International Journal of GEOMATE, Nov., 2020, Vol.19, Issue 75, pp.209-214
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2020.75.26880
Geotechnique, Construction Materials and Environment

RATIONALIZATION OF DISCHARGE DATA FLUCTUATIONS USING MINIMUM RESIDUALS MODEL OF MOVING AVERAGE METHODS IN CITARUM CASCADE RESERVOIRS

Eddy Iskandar Muda Nasution[1], Arwin Sabar[2], Indah Rachmatiah Siti Salami[3], Dyah Marganingrum[4]

[1] Doctoral Student, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Indonesia
[2] Professor, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Indonesia
[3] Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Indonesia
[4] Geotechnology Research Center, Indonesian Institute of Sciences, Indonesia

*Corresponding Author, Received: 13 July 2020, Revised: 17 Aug. 2020, Accepted: 24 Sept. 2020

Abstract: Since 1981, a cascade of dams consisting of the Saguling, Cirata, and Jatiluhur reservoirs has been constructed along the Citarum River in order to fulfill demands for flood control, water supply, and electrical power generation, and this requires the implementation of an appropriate hydrological approach. The efficient operation of a water resource system requires an accurate water flow analysis that approximates the actual fact as closely as possible to minimize risks in decision-making. Discharge is a random and stochastic component in hydrology, influenced by rainfall variations with a maximum in the tropical wet season and a minimum in the dry season. Thus, discharge analysis requires a statistical approach especially minimum residuals. Values for average annual inflows has constructed through four different calculation techniques for the moving average method to rationalize the fluctuations in discharge data Citarum cascade reservoirs by reducing randomness in the data, with the eventual aim of obtaining more rational trend lines. An analysis of minimum residuals between the results of the calculation and those of model regressions has demonstrated that the cumulative moving average calculation technique provides the best fit among all the models with the smallest residual value, as shown in its lowest mean squared error values of 0.83, 2.04, and 1.77 for the three reservoirs.

Keywords : Discharge fluctuation, Citarum cascade reservoirs, Minimum residual, Moving average.

1. INTRODUCTION

The Citarum is one of the most important rivers in West Java with a watershed covering approximately 6,600 km^2, extending for 270 km from its source on Mt. Wayang in the Bandung Regency to its mouth on the Java Sea in the Bekasi Regency and carrying an average annual flow of 5.5 billion m^3 [1]. A number of reservoirs have been built on this river, namely the Saguling, Cirata, and Jatiluhur reservoirs, each reservoir at a lower elevation receives water that has passed through another at a higher elevation, an arrangement known as a serial or cascade of reservoirs [1]. The principal function of these three reservoirs is to provide flood control, electrical power, and water supply to their surroundings, with the Jatiluhur reservoir carrying the specific additional function of supplying water to the Jakarta Capital Region [1]. The fulfillment of these water demands requires the appropriate hydrological approach in the management of water resources [2-5].

Rain and discharge are hydrological components which are random variables with stochastic characteristics [2]. Discharge is influenced by rainfall variations [5], with resulting fluctuations that lead to issues such as flooding during its maximum in the tropical wet season and water supply crises during its minimum in the dry season [2]. These fluctuations lead to phenomena such as discharge extremes, the occurrence of which is influenced by topography, morphology, and climate [2], an analysis thereof requires an empirical method that suits the actual hydrological characteristics [2,3,6]. An accurate analysis of discharge would help minimize risks in decision-making [4]. One of the methodological approaches is mathematically based, namely statistics [2,5,7], which has been empirically found to be an accurate and powerful tool in analyzing hydrological time series data in order to detect and quantitatively describe every process that underlies a particular set of observations [7].

Hydrology utilizes time series analysis to build a mathematical model in order to produce a hydrological synthesis, predict hydrological occurrences, detect trends and movements in hydrological data, fill in missing data, ignore data that interfere with frequency stability, and rationalize hydrological data [7]. Mathematical models can be expressed in either linear or nonlinear manners, but linear models are far more common in hydrology [7] due to the ability to extend the dataset up to the limitations of expansion time by the first-order stability of Taylor series according to the perturbation theory [8].

