristics from 15 gauged stations are used to prepare the flow duration curves (FDC). The predictable variables as Q_{ext}, Q_{max}, $Q_{0.05}$, $Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$, $Q_{0.95}$, Q_{min} and Q_{mean} were extracted from the FDC to develop the models. In order that the models may be reliably used for flow simulations in some other catchments within the study area, the accuracies of the models using standard statistical procedures such as; NSE, RMSE, R^2 and goodness of fit from the software were measured for both calibrated and validated flows. These indicate very good performance.

ABSTRAK

Kajian hidrologi dipermudahkan dengan rekod yang banyak dan berterusan streamflows dan tidak langsung ukuran menunaikan puncak. Ini berfungsi sebagai asas bagi reka bentuk struktur hidraulik, sumber air perancangan dan pengurusan, operasi kuasa hidro, pengurusan risiko bencana hidrologi serta dalam menilai kesan perubahan alam sekitar. Data hujan, suhu, kelembapan, kelajuan angin adalah sebahagian daripada data meteorologi penting yang diperlukan untuk kajian sesuai. Tulang belakang data hidrologi untuk jenis pengajian adalah rekod berterusan tolok aliran sungai. Walau bagaimanapun, di mana sungai adalah ungauged, tindakan perlu dibuat dengan taburan hujan - air larian memproses berwibawa untuk mensimulasikan keadaan aliran di dalam kawasan tadahan yang menarik. Data penting lain yang diperlukan termasuk geomorphologic dan tanah ciri-ciri kawasan tadahan serta penggunaan tanah dan perlindungan. Episod berulang banjir dan keperluan untuk mempunyai wawasan yang lebih baik mengalir kebolehubahan di negeri-negeri Kelantan dan Terengganu (beberapa bahagian wilayah pantai timur Semenanjung Malaysia) telah menjadi penunjuk kepada keperluan untuk pembangunan model yang boleh dijadikan sebagai alat untuk simulasi aliran dalam mana-mana kawasan tadahan di dalam kawasan kajian. Oleh itu, objektif utama kajian ini adalah untuk meramalkan luahan sungai di lembangan sungai ungauged di kawasan kajian. Bagi tujuan ini, satu set persamaan multivariat dibangunkan; menggunakan Expression genetik Programing (GEP) model yang terdapat di lembut GeneXProTools perisian pengkomputeran 4.0 menggunakan 4-7 pembolehubah penerangan. Ini adalah: hujan, kawasan, perimeter, panjang aliran utama, cerun, kepadatan saliran dan nombor lengkung. Oleh itu; data aliran sungai ada bersama-sama dengan ciri-ciri kawasan tadahan lain dari 15 stesen diukur digunakan untuk menyediakan keluk Tempoh aliran (FDC). Pembolehubah diramalkan sebagai Q_{ext}, Q_{max}, $Q_{0.05}$, $Q_{0.10}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$, $Q_{0.90}$, $Q_{0.95}$, Q_{min} dan Q_{mean} dipetik dari FDC untuk membangunkan model. Bagi bahawa model boleh pasti digunakan untuk simulasi aliran di beberapa kawasan tadahan lain dalam kawasan kajian, ketepatan model menggunakan prosedur statistik standard seperti; NSE RMSE, R^2 dan kebaikan penyuaian daripada perisian diukur untuk kedua-dua aliran ditentukur dan disahkan. Ini menunjukkan prestasi yang sangat baik.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Q_{ext}	-	Annual extreme flow
Q_{max}	-	Annual maximum flow
$Q_{0.05}$	-	Five percentile flow
$Q_{0.10}$	-	Ten percentile flow
$Q_{0.25}$	-	Twenty five percentile flow
$Q_{0.50}$	-	Fifty percentile flow
$Q_{0.75}$	-	Seventy five percentile flow
$Q_{0.90}$	-	Ninety percentile flow
$Q_{0.95}$	-	Ninety five percentile flow
Q_{min}	-	Annual minimum flow
Q_{mean}	-	Mean annual flow
$Q_{(p),1}$	-	Nonparametric quantile estimate
P_i	-	Plotting position
q_i	-	Ordered observation

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Runoff hydrographs are key requirements for designing hydraulics structures, water resources planning and management, hydropower operation, hydrological disaster risk management as well as in assessing the effects of environmental changes (Yaşar and Baykan, 2013). Hydrologic studies are facilitated by abundant and continuous records of streamflows and indirect peak discharge measurements. Pertinent meteorological data needed include precipitation, temperature, humidity and wind speed, which are collected by meteorological stations. The best hydrological data are obtained from continuously recording gauges. Where the streams are ungauged, recourse has to be made to modeling of the rainfall – runoff processes, using available precipitation data and detailed field study of the drainage characteristics of the watershed for appropriate predictions.

