JOURNAL OF
ENVIRONMENTAL HYDROLOGY
Open Access Online Journal of the International Association for Environmental Hydrology

VOLUME 24 2016

PREDICTION AT UNGAUGED SITE WITH TOPOLOGICAL KRIGING AND MODIFIED GROUP METHOD OF DATA HANDLING

Basri Badyalina
Ani Shabri
Nur Amalina Mat Jan

Department of Mathematics
Faculty of Science
Universiti Teknologi Malaysia
Johor Bahru, Johor, Malaysia

Since the number of ungauged basins is increasing over the world, there is a need for reliable streamflow prediction. This research aims to develop an accurate and reliable instrument for streamflow prediction in ungauged basins. Modified Group Method of Data Handling (MGMDH) is proposed for flood streamflow prediction. The MGMDH is an enhancement from conventional Group Method of Data Handling (GMDH) model. Previously, GMDH model is using second order polynomial as a function to describe the input and output variable relationship. In MGMDH model, tangent sigmoid function is used as a function to describe the relationship between input and output variable. Topological kriging, a geostatistical technique is then used to build a region in Peninsular Malaysia based on rainfall data on gauged station. The result indicates that the proposed approach MGMDH based top-kriging delivers the best performance in terms of prediction accuracy.

INTRODUCTION

The availability of hydrological data is important and demanded by water resources planning and management which is enabled to be prepared when flood happens. Lack of hydrological data produces an uncertainty in both the design and management of water resources systems. In Peninsular Malaysia, the number of ungauged catchment is increasing from year to year (Mamun et al., 2011). Ungauged catchments are those catchments without hydrological data. In other words, hydrometric stations do not exist in ungauged basins or they became inactive. Sivapalan et al. (2003) defines ungauged catchments as the ones with no available records (in terms of both data quantity and quality) of hydrological observations. Although the needs for hydrological information are increasing, their technical and human capacities are declining which make it harder for data collection. If hydrometric station were to be made accessible for the extension of hydrometric networks, it will take 10 to 30 years before enough data are collected. Generally, long term historical hydrological data were required to yield a more reliable prediction compared to short term hydrological data and could also reduce the risk. The UK Flood Estimation Handbook (FEH) notes "many flood prediction problems arise at ungauged sites that there are no flood peak data" (Reed and Robson, 1999). The International Association of Hydrological Sciences (IAHS) acknowledge this need in 2002, and implemented the Prediction of Ungauged Basins (PUBS) as a research agenda for the coming decade (Sivapalan et al., 2003). Prediction of flow characteristics of ungauged catchments is usually based on transferring information from gauged to ungauged sites, a process called regionalization (Wagener and Wheater, 2006). Regionalization refers to a process of transferring hydrological information from gauged to ungauged or poorly gauged catchments to make prediction of the streamflow. Streamflow regionalization can be done through rainfall-runoff models. In previous scenario, model parameters are used as tools to transfer hydrological information from gauged to ungauged catchments. Regionalization process also includes fitting a probability distribution function to streamflow data and then connecting the relationship to catchment characteristics. In relating flood quantile at site of interest to catchment characteristics a power form equation is mostly used (Seckin, 2011). At ungauged sites, linear regression (LR) model is always worthy of depended model to estimates flow statistics or flood quantiles quantiles (Shu & Ouarda, 2008; Pandey & Nguyen, 1999). Mamun et al. (2011) used linear regression of various return periods in ten flood regions in Peninsular Malaysia. Badyalina and Shabri (2015) introduced one of sub neural network models that is Group Method of Data Handling (GMDH) model which is used to make prediction of flood quantile at ungauged site. GMDH model was initially established by Ivakhnenko (1971) for identification and modeling of complex system. The advantages of GMDH model are; choosing significant input for the model automatically, short time model development, instinctive arrangement of model structure and enhanced prediction performance (Onwubolu, 2008). In MGMDH model, the model has the same architecture but different transfer function that is tangent sigmoid function. Using other transfer function rather than conventional polynomial transfer function can improve the model prediction accuracy. Topological Kriging is used to form region in Peninsular Malaysia according to rainfall characteristics. The MGMDH model is then used to establish the nonlinear relationships between the site physiographical space and at site characteristics is estimated separately based on region produced by topological kriging (TK).

CASE STUDY

The hydrometric station network of Peninsular Malaysia is chosen as the case study of this work. According to the following criteria, 60 hydrometric stations located in Peninsular Malaysia are selected (Figure 1).

