




Comparing automatic baseflow separation filters on three watersheds in the urbanised area of East Java, Indonesia

Indarto Indarto  , Mujiono Hardiansyah, Mohamad Wawan Sujarwo 

University of Jember, Faculty of Agricultural Technology, Jl kalimantan No. 37 Kampus Tegalboto, 68121, Jember, Jawa Timur, Indonesia

RECEIVED 25.05.2020

REVIEWED 06.10.2020

ACCEPTED 15.12.2020

Abstract: Baseflow is the primary source of water for irrigation and other water needs during prolonged dry periods; accurate and rapid estimation of baseflow is therefore crucial for water resource allocation. This research aims to estimate baseflow contribution during dry periods in three small watersheds in East Java: Surabaya-Perning (114 km²), Lamong-Simoangrok (235 km²), and Bangsal-Kedunguneng (26 km²). Six recursive digital filters (RDFs) algorithms are explored using a procedure consisting of calibration, validation, evaluation and interpretation. In this study, the period of July to September is considered as the peak of the dry season. Moreover, data for the period 1996 to 2005 is used to calibrate the algorithms. By yearly averaging, values are obtained for the parameters and then used to test performance during the validation period from 2006 to 2015. Statistical analysis, flow duration curves and hydrographs are used to evaluate and compare the performance of each algorithm. The results show that all the filters explored can be applied to estimate baseflow in the region. However, the Lyne–Hollick (with $RMSE = 0.022, 0.125, 0.010$ and $R^2 = 0.951, 0.968, 0.712$) and exponentially weighted moving average or EWMA (with $RMSE = 0.022, 0.124, 0.009$ and $R^2 = 0.957, 0.968, 0.891$) for the three sub-watersheds versions give the best performance.

Keywords: baseflow, East Java, recursive digital filter (RDF), separation, urbanised area, water management, watershed

INTRODUCTION

Predicting the portion of the baseflow component of total streamflow is a crucial activity in water management, and this is mainly the case in the East Java region. This region is subject to a specific tropical climate regime featuring two different monsoon seasons (dry and wet seasons). The dry season is usually from May to September, and the region receives little or no rainfall during the peak of this dry season. As a result, lower streamflow can limit water resources available for irrigation, industrial and residential uses.

During the dry season, low or absent rainfall will reduce the quickflow component of streamflow. The remaining streamflow present in rivers during these dry periods is therefore contributed by baseflow, and at the peak of the dry season, this baseflow component will be dominant. In contrast, in the wet season (usually from October to April), the streamflow of most rivers in East Java is significantly increased, and the magnitude (intensity and duration) of rainfall can lead to flood events.

The prediction of baseflow in the region will benefit water resource management decisions both for dry and wet seasons. In the dry season, the calculated baseflow serves as the water available for irrigation and other purposes. In the wet season, the separation of baseflow and quickflow components will help water resource managers to estimate the streamflow level that may cause flood events.

Another benefit of baseflow prediction is the calculation of minimum flows available for environmental and conservation-related purposes. All habitats (for both flora and fauna species) exist in river environments in which their sustainability mainly depends on the amount of water available in the river. By predicting the baseflow component of the river streamflow, the water resources manager can maintain favourable streamflow levels for many linked habitats. More detail on the benefit of baseflow prediction has been presented in many works [BRODIE, HOSTETLER 2005; MURPHY *et al.* 2009; SMAKHITIN 2001a, b].

The recursive digital filter (RDF) method is one of the techniques used for baseflow separation. RDF algorithms use the same filter principle as is applied in signal processing techniques. In the signal processing domain, the filters are used to separate high-pulse from low-pulse signals. RDF algorithms in baseflow separation use the same principle. High-pulse signals correspond to high flow (quickflow) filtered from existing low-pulse or baseflow according to particular criteria. Many researchers have been working on this method. They tried to use one parameter [CHAPMAN, MAXWELL 1996], two parameters [BOUGHTON 1993], or three parameters [JAKEMAN, HORNBERGER 1993], the latter being known as the IHACRES (acronym for identification of unit hydrographs and component flows from rainfall, evaporation and streamflow data) three-parameter algorithm. Other algorithms developed by the same principles are the Lyne–Hollick algorithm [LYNE, HOLLICK 1979], the Chapman algorithm [CHAPMAN 1991], and the exponentially weighted moving average (EWMA) filter [TULARAM, ILAHEE 2008].

Research works on baseflow separation have been published by several authors, including TALLAKSEN [1995], SMAKHITIN [2001a, b], FUREY, GUPTA [2003], ECKHARDT [2005], TULARAM and ILAHEE [2008], GONZALES *et al.* [2009], KISSEL and SCHMALZ [2020], and SHAO *et al.* [2020].

Likewise, TAN *et al.* [2009] have investigated and compared the separation characteristics of two single-parameter digital filters, i.e., one-parameter algorithms and the conceptual method. Additionally, PARTINGTON *et al.* [2012] evaluated the elasticity of baseflow with climate variability and agricultural land-use change. The baseflow was derived from daily streamflow records using an RDF method.

Moreover, XIE *et al.* [2020] have applied the evaluation criterion to 1815 catchments across the United States. The study evaluates nine baseflow separation (four graphic and five digital filters) methods. Their results show that the digital filter method proposed by ECKHARDT [2005] has the best performance and that its performance is independent of catchment characteristics.

These research results from around the world reveal the importance of baseflow separation techniques for water resource management. This research aims to test filters in more specific regions (i.e. in small watersheds in tropical urbanised area) and by using a series of streamflow data for calibration and validation. The objectives of this research are to (1) determine the parameter values for the six algorithms (filters), (2) compare the performance of the algorithms, and (3) interpret the baseflow index (*BFI*) for the study area.

In East Java, the only available flow data comes from conventional measurements. However, the measured flow both by manual level meters and by automatic water level recorders are limited to only a few watersheds and limited periods of recording. Furthermore, many watersheds are still ungauged. It is therefore supposed that investigation and testing of baseflow separation algorithms could provide an alternative practical solution for local water resource management.

MATERIALS AND METHODS

STUDY AREA

Three small watersheds, Lamong-Simoangrok (1 – Lamong), covering an area of 235 km², Surabaya-Perning (2 – Surabaya), with an area of 114 km², and Bangsal-Kedunguneng (3 – Bangsal)

of 26 km² were selected for this study (Fig. 1). These three small watersheds are located in the most populated region of East Java (Fig. 1). The classified Landsat-8 map shows largely paved or urban areas present in the region, including the cities of Surabaya (provincial capital), Sidoarjo, Gresik, Mojokerto, Jombang and Lamongan. The region is part of the Brantas watershed. The Brantas River (“Kali Brantas” in Fig. 1) is divided into two outlets, one in the direction of Surabaya (“Kali Surabaya”) and the other to the east of Sidoarjo (“Kali Porong”).

The main source of water supply for this area is, therefore, the Brantas River. However, the three small watersheds identified play an essential role as additional supplies of water for the rest of the area that is not covered by the Brantas River. Water resources from the three watersheds are used to supply water for irrigation, households and industrial purposes.

