11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

PREDICTING STREAM BASEFLOW USING GENETIC PROGRAMING

ALI MESHGI (1), PETRA SCHMITTER (2), VLADAN BABOVIC (1), TING FONG MAY CHUI (3)
[1]*Department of Civil and Environmental Engineering, National University of Singapore*
[2]*Singapore-Delft Water Alliance, Faculty of Engineering, National University of Singapore*
[3]*Department of Civil Engineering, The University of Hong Kong*

Abstract

Developing reliable methods to estimate stream baseflow has been a subject of research over the past decades due to its importance in catchment response and sustainable watershed management (e.g. ground water recharge vs. extraction). Limitations and complexities of existing methods have been addressed by a number of researchers. For instance, physically based numerical models are complex, requiring substantial computational time and data which may not be always available. Artificial Intelligence (AI) tools such as Genetic Programming (GP) have been used widely to reduce the challenges associated with complex hydrological systems without losing the physical meanings. However, up to date, in the absence of complex numerical models, baseflow is frequently estimated using statistically derived empirical equations without significant physical insights. This study investigates the capability of GP in estimating baseflow for a small monitored semi-urban catchment (0.021 km^2) located in Singapore. A Recursive Digital Filter (RDF) is first adopted to separate the baseflow from observed streamflow. GP is then used to derive an empirical equation to relate the filtered baseflow time series particularly with groundwater table fluctuations which are relatively easy to be measured and are physically related to baseflow generation. The equation is then validated with a longer time series of baseflow data from a groundwater numerical model. These results indicate that GP is an effective tool in determining baseflow.

Keywords: Baseflow, Genetic Programing, Recursive Digital Filters, Numerical modeling

1. INTRODUCTION

Developing reliable methods to estimate stream baseflow has been a subject of research interest over the past decades [1]. Various graphical baseflow separation methods have been developed by assuming the baseflow to be equal to the streamflow between distinct and consecutive rainfall events [2]. According to Linsley et al [3] this method is not appropriate for long continuous streamflow records.

RDFs are signal processing techniques that remove the high-frequency quick flow signal from a streamflow time series in order to obtain the low-frequency baseflow signal. Numerous RDFs exist for baseflow separation such as one-parameter algorithm [4], two-parameter algorithm [5, 6] and three-parameter algorithm [5]. These approaches are often computationally efficient and also overcome the limitations associated with graphical based methods when applied to long continuous streamflow records. Therefore, RDFs are currently the most widely adopted method for baseflow separation. However, RDFs are based on statistically derived equations without significant physical information.

Application of physically based numerical modelling for baseflow quantification has been recently explored by Partington et al. [7]. In this method, flow solutions obtained from numerical models would be processed by a hydraulic mixing-cell method to quantify hydrograph flow components. This method overcomes many of the limitations of other methods mentioned above. However, up to date it has only been tested for a hypothetical catchment. Furthermore, such models are complex, requiring significant computational time and sufficient data which may not always be available.

On the other hand, Artificial Intelligence (AI) tools such as Genetic Algorithms (GA) have been used widely in hydrology [e.g. 8, 9-11]. Genetic Programming (GP), a specialization of Genetic Algorithms (GA), has been also employed over the past decades to simplify complex hydrological problems such as the development of rainfall-runoff models based on meteorological data [12], predicting the flood routing in natural channels [13], estimating saturated hydraulic conductivity [14], evapotranspiration [15] and groundwater levels [16]. As GP has been successful in solving a number of complex hydrological problems, it can potentially be used to estimate baseflow. A GP model requires significantly less computational time as well as data for calibration when compared to numerical hydrological models. However, up to date, no equation has been derived using GP for determining baseflow based on physical catchment parameters. Therefore, this study assess whether GP can be adopted to obtain an empirical equation for baseflow estimation.

2. METHODLOGY

2.1 Description of Study Site

The Kent Ridge catchment (0.021 km^2), a small catchment located inside the Kent Ridge campus of National University of Singapore (NUS), is selected for the current study. The land-use consists of bushes, grass and paved area. Extensive information on discharge, soil type and climatic parameters are also available for this catchment. The rainfall pattern varies over the year with two monsoons (mid-November to early March and mid-June to September). There are moderate rainfall events to intense thunderstorm activity during the monsoon period, while short shower events interrupted by thunderstorms in the inter-monsoon period. According to the weather station maintained by the NUS Department of Geography located nearby the study catchment, the mean annual precipitation is 2500 mm and the daily mean temperature varies between a minimum of 23.9°C and maximum of 32.3°C. The mean annual relative humidity is 84.2% and reaches 100% during periods of rain, while the mean annual wind velocity is 15km/hour.