Usually, hydrological information transfer from gauged to ungauged catchments has been fundamental approach used for streamflow predictions the world over, using a number of methods such as; hydrological model simulation, observation by remote sensing and integrated meteorological and hydrological modeling for water resources applications (Goswami et al., 2007). Complete gauging of the whole catchments may not be possible particularly, in developing countries.

Thus; runoff data are usually not available in many catchments of interest. Therefore, it is often required to predict runoff hydrographs of ungauged catchments from other information within that catchment or from other catchments (Salinas et al., 2013). Many methods have been developed and applied in different parts of the world for this purpose (Lacombe et al., 2014; Kovacs et al., 2012; Yaşar and Baykan, 2013); however, prediction in ungauged basins remains a major challenge in hydrology.

The challenges can be more in tropical regions where most of the catchments are ungauged, as such; the need for improved knowledge of flow variability in such regions may be necessary, especially in the context of changing hydrological processes and growing hydrological disasters due to climate change. Increasing severity and frequency of floods due to changing rainfall pattern is a growing concern in the east coast of Peninsular Malaysia (Basarudin et al., 2014; Awadalla and Noor, 1991; Pradhan, 2009) which justifies the need for streamflow prediction in the region. The major challenge in hydrological studies in the area is the unavailability of reliable and long-term streamflow data in most catchments of interest.

For this purpose, a symbolic regression method based on Gene Expression Programming (GEP) available in powerful soft computing software; GeneXPro Tools 4.0 (Fernando et al., 2009; Ferreira, 2001) was used to develop a GEP model using 4 - 7 candidate explanatory variables prepared from climatic, geomorphologic, geographic characteristics, soil properties, land use and land cover of the area under study. The streamflow data are the predictable variables. These are prepared using flow duration curve (FDC). FDC presents the probability of flood of a particular magnitude to be equaled or exceeded over a historical period. With FDC, a comprehensive graphical view of the historical change in the overall flow event is possible, for the catchment of interest. Because of highly skewed nature of daily streamflow data, nonparametric approach was used for the FDC framework rather than the parametric to avoid high tendencies to biasness. Median annual FDC being

most appropriate method for this type of study have been chosen (Vogel and Fennessey, 1994).

1.2 Problem Statement

In recent years, there have been records of frequent occurrence of floods of higher magnitudes, that have continue to constitute danger to life and property in the east coast of Peninsular Malaysia. Better understanding of flow variability, for various water resources management and planning for ungauged catchments are desired. It is unarguable fact, that most of the catchments in the area are ungauged or poorly gauged, streamflow prediction for the ungauged catchments are therefore of paramount importance. The required and necessary information for the ungauged catchments can be extrapolated from gauged catchments using appropriate hydrological model to proffer long-term engineering solutions. This research attempts to develop models that can be useful in this regards thus; formed the rationale for the study.

1.3 Objectives of the Work

The major objective of this study is to predict streamflow in ungauged catchments of the study area using available records from the gauged catchments and other catchment's characteristics. The specific objectives are:

i) To prepare all – year specific median annual flow duration curves (FDCs) for the gauged catchments of the northeast coastal region of Peninsular Malaysia.

ii) To develop a set of multivariate equations which are expressed in the forms of GEP models for the area under study and ultimately;

iii) To apply the developed equations for streamflow predictions in ungauged catchments within the study area.

1.4 Scope of Work

The following are the scopes of the study:

i) Watersheds definitions and delineation.

ii) Data collection and data analysis.

iii) Preparation of annual flow duration curves for the gauged catchments in the study area.

iv) Development of multivariate equations for flow predictions of ungauged catchments and subsequent test of applicability.

1.5 Significance of Study

The flooding episode in the east coast of Peninsular Malaysia is becoming more frequent in recent time than ever, particularly during the north east monsoon season. The two worst case scenarios being in 1926 and 1971 have been reported. The 2014 flooding event has been the most devastating in recent time where days of rainfall has resulted in to damage of hydraulic structures, social lives of people and ultimate loss of lives.

There is no doubt that the planning and design of hydraulic structures, water resources management, hydrological disaster risk management depend largely on the

accurate estimate of the magnitude and volume of flow in the region under consideration. However, the non-availability and inadequate streamflow data have continued to hamper the process in most cases, as most of the upper catchments of the in the region are ungauged. Thus, this obviates the need for improved knowledge of flow variability.