STREAMFLOW DATA

Table 1 displays the summary of daily streamflow data available from the three watersheds.

Table 2 presents a statistical summary of daily streamflow characteristics for the three watersheds.

The streamflow data from the three outlets were analysed to derive the statistical summary of daily streamflow data, i.e., maximum (Max.), mean, median (Med.), standard deviation (*SD*), percentiles of 10, 20, 25, 33, 66, 70, 75, 90 and 100 (P_{10} , P_{20} , P_{25} , P_{33} , P_{66} , P_{70} , P_{75} , P_{90} , and P_{100}). Measurements are in m³·s⁻¹. Mean is calculated as the average of the records (sum of values/number of days). The percentile value (P_{10} – P_{100}) is the value that is exceeded for a certain percentage of time. For instance, the 10th percentile is the value that is exceeded by 10% of the records.

Figure 2 shows the master *FDC* (flow duration curve) for the three watersheds. The master *FDC* is produced using all daily flow data available at each watershed plotted as a single curve.

The slope of flow duration curve (*SFDC*) is calculated by using the segment of the *FDC* from percentile 33 (P_{33}) to percentile 66 (P_{66}), assuming that this segment is relatively linear (Eq. 1).

$$SFDC = \frac{\ln(P_{33}) - \ln(P_{66})}{(0.66 - 0.33)} \quad (1)$$

where: *SFDC* = slope of *FDC*, Q_{33} = flow at percentile 33, and Q_{66} = flow at percentile 66.

The steepest slope in the *FDC* curve shows 1 – that the river is subject to high variation in its flow regime. In contrast, the gentle slopes in the *FDC* curves for 2 – Surabaya and 3 – Bangsal indicate that their flow regimes are relatively stable during one-year periods. The stability of the flow regime is formed by the combination of rainfall events that are distributed spatially around parts of the watershed area and rainfall events that consistently occur during the year. The gentle slope of the *FDC* also indicates that the contribution of the groundwater portion to the river flow is significant [INDARTO 2013; MARSH 2003].

RAINFALL DATA

Daily rainfall data series from 61 locations were available for this study. Recording periods available for each rainfall station vary and span from 5 to 60 years, with the average period of recordings being 44 years.

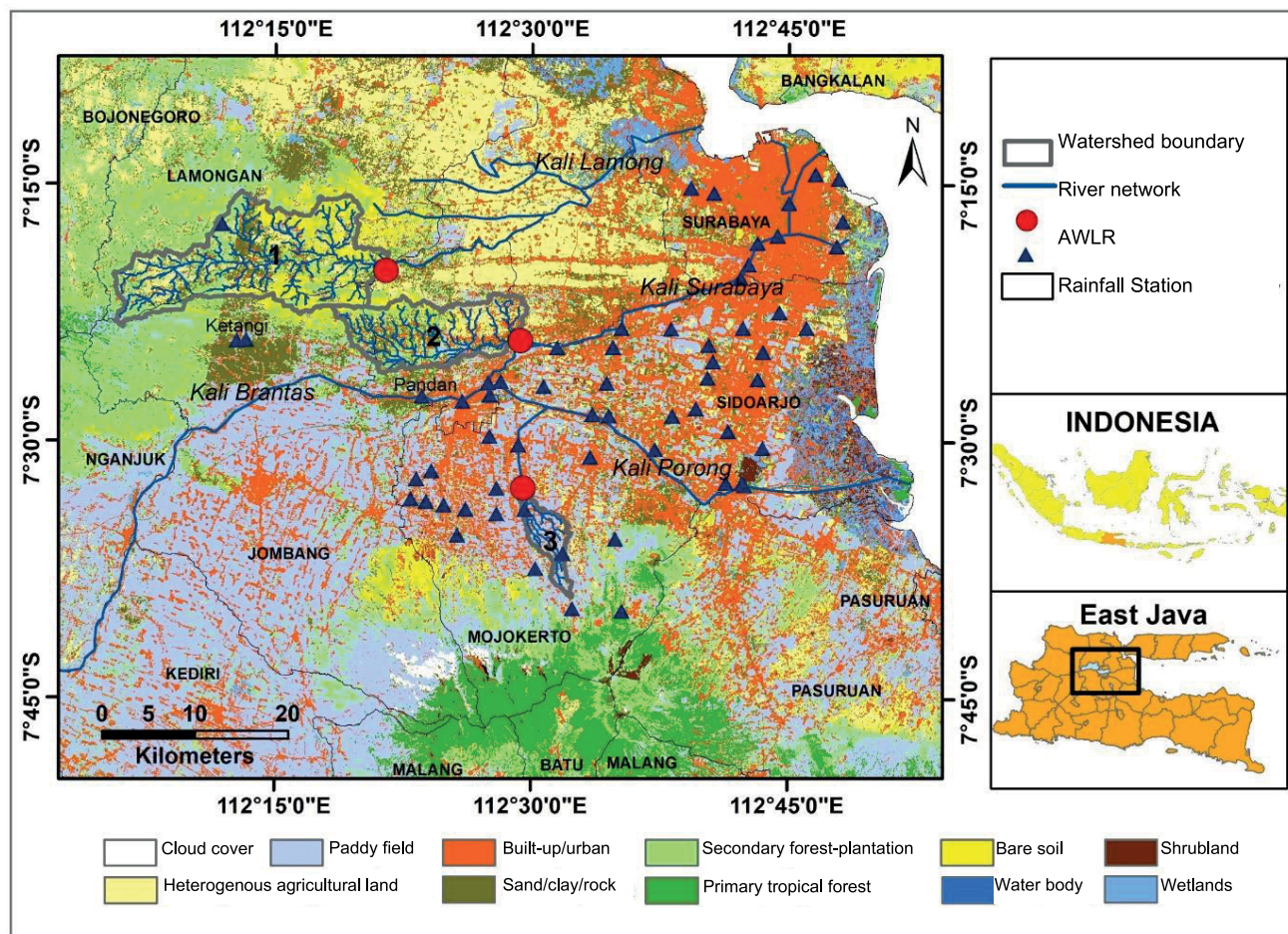


Fig. 1. Study location; source: own elaboration

Table 1. Streamflow data series

Item	1 – Lamong	2 – Surabaya	3 – Bangsal
Start	1996	1996	2002
End	2015	2015	2013
Length (years)	20	20	12
Missing data	03.10–24.11.2002 28.07–03.10.2003 01.10–31.12.2004	01.10– 31.12.2004	01.12– 31.12.2009

Source: own elaboration.

The average annual rainfall ranges from 741 to 2 839 mm·y⁻¹. However, during extremely dry seasons, this can be as low as 442 mm·y⁻¹. In contrast, in extreme wet seasons, the maximum value can reach 10 075 mm·y⁻¹. Moreover, the average monthly rainfall data spans from 6 to 577 mm per month, with the lowest value of 0 mm per month and the highest monthly rainfall of 1 985 mm per month.