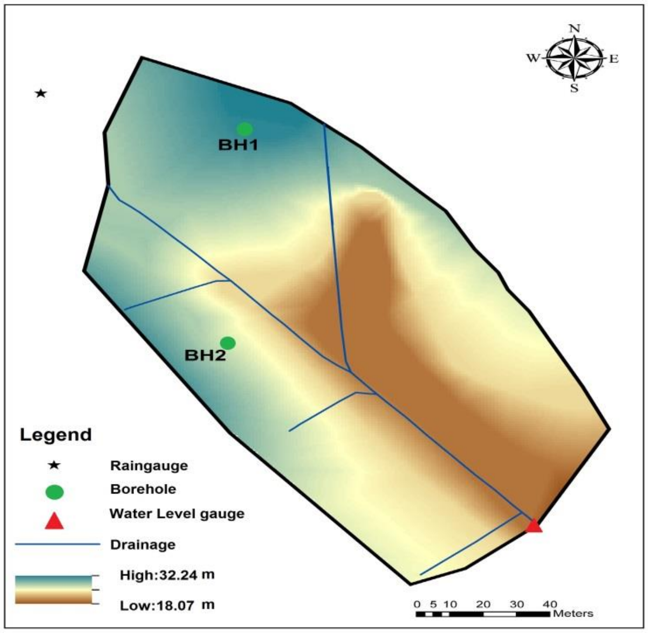

Figure 1. Location of monitoring stations, drainage network and DEM of Kent Ridge catchment, Singapore

A water level gauge for discharge measurement and a rainfall gauge operated simultaneously from September 2011 to August 2012 and January to June 2013 at 1-minute intervals (Figure 1). In January 2012, pressure transducers and loggers (i.e., Mini-Divers) are installed in two boreholes (BH1 and BH2) to record groundwater table elevations at 15-minute intervals (Figure 1). To eliminate the fluctuations in atmospheric pressure from the pressure transducers submerged in water, another pressure transducer (i.e., a Baro-Diver) is installed but suspended in the air.

2.2 Recursive Digital Filters

In this study, a recursive digital filters tool developed by Willems [18] is used to separate the baseflow from streamflow in our case study. The Water Engineering Time Series Processing tool (WETSPRO) is a generalization of the original Chapman-filter [17] and can be used to filter subflow.

Discharge data from September 2011 until August 2012 is used to calibrate the Chapman-filter parameters proposed by Willems, while the data from January until June 2013 is adopted for validation.

2.3 Genetic Programing

GP evolves symbolic relationships to relate the input information to the output information, to solve a specific problem and develop a data-based model. GP evolves function trees techniques with two different types of nodes including inner nodes and terminal ones. Inner nodes consume one or more input values and produce a single output value (e.g. -,*, /, +) to define a function set for the problem. While external inputs, constants, and zero augment functions are represented by terminal nodes. These trees can be created randomly in GP using different methods such as full, grow, ramped half-and-half and exact uniform initialization. Afterwards, a fitness function is constructed to select the models (trees) which have better performance for reproduction in a probabilistic manner. In fact, models with low fitness have less chance to be selected for reproduction than those of higher fitness. In the next stage, three genetic operators including crossover, mutation, and reproduction may be applied to create subsequent generations from selected models. After creating new generation (offspring) from parents, a decision must be made regarding the models which must be rejected from the population. GP then continues creating new generation from the selected population. The program is usually terminated by a pre-specified number of generations.

In the current research, a GP software called GPKERNEL [19] is used to relate catchment baseflow with its hydrological and physical parameters (Table 1). An overview of the evolutionary algorithm setup in this study is presented in Table 2. One experiment is set up in GP, to relate the baseflow estimated by the RDF with catchment characteristic and time series of groundwater table. In this experiment, observed pressure head, precipitation and evapotranspiration data from January until August 2012 are used as input parameters in GP. In addition, baseflow data filtered from observed discharge data by WETSPRO is defined as target parameter.

