App. Envi. Res. 36 (4): 53-65

Applied Environmental Research

Journal homepage : http://www.tci-thaijo.org/index.php/aer

A Simplified Rainfall-Streamflow Network Model on Multivariate Regression Analysis for Water Level Forecasting in Klong Luang (KGT.19 Station) Sub-watershed, Chon Buri Province, Thailand

Charoen Panant* and Samakkee Boonyawat**

Department of Conservation, Faculty of Forestry, Kasetsart University
Pahonyothin Rd., Jatuchak, Bangkok 10900
* Corresponding author: Email: *chpanant@gmail.com, **fforskb@ku.ac.th;
Phone: 03-815-2156 Fax. 03-815-2157

Article History
Submitted: 7 March 2014/ Accepted: 3 July 2014/ Published online: 17 October 2014

Abstract

A simplified rainfall-streamflow network model based on multivariate linear regression (MLR) analysis has been proposed. To determine significant coefficients of streamflow network, eleven MLR models were examined. The study's three objectives were 1) to develop a novel a mathematical model based on MLR analysis for forecasting optimal water levels; 2) to determine the most significant coefficient of rainfall-streamflow network among in the area of interest in the vicinity of Klong Luang sub-watershed KGT.19 station; and 3) to apply the optimal MLR model for water level and flood forecasting maps in Klong Luang Sub-watershed. We used Geographic Information System (GIS) and Remotely Sensed Data (RS) data recorded from Klong Luang (KGT.19 Station) sub-watershed, and Phanat Nikhom, Chonburi, Ban Bueng and Phan Thong districts, in Chonburi Province, Thailand. The findings indicated that the MLR based Model No. 8 is the most applicable and effective. The proposed model also could be applied in water level forecasting, water resource management, flood hazard planning, and flood early warning.

Keywords: Water level forecasting; multiple regression analysis; Streamflow Network model; Klong Luang Sub-watershed

Introduction

Geographic Information System (GIS) and Remote Sensing (RS) are tools that use the Digital Elevation Model (DEM) to create interactive queries, analyze spatial information, edit data, map and present the results of operations. The GIS and hydrologic model can both predict flow rate and the flood prone area. On the other hand, basic water management theory can be applied to develop an integrated model from meteorology, geography and hydrography in terms of a spatial information system. Fotheringham and Rogerson [1] proposed that the topography, land use, soil properties and the moisture status of the study area could be coded as a hydrologic response unit named as the curve number, as referred to by U.S. Department of Agriculture and Soil Conservation Service.

In Thailand, the Department of Water Resources reported that Klong Luang sub-watershed covers an area of 1,897 square kilometers which includes the entire districts of Phanat Nikhom and Ban Bueng, some parts of Phan Thong, Nong Yai, Bo Thong, Ko Chan, and Muang Chonburi districts. This sub-watershed is the most important in Chonburi Province, which is the development centre of the Eastern Seaboard of Thailand. The watershed drains from the South-West Mountain in Bo Thong district and floods over Phan Thong district every year [2].

The flood disaster area covers 7 districts, 17 sub-districts, 66 villages, and a population of 41,807 persons in 14,075 households, with damage to 2,871 houses and 36.96 km^2 (23,103 hectares) of agricultural land. It also causes major economic and social damage and disruption to trade, industry and infrastructure across the eastern region of Thailand, especially in Mueang Chonburi, Phanat Nikhom, Ban Bueng and Phan Thong districts.

The use of global ensemble streamflow forecasting and flood early warning is therefore an essential tool in flood disaster mitigation and management. In some cases [3-6], conceptual rainfall-streamflow models e.g. the neural network model, are used even though these models are not intended for forecasting purposes, An optimal neural network requires significant training and collection of large quantities of data.

Recent studies have shown that the use of real-time precipitation and streamflow data in rainfall and streamflow routing models can be utilized as a parameter for water level forecasting [8-10]. Flood forecasting is a component of flood warning, where the distinction between the two is that the outcome of flood forecasting is a set of forecast time-profiles of channel flows or river levels at various locations, while flood warning refers to the task of making use of these forecasts to make decisions about whether warnings of floods should be issued to the general public or whether previous warnings should be rescinded or retracted.

In this study, we propose the streamflow network model based on multivariate regression analysis for flood hazard planning and simple flood forecasting method without a training process. The Geological survey has collected streamflow data at Klong Luang sub-watershed from 1965 to 1990. Flood prevention can be successful by planning land use both in the watershed area and in the flood area. Flood prospecting by proper model and input data may yield the flood prone area due to different amounts of rainfall. The proposed model is able to estimate the output magnitude of streamflow for water resources management and minimizing flood areas in Klong Luang sub-watershed.