This study therefore, can contribute significantly to better understanding of flows in the catchments, by providing approaches that can be used to extrapolate by hydrological information transfer from gauged to ungauged catchments of various flow metrics in the states of Kelantan and Terengganu.

1.6 Definitions of Terms

- Gene Expression Programming (GEP): Is one of the powerful optimization processes that are used to solve symbolic regression problems for developing models or create a computer programs which basic principle is that of biological evolution similar to other evolutionary optimization algorithms.
- Symbolic regression: Is nonparametric regression method that is used to establish a functional relationship between dependent and independent variables similar to the traditional parametric regression, except that the function is not specified a *priori*.
- Streamflow: This is the flow of water in streams, rivers, channels etc. it is an essential component of hydrologic cycle as such; it plays an important role in any hydrologic studies.
- Ungauged River Basin: This is an area of land contributing flow to a river or its tributaries where there is no mechanism in place for recording continuous flow.

- Flow Duration Curve: FDC presents the probability of flood of a particular magnitude to be equaled or exceeded over a historical period.

REFERENCES

Abdullahi, M.G., Gasim, M.B. and Juahir, H. (2015). Determination of Groundwater Level Based on Rainfall Distribution: Using Integrated Modeling Techniques in Terengganu, Malaysia. *Journal of Geology & Geosciences.* 4(1)**,** 1-4.

Adnan, N.A. and Atkinson, P.M. (2011). Exploring the impact of climate and land use changes on streamflow trends in a monsoon catchment. *International Journal of Climatology*. 31**,** 815–831.

Adnan, N.A., Atkinson, P.M., Yusoff, Z.M. and Rasam, A.R.A. (2014). Climate Variability and Anthropogenic Impacts on a Semi-Distributed Monsoon Catchment Runoff Simulations. 1-6.

Awadalla, S. and Noor, I.M. (1991). Induced climate change on surface runoff in Kelantan Malaysia. *International Journal of Water Resources Development.* 7(1)**,** 53-59.

Basarudin, Z., Adnan, N.A., Latif, A.R.A., Tahir, W. and Syafiqah, N. (2014). Event-based rainfall-runoff modelling of the Kelantan River Basin. *IOP Conference Series: Earth and Environmental Science.* 18**,** 012084.

Billa, L. (2006). Application of hydro-metrological model and GIS in short range severe flood forecasting and mapping. *Engineering.* Doctorate**,** 21.

Cuo, L., Zhang, Y., Gao, Y., Hao, Z. and Cairang, L. (2013). The impacts of climate change and land cover/use transition on the hydrology in the upper Yellow River Basin, China. *Journal of Hydrology.* 502**,** 37-52.

Droogers, P. and Allen, R.G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and Drainage Systems.* 16**,** 33–45.

Fernando, D.a.K., Shamseldin, A.Y. and Abrahart, R.J. (2009). Using gene expression programming to develop a combined runoff estimate model from conventional rainfall-runoff model outputs. *18th World IMACS / MODSIM Congress*. Cairns, Australia.

Ferreira, C. (2001). Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. *Complex Systems.* 13(2)**,** 87-129.

Goswami, M., O'connor, K.M. and Bhattarai, K.P. (2007). Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology.* 333(2-4)**,** 517-531.

Gupta, H.V. and Kling, H. (2011). On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. *Water Resources Research.* 47**,** 1-3.

Hassan, A.a.G. (2004). *Growth, Structural Change and Regional Inequality in Malaysia*. 2). Ashgate, Aldershot.

Helsel, D.R. and Hirsch, R.M. (2002). Techniques of Water-Resources Investigations of the United States Geological Survey. *In:* Survey, U. S. G. (ed.) *Hydrologic Analysis and Interpretation.* U.S.A.

Jabloun, M. and Sahli, A. (2008). Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited data available to Tunisia evapotranspiration using limited climatic data Application to Tunisia. *Agricultural water management.* 95**,** 707–715.

Jones, D.A. and Kay, A.L. (2007). Uncertainty analysis for estimating flood frequencies for ungauged catchments using rainfall-runoff models. *Advances in Water Resources.* 30(5)**,** 1190-1204.

Khan, M.M.A., Shaari, N.a.B., Bahar, A.M.A., Baten, M.A. and Nazaruddin, D.a.B. (2014). Flood Impact Assessment in Kota Bharu, Malaysia: A Statistical Analysis. *World Applied Sciences Journal.* 32 (4)**,** 626-634.

Kovacs, A., Honti, M., Zessner, M., Eder, A., Clement, A. and Bloschl, G. (2012). Identification of phosphorus emission hotspots in agricultural catchments. *Sci Total Environ.* 433**,** 74-88.