The monthly rainfall histogram, as shown in Figure 3, is calculated from all available rainfall data series, and so it summarises monthly average rainfall data from 61 stations. It can be seen that the dry season runs from June to October with its peak between July and September and receives rainfall of <100 mm per month. The wet season period runs from November to April. As an example, the histograms presented in

Table 2. Statistical summary of daily streamflow characteristics for the three watersheds

Parameter	1 – Lamong	2 – Surabaya	3 – Bangsal
	m ³ ·s ⁻¹		
Maximum	346	358.4	8.80
Mean	5.76	47.3	1.30
Median	1.80	37.1	0.95
P ₁₀	0.01	18.4	0.40
P ₂₀	0.09	22.8	0.55
P ₂₅	0.24	23.8	0.61
P ₃₀	0.41	25.9	0.67
P ₃₃	0.54	27.1	0.71
P ₅₀	1.78	37.1	0.95
P ₆₆	3.83	48.6	1.35
P ₇₀	4.37	52.0	1.50
P ₇₅	5.38	57.32	1.69
P ₈₀	7.01	64.4	1.96
P ₉₀	13.8	90.6	2.79
P ₁₀₀	346	358.4	8.81
SD	13.6	36.3	1.03
Skewness	3.24	1.27	1.37
SFDC	5.93	1.77	1.95

Explanations: P₁₀–P₁₀₀ = percentiles of 10–100, SD = standard deviation; SFDC = slope of flow duration curve.

Source: own elaboration.

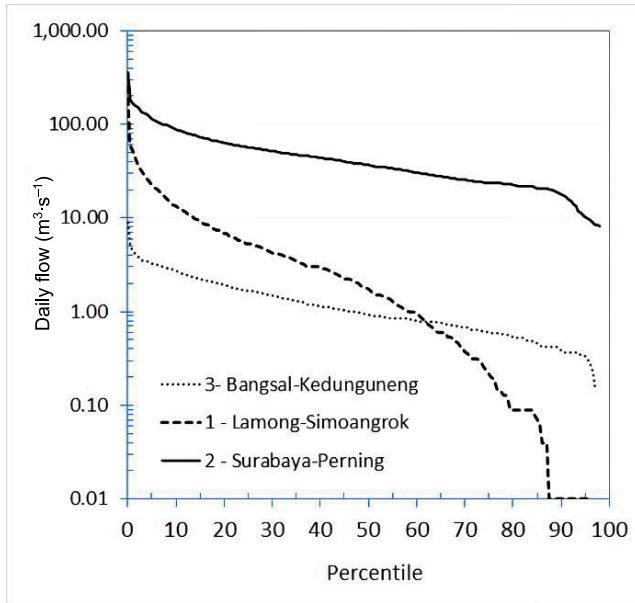


Fig. 2. Master flow duration curve of three watersheds; source: own elaboration

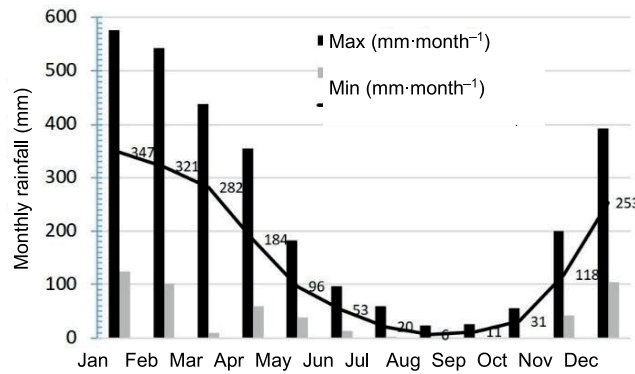


Fig. 3. Histogram of monthly rainfall; source: own elaboration

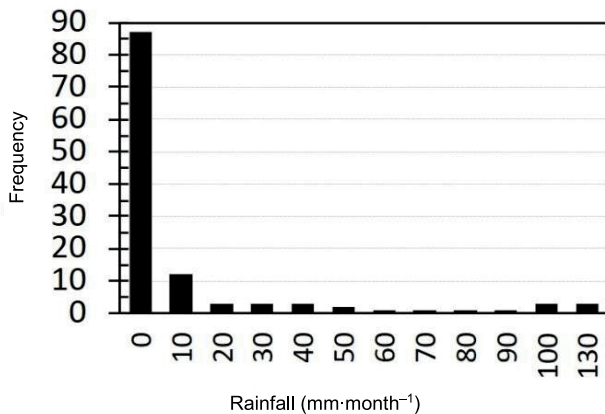


Fig. 4. Frequency distribution of rainfall in Ketangi; source: own elaboration

Figure 4 display the frequency distribution of monthly rainfall from July to September for Ketangi (using 35 years of sample data).

The location of stations is shown in Figure 1. Figure 4 enables us to interpret that the study area is subject to a typical tropical climatic regime which is strongly characterised by

distinct dry and wet seasons. In the wet season, the area receives more rainfall, while in the dry season, the area receives less. Additionally, the peak of the dry season occurs from July to September.

During the last 35 years of dry seasons, very little or no rain fall during July, August and September. This specific climatic regime has a strong influence on hydrological processes in these areas. It is assumed that during the peak of the dry season, the streamflow of rivers will be dominated by baseflow contribution and that the portion of baseflow will be significant.

Significant land resources in the three watersheds are occupied by paddy fields, heterogeneous agricultural land, secondary forest-plantations and built-up/urban areas. Moreover, change in land occupation is observed in the increase of heterogeneous agricultural land, bare soil and abandoned land (sand/clay/rock).

In this study, the digital elevation model (DEM) is applied to derive river networks and to delineate watershed boundaries (Fig. 5).

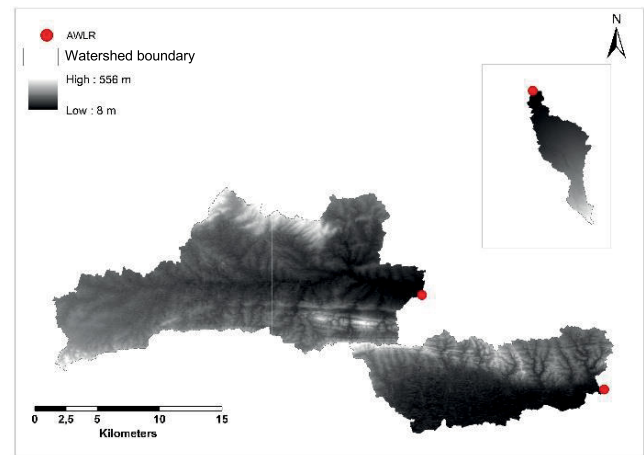


Fig. 5. DEM of the three watersheds; source: own elaboration

The altitude varies among watersheds, ranging from 28 m to 190 m in 1 – Lamong. In the 2 – Surabaya watershed, altitude spans from 8 to 131 m and thus these two watersheds can be seen as being located in relatively flat areas. In comparison, the 3 – Bangsal watershed is located on steeper slopes with altitude varying from 51 to 556 m.