Rationalizing the fluctuations in discharge data for the Citarum cascade reservoirs using the moving average method to identify a pattern in rising or falling level by smoothing of the data has the advantage that is can be adjusted as needed to cope with noise in the data. Moving averages is a highly important statistical method in modeling, analyzing, and predicting time series data, whether graphically or numerically, by transforming and set up the characterization of the data. The performance of the model can be monitored by comparing the calculation results with historic time series data, the difference is known as the residual [9,10].

This study's aim is to identify the appropriate moving average model for the calculation of discharges for the Citarum cascade reservoirs. Some of the calculation techniques for the moving average method, are simple moving average (SMA), cumulative moving average (CMA), weighted moving average (WMA) [11], and exponential moving average (EMA) [12,13], so it is necessary to find out which one provides the best-fitting model with the smallest residual value. The time series data utilized for this purpose consists of historic discharges for 1994-2019 from the Water Resources Research and Development Centre in the Ministry of Public Works and Housing, taken from monitoring posts at the Saguling, Cirata, and Jatiluhur reservoirs.

2. METHOD

2.1. Moving Average Calculation Techniques

The moving average method is a process for calculating averages that move through the time series at each observation by taking averages across a specific time period, thus producing a time series of the averages from another set of time series data [13]. This method is also known as the rolling mean [11], running means or rolling averages [13], and moving median or middle smooth [14]. The averaging removes randomness in the data and produces trends with a minimum of residuals [13].

The most basic calculation technique in the moving average method is the SMA model [11,15], where the average x_i that moves at time t to the previous time period over a time span of N. The output in this model is a series of equally weighted averages with no influence from values outside the specified span of time, as expressed in the Eq. (1):

$$SMA = \frac{\sum_{i=t}^{t=N-1} x_i}{N} \quad (1)$$

In the CMA model, the average x_i that moves from time $t+1$ into the next period over a span of N by accumulating the average over the time period t over the same time span [11]. The most recent output in this model is influenced by earlier outputs as expressed in the following Eqs. (2) and (3):

$$CMA_t = \frac{\sum_{i=1}^{N} x_i}{N} \quad (2)$$

$$CMA_{t+N} = \sum_{t+1}^{t=N} CMA_t + \frac{x_{t+1} - CMA_t}{N+1} \quad (3)$$

On the other hand, the WMA and EMA models implement a weighting factor ω with greater weight for more recent data [11,16]. The EMA model in particular accumulating of earlier averages over the time period for the same time span [12] as expressed in the Eqs. (4) and (5):

$$WMA = \sum_{j=i-\omega+1}^{i} \omega_j . x_j \quad (4)$$

$$EMA = \sum_{i=t}^{t=N} \left(X_i \left(\frac{2}{1+N} \right) \right) + EMA_{t-1} \left(1 - \left(\frac{2}{1+N} \right) \right) \quad (5)$$

The longer the time span N is, the smoother the resulting trend will become. However, even though the data series can be extended, there is a boundedness to the acceptable expansion of the time span, determined by the first-order stability of the Taylor series that constitutes the polynomial approach in the perturbation theory introduced by Rayleigh and Schrodinger to provide a solution for the approximation of eigen values in a linear model [8,9]. Perturbation theory comprises mathematical methods for finding an approximate solution to a problem, by starting from the exact solution of a related, simpler problem. A critical feature of the technique is a middle step that breaks the problem into "solvable" and "perturbation" parts. Perturbation theory is applicable if the problem at hand cannot be solved exactly, but can be formulated by adding a "small" term to the mathematical description of the exactly solvable problem [8].

2.2. Calculation of Residuals

An evaluation of model performance requires the calculation of the resulting residual [17]. Residual testing is particularly crucial in regression diagnostics, whose graphical presentation is one of the most informative methods in statistics [10].

