Sustainable Water Resources Management (2022) 8:97
https://doi.org/10.1007/s40899-022-00694-1

ORIGINAL ARTICLE

Development of hybrid baseflow prediction model by integrating analytical method with deep learning

Wondmagegn Taye Abebe[1] · **Demeke Endalie**[2] · **Getamesay Haile**[2]

Received: 24 December 2021 / Accepted: 4 June 2022

Abstract
In recent years, the success of deep learning in many different fields of Engineering has attracted attention. Baseflow separation is one of the Engineering problems which remains difficult due to different hydro-climatic circumstances. In this study, we proposed a hybrid baseflow prediction model by combining analytical methods and deep learning algorithms. Six analytical methods were chosen and their performance was compared by different metrics. Baseflow-Lyne and Hollick algorithm (BFLOW-LHA) outperforms the others in terms of R^2, Mean Absolute Error (MAE), BIAS, Nash–Sutcliffe Efficiency (NSE), and Root Mean Squared Error (RMSE) metrics. The proposed model was trained using streamflow and baseflow data generated by the BFLOW-LHA with the Dawa Melka Guba dataset and then tested on prediction for the basin's remaining three watersheds. The experimental results show that the proposed model improves the prediction of baseflow as compared with BFLOW-LHA and can be used for watersheds with similar characteristics.

Keywords Baseflow · Deep learning · Graphical methods · Recursive digital filter · Streamflow

Introduction

Following a rainfall event, streamflow is typically assumed to be made up of three components: rapid flow, interflow, and baseflow (Tallaksen 1995). Baseflow has been defined as the component of streamflow that is always the slowest to respond and lasts the longest groundwater aquifer flow, flow from groundwater storage or other delayed sources, as well as groundwater contribution to streamflow (Tallaksen 1995; Aboelnour et al. 2020; Murphy et al. 2011; Lott and Stewart 2016). Baseflow sustains streamflow during low rainfall periods (Pielke et al. 2007). The analysis of flow hydrographs, comprising base flow, flood peaks and volume, and groundwater contributions in the base flow, is one of the most important aspects of engineering hydrology (Sivapalan 2018).

Baseflow separation is important in hydrological research such as water resource management, basin hydrology studies, rainfall-runoff modeling, water quality modeling and flow condition, comparing alternative land use management effects on groundwater, calibration and validation of hydrological models, and determining an index to compute rainfall excess and rainfall losses in a watershed (Spongberg 2000; Stewart et al. 2007). Baseflow behavior offers information on groundwater quality, seasonal low flows, and instream ecology (Duncan August 2019). As a result, a baseflow estimation can be used for irrigation planning, agriculture, drought management, groundwater recharging, and reducing water losses (Sivapalan 2018).

One of the major uses of baseflow separation, particularly in numerical modeling and analysis of the water balance at the basin-wide, is evaluating the contribution of groundwater to river flow which can be expressed by Base Flow Index (BFI) (Sivapalan 2018). The BFI (a dimensionless variable that always ranges from 0 to 1) is the ratio of cumulative baseflow and cumulative total discharge throughout the record of analysis. It is a hydrogeological parameter that can be used to define un-gauged basins, and it is supposed to illustrate geology's impact on basin low flows (Young et al. 2000; Roy and Mistri 2013). BFI has a variety of uses, including rainfall-runoff modeling, and can be used to compare the flow characteristics of different catchments and as an indicator of watershed features (Sivapalan 2018).