RECURSIVE DIGITAL FILTERS (RDFS)

Equations (2)–(7) show six RDF algorithms used in this study and adopted from GREGOR and MALIK [2012]. The one-parameter algorithm proposed by CHAPMAN and MAXWELL [1996] is, as stated in Equation (2):

$$q_{b(i)} = \frac{km}{2 - km} q_{b(i-1)} + \frac{1 - km}{2 - km} q_{(i)} \tag{2}$$

where: $q_{b(i)}$ and $q_{b(i-1)}$ are baseflow at day i and day $i - 1$ respectively, $q_{(i)}$ is total flow (observed streamflow or observed discharge) at day i , and km is the constant.

The two-parameter RDF algorithm, as published by BOUGHTON [1993] and CHAPMAN and MAXWELL [1996], is shown in Equation (3):

$$q_{b(i)} = \frac{kb}{1 + Cb} q_{b(i-1)} + \frac{Cb}{1 + Cb} q_{(i)} \quad (3)$$

where: $q_{b(i)}$, $q_{b(i-1)}$, $q_{(i)}$ are as in Equation (2), and kb and Cb are the filter constants or parameters.

The IHACRES three-parameter algorithm [JAKEMAN, HORNBERGER 1993], an extension of the Boughton two-parameter algorithm, is shown in Equation (4):

$$q_{b(i)} = \frac{ki}{1 + Ci} q_{b(i-1)} + \frac{Ci}{1 + Ci} (q_{(i)} + \alpha q_{(i-1)}) \quad (4)$$

where: $q_{b(i)}$, $q_{b(i-1)}$, $q_{(i)}$ are as in Equation (2), and ki , Ci , and α are the parameters.

The Lyne–Hollick filter [LYNE, HOLLICK 1979; NATHAN, MCMAHON 1990] uses only one filter constant α , as presented in Equation (5):

$$q_{f(i)} = \alpha q_{f(i-1)} + (q_{(i)} - q_{(i-1)}) \frac{1 + \alpha}{2} \quad (5)$$

where: $q_{f(i)} \geq 0$. In this case, $q_{f(i)}$ and $q_{f(i-1)}$ are the filtered quickflow for day i and day $i - 1$. The α value of 0.925 is recommended for daily stream data and recommended to be applied in three passes [GREGOR, MALIK 2012].

Then, baseflow is calculated using $q_b = q - q_f$. The Chapman algorithm [CHAPMAN 1991] uses one filter constant, as shown in Equation (6):

$$q_{f(i)} = \frac{3\beta - 1}{3 - \beta} q_{f(i-1)} + \frac{2}{3 - \beta} (q_{(i)} - \beta q_{(i-1)}) \quad (6)$$

where: β is the filter constant, the baseflow is calculated using $q_b = q - q_f$.

The exponential smoothing method of baseflow separation or EWMA [TULARAM, ILLAHEE 2008] uses only one filter parameter, as shown in Equation (7):

$$q_{b(i)} = \mu q_{(i)} + (1 - \mu) q_{b(i-1)} \quad (7)$$

where μ is the filter parameter.

Cb and Ci are parameters that allow the shape of the separation to be altered [GREGOR, MALIK 2012].

CALIBRATION PROCEDURE

Firstly, the available streamflow data is divided into two periods. Streamflow data from 1st January 1996 to 31st December 2005 are used as the calibration and streamflow data from 2006 to 2015 are used for validation. This separation is conducted to 2 – Surabaya and 1 – Lamong, while for the 3 – Bangsal watershed, the calibration period is set from 2002 to 2005 and the validation period from 2006 to 2013.

Secondly, for the calibration period, the observed streamflow data from July to September are selected to fit the values of the parameters. The streamflow period from July to September corresponds to the peak of the dry season. During these periods, rainfall is very low (Figs. 2, 3 and 4), and the major component of streamflow is supplied by baseflow. We assume that quickflow or direct runoff (DRO) is very low or tends to zero during this period and therefore observed streamflow of the river is contributed by baseflow.

Thirdly, the fitting of parameter values is guided by streamflow data from 1996 to 2005. The value of parameters for each algorithm has been entered manually by trial and error. This calibration is conducted on the baseflow index (BFI³⁺) module [GREGOR, MALIK 2012]. The BFI³⁺ module shows the calibration processes for each year. The trial is started for the year 1996 by entering a specific value of parameters (km , kb , ki , α , α_q , β , μ , Cb , Ci) and then running the filter.

The trial has continued by entering individual values of the parameters, and then the graphic is visualised. Figure 6 visualises a zoom of the calibration process for 10.04.2001 to 10.10.2001 using the Lyne–Hollick filter for the 1 – Lamong watershed with parameter $\alpha = 0.975$. The trial is stopped when the red-line curve (baseflow) fits the blue-area curve (discharge) for the dry-periods.

The best-fitting values appear accurately for the dry period of July, August and September. This is because at this period the observed baseflow is maximal, and runoff (quickflow or direct runoff) tends towards 0. When this condition is achieved, the trial for the one-year data is stopped, and the parameter value obtained for the year is noted manually.

The trials are repeated for the next year until the two curves again fit. The best value of the parameter (the optimal parameter) is obtained by yearly averaging of values. The baseflow level varies

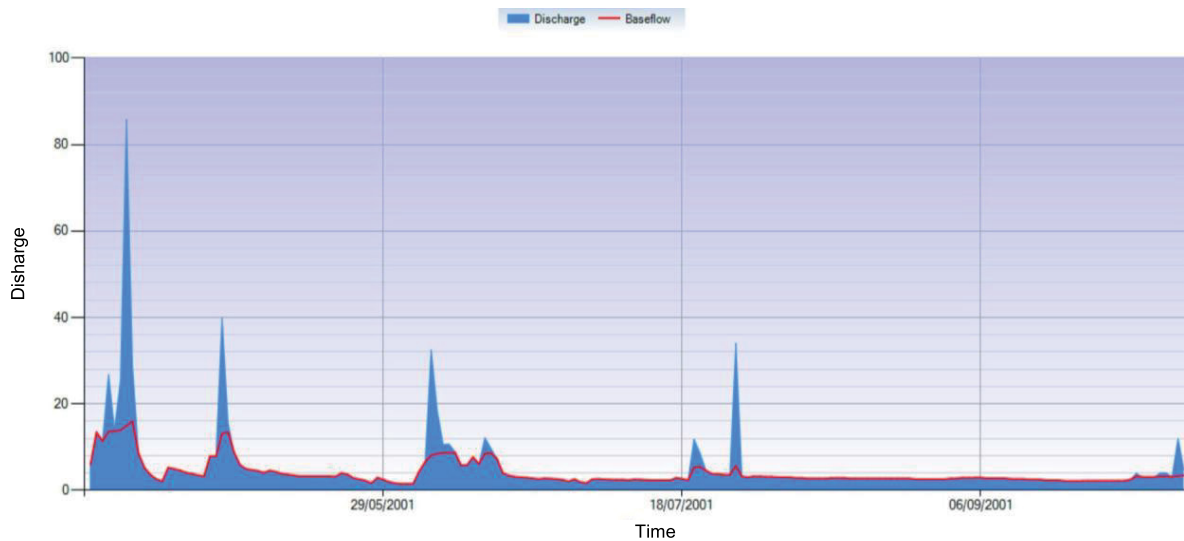


Fig. 6. Zoom of a calibration process; source: own elaboration

from year to year, and the optimal value is fixed as the average value. It is presumed that using the average value for all the periods of calibration will reduce calculation error. Furthermore, after the first algorithm is calibrated, the trials are re-started for the other algorithms until all six are calibrated. The results of the calibration process are presented in Tables 3 and 4.