The geometric residual r_i is the vertical distance on the y axis between observed data y_i and the observational data's regression line $\hat{y}_i$ as expressed in the following Eqs. (6) and (7):

$$r_i = y_i - \hat{y}_i \quad (6)$$

$$\hat{y}_i = a_0 + a_1 . x_i \quad (7)$$

where a_0 is the intercept and a_1 the slope, while its implementation in the model is expressed in the following Eq. (8):

$$y_i = \alpha_0 + \alpha_1.x_i + \varepsilon_i \quad (8)$$

where ε_i represents *random error*, while a_0 and a_1 are estimations of the actual values α_0 and α_1, from which is obtained the sum square of the minimum residual as expressed in the Eq. (9) [10]:

$$Q_{min} = [\sum r_i^2]_{min} \quad (9)$$

Therefore, the residual is the difference between the observed data and the output value of the model's equation, which displays as the observed error in an accurate model by showing the existing asymmetry in the response function and the independent variable in regressions [10].

The calculation of residuals is shown as $(y_i - \hat{y}_i)^2$, being the Sum Squared Error (SSE), becoming the Mean Squared Error (MSE) once divided by the number of data points, and then turning into the Root Mean Squared Error (RMSE) by way of extracting its root. The smaller the residual, the closer the model is to the actual condition [18].

3. RESULTS

3.1. Concept for Rationalizing Discharge Fluctuations

The components of hydrology are influenced by global warming [19] and land use changes [5,6,20]. Successive land use changes in upstream regions, from forests into agriculture lands, then into rural and eventually urban settlements, have led to the degradation of hydrological functions. This causes phenomena such as extreme changes in discharge due to increased surface runoff and decreased base flow, thus increasing the risk of flooding in the wet season and drought in the dry season [5]. Land use changes can also affect the local microclimate by increasing local temperatures and reducing the incidence of rain [5]. This hydrological uncertainty stems from climate change [4,6,21,22,23], consists of the main components of temperature and rainfall, which hits especially hard in the tropics where there is little variation in temperature but large variations in rainfall [24].

Rainfall and discharge are random components with stochastic characteristics in a statistical hydrology watershed model [5]. Flow rate is influenced by variations in rainfall [8], which fluctuates greatly with a maximum in the wet season and a minimum in the dry season [5], thus giving discharge a periodic character in hydrological time series. This periodicity must be visible within a temporal span shorter than a year, such as monthly or biannual. If the periodicity occurs over a span longer than a year, it would be impractical to use monthly or biannual time series data to analyze it [9].

The explanation above shows that it is very important to accurately rationalize discharge data fluctuations for the Citarum cascade reservoirs in order to inform the efficient operation of water resource systems, the management and control of extreme events such as floods and water supply crises, effective design for hydraulic structures such as dams and bridges, and the optimization of water supply for domestic, industrial, irrigation, and power generation purposes [25,26,27], all of which would benefit from the minimization of risk in decision-making [7].

3.2. Model Performance

Discharge time series data for 1994–2019 produced average annual values for inflow into the Saguling, Cirata, and Jatiluhur reservoirs as shown in Fig. 1.

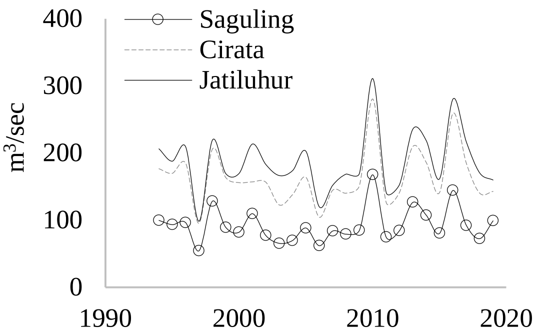

Fig. 1 Average annual inflows into the Saguling, Cirata, and Jatiluhur reservoirs in 1994 -2019

The fluctuations in these average annual inflows has rationalized with four moving average calculation techniques over 5-year time periods, as specified in Government Regulation No. 37 of 2010 on Dams, Article 44, which mandated that the operation of a reservoir should be evaluated every 5 years. This produces SMA, CMA, WMA, and EMA models for the Saguling, Cirata, and Jatiluhur Reservoirs as shown in Figs. 2-4 below.