Methods

We propose a multiple linear regression (MLR) to model the relationship between the explanatory and response variables. The MLR uses several explanatory variables to forecast the outcome of a response variable. The study's hypothesis is that the water level depends on the rainfall and streamflow variables [8]. Therefore, the general forecasting equation based on MLR can be expressed as follows:

$$\mathbf{W}_{t+\tau} = \alpha_0 + \beta_t \cdot \sum \mathbf{B}_t \quad \text{for } t = 1, 2, \ldots, n \qquad (1)$$

where

$\mathbf{W}_{t+\tau}$ is a output variable of responses

$\mathbf{B}_t$ is a design matrix of predictor variables

β_t is vector or matrix of regression coefficients

α_0 is a constant, with multivariate normal distribution

Using the ordinary multivariate normal maximum likelihood estimation [7], the regression coefficients can readily be obtained. In this preliminary study, we examine the possible MLR based model by using rainfall database as dependent variable. After analysis, we focus on the average monthly streamflow volume in the enclosed drainage area. This section presents the MLR based streamflow and rainfall results. To show the results in terms of the MLR model, the linear regression line and its 95% confidence interval are determined. This suggests the forecasted streamflow at KGT.19 station by the rainfall periods from 1965–1990. The general scheme of water level forecasting model based on MLR model is depicted in Figure 1. We collected the following data: Topographic map from Royal Thai Survey Department (RTSD, 1997) series L 7018 scale 1:50,000 (sheet 5235I- IV 5236 II 5236III 5335III and 5335IV), and remote sensing data by LANDSAT-5 SATELLITE 2008. Data were analyzed using MATLAB R2010B.

Due to major economic and social damage to trade, industry and infrastructure in the affected eastern region of Thailand, we focus mainly on four severely-impacted districts in the vicinity of Klong Luang (KL) KGT.19 station sub-watershed: Phanat Nikhom (PN), Mueang Chon Buri (MC), Ban Bueng (BB), and Phan Thong (PT) as shown in Figure 2. The latitude and longitude coordinates were provided by the Royal Thai Survey Department and Meteorological Station Information (Chon Buri Province).

Results and Discussions

The data are presented in Figure 3, while regression parameters of annual rainfall against annual streamflow at Klong Luang sub-water shed are tabulated in Tables 1 and 2. Based on the results shown in Table 1, the highest regression and best correlation coefficients can be found at MC and PT, respectively. These results shown that the water rainfall at MC, PN and PT can statistically forecast the water streamflow at KL, whereas the results from BB are less clearly correlated. The positive regression coefficient indicates the increasing trend of rainfall. Since the data used in this study covered a relatively short period from 1965 to 1990, the effectiveness of forecasting can be improved by using a larger data set covering a longer duration. In order to apply the proposed model in practice, the 95% confidence interval was determined as shown in Table 2. The equation (1) can be extended to provide a simplified MLR model in this study, and can be expressed as follows (Equation 2): where β is the regression coefficient. We then applied equation (2) to forecast streamflow over the period 1965 to 1990 ($m = 26$). It can be solved by the ordinary multivariate normal maximum likelihood estimation [7]. We used the meregress, the multivariate linear regression functions in MATLAB's Statistics Toolbox, function to determine coefficients for a multivariate normal regression of the *d*-dimensional responses. The two optimal MLR based models for streamflow network have been found as shown in Figure 4. All forecasted results were determined using the 11 models, as shown in Table 3. After forecasting, we employed the MLR model in equation (2) to show the different results predicted by the 11 models. Figure 5 shows the obtained results, including an outlier due to overshoot rainfall data in 1974. We applied the optimal MLR models in Figure 4 to simulate forecasting streamflow for the upcoming years and for the future; the results are presented in Table 4. The forecasts predicted a linear increase in streamflow at KL sub-watershed. In equation (1) we assume that the rainfall data are independent, meaning that the rainfall records are independent of other variables. In this study, although the MLR model cannot itself prove causation, we found that the streamflow variable is usually affected by the rainfall variables. Moreover, the presence of outliers as shown in Figure 5 can seriously affect the results. In this case, the MLR might not give a high-precision result due to the linear combination in equation (2). We therefore suggest three possible solutions to deal with this problem: 1) by excluding it while determining the regression coefficient; 2) by subdividing data over shorter periods; and 3) by using nonlinear or other intelligent models. However, the process of nonlinear or intelligent models is iterative and requires an initial value and greater computational process, complicating the task of identifying an optimal model [7]. We can obtain better results using other perspective, for example in terms of the normal probability distribution. The proposed MLR model and the original water streamflow are compared to reveal their respective capabilities in forecasting streamflow, as shown in Figure. 6. We also found that expected values, μ, are similar for both models (μ_0=115.9619, μ=111.0704 (in million cubic metre: mcm) for the proposed MLR model and the original water streamflow, respectively). After excluding outliers, the probability curve reverted to follow close to the proposed MLR model. Therefore, the MLR model is identified as the most suitable and practical approach to preliminary investigation, presenting the fastest computation without requiring a training process. We applied the optimal model (Model 8) to forecast streamflow for the period 1991-2004. The results (Figure 7) indicate a very high correlation coefficient ($R = 0.8240$) and can be discussed in terms of hydrological characteristics. The mean annual streamflow at KL sub-watershed are shown in Table 5. Figure 8 shows average monthly streamflow and the results forecast by the proposed model. We found that the bulk of the streamflow occurred from May to October, whilst maximum monthly rainfall was in September and October. A map of flood stage forecast is able to identify the location related to an expected inundation area obtained from topographical water level forecasting. After applying the optimal rainfall-streamflow network model,the map shows that streamflow increased in accordance to the rainfall. This was probably an artefact of the MLR model itself, as the value of regression coefficients generally adjusts output water level because there is more location in the model to be determined; however, in this experiment, the results obtained from the optimal model explained the greatest variation in rainfall of MC, PT, and PN. The reason is that the coefficient of BB have no an effect on the model due to the lowest correlation in Tables 1 and 2. Thus, when rainfall data were high in individuality, an attribution is made to the streamflow result in terms of consistency of the provided information, and the optimal model was supported, because streamflow processes (especially daily streamflow) are generally accepted as nonlinear. One type of commonly used nonlinear time series model is the threshold autoregressive (TAR) model [17]. However, the model uses the threshold value instead of using any observed values. The rainfall to streamflow process is non-linear in nature and it is this that proves to be the main difficulty in streamflow modeling. Nevertheless, the rainfall-streamflow model can be reasonably approximated as a linear process [17], provided backwater effects are negligible. Therefore, we hypothesized the relationship between the rainfall and streamflow processes on the basis of general forecasting equation. In terms of the tabulated values of the performance statistics, a model may perform better in the testing stage than in the training stage. The data set used for testing happens to be more appropriate for that model [18]. However, in this study, we propose the streamflow network model that the simplified MLR model can be applicable, with the simplest computing requirement and no training process.