VALIDATION PROCEDURE

The optimal parameter obtained from calibration periods is then used to test the algorithms for the validation period of 1st January 2006 to 31st December 2015 (Tab. 5).

STATISTICAL EVALUATION

The performance of the six algorithms used in this research are judged statistically using root mean square error (*RMSE*) (Tab. 5) and coefficient determined (R^2) (Tab. 6) by scatter plot methods. statistical analyses of calibration and validation results are conducted by comparing the calculated or simulated baseflow drawn from the six algorithms and the measured total flow for the three driest months (July, August, September). In this case, *RMSE* (Eq. 8) is used to evaluate the goodness of fit between measured and calculated baseflow.

$$RMSE = \sqrt{\frac{\sum (Q_c - Q_o)^2}{n}} \quad (8)$$

where: Q_c is the calculated baseflow ($m^3 \cdot s^{-1}$), Q_o is the observed total flow ($m^3 \cdot s^{-1}$), and n is the number of samples.

The n value is the number of samples. It is the sum of daily baseflow events calculated from the 1st July to 30th September data for 1996 to 2005 (3 months \times 30 days \times 9 years) = 810 events for the calibration period, and from 2006 to 2015 (3 months \times 30 days \times 9 years) = 810 events for the validation period. Zero values of *RMSE* ($RMSE = 0$) show the goodness of fit between measured and calculated baseflows. Evaluation is also performed using a scatter plot to obtain the determined coefficient (R^2) between measured and calculated baseflow. If the value of R^2 is close to 1, there is a strong correlation between calculated baseflow and observed streamflow during the dry periods.

VISUALISATION AND INTERPRETATION

The calculated baseflows from the six filters are compared and visualised using *FDCs* and hydrographs. The *FDC* is used to compare streamflow (observed or total flow) and the calculated baseflow provided by the six algorithms. Usually, in *FDCs*, the baseflow and the total flow (streamflow) for the dry season will tend to fit each other. In contrast, for the wet season, the baseflow will be separated from the total flow and appears below the observed streamflow curve.

The result of this separation is also visualised using a hydrograph to compare calculated baseflow with the total flow. The separation result is also displayed by the baseflow index (*BFI*). *BFI* is defined as baseflow per total flow. In this study, the algorithms calculate daily *BFI* from 1996 to 2015, displayed in tabular form (Tab. 8). Finally, the interpretation of baseflow at the watersheds is summarised in the last paragraph of this study.

RESULTS AND DISCUSSION

CALIBRATION AND VALIDATION

PARAMETER VALUES

Parameter values are entered manually during the trial of calibration processes, and the results are presented in Table 3. The table shows, for example, that the accepted range value for km is between 0.980 and 0.997 for the three watersheds. The value of α for the Lyne–Hollick filter ranges from 0.960 to 0.998. While for the *EWMA* filter, the value of parameter μ ranges between 0.005 and 0.020.

The range in values is narrower for parameters km , kb , and ki for the three algorithms (i.e., one parameter, two parameters, and IHACRES, respectively). Although the number of parameters for the six algorithms is different (from one to three parameters), the algorithms tested in this study show a similar range in parameter values. The six RDF methods are tested in the same climatic regime, and therefore a similar range of parameter values may be acceptable.

OPTIMAL VALUES

Table 4 summarises the optimal value of each parameter. The optimal value for ki is strictly between 0.948 and 0.985. The same result for parameter α_q ranges from 0.921 to 0.930. The role of parameter μ in the *EWMA* filter is similar to the role of Cb in two-parameter methods and Ci in IHACRES methods.

It is noted that each parameter contributes to specific tasks. Parameter α in Lyne–Hollick and β in Chapman algorithms have the same role as km in one-parameter algorithms, kb in two-parameter algorithms and ki in IHACRES algorithms.

STATISTICAL ANALYSIS

The results of statistical analysis using *RMSE* are presented in Table 5 for the calibration and validation period.

Table 5 shows that all values of *RMSE* tend to be close to zero. The calculated baseflows for each of the six methods are close to the measured flow for dry periods (July, August and September). The *RMSE* results for four of the methods (one parameter, two parameters, IHACRES, and Chapman) at 2 – Surabaya watershed are less accurate due to the anomaly in measured streamflow data for specific periods in this watershed.

The *RMSE* values obtained from the validation periods (Tab. 5) are shown to be comparable to the calibration period. This means that all the filters calculate the baseflow consistently for the three driest months. The value of parameters set up at the calibration period can be transferred to the validation period. The six methods perform well both for calibration and validation periods. The quality of recorded streamflow data at the 2 – Surabaya watershed influences the accuracy of *RMSE* calculation both for calibration and validation periods. The best performing filters in this experiment are the Lyne–Hollick and *EWMA* filters. These two filters produced better *RMSE* values than the others both for calibration and validation.

Table 6 show the coefficient of determination (R^2) obtained from scatter plots of calculated baseflow from the six methods, and the measured streamflow observed for dry periods.

Table 6 shows that the value of R^2 for the calibration period is ≥ 0.70 for all filters at all watersheds. It justifies the strong correlation between calculated baseflow and observed streamflow

Table 3. The range of parameter values entered during calibration

Algorithm/filter	Parameter	Values for			Range of values
		1 – Lamong	2 – Surabaya	3 – Bangsal	
One parameter	<i>km</i>	0.988–0.997	0.980–0.995	0.980–0.993	0.980–0.997
Two parameters	<i>kb</i>	0.980–0.995	0.975–0.995	0.990–0.995	0.975–0.995
	<i>Cb</i>	0.008–0.015	0.017–0.030	0.012–0.018	0.008–0.030
IHACRES	<i>ki</i>	0.935–0.960	0.960–0.970	0.980–0.988	0.935–0.988
	<i>Ci</i>	0.007–0.015	0.015–0.025	0.009–0.012	0.007–0.025
	α_q	0.915–0.930	0.920–0.940	0.920–0.935	0.915–0.940
Lyne–Hollick	α	0.995–0.998	0.985–0.997	0.960–0.990	0.960–0.998
Chapman	β	0.998–0.999	0.965–0.999	0.985–0.998	0.965–0.999
EWMA	μ	0.007–0.013	0.005–0.011	0.012–0.020	0.005–0.020

Explanations: IHACRES = identification of unit hydrographs and component flows from rainfall, evaporation and streamflow data, parameters used in Tab. are explained in p. 133–134.
Source: own study.