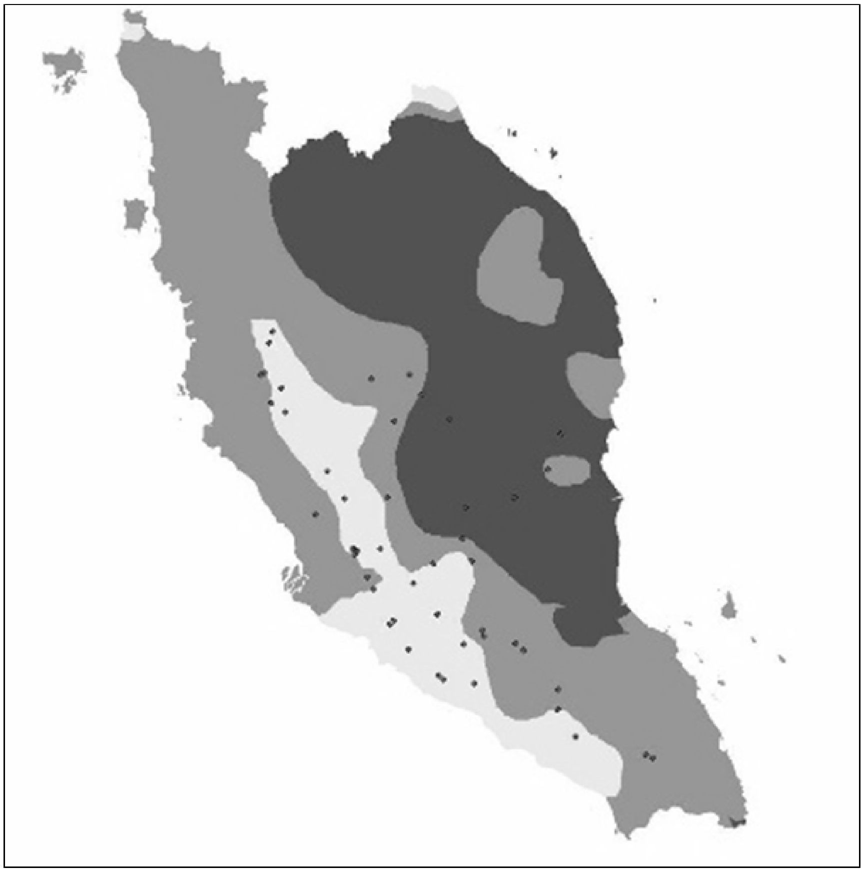

Figure 1. Map showing location stream flow station used in this study and region created by Topological Kriging.

1. To get reliable at-site estimation, a historical flood record of 15 years or longer are needed.
2. The gauged river should present natural flow regime.

The historical data of 60 catchments in the province of Peninsular Malaysia were implemented in this study. They are located within latitude and longitude of. The areas of these catchments are ranging between 16.3 km^2 to 19,000 km^2. The locations of these catchments are shown in Fig. 1. Three types of data, physiographical, meteorological and hydrological are used in this study. The variables selected in this study on the basis of previous study by Seckin (2011) and by Shu and Ouarda (2008). Four physiographical variables are the catchment area, elevation, mean river slope and longest drainage path. The meteorological variable is mean annual total rainfall.

MODIFIED GROUP METHOD OF DATA HANDLING

The conventional GMDH used a second order polynomial function as a partial description (PD). In modified GMDH, tang-sigmoid function is proposed to construct GMDH.

Step 1: Select input variables $X = x_1, x_2, \ldots, x_M$ where M is the total number of input. The data are separated into training and testing data sets. The training data set is used to construct a GMDH model and the testing data set is used to evaluate the estimated GMDH model.

Table 1. Descriptive statistics of hydrologic, physiographical and meteorological variables

Variables	Min	Mean	Max	STD
AREA [km^2]	30	2453.91	19000	5010.91
ELV [m]	4	35.58	1450	26.10
LDP [m]	3800	38457.97	280000	59553.88
SLP [%]	0.01	0.40	2.56	0.50
AMR [mm]	314.30	2099.75	4678.70	717.26
Q_{10} [m^3/s]	12.87	891.60	7256.76	1702.28
Q_{50} [m^3/s]	29.54	1386.32	10089.80	2477.91
Q_{100} [m^3/s]	43.82	1194.17	1656.02	2895.21

Step 2: Construct $L = M(M-1)/2$ new variables $Z = z_1, z_2, \ldots, z_L$ in the training data set for all independent variables and choose a PD of the GMDH. Conventional GMDH is developed using the polynomial as PD. In this study, a PD structure, namely tansigmoid function using the polynomial function is proposed in construct the GMDH. The tan sigmoid model is used in the form

$$z_l = G(x_i, x_j) = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \qquad \text{for} \quad l = 1, 2, \ldots, L \tag{1}$$

$$z_l = \left(\frac{2}{1+e^{-2g_l}}\right) - 1 \qquad where\ g_l = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \tag{2}$$

Step 3: Estimate the coefficient of the PD. The vectors of coefficients of the PDs are determined using the least square method.