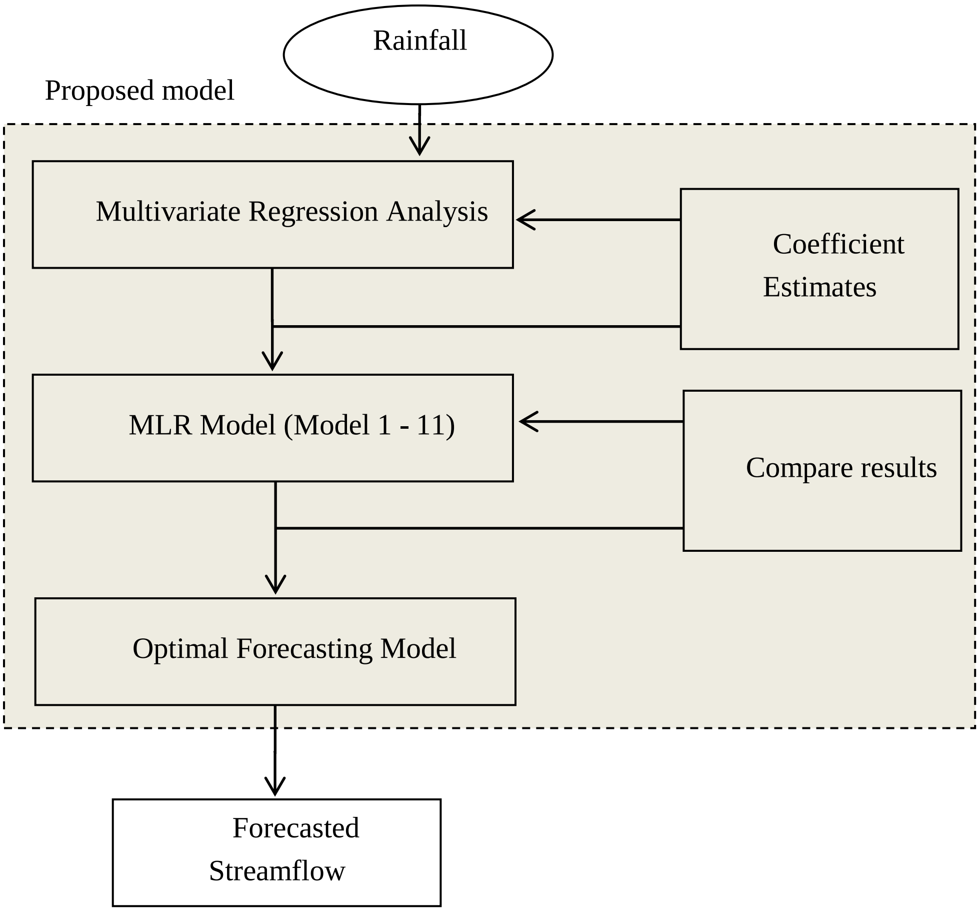


Figure 1 Proposed general scheme of early flood forecasting model

Table 1 Regression parameters of annual rainfall against annual streamflow at KL

District	Regression coefficient	Correlation coefficient
Mueang Chon Buri (MC)	0.1582*	0.6062
Phanat Nikhom (PN)	0.1198	0.5037
Ban Bueng (BB)	0.0402	0.2024
Phan Thong (PT)	0.1565	0.6532**

Note: * = highest positive slop, ** = best correlation

Table 2 Regression parameters of annual rainfall against annual streamflow at KL within 95% confidence interval for the coefficient estimates

District	Regression coefficient (upper)	Regression coefficient (lower)
Mueang Chon Buri (MC)	0.2457	0.0708
Phanat Nikhom (PN)	0.2064	0.0332
Ban Bueng (BB)	0.1221	-0.0417
Phan Thong (PT)	0.2329	0.0801

Table 3 Comparison result of statistically different MLR models based on Eq. (2)

No.	Model	Regression coefficient, β	Correlation coefficient
1	[MC, PT]	[0.0369, 0.0511]	0.7470
2	[MC, BB]	[0.1231, -0.0368]	0.6041
3	[MC, PN]	[0.0642, 0.0304]	0.6354
4	[PT, BB]	[0.1076, -0.0298]	0.6361
5	[PT, PN]	[0.0579, 0.0326]	0.7455
6	[PN. BB]	[0.0905, 0.0101]	0.5159
7	[MC, PT, BB]	[0.0699, 0.0694, -0.0608]	0.7452
8	[MC, PT, PN]	[0.0236, 0.0484, 0.0181]	0.7833*
9	[MC, BB, PN]	[0.0986, -0.0342, 0.0249]	0.6213
10	[BB, PN, PT]	[-0.0375, 0.0400, 0.0815]	0.7035
11	$\begin{bmatrix} \mathrm{MC} & \mathrm{PT} \\ \mathrm{BB} & \mathrm{PN} \end{bmatrix}$	$\begin{bmatrix} 0.0672 & 0.0688 \\ -0.0602 & 0.0032 \end{bmatrix}$	0.7650