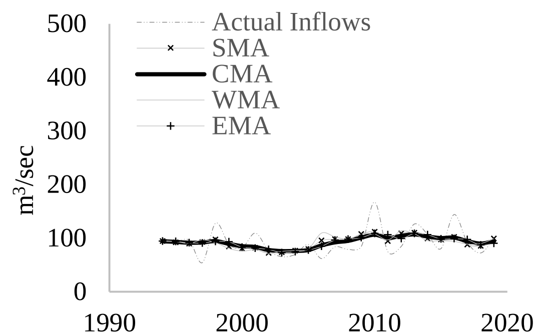

Fig. 2 *SMA*, *CMA*, *WMA*, and *EMA* models compared to actual average annual inflows to the Saguling reservoir in 1994 – 2019

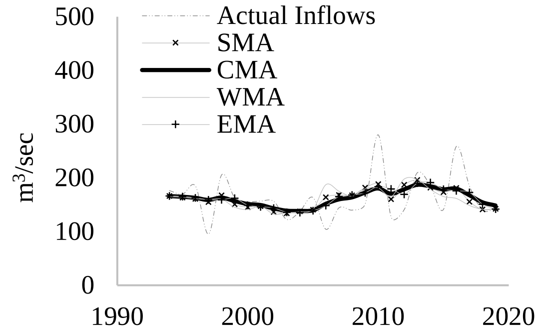

Fig. 3 *SMA*, *CMA*, *WMA*, and *EMA* models compared to actual average annual inflows to the Cirata reservoir in 1994 – 2019

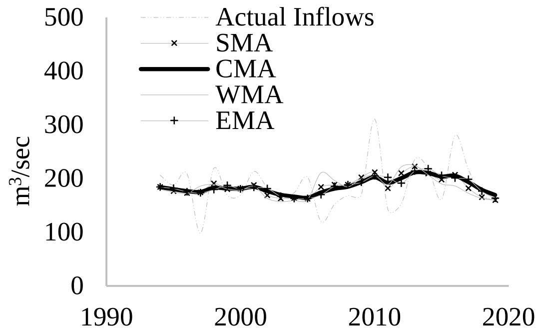

Fig. 4 *SMA*, *CMA*, *WMA*, and *EMA* models compared to actual average annual inflows to the Jatiluhur reservoir in 1994 – 2019

Figure 1 shows that the average annual inflows into the Saguling, Cirata, and Jatiluhur reservoirs exhibit a high level of fluctuation, and this can be attributed to the swings in rainfall, which lead to phenomena as discharge extremes from a maximum in the wet season to a minimum in the dry season as an effect of climate change as a result of global warming. Meanwhile, Figs. 2-4 show that the technique of constructing a time series of averages out of the historic time series data – with four different moving average calculation techniques – has successfully rationalized fluctuations in average annual inflow by reducing the randomness of the data, thus producing more rational trend lines.

3.3. Analysis of Residuals

The analysis of residuals between the calculation and regression results of the *SMA*, *CMA*, *WMA* and *EMA* models for the Saguling, Cirata, and Jatiluhur reservoirs is presented in Figs. 5-7, while the *SSE*, *MSE* and *RMSE* values are presented in Table 1.