✉ Wondmagegn Taye Abebe
wondmagegn.abebe@ju.edu.et

[1] Faculty of Civil and Environmental Engineering, Jimma University Institute of Technology, Jimma, Ethiopia

[2] Faculty of Computing and Informatics, Jimma University Institute of Technology, Jimma, Ethiopia

Published online: 28 June 2022

It has been challenging to define baseflow accurately or to consistently distinguish it from other streamflow components (Nathan and McMahon 1990). During a rainfall event, when groundwater is being replenished and the total hydrograph is dominated by the quicker components of runoff, baseflow is more difficult to describe or measure (Duncan 2019). The baseflow component of streamflow is generally determined by separating a stream hydrograph into two components: baseflow and runoff (Lott and Stewart 2016). The very smooth decline observed over extended periods of little or no rain is the most recognized phase of baseflow and the most accessible to quantitative examination (Nathan and McMahon 1990).

Baseflow separation in an acceptable manner has long been a desirable but elusive aim (Duncan 2019). Because direct measurement of baseflow is difficult, comparing the relative accuracy of baseflow methods is difficult (Lott and Stewart 2016). The use of chemical and radioactive tracers is the most accurate means of estimating the baseflow contribution in the watershed discharge due to the complexity of the base flow characteristics (Chapman 1999). However, more qualitative procedures are frequently used because chemical and radioactive tracers are expensive and require time and complex understanding, and also detailed tracer data is rarely available (Galli et al. 2021). Many approaches for separating the base flow component from total streamflow have been presented.

The most commonly used baseflow separation methods are analytical methods and mathematical functions or algorithms that determine baseflow directly from discharge (Eckhardt 2005; Huyck et al. 2005). In analytical filtering or smoothing methods, baseflow hydrograph is assumed to be a lower amplitude, a lower frequency component of the entire streamflow hydrograph. Their benefits include ease of automation, repetition of outcomes, and the ability to use the whole discharge record (Huyck et al. 2005). But they are frequently used without calibration to the basin or gage-specific characteristics other than basin area (Lott and Stewart 2016).

In Kouanda et al. (2018) four Recursive Digital Filter (RDF) methods (Chapman, Chapman and Maxwell, Lyne and Hollick and Eckhardt) have been examined on a watershed in the Sudano-Sahelian zone, West Africa, using the daily streamflow of the Mouhoun River. The RDF and Chloride Mass Balance (CMB) approaches were compared using statistical analysis. It was found that among the RDF methods, the Eckhardt method was successfully calibrated using the CMB method and the parameter BFImax of the Eckhardt method was adjusted to 0.32 in the study area context. The calibrated results demonstrate that the Eckhart approach has improved significantly.

The authors of Liu et al. (2019) applied a digital filter method to separate baseflow from local daily streamflow records for 1983–2014 using different values of filtering parameter (β) and filtering times (T) across the small watershed of Pengchongjian in Jiangxi Province, southern China. They validated the separation results by the baseflow index (BFI) method and obtain an optimal value of $\beta = 0.90$ and $T = 2$. According to their study, the results of the baseflow separation study matched with the real field situation in the watershed.

The authors of Stadnyk et al. (2015) compare two RDF methods and one hydrologic model in two baseflow dominant sub-basins of the Grand River Basin in southern Ontario, Canada and they verified the methods by applying Stable water isotopes (SWIs). They obtain an average baseflow contribution of 47% for Eramosa River and 74% for Whiteman's Creek by using the WATFLOOD-based model and indicated that the variation is a result of physiographic differences. The study's findings support the use of SWIs for regional hydrograph separation, as well as indicating that the RDF approach agrees with similar low-flow observed data.

In Mazvimavi et al. (2004), the authors applied linear regression and artificial neural networks to predict the baseflow index (BFI) from basin characteristics for 52 basins in Zimbabwe. The study found that both methods were suitable for the prediction of BFI value and they used this value to derive flow duration curves with R^2 being 0.89–0.99.

Although there are several baseflow separation methods in the literature, finding an appropriate method for a given watershed and hydro-climatic circumstances remains a difficult issue (Kouanda et al. 2018). The goal of this research is to find the best fit baseflow separation method for the Genale Dawa river basin and to combine it with deep learning to improve its accuracy and for future use. The Genale Dawa river basin is bounded by latitudes of 3°40′N and 7°43′N, and longitudes of 37°04′E and 43°28′E, and is located in Ethiopia's southern area, bordering Kenya and Somalia. It is the country's third-largest river basin, after the Abay and Wabi Shebele river basins, with an estimated area of 17,6705 km^2. The Genale Dawa River Basin is one of Ethiopia's most drought-prone areas.