Note: * = best correlation

Urban areas were outlined to superimpose a forecasted water level map as shown in Figure 9. The results suggest that urban areas along the stream must be monitored carefully for the next century, and fold forecast must be incorporated into policies and strategies to stimulate local economic growth and development. In Figure 9, the simulated result shows the geological landscape map with forecasted flood at return periods of 500 years over the urban area in KL sub-watershed. The model forecasts a flooding area covering 57.32 % of urban area of 7 km^2 which was above MSL. The accuracy of the proposed method depends on the mathematical model (MLR) in which the most important factor is the set of regression coefficients [4]. In our experiment, we applied Model 8 to be considered as an optimal rainfall-streamflow network model.

However, the general model can readily be adapted to a variety of other networks. We suggest applying the obtained results to maintain the drainage basin to limit flood damage to lower elevation land, e.g., along the banks of the Bang Pakong river, and the building of the Klong Luang dam. Based on the forecast results up to 500 years, flood management measures need to be evaluated to determine appropriate use of each area as part of a large-scale flood protection project starting from selection of development areas, taking into account the flood risk and impact on the environment as mentioned above.

Table 4 Results of forecasted streamflow at KL sub-watershed

AD Year (from 2014)	Forecasted by Model 8 (mcm)	Forecasted by Model 11 (mcm)
2019 (+5)	121.92	163.59
2024 (+10)	122.63	169.34
2064 (+50)	128.31	215.32
2114 (+100)	135.41	272.78

Table 5 Summary of water level forecasting result at KL sub-watershed by the return periods

Return period (year*)	Max. Daily rainfall (mm)	Max. Inlet flow (m^3/sec)	Water level from MSL (m)	Surcharge depth (m)
5	91.70	227.13	26.05	0.58
10	103.70	269.45	57.34	0.66
50	119.00	373.85	85.97	0.85
100	141.50	422.83	148.36	0.94
500	167.40	547.59	201.85	1.16

Note: * = AD Year (from 2014)

In terms of landscape planning (which is directly related to regional planning), typically, the result from forecasted mapping is very useful for a landscape architect as well because it can inform the planning process through a basic understanding of the linkages between development and environment. To mitigate flooding, all projects such as condominium or house building should use pervious materials that allow permeation of water into the ground, and maximize the planting of trees [11]. Furthermore, the choice of drainage system introduced the concept of nature to be used such as drainage ditches on the mixing tube system or pervious pavement adjusted to the curb-less streets, through which water can flow. This offers an alternative to the traditional stormwater drain design widely used in developed countries, and fits with the concept of low impact development (LID) technique since it is applicable to all areas and dimensions [12, 13]. The effects of traditional development practices on the hydrologic cycle must be monitored as well. The optimal rainfall-streamflow network model indicated the forecasted flood changes in terms of hydrological model by an expected return period. Increased use of impervious surfaces associated with urbanization have resulted in increased surface streamflow [14, 15], increased streamflow velocity, decreased time of concentration, and decreased water quality [16].