Table 4. Optimal (average) of parameter values

Method	Parameter	Values for		
		1 – Lamong	2 – Surabaya	3 – Bangsal
One parameter	<i>km</i>	0.988	0.995	0.988
Two parameters	<i>kb</i>	0.984	0.966	0.992
	<i>Cb</i>	0.024	0.012	0.015
IHACRES	<i>ki</i>	0.966	0.948	0.985
	<i>Ci</i>	0.019	0.010	0.010
	α_q	0.930	0.921	0.928
Lyne–Hollick	α	0.992	0.997	0.979
Chapman	β	0.991	0.999	0.991
EWMA	μ	0.008	0.009	0.016

Explanations: IHACRES = identification of unit hydrographs and component flows from rainfall, evaporation and streamflow data, EWMA = exponentially weighted moving average, other parameters used in Tab. are explained in p. 133–134.
Source: own study.

Table 5. Root mean square error (RMSE) for calibration and validation periods

Algorithm	Values for		
	1 –Lamong	2 – Surabaya	3 – Bangsal
Calibration periode (1996–2005)			
One parameter	0.022	0.272	0.024
Two parameters	0.034	0.367	0.018
IHACRES	0.033	0.188	0.014
Lyne–Hollick	0.013	0.070	0.008
Chapman	0.023	0.254	0.024
EWMA	0.013	0.070	0.008
Validation periode (2006–2015)			
One parameter	0.032	0.407	0.031
Two parameters	0.052	0.574	0.022
IHACRES	0.051	0.295	0.016
Lyne–Hollick	0.022	0.125	0.010
Chapman	0.033	0.379	0.031
EWMA	0.022	0.124	0.009

Explanation: IHACRES, EWMA as in Tab. 4.
Source: own study.

during the driest periods and also proves that all filters are adaptable to calculate baseflow in the region.

Table 6 also show that for all values, R^2 is better for calibration than the validation. This is normal because the value of the parameter is set up during the calibration period (Tab. 6). Moreover, table 6 also shows that the EWMA and Lyne–Hollick filters provide the best R^2 values.

Considering the result of statistical evaluations using RMSE and R^2 , it can be stated that, in principle, all methods are adaptable to the calculation of baseflow for the region. However, the best performance is obtained by the EWMA and Lyne–Hollick filters. It can also, therefore, be stated that algorithms with only one parameter (EWMA and Lyne–Hollick) may perform better than other algorithms using two or three parameters.

VISUALISATION

FLOW DURATION CURVE (FDC)

Figure 7 shows the long-term FDC for the 1 – Lamong watersheds. Long-term FDCs are graphics constructed from all data series that are available from 1996 to 2015. Moreover, all of the calculated baseflows obtained from the six filters are plotted in one chart (Fig. 7) that visualises the FDC for the 1 – Lamong watershed.

Table 6. Coefficient of determination (R^2) for the calibration and validation periods

Algorithm	Values for		
	1 - Lamong	2 - Surabaya	3 - Bangsal
Calibration periode (1996–2005)			
One parameter	0.863	0.777	0.734
Two parameters	0.704	0.905	0.758
IHACRES	0.721	0.895	0.796
Lyne-Hollick	0.951	0.968	0.882
Chapman	0.848	0.726	0.712
EWMA	0.957	0.968	0.891
Calibration periode (2006–2015)			
One parameter	0.702	0.777	0.868
Two parameters	0.516	0.683	0.885
IHACRES	0.551	0.755	0.897
Lyne-Hollick	0.872	0.918	0.943
Chapman	0.692	0.561	0.867
EWMA	0.868	0.918	0.947

Explanation: IHACRES, EWMA as in Tab. 4.
Source: own study.

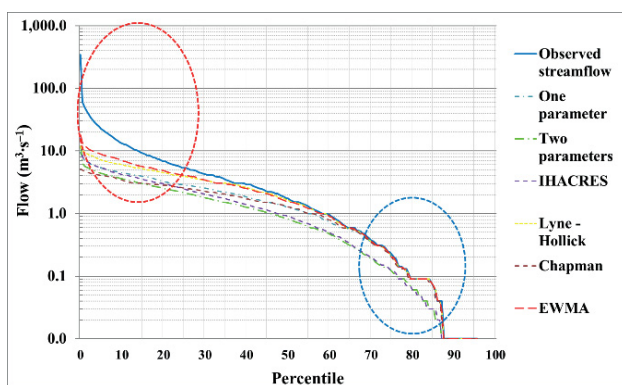


Fig. 7. Long-term flow duration curve (*FDC*) for 1 - Lamong; IHACRES, EWMA as in Tab. 4; source: own study

The blue line plots the *FDC* of observed streamflow (total flow in the river) (Fig. 7) also compares the different *FDC*s constructed from baseflow which is calculated using the six filters (local minimum, fixed interval, sliding interval, one parameter, two parameters, IHACRES, Lyne-Hollick, Chapman and EWMA). It can be seen that in the top left of Figure 7 the observed flow is higher than the other calculated baseflows, and the separation results of baseflow from the streamflow (total flow) are visualised more clearly in this segment.

In contrast, at the bottom left of Figure 7 (area in the blue dashed-line circle), the calculated baseflow curves are close to the observed streamflow. This segment plots flow during the dry season. In this period, the estimated baseflows are relatively similar to observed streamflow because the baseflow component mainly supplies the flow in the river (streamflow). The quickflow or direct-runoff component tends towards zero. However, the conditions of each watershed are specific. Physical properties of the watersheds (land use, soil type and geology) and rainfall may influence the rate of baseflow contribution.

Figure 7 indicates that visualisation in the form of *FDC* guided the separation results. At times of high flow (during rainy seasons) (percentiles 0–30%) the observed streamflows are elevated compared to all baseflows calculated from the six algorithms. Meanwhile, at the peak of the dry seasons (percentiles 80–100%) the graphics are closer to each other.

HYDROGRAPHS

Figure 8 shows the portion of the hydrographs that plots the calculated baseflows compared to observed streamflow during the driest periods (from 1st July to 30th September 2005). During the sample of dry periods (Fig. 8), it is shown that EWMA and Lyne-Hollick filters calculated baseflow closer to the observed streamflow (total river flow).