Table 1: Definition of terminal set parameters

Parameter name	**Parameter definition**	**Unit**	**Type**
R	Daily precipitation	[L]	Input
ET	Daily evapotranspiration	[L]	Input
Δh_p	Normalized daily average of pressure head	[L]	Input
ET_{av}	Annual daily average of evapotranspiration	[L]	Constant
R_{av}	Annual daily average of precipitation	[L]	Constant
B_{min}	Average of minimum daily baseflow volume	[L^3]	Constant
B_{max}	Average of maximum daily baseflow volume	[L^3]	Constant
A	Area of catchment	[L^2]	Constant

Table 2: An overview of the evolutionary algorithm setup

Parameter	**Value**
Objective	Find the daily baseflow volume (B)
Population size	250
Number of children to produce	500
Number of generations	500
Tournament size	3
Brood size	2 (culling function on unit error)
Crossover probability	0.4
Mutation probability	0.05
Crossover method	Random subtree crossover
Objective Functions	RMSE and unit error
Function set	*, +,-, %, - x , sqrt, power
Maximum size at initialisation	15
Maximum size	41
Probability of selecting a constant vs. a variable	0.05
Constant mutation probability	0.05
Stopping criteria	500 generations

2.4 Numerical Modeling

To provide a longer time series of baseflow data in Singapore and validate the empirical equation derived by GP, a numerical groundwater model, HYDRUS-2D/3D, is adopted. The calibrated HYDRUS3D provides additional groundwater table and baseflow data from January 2011 until June 2013. Baseflow is extracted from the simulation by integrating the flux across the seepage face boundary.

2.5 Statistical Tests of Accuracy

Performance of the established equation in GP is tested using three commonly used error functions: Relative Root Mean Squared Error (RRMSE), Correlation Coefficient (CC) and the Nash–Sutcliffe Efficiency (NSE) statistic [20].

3. RESULTS AND DISCUSSION

3.1 Separating Baseflow from Observed Streamflow using a Recursive Digital Filter

By visually inspecting the plots of filtered results in WETSPRO, the filtering parameters 'k' and 'w' for baseflow separation are found to be 4 days and 0.7 respectively. Such filtering suggests that 30% of the total discharge is contributed by baseflow. Analysis on rainfall events from September 2011 to August 2012 indicates that the mean contribution of baseflow during the rainfall events is about 12%. Moreover, the contribution during the rainfall events varies from a minimum of 2% to a maximum of 56% in June (dry season) and November (wet season), respectively. In other words, the groundwater table is shallower during the wet season due to heavy rainfall events, creating a higher contribution of baseflow.

3.2 Approximating Baseflow Timeseries Using Genetic Programing

Based on the time series baseflow filtered from the observed discharge data using WETSPRO, GP is set up to derive the empirical equation. The following equation is obtained:

$$\boldsymbol{B}_{(\boldsymbol{t})} = \mathrm{B_{min}} + \sqrt{0.29\ \mathrm{A}}\ \Delta \mathrm{h}^2_{\mathrm{p(t)}} \qquad (1)$$

where $\mathrm{B_{(t)}}$ presents the daily baseflow (m^3/day), $\mathrm{B_{min}}$ is the minimum daily baseflow in dry period (m^3/day), A is the catchment area (m^2), $\Delta\mathrm{h_{p(t)}}$ is the normalized daily average of pressure head ($\Delta\mathrm{h_{p(t)}} = \mathrm{h_{(t)}} - \mathrm{h_{min}}$ in which $\mathrm{h_{(t)}}$ is the daily average of pressure head (m/day) and $\mathrm{h_{min}}$ is the minimum daily average of pressure head (m) in dry period). Figure 2 compares the baseflow estimated by the empirical equation and that filtered from WETSPRO. In addition, Table 3 presents the error criteria associated with the baseflow estimated by the empirical equation and WETSPRO. According to these results, differences between the baseflow obtained by WETSPRO and empirical equation are minimal in both the training and testing periods.

The first term in the empirical equation is the minimum baseflow corresponding to the deepest groundwater table in the dry period, while the second term approximates the additional baseflow due to the rise in groundwater table. In this equation, pressure head (h) is the only variable and baseflow is correlated with h^2. This is similar to Darcy's Law ($q = kh(\frac{\partial h}{\partial x})$) that relates the discharge through an unconfined aquifer to h^2. It shows that the empirical equation derived by GP for estimating baseflow retains physical information.