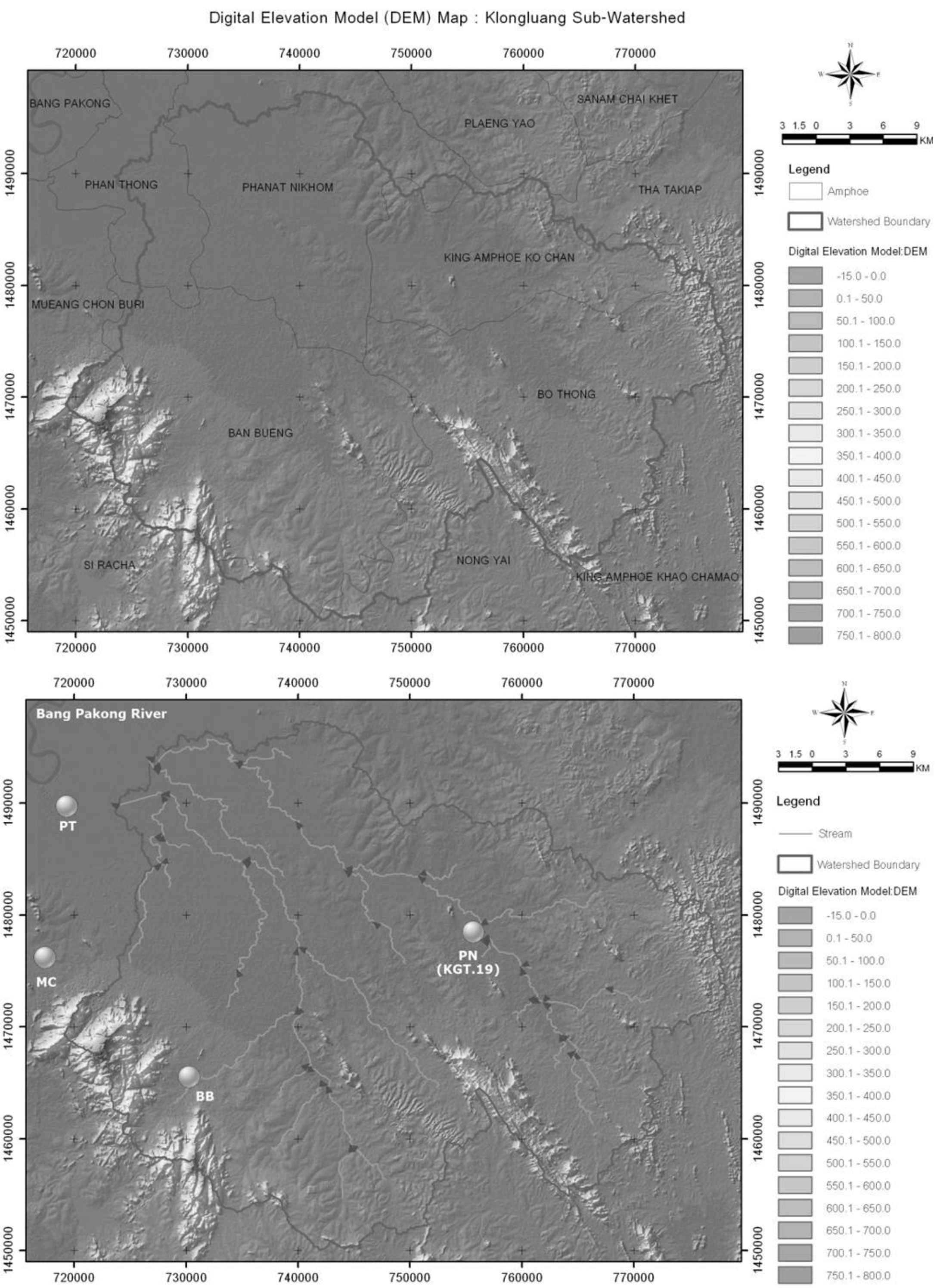


Figure 2 DEM Map and coordinate of PT (13°47'95", 101°00'73")
MC (13°22'00", 100°59' 00") BB(13°14'38", 101°7'26") PN(13°23'17", 101°20'40") in this study
(KL = Klong Luang, PN = Phanat Nikhom, MC = Mueang Chon Buri, BB = Ban Bueng, PT = Phan Thong)