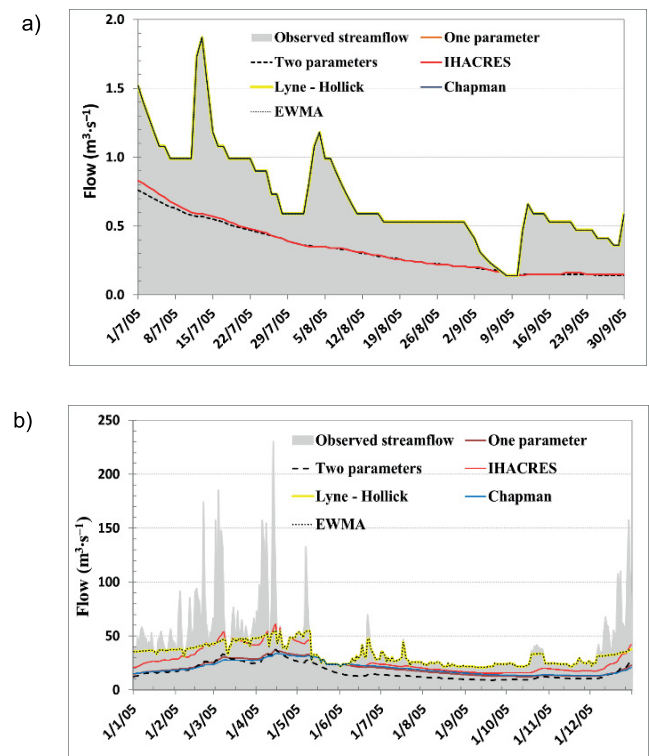


Fig. 8. Hydrograph of two investigated watersheds for the year 2005: a) 1 - Lamong, b) 2 - Surabaya; IHACRES, EWMA as in Tab. 4; source: own study.

The other filters (one parameter, two parameters, IHACRES and Chapman) tend to estimate baseflow below the observed streamflow. Therefore, using assumptions as stated in the introduction to this paper, we may state that the Lyne-Hollick and EWMA filters give the best performance in simulating baseflow in this region.

BASEFLOW STATISTICS

Tables 7 displays the statistical summaries of the calculated daily baseflow from the six filters. Table 7 shows that the 2 - Surabaya watershed has more significant mean daily baseflow (MDBF) than the 1 - Lamong watershed. Although the 1 - Lamong catchment area is the largest, its MDBF is small compared to the second watershed, 2 - Surabaya.

Water resources are minimal at 1 - Lamong during prolonged dry periods. This may be caused by over-exploitation

Table 7. Summary statistics of baseflow of two investigated watersheds

Filter	1996–2005			2006–2015		
	max.	min.	mean	max.	min.	mean
1 – Lamong						
Streamflow	104.2	0.01	4.97	346.0	0.01	6.94
One parameter	7.21	0.00	1.49	9.53	0.01	2.33
Two parameters	8.45	0.01	1.16	11.91	0.00	1.71
IHACRES	11.11	0.00	1.36	16.17	0.00	2.00
Lyne–Hollick	7.11	0.00	1.99	11.74	0.01	3.08
Chapman	3.40	0.00	1.24	5.18	0.01	1.93
EWMA	13.52	0.00	2.22	17.93	0.01	3.39
Mean of filters	8.47	0.00	1.58	12.08	0.01	2.41
Standard deviation	3.20	0.00	0.39	4.19	0.00	0.62
2 – Surabaya						
Streamflow	236.0	7.9	44.5	358.39	10.70	51.20
One parameter	48.0	0.6	21.7	61.17	1.61	25.24
Two parameters	53.1	1.2	18.2	61.29	3.18	21.08
IHACRES	86.6	0.9	29.5	99.76	2.53	33.82
Lyne–Hollick	69.9	0.2	33.6	92.98	0.54	38.15
Chapman	42.9	0.2	21.6	57.83	0.61	25.12
EWMA	69.9	0.4	33.7	93.14	1.09	38.20
Mean of filters	61.75	0.58	26.39	77.70	1.59	30.27
Standard deviation	16.55	0.39	6.74	19.47	1.07	7.40

Explanations: IHACRES, EWMA as in Tab. 4.
Source: own study.

of water for irrigation purposes in these two watersheds. Furthermore, the EWMA and Lyne–Hollick filters tend to estimate MDBF as higher than the other filters.

BASEFLOW INDEX (BFI)

Table 8 displays the statistical summaries of maximum and mean daily baseflow index (MDBFI) calculated by the six filters and indicate that all filters consistently calculated BFI. This means that both for calibration and validation periods, the filters produce nearly similar values of maximum and mean BFI.

The Table 8 also shows that the Lyne–Hollick and EWMA filters always estimate the BFI as higher than the other four filters.

MONTHLY BASEFLOW INDEX (MBFI)

Figure 9 shows the variation of MBFI, calculated as an average of daily BFI values for one month. The graphic visualise MBFI for the period 1st Jan. 2005 to 1st Dec. 2008. The MBFI from all six filters is high during the dry season and low during the wet season. At the peak of the dry season, MBFI reaches 1, meaning that surface runoff is calculated at zero.

Table 9 shows the summary of MBFI statistical values, in which the maximum monthly value reaches 1. In contrast, the average value of MBFI ranges from 0.47 to 0.90. The Lyne–Hollick and EWMA filters always estimate the MBFI value higher than the other filters.

Table 8. Summary baseflow index (BFI) for two investigated watersheds

Filter	1996–2005		2006–2015	
	max.	mean	max.	mean
1 – Lamong				
One parameter	1.00	0.76	1.00	0.78
Two parameters	1.00	0.55	1.00	0.58
IHACRES	1.00	0.57	1.00	0.59
Lyne–Hollick	1.00	0.84	1.00	0.87
Chapman	1.00	0.74	1.00	0.75
EWMA	1.00	0.86	1.00	0.88
2 – Surabaya				
One parameter	1.00	0.58	1.00	0.60
Two parameters	1.00	0.46	1.00	0.47
IHACRES	1.00	0.74	1.00	0.76
Lyne–Hollick	1.00	0.88	1.00	0.88
Chapman	1.00	0.59	1.00	0.61
EWMA	1.00	0.88	1.00	0.88

Explanations: IHACRES, EWMA as in Tab. 4.
Source: own study.

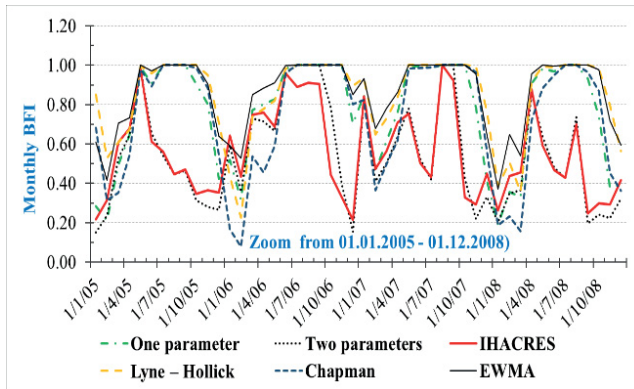


Fig. 9. Monthly baseflow index (MBFI) values for 1 – Lamong; IHACRES, EWMA as in Tab. 4; source: own study source: own study

Table 9. Monthly baseflow index (MBFI) for 1 – Lamong

Recursive digital filter	Max.	Mean	Standard deviation	Variants
One parameter	1.00	0.77	0.24	0.06
Two parameters	1.00	0.56	0.24	0.06
IHACRES	1.00	0.58	0.21	0.04
Lyne-Hollick	1.00	0.86	0.18	0.03
Chapman	1.00	0.74	0.27	0.07
EWMA	1.00	0.87	0.16	0.03

Explanations: IHACRES, EWMA as in Tab. 4. Source: own study.