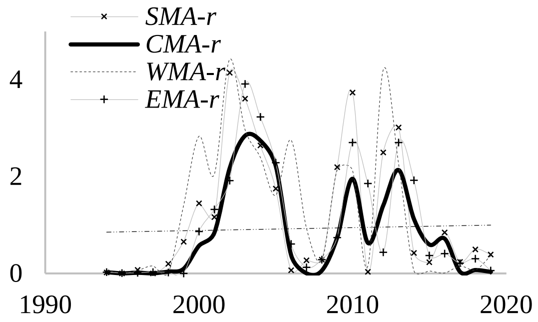

Fig. 5 Residuals from *SMA*, *CMA*, *WMA* and *EMA* relative to the regression model for the Saguling reservoir in 1994 – 2019

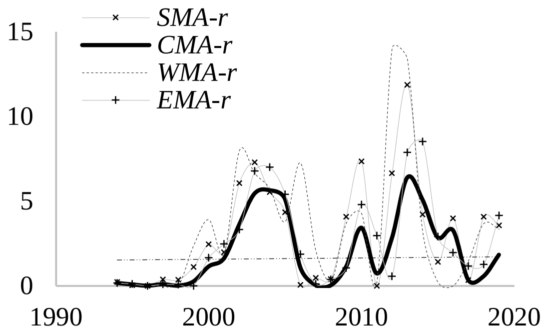

Fig. 6 Residuals from *SMA*, *CMA*, *WMA* and *EMA* relative to the regression model for the Cirata reservoir in 1994 – 2019

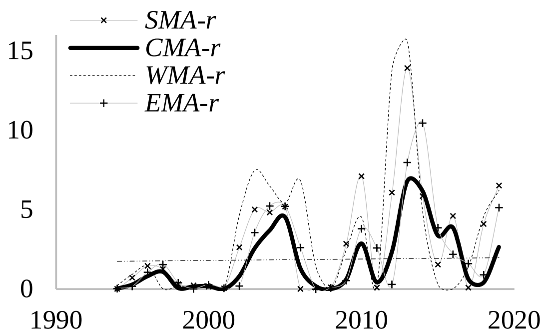

Fig. 7 Residuals from *SMA*, *CMA*, *WMA* and *EMA* relative to the regression model for the Jatiluhur reservoir in 1994 – 2019

Table 1 *SSE*, *MSE*, and *RMSE* values for *CMA*, *EMA*, *SMA*, and *WMA* models of the Saguling, Cirata, and Jatiluhur reservoirs

Reservoir	Model	*SSE*	*MSE*	*RMSE*
Saguling	*CMA*	21.46	0.83	0.91
	EMA	26.42	1.02	1.01
	SMA	30.48	1.17	1.08
	WMA	33.62	1.29	1.14
Cirata	*CMA*	52.94	2.04	1.43

	EMA	66.71	2.57	1.60
	SMA	78.37	3.01	1.74
	WMA	89.93	3.46	1.86
Jatiluhur	*CMA*	45.94	1.77	1.33
	EMA	59.86	2.30	1.52
	SMA	74.99	2.88	1.70
	WMA	89.74	3.45	1.86

Figures 5-7 show that the residuals from the *CMA* model deviate the least from the regression trend line compared to the other models, so the CMA model can be seen as the best-fit model for the Saguling, Cirata, and Jatiluhur reservoirs with the lowest residual values, which can be seen in the MSE values of 0.83, 2.04, and 1.77 respectively for the three reservoirs as shown in Table 1.

4. DISCUSSION AND CONCLUSION

4.1. Discussion

The analysis of time series data for discharge of Citarum cascade reservoirs shows that the CMA model displays the most rational performace due to its ability to accumulate of earlier averages over the time period in assessing the most recent average value. The EMA model similarly techniques for earlier averages over the time period in calculating the present average, although it assigns greater weighting to more recent periods. By comparison, the SMA model by calculating averages without weighting or influence from earlier averages over the time period. The WMA model, which only assigns greater weight to more recent data, exhibits the poorest performance.

The largest residual value produced by the WMA model indicates that this model suffers from severe limitations when it is applied to discharge data as a hydrological component due to the variable's random and stochastic nature. These discharge fluctuations are influenced by rainfall variations between the wet and the dry season, which takes place within a span of less than 1 year, such that it exhibits periodicity that allows to infer future discharge patterns and characteristics from data on past discharge patterns and characteristics.