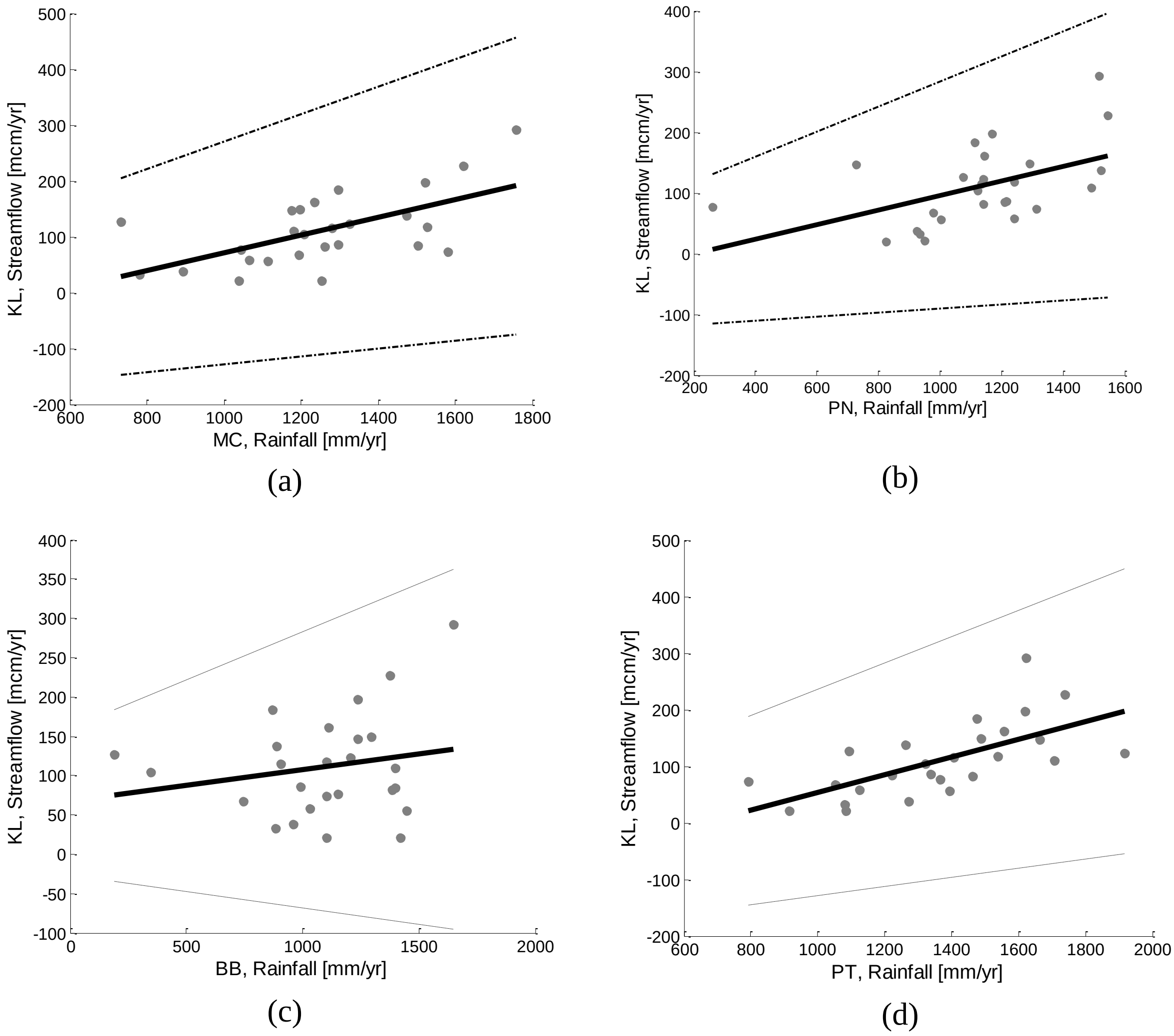
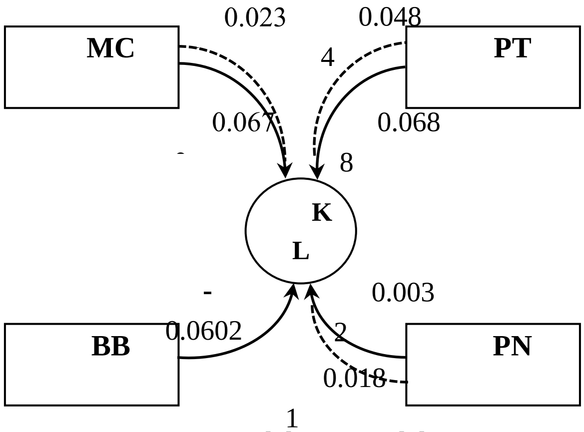
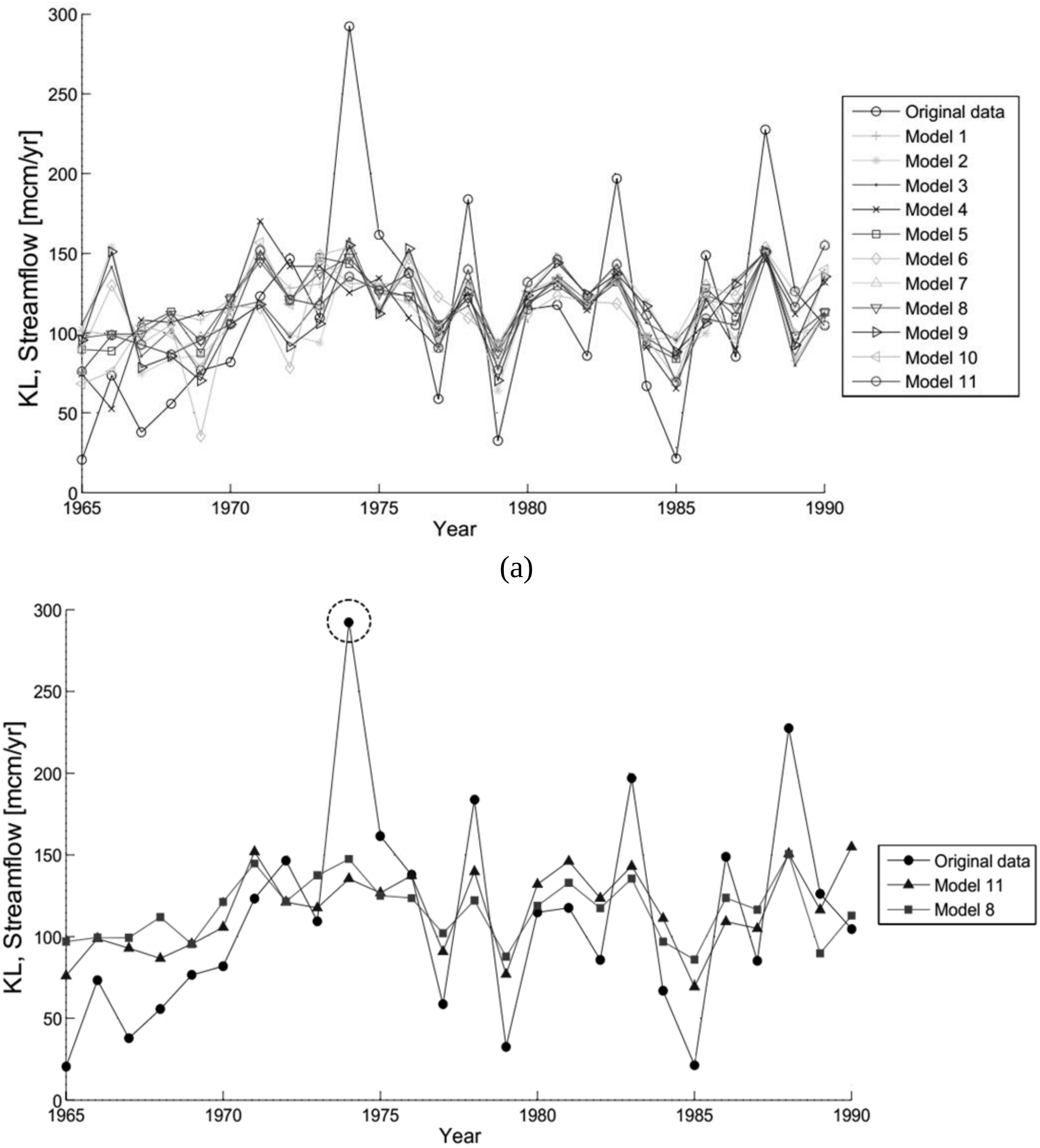


Figure 3 Results of MLR with 95% confidence interval for
(a) MC vs KL, (b) PN vs KL, (c) BB vs KL, and (d) PT vs KL



-- Model 8, - Model 11

Figure 4 Two optimal MLR based rainfall-streamflow network models



(b)

Figure 5 (a) Comparison of forecasted streamflow determined by all models, (b) Selected MLR based models 8 and 11 at KL and outlier marked in dotted circle (note: 'mcm' stands for million cubic metres)

Conclusions

In this paper, the MLR based model has been proposed for water level and flood forecasting in Klong Luang sub-watershed in eastern Thailand. The proposed model offers utility for strategic planning and water resource management, and we have proposed an optimal MLR based model (Model 8) for constructing the streamflow network, and offers the most suitable approach as a benchmark for water level forecasting in Klong Luang sub-watershed KGT.19 station; this model presents the fastest computation without a training process. This network may also be readily applicable to other flood forecasting models.