Generally, we can conclude that in these three watersheds, the contribution of baseflow to streamflow is significant. The watersheds are classified as perennial rivers with substantial baseflow contribution. However, exploitation of flow during dry seasons will decrease the baseflow, and extraction of flow for irrigation purposes during dry periods will significantly reduce the baseflow of the rivers.

CONCLUSIONS

This paper shows the calibration and validation of six baseflow separation methods based on the recursive digital filters (RDFs) for three small watersheds. The three watersheds are characterised by a specific tropical climatic regime marked by a significant difference between wet and dry seasons. The calibration period is determined by assuming that in dry periods, the contribution of baseflow to streamflow is more dominant. The results show that in principle, all of the six filters tested are adaptable to predicting baseflow in the region. However, based on statistical evaluation using root mean square error (RMSE) and determination coefficient (R^2) the Lyne-Hollick and exponentially weighted moving average (EWMA) filters perform better than the others. Evaluation of baseflow index (BFI) shows that the three watersheds are subject to climatic conditions that produce substantial baseflow contribution. The paper also identifies the importance of baseflow separation in determining the portion of the flow that potentially produces runoff.

FUNDING

The publication of this article is supported by a Reworking Thesis – Grant from the Research Institute (LP2M), University of Jember, 2019.

REFERENCES

BOUGHTON W.C. 1993. A hydrograph-based model for estimating water yield of ungauged catchments. Institute of Engineers Australia National Conference. Vol. 93. Iss. 14 p. 317–324.

BRODIE R.S., HOSTETLER S. 2005. A review of techniques for analysing baseflow from stream hydrographs. Components. Vol. 28 p. 1–13.

CHAPMAN T.G. 1991. Comment on the evaluation of automated techniques for base flow and recession analyses by R.J. Nathan and T.A. McMahon. Water Resources Research. Vol. 27. Iss. 7 p. 1783–1784.

CHAPMAN T.G., MAXWELL A.I. 1996. Baseflow separation – comparison of numerical methods with tracer experiments. Institute of Engineers Australia National Conference. Vol. 96. Iss. 5 p. 593–545.

ECKHARDT K. 2005. How to construct recursive digital filters for baseflow separation. Hydrological Processes. Vol. 19. Iss. 2 p. 507–515. DOI 10.1002/hyp.5675.

FUREY P.R., GUPTA V.K. 2003. Tests of two physically based filters for base flow separation, Water Resources Research. Vol. 39(10), 1297. DOI 10.1029/2002WR001621.

GONZALES A.L., NONNER J., HEIJKERS J., UHLENBROOK S. 2009. Comparison of different base flow separation methods in a lowland catchment. Hydrology and Earth System Sciences. Vol. 13. Iss. 11 p. 2055–2068. DOI 10.5194/hess-13-2055-2009.

GREGOR M., MALIK P. 2012. RC 4.0 user’s manual [online]. Hydro Office Software Water Science. [Access 15.04.2018]. Available at: <https://hydrooffice.org/Files/UM%20RC.pdf>

INDARTO I. 2013. Studi Tentang Karakteristik Fisik dan Hidrologi pada 15 DAS di Jawa Timur [Study on the physical characteristics and hydrology of 15 watershed in East Java]. Forum Geografi. Vol. 27. Iss. 2 p. 163–182. DOI 10.23917/forgeo.v27i2.2374.

JAKEMAN A.J., HORNBERGER G.M. 1993. How much complexity is warranted in a rainfall-runoff model?. Water Resources Research. Vol. 29. Iss. 8 p. 2637–2649. DOI 10.1029/93WR00877.

KISSEL M., SCHMALZ B. 2020. Comparison of baseflow separation methods in the German Low Mountain range. Water. Vol. 12. Iss. 6, 1740. DOI 10.3390/w12061740.

LYNE V., HOLLICK M. 1979. Stochastic time-variable rainfall-runoff modelling. Institute of Engineers Australia National Conference. Vol. 79(10) p. 89–93.

MARSH N. 2003. River analysis package – Users guide [online]. CRC for Catchment Hydrology-Australia. [Access 16.04.2018]. Available at: <https://toolkit.ewater.org.au/Tools/RAP/documentation>

MURPHY R., GRASZKIEWICZ Z., HILL P., NEAL B., NATHAN R., LADSON T. 2009. Australian rainfall and runoff revision. Revision Projects. Project 7: Baseflow for catchment simulation. Stage 1 report. Vol. 1. Selection of approach. Engineers Australia Water Engineering. ISBN 978-085825-9218 pp. 111.

NATHAN R.J., MCMAHON T.A. 1990. Evaluation of automated techniques for base flow and recession analyses. Water Resources Research. Vol. 26. Iss. 7 p. 1465–1473. DOI 10.1029/WR026i007p01465.

PARTINGTON D., BRUNNER P., SIMMON C.T., WERNER A.D., THERRIER R., MAIER H.R., DANDY G.C. 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. Journal of Hydrology. Vol. 458–459 p. 28–39. DOI 10.1016/j.jhydrol.2012.06.029.

- SHAO G., ZHANG D., GUAN Y., ANWAR S.M., HUANG F. 2020. Application of different separation methods to investigate the baseflow characteristics of a semi-arid sandy area, Northwestern China. *Water*. Vol. 12. Iss. 2, 434. DOI 10.3390/w12020434.
- SMAKHTIN V.U. 2001a. Estimating continuous monthly baseflow time series and their possible applications in the context of the ecological reserve. *Water SA*. Vol. 27 Iss. 2 p. 213–217. DOI 10.4314/wsa.v27i2.4995.
- SMAKHTIN V.U. 2001b. Low flow hydrology: A review. *Journal of Hydrology*. Vol. 240 p. 147–186. DOI 10.1016/S0022-1694(00)00340-1.
- TALLAKSEN L. 1995. A review of baseflow recession analysis. *Journal of Hydrology*. Vol. 165. Iss. 1–4 p. 349–370. DOI 10.1016/0022-1694(94)02540-R.
- TAN S.B., LO E.Y., SHUY E.B., CHUA L.H., LIM W.H. 2009. Hydrograph separation and development of empirical relationships using single-parameter digital filters. *Journal of Hydrologic Engineering*. Vol. 14. Iss. 3 p. 271–279. DOI 10.1061/(ASCE)1084-0699(2009)14:3(271).
- TULARAM G.A., ILAHEE M. 2008. Exponential smoothing method of base flow separation and its impact on continuous loss estimates. *American Journal of Environmental Sciences*. Vol. 4. Iss. 2 p. 136–138. DOI 10.3844/ajessp.2008.136.144.
- XIE J., LIU X., WANG K., YANG T., LIANG K., LIU C. 2020. Evaluation of typical methods for baseflow separation in The Contiguous United States. *Journal of Hydrology*. Vol. 583, 124628. DOI 10.1016/j.jhydrol.2020.124628.