Step 4: Determine new input variables for the next layer. There are several specific selection criteria to identify the input variables for the next layer. After completing the process, the algorithm has constructed L number of new input variable but only one from L is chosen for the new input of GMDH based on RMSE value. After determining the new input, the whole GMDH process is repeated again. If $\mathrm{RMSE}_k \leq \mathrm{RMSE}_{k-1}$, set new input variables and repeat the GMDH process, otherwise if RMSE show an improvement, the process is stopped and the results from the previous minimum value of RMSE are used.

$$RMSE_k = \sqrt{\frac{1}{p}\sum_{i=1}^{p}\left(y_i - z_{i,k}\right)^2}\, for\, k = 1, 2 \ldots, L \tag{3}$$

TOPOLOGICAL KRIGING (TK)

TK applies kriging methods over a geographical space and combines two groups of forcing for hydrological variability (Skøien et al., 2005; Archfield et al., 2013) . The first group consists of variables that are continuous in space such as rainfall, evapotranspiration and soil characteristics, which are related to local runoff generation. The second group of forcing is related to aggregation and routing in the stream network. The resulting stream flow variables are only defined for points on the stream network. In TK, the aggregation effects that lead to these groups of variables are represented by the catchment boundaries associated with each point on the stream network. TK is then applied to build region on Peninsular Malaysia based on rainfall data.

RESULTS AND DISCUSSION

Table 1. Jackknife validation result in term of prediction accuracy

	Hydrological Variables	LR	GMDH	Top-Kriging MGMDH
NASH	q10	0.6831	0.8023	0.8311
	q50	0.6722	0.7814	0.810741
	q100	0.6324	0.7439	0.7912
RMSE	q10	825.9714	634.2316	511.3202
	q50	874.1341	687.3211	572.7238
	q100	912.2361	724.6173	612.1327
MAE	q10	398.1529	235.8594	173.2312
	q50	437.8761	262.7143	192.0452
	q100	481.2317	290.2147	229.1360

Table 1 shows the performance of each model that is used in this study. The proposed model, modified group method of data handling (MGMDH) outperformed comparison model that are LR and GMDH model in term of prediction accuracy. Topological Kriging had created a region that has similar hydrological behavior based on rainfall data at gauged site. By using integration of both Topological Kriging and MGMDH, the prediction of model has improved and outperformed the other two comparison models. Other than that, the result indicated by changing the transfer function of GMDH model using tangent sigmoid also increases the prediction accuracy. Overall, the Topological Kriging GMDH model leads to a much better performance with CE, RMSE and MAE indices compared to GMDH, and LR model. Thus, the Topological Kriging GMDH model is better than benchmark model that is traditional LR model.

CONCLUSION

The integration of Topological Kriging and MGMDH for flood quantile prediction at ungauged site are proposed in this paper. Topological Kriging used rainfall extreme event and interpolate to produce a region in Peninsular Malaysia. Then MGMDH model is used to build the functional relationship between flood quantiles and characteristics of catchment inside the region. Three various return period in this study were used to see the capability of the model to estimate for short term and long term. Topological Kriging and MGMDH model was compared to two other models that are LR and GMDH models. The result shows Topological Kriging with MGMDH outperformed the comparison model in relative accuracy in prediction of flood quantile.

ACKNOWLEDGMENTS

The paper was reviewed by Dr. Zaharatul Amani Zakaria, Faculty of Informatic & Computing, Universiti Sultan Zainal Abidin and Mrs. Nurul Izzaty Yunus, Department of Mathematics, Universiti Teknologi Malaysia.

REFERENCES

Archfield, S.A., A. Pugliese, A. Castellarin, J.O. Skøien, and J.E. Kiang. 2013. Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach? Hydrology and Earth System Sciences, Vol. 17(4), pp. 1575-1588.

Badyalina, B. and A. Shabri. 2015. Flood Estimation at Ungauged Sites Using Group Method of Data Handling in Peninsular Malaysia. Jurnal Teknologi., Vol. 76(1).

Ivakhnenko, A.G. 1971. Polynomial theory of complex systems. IEEE Transactions on Systems, Man, and Cybernetics., Vol. 4, pp. 364-378.

Mamun, A.A., A. Hashim and Z. Amir. 2011. Regional Statistical Models for the Estimation of Flood Peak Values at Ungauged Catchments: Peninsular Malaysia. Journal of Hydrologic Engineering. Vol., 17(4), pp. 547-553.

Onwubolu, G.C. 2008. Design of hybrid differential evolution and group method of data handling networks for modeling and prediction. Information Sciences, Vol. 178(18), pp. 3616-3634.

Robson, A. and D. W. Reed. 1999. Flood estimation handbook. Institute of Hydrology, Wallingford.

Sivapalan M, K. Teaches. S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, et al. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal. Vol. 48(6), pp. 857-80.

Skøien, J. O., R. Merz and G. Blöschl. 2005. Top-kriging? geostatistics on stream networks. Hydrology and Earth System Sciences Discussions. Vol. 2(6), pp. 2253-2286.

Shu, C. and T. B. M. J. Ouarda. 2008. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system., Journal of Hydrology. Vol. 349(1), pp.31-43.

Seckin, N. 2011. Modeling flood discharge at ungauged sites across Turkey using neuro-fuzzy and neural networks. Journal of hydroinformatics., Vol. 13(4), pp. 842-849.

Wagener, T. and H. S. Wheater. 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology., Vol. 320(1), pp.132-154.

ADDRESS FOR CORRESPONDENCE

Basri Badyalina
Department of Mathematics
Faculty of Science
Universiti Teknologi Malaysia
81310, Skudai, Johor Bahru, Johor
Malaysia

Email: basribadyalina@gmail.com