This investigation concludes with recommendations as a choice of methods of determining the significance of rank different coefficients. The MLR based model 8 is suggested based on its highest correlation level. However, extreme factors such as high rainfall, low infiltration rates, high tides, flow regulation schemes, domestic and urban land use patterns, temperature, and relative humidity have also to be taken into account as part of the forecasting exercise. In this study, we proposed preliminary results using the rainfall-streamflow network model

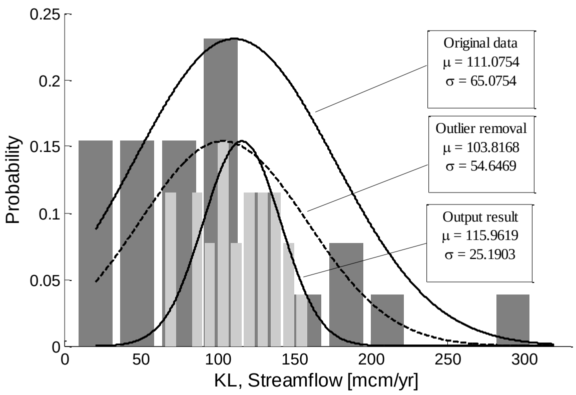


Figure 6 Comparison of real KL streamflow and forecasted normal distributions

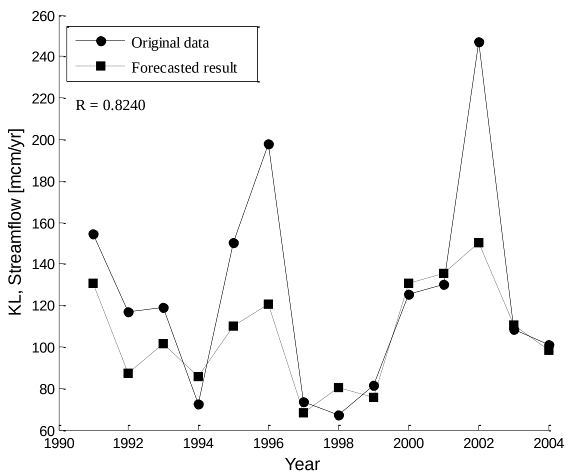


Figure 7 Applying optimal model 8 to forecast streamflow period 1991 – 2004

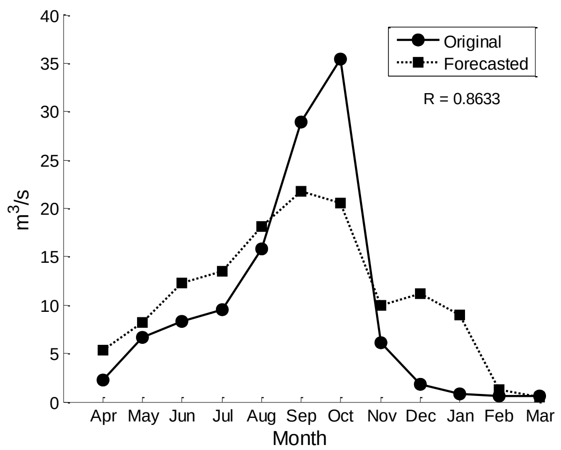


Figure 8 Actual and forecasted results in an average monthly streamflow (The original data provided by the hydrology and water management center for the Eastern region Chon Buri, Thailand, May, 2012)

Acknowledgements

The authors acknowledge the Royal Thai Survey Department for supplying GIS and RS data in this experiment, and sincerely appreciate the referees for providing valuable comments and/or suggestions that benefit our manuscript.

References

[1] Fotheringham, S., and Rogerson, P. 1994. Spatial analysis and GIS. NCGIA /Department of Geography. University of Buffalo, Texas, USA.

[2] Office of Natural Resources and Environment Chon Buri. 2004. Royal Irrigation Department. Rainfall and stream flow data at KGT.19 station. Ministry of Agriculture and Coope-ratives, Thailand.

[3] Danh, N.T., Phien, H.N., and Dasgupta, A. 1999. Neural network models for river flow forecasting. Water SA. 25 (1):33–39.