In this study, we used a deep learning model to improve baseflow prediction using the calibration techniques described by Stewart et al. (2007). This is because deep learning improves the regression of the curve from the hydrograph by retaining both short and long-term memories (Hochreiter and Schmidhuber 1997).

While reviewing the preceding studies, it became clear that parameters adjusted for HYSEP and RDF that worked on one station would not work on another. In this study, however, we develop a baseflow prediction model that combines analytical methods with a deep learning model. We examine three HYSEP and three RDF methods of baseflow separation and identify the best method for the Genale Dawa river basin in Ethiopia, and then we integrate the selected method with deep learning algorithms. The study's contribution can be summarized as follows:

1. Calibration and evaluation of six baseflow separation methods based on the basin or gage-specific characteristics.
2. Integration of the selected method with a Deep learning model to use for watersheds within the study area.
3. An extensive experiment is used to present a detailed analysis of the integrated model.

Materials and methods

The proposed baseflow prediction model architecture used in this study is divided into two phases: training and testing. During the training phase, streamflow datasets are obtained from the Ministry of Water, Irrigation, and Energy (MoWIE) and preprocessed. Following that, we generate baseflow for the study area using HYSEP and RDF algorithms and calibrate the results (Stadnyk et al. 2015). The best baseflow separation method in terms of R^2, MAE, BIAS, NSE, and RMSE metrics is then chosen from the analytical methods and the best method is used to separate the baseflow.

A deep learning model is created to predict baseflow from streamflow which, is capable of working on different gauging stations once it is well trained. The layers of the deep learning model and their functions are described below. The proposed deep learning model employs input, Long Short-Term Memory (LSTM), dens, dropout, and output layers. The input layer delivers the preprocessed streamflow data to the subsequent layers. LSTMs were developed specifically to address the problem of long-term dependency. They do not have to work hard to remember things for long periods of time; it comes naturally to them. Dense layers were then used to improve the interconnectedness of neurons specified in LSTM layers (Lu et al. 2020).

A dropout layer is added to the model to prevent overfitting during training. Dropout was applied to all layers of the network, with the probability of retaining the unit $p=(0.9, 0.75, 0.75, 0.5, 0.5, 0.5)$ for each layer (going from input to convolutional layers to fully connected layers) (Nandini et al. 2021). Finally, the output layer produces the respective streamflow's baseflow.

The trained deep learning model is tested for baseflow prediction in other watersheds. The proposed model's performance is then tested using streamflow data that were not included in the training phase. Figure 1 depicts the overall block diagram of the hybrid baseflow predictive model.

Dataset description

Taking into account the length of the record, continuity of data and concurrent period of observation, the dataset used in this study is the streamflow data of 22 years (1995–2016) from the Genale Dawa river basin, which is one of the major river basins in Ethiopia. Data of streamflow were collected from MoWIE of Ethiopia for four gauging stations shown in Fig. 2 and Table 1 below. The dataset contains two parameters: day and streamflow. We used Dawa at Melka Guba over a 22-years period to train the proposed deep learning model. We used the remaining watersheds to test the performance of the designed model. Before use, the streamflow data were checked for its consistency.