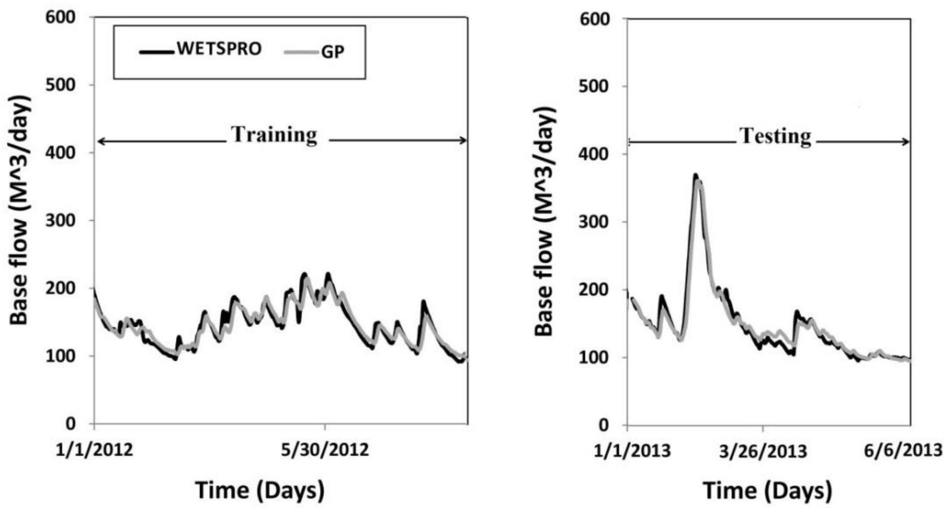

Figure 2. Comparison between baseflow estimated by the empirical equation and WETSPRO in Kent Ridge Catchment, Singapore

Table 3: Error functions associated with baseflow estimated by the empirical equation

Method	Data Set	Error criteria		
		RRMSE	NSE	CC
WETSPRO	Train	0.056	0.941	0.963
	Test	0.054	0.977	0.981
HYDRUS	Train	0.055	0.958	0.975
	Test	0.061	0.962	0.989

3.3 Verifying Proposed Equation with Simulated Baseflow from HYDRUS 3D

Figure 3 compares the baseflow estimated by the empirical equation and those simulated by HYDRUS3D. Error criteria including NSE, CC and RRMSE between baseflow simulated by HYDRUS3D and the empirical equation are listed in Table 3. According to these results, differences between the baseflow simulated by HYDRUS3D and empirical equation are minimal, confirming that the empirical equation can accurately estimate the baseflow in the absence of discharge measurements.

4. SUMMARY AND CONCLUSION

This study assesses the capability of GP in estimating stream baseflow. First, an RDF is adopted to separate the baseflow from observed discharge for a small semi-urban catchment in Singapore. An empirical equation is then derived using GP to relate the filtered baseflow with minimum baseflow in dry period, area of the catchment and time series of groundwater table fluctuations. The baseflow estimated by the empirical equation matches very well with those from the RDF and numerical groundwater model. Overall, this study proposes a new approach to predict baseflow with only three parameters. It serves as an alternative approach for baseflow estimation when groundwater table information is available. This method is an alternative to other methods (e.g., digital filter method) when discharge measurements are not available.

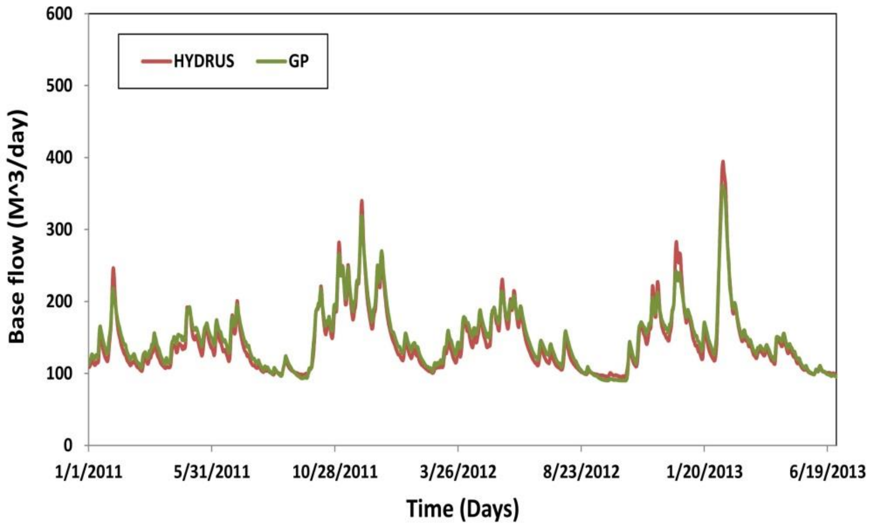

Figure 3. Comparison between baseflow estimated by the empirical equation and HYDRUS3D in Kent Ridge Catchment, Singapore