4.2. Conclution

Average annual inflows into the Saguling, Cirata, and Jatiluhur reservoirs fluctuate greatly due rainfall variations lead to phenomena as discharge extremes, which in turn can be attributed to climate change as a result of global warming. Discharge analysis requires an empirical method that suits actual hydrological characteristics in order to rationalize the fluctuations by reducing the randomness of the data, thus producing more rational trend lines and minimizing residuals. An accurate analysis of discharge that closely approximates the facts is crucial to the efficient operation of water resource systems in fulfilling water use demands.

Moving averages is a highly important statistical method in the modeling, analysis, and prediction of hydrological time series data. By analyzing the residual between calculation results and model regression results, can conclude that the CMA model is the best-fitted model for the Saguling, Cirata, and Jatiluhur reservoirs with the smallest residual values. The performance of the other models can be presented in descending order as *EMA*, *SMA*, and finally *WMA* in the last place.

5. REFERENCES

[1] Corsita L., Arwin, Muntalif B. S. and Salami I. R. S., The Hydrological Regime, Water Quality Assesment and Trophic Status of The Reservoir, Case Study: Jatluhur Cascade Citarum Reservoir, International Conference on Latest Trends in Food, Biological and Ecological Sciences, ICLTFBE, Phuket, Thailand, 2014, pp. 18-24. DOI: 10.17758/IAAST.A0714022

[2] Sabar A., Perubahan Iklim, Konservasi Lahan dan Ancaman Banjir, dan Kekeringan di Kawasan terbangun, Pidato Ilmiah Guru Besar Institut Teknologi Bandung, Majelis Guru Besar ITB, CV. Sentama Wikarya, Bandung, 2009, pp. 1-58.

[3] Casadei S., Liucci L., and Valigi D., Hydrological Uncertainty and Hydro-power: New Methods to Optimize the Performance of the Plant, Journal Energy Procedia, Vol. 59, 2014, pp. 263-269. DOI: 10.1016/j.egypro.2014.10.376

[4] Salleh N., Aziz M., and Adzhar N., Optimal Design of a Rain Gauge Network Models: Review Paper, Journal of Physics: Conference Series, Vol. 1366, 2019, 012072. DOI: 10.1088/1742-6596/1366/1/012072

[5] Afifah E., Sabar A., Wulandari S. and Marselina M., The Reliability Study of Raw Water Sources in the Development of Potable Water Supply Systems in Indonesia, International Journal of Geomate, Vol. 16, Issue 54, 2019, pp. 209-216. DOI: 10.21660/2019.54.75034

[6] Shu L., Ullrich P. and Duffy C., Solver for Hydrologic Unstructured Domain (SHUD): Numerical Modeling of Watershed Hydrology with The Finite Volume method, Geoscientific Model Development, 2020, pp. 1-33. DOI: 10.5194/gmd-2019-354

[7] Machiwal D. and Jha M., Hydrologic Time Series Analysis: Theory and Practise, Capital Publishing Company, New Delhi, India, 2012, pp. 1-303.

[8] Kato T., Perturbation Theory for Linear Operators, 2nd Ed., Reprint of The 1980 Ed.,