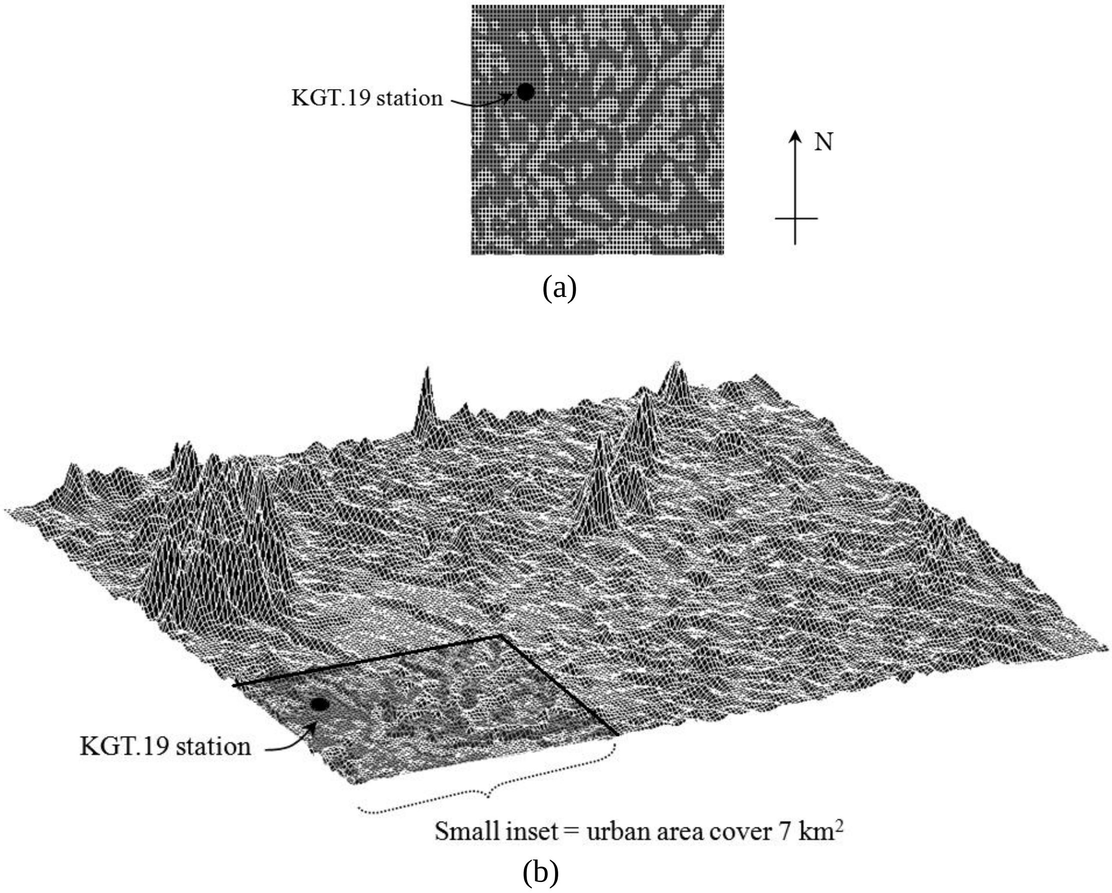


Figure 9 Typical simulated map with forecasted water level at a return period of 500 years. (a) Area of interest covers KGT.19 station by cell size of 100 m^2 in top view, (b) Corresponding result in 3D map

[4] Chau, K. W., Wu, C. L., and Li, Y.S. 2005. Comparison of several flood forecasting models in Yangtze River. Journal of *Hydrologic Engineering*. 10 (6): 485–491.

[5] Phien, H.N., and Kha, N.D.A. 2003. Flood forecasting for the upper reach of the Red River Basin North Vietnam. Water SA. 29: 267–272.

[6] Hossain, M.M., Neaupane, K., Tripathi, N., and Piantanakulchai, M. 2013. Prediction of groundwater arsenic contamination using geographic information system and artificial neural network. EnvironmentAsia. 6(1):38–44.

[7] Little, R.J.A., and Rubin, D.B. 2002. Statistical analysis with missing data (2nd Ed.). John Wiley & Sons, NJ, USA.

[8] Mwale, F.D., Adeloye, A.J., and Rustum, R. 2014. Application of self-organising maps and multi-layer perception-artificial neural networks for streamflow and water level forecasting in data-poor catchments: the case of the Lower Shire floodplain, Malawi. (in press).

[9] Tokar, A.S., and Johnson, P.A. 1999. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering. 4(3): 232–239.

[10] Smith, J. and Eli, R.N. 1995. Neural-network models of rainfall-runoff process. Journal of Water Resources Planning and Management. 121(6): 499–508.

[11] Rushton, B. 2001. Low-impact parking lot design reduces runoff and pollutant loads. Journal of Water Resources Planning and Management. 127(3): 172–179.

[12] Jia, H., Yao, H., and Yu, S.L. 2013. Advances in LID BMPs research and practice for urban runoff control in China. Frontiers of Environmental Science and Engineering. 7(5):709–720.

[13]Dietz, M.E. 2007. Low Impact Development Practices: A review of current research and recommendations for future directions, Water, Air, & Soil Pollution. 186: 351–363.

[14] Hollis, G.E. 1977. Water yield changes after the urbanization of the Canon's Brook catchment, Harlow, England. Hydrological Sciences Bulletin. 22: 61–75.

[15] Jennings, D.B., and Jarnagin, S.T. 2002. Changes in anthropogenic impervious surfaces, precipitation and daily streamflow discharge: A historical perspective in a mid-Atlantic subwatershed. Landscape Ecology. 17:471–489.

[16] Makepeace, D.K., Smith, D.W., and Stanley, S.J. 1995. Urban stormwater quality: Summary of contaminant data. Critical Reviews in Environmental Science and Technology. 25(2): 93–139.

[17] Todini, E., and Bossi, A. 1986. PAB (Parabolic and Backwater) an unconditionally stable flood routing scheme particularly suited for real time forecasting and control, Journal of Hydraulic Research. 24:405–424.

[18] Phien, H.N., and Kha, N.D.A. 2003. Flood forecasting for the upper reach of the Red River Basin North Vietnam. Water SA. 29: 267–272.