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and contributions of Singapore-Delft Water Alliance (SDWA). The research presented in this work is carried out as part of the SDWA's Multi-objective Multiple Reservoir Management research program (R-303-001-005-272).

REFERENCES

1. Gonzales, A.L., et al., *Comparison of different base flow separation methods in a lowland catchment.* Hydrology and Earth System Sciences, 2009. **13**(11): p. 2055-2068.
2. Furey, P.R. and V.K. Gupta, *A physically based filter for separating base flow from streamflow time series.* Water Resources Research, 2001. **37**(11): p. 2709-2722.
3. Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus, *Hydrology for engineers*1982, London: McGraw Hill Book Company.
4. Chapman, T.G. and A.I. Maxwell, *Baseflow separation - comparison of numerical methods with tracer experiments*, in *In: Proceedings of the 23rd Hydrology and Water Resources Symposium*1996: Hobart Australia.
5. Chapman, T., *A comparison of algorithms for stream flow recession and baseflow separation.* Hydrological Processes, 1999. **13**(5): p. 701-714.
6. Eckhardt, K., *How to construct recursive digital filters for baseflow separation.* Hydrological Processes, 2005. **19**(2): p. 507-515.
7. Partington, D., et al., *A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water–groundwater flow models.* Environmental Modelling & Software, 2011. **26**(7): p. 886-898.
8. Anctil, F., et al., *Improvement of rainfall-runoff forecasts through mean areal rainfall optimization.* Journal of Hydrology, 2006. **328**(3–4): p. 717-725.
9. Sedki, A., D. Ouazar, and E. El Mazoudi, *Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting.* Expert Systems with Applications, 2009. **36**(3, Part 1): p. 4523-4527.

10. Kim, S. and H.S. Kim, *Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling.* Journal of Hydrology, 2008. **351**(3–4): p. 299-317.
11. Babovic, V., *Data mining in hydrology.* Hydrological Processes, 2005. **19**(7): p. 1511-1515.
12. Babovic, V. and M. Keijzer, *Rainfall-Runoff Modeling Based on Genetic Programming*, in *Encyclopedia of Hydrological Sciences*2006, John Wiley & Sons, Ltd.
13. Sivapragasam, C., R. Maheswaran, and V. Venkatesh, *Genetic programming approach for flood routing in natural channels.* Hydrological Processes, 2008. **22**(5): p. 623-628.
14. Parasuraman, K., A. Elshorbagy, and B.C. Si, *Estimating Saturated Hydraulic Conductivity Using Genetic Programming* Soil Science Society of America Journal, 2007. **71**(6): p. 1676-1684.
15. Izadifar, Z. and A. Elshorbagy, *Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models.* Hydrological Processes, 2010. **24**(23): p. 3413-3425.
16. Fallah-Mehdipour, E., O. Bozorg Haddad, and M.A. Mariño, *Prediction and simulation of monthly groundwater levels by genetic programming.* Journal of Hydro-environment Research. doi:http://dx.doi.org/10.1016/j.jher.2013.03.005, 2013.
17. Chapman, T.G., *Comment on "Evaluation of automated techniques for base flow and recession analyses" by R. J. Nathan and T. A. McMahon.* Water Resources Research, 1991. **27**(7): p. 1783-1784.
18. Willems, P., *A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models.* Environmental Modelling & Software, 2009. **24**(3): p. 311-321.
19. Babovic, V. and M. Keijzer, *Genetic programming as a model induction engine.* Journal of Hydroinformatics, 2000. **2**(1): p. 35-60.
20. Nash, J.E. and J.V. Sutcliffe, *River flow forecasting through conceptual models part I — A discussion of principles.* Journal of Hydrology, 1970. **10**(3): p. 282-290.