Springer-Verlag Berlin Heidelberg, Printed in Germany, 1995, pp. 1-619.
[9] Montgomery D. C., Jennings C. L. and Kulahci M., Introduction to Time Series Analysis and Forecasting, a John Wiley and Sons, Inc. Publication, 2008, pp. 1-445.
[10] Martin J., Adana D. D. R. and Asusero A. G., Fitting Models to Data: Residual Analysis, a Primer, 7th Chapter in a Book, Uncertainty Quantification and Model Calibration, Edited by Jan Peter Hessling, IntechOpen, 2017, pp. 133-173. DOI: 10.5772/68049.
[11] Parekh P. and Ghariya V., Analysis of Moving Average, International Journal of Engineering and Technical Research (IJETR), Volume 3, Issue 1, 2015, pp 178-179.
[12] Klinker F., Exponential Moving Average versus Moving Exponential Average, Journal of Math. Semesterber, Vol. 58, No. 1, 2011, pp. 97-107, DOI: 10.1007/s00591-010-0080.
[13] Hyndman R. J., Moving Average, International Encyclopedia of Statistical sciences, 2010, pp 866-869. DOI: 10.1007/978-3-642-04898-2_380
[14] Helsel D. R. and Hirsch R. M., Statistical Methods in Water Resources, Techniques of Water Resources Investigations of The United States Geological Survey, Hydrologic Analysis and Interpretation, Book 4, Chapter A3, United States Geological Survey, 2002, pp. 1-510.
[15] Svetunkov I. And Petropoulos F., Old Dog, New Tricks: A Modelling View of Simple Moving Averages, International Journal of Production Research, Vol. 56, Issue 18, 2018, pp.6034-6047.
DOI: 10.1080/00207543.2017.1380326
[16] Perry M. B., The Weighted Moving Average Technique, Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran, 2010, John Wiley and Sons, Inc.
DOI: 10.1002/9780470400531.eorms0964
[17] Makridakis S. G, Wheelwright S. C. and Hyndman R. J., Forecasting: Methods and Application, 3rd edition, Wiley, New York, 1997, pp. 1-632.
[18] Faraway J. J., Linear Models with R, Chapman and Hall /CRC, a CRC Press Company, 2005, pp. 1-241.
[19] Perminov V. and Goudov A., Mathematical Modeling of Forest Fires Initiation, Spread and Impact on Environment, International Journal of Geomate, Vol. 13, Issue 35, 2017, pp. 93-99. DOI: 10.21660/2017.35.670.
[20] Permatasari R., Sabar A., Natakusumah D. K. and Samaulah H., Effects of Watershed Topography and Landuse on Baseflow Hydrology in Upstream Komering South Sumatera, Indonesia, International Journal of Geomate, Vol. 17, Issue 59, 2019, pp. 28-33. DOI: 10.21660/2019.59.4695.
[21] Jayanti M., Arwin, Hadihardadja I. K., Ariesyadi H. D. and Messakh J. J., Climate Change Impacts on Hydrology Regime and Water Resources Sustainability in Cimanuk Watershed, West Java, Indonesia, International Journal of Geomate, Vol. 19, Issue 71, 2020, pp. 90-97. DOI: 10.21660/2020.71.9215.
[22] Elkarim A. B., Awawdeh M. M., Alogayell H. M. and Al-Alola S. S. S., Integration Remote Sensing and Hydrologic, Hydraulic Modelling on Assessment Flood Risk and Mitigation: Al-Lith City, KSA, International Journal of Geomate, Vol. 18, Issue 70, 2020, pp. 252-280. DOI: 10.21660/2020.70.68180.
[23] Ghosh S. and Misra C., Assessing Hydrological Impacts of Climate Change: Modeling Techniques and Challenges, The Open Hydrology Journal, Vol. 410, 2010, pp. 115-121.
DOI: 10.2174/1874378101004010115
[24] Hermawan, E., Classification of Rainfall Pattern in Several Areas Sumatra Island Based on The Analysis of Spectral Techniques, Journal of Meteorology and Geophysics, Vol. 11, No. 2, 2010, pp. 75-85. DOI: 10.31172/jmg.v11i2.67
[25] Londhe S. and Charhate S., Comparison of Data-driven Modelling Techniques for River Flow Forecasting, Hydrological Sciences Journal – Journal des Sciences Hydrologiques, Vol. 55, Issue 7, 2010, pp. 1163-1174. DOI: 10.1080/02626667.2010.512867
[26] Marganingrum D., Carrying Capacity of Water Resources in Bandung Basin, IOP Conference Series: Earth and Environmental Science, Vol. 118, 2018, 012026. DOI: 10.1088/1755-1315/118/1/012026
[27] Huang Y. F., Mirzaei M. and Yap W. K., Flood Analysis in Langat River Basin using Stochastic Model, International Journal of Geomate, Vol. 11, Issue 27, 2016, pp. 2796-2803. DOI: 10.21660/2016.27.1143.