Lacombe, G., A. Pierret, Hoanh, C.T., Sengtaheuanghoung, O. and A.D.Noble (2010). Conflict, migration and land-cover changes in Indochina a hydrological assessment. *Journal of Ecohydrology.* 3**,** 382–391.

Lacombe, G., Douangsavanh, S., Vogel, R.M., Mccartney, M., Chemin, Y., Rebelo, L.M. and Sotouke, T. (2014). Multivariate power-law models for streamflow prediction in the Mekong Basin. *Journal of Hydrology.* 2**,** 35–48.

Legates, D.R. and Mccabe, G.J. (1999). Evaluating the use of "goodness - of - fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research.* 35(1)**,** 233 - 241.

Mccuen, R.H. (1989). *Hydrologic Analysis and Design* In: Stenquist, B. (ed.) 833), 2nd Edition ed.: A Simon and Schuster/Aviacom upper saddle river, New Jersey, 07458.

Merwade, V. (2012). Watershed and Stream Network Delineation using ArcHydro Tools. School of Civil Engineering, Purdue University.

Milad, J. (2012). Modeling daily stream flow using plant evapotranspiration method. *International Journal of Water Resources and Environmental Engineering.* 4(6).

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers.* 50(3)**,** 885-900.

Moten, S., Yunus, F., Ariffin, M., Burham, N., Yik, D.J., Adam, M.K.M. and Sang, Y.W. (2014). Statistics of Northeast Monsoon Onset, Withdrawal and Cold Surges in Malaysia. *In:* Department, M. M. (ed.). Kuala Lumpur, Malaysia.

Niehoff, D., Fritsch, U. and Bronstert, A. (2002). Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. *Journal of Hydrology.* 267**,** 80–93.

O'connell, E., Ewen, J., O'donnell, G. and Quinn, P. (2007). Is there a link between agricultural land-use management and flooding. *Hydrology and Earth Sciences.* 11(1)**,** 96-107.

Paramananthan, S. (2012). Keys to the identification of the Malaysian soils using parent materials. *In:* Bhd, P. a. S. S. M. S. (ed.). 46400 Patailing Jaya, Selangor, Malaysia.

Pour, S., Harun, S. and Shahid, S. (2014). Genetic Programming for the Downscaling of Extreme Rainfall Events on the East Coast of Peninsular Malaysia. *Atmosphere.* 5(4)**,** 914-936.

Pradhan, B. (2009). Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology.* 9**,** 1-18.

Pradhan, B. and Youssef, A.M. (2011). A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia. *Journal of Flood Risk Management.* 4(3)**,** 189-202.

Ritter, A. and Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*. 480**,** 33-45.

Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M. and Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies. *Hydrology and Earth System Sciences.* 17(7)**,** 2637-2652.

Sani G. D/Iya, M.B., Ekhwantoriman, M. and Abdullahi, M.G. (2014). Flood in Malaysia Historical Reviews, Causes, Effects and Mitigations Approach. *International Journal of Interdisciplinary Research and Innovations.* 2(4)**,** 59-65).

Tabari, H., Grismer, M.E. and Trajkovic, S. (2011). Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science.* 31(2)**,** 107-117.

Tokar, A.S. and Johnson, P.A. (1999). Rainfall - Runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering.* 4(3)**,** 232-239.

Veldkamp, A. and Fresco, L.O. (1996). CLUE: a conceptual model to study the Conversion of Land Use and its Effects.pdf>. *Ecological Modelling.* 85**,** 253-270.

Vogel, R.M. and Fennessey, N.M. (1994). Flow Duration Curve: New interpretation and confidence intervals. 485-503.

Wong, C.L., Venneker, R., Uhlenbrook, S., Jamil, A.B.M. and Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia. *Journal of Hydrology and Earth System Sciences Discussions.* 6**,** 5471–5503,.

Xu, C.Y. and Singh, V.P. (2002). Cross Comparison of Empirical Equations for Calculating Potential Evapotranspiration with Data from Switzerland. *Water Resources Management.* 16**,** 197–219.

Yaşar, M. and Baykan, N.O. (2013). Prediction of Flow Duration Curves for Ungauged Basins with Quasi-Newton Method. *Journal of Water Resource and Protection.* 5**,** 97-110.

Yik, D.J., Adam, M.K.M., Sang, Y.W. and Moten, S. (2014). Anomalous Winter Monsoon Season of 2012/2013 Over the Malaysian Region. *In:* Department, M. M. (ed.). Kuala Lumpur, Malaysia.