Baseflow separation is done using different methods by considering basin or gage-specific characteristics. Based on the quality of streamflow data, geographical location, area coverage, and physical characteristics of the study area, the Dawa Melka Guba watershed is used as a representative watershed. Due to the non-existence of direct measurement to confirm baseflow contributions to total streamflow on the study area, parameters of different baseflow separation methods are calibrated for the selected watershed by considering mostly dry periods (December–February) of Ethiopia

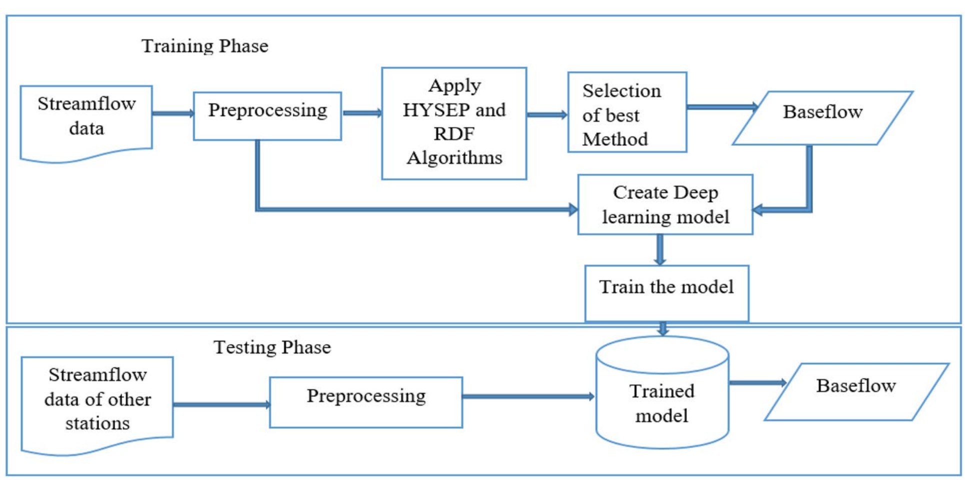

Fig. 1 Block diagram of the proposed baseflow separation model

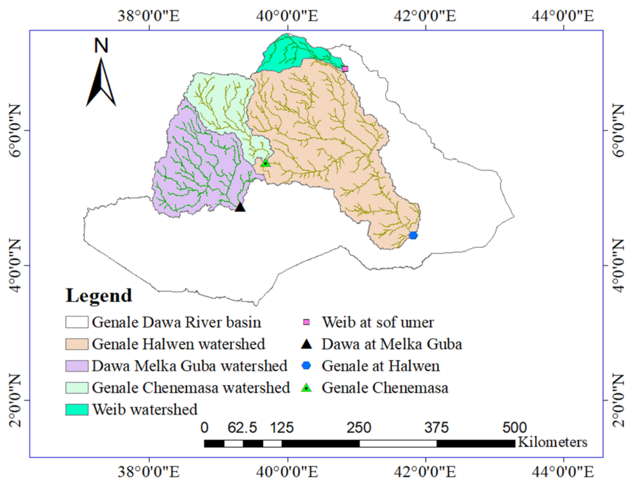

Fig. 2 Map of the study area with sub-watersheds and gauging stations

Table 1 Gauging stations used in this study

Station name	Latitude	Longitude	Drainage area (km^2)
Weib at Sof Umer	6.90	40.83	3792.7
Dawa at Melka Guba	4.87	39.32	19,611
Genale at Halwen	4.43	41.83	54,093
Genale Chenemasa	5.52	39.68	10,574

as well as the study area. Dry periods were selected based on the assumption that during this period there is no or less rainfall in the study area and resulting streamflow can be assumed as contributed from baseflow (i.e., during this period, baseflow is approximately equal to streamflow).

Baseflow separation methods

Three graphical methods (HYSEP) and three RDF methods (Chapman, Eckhardt Filter, and Lyne and Hollick) are selected and evaluated for this study. The three approaches in HYSEP (Sloto and Crouse 1996) are fixed interval, sliding interval, and local minimum, which reflect three alternative algorithms for drawing lines between hydrograph local minima. All three techniques are automated adaptations of prior graphical techniques. They provide a uniform response between users and are compatible with previous manual methods (Duncan 2019). The time interval in which the minima are chosen, N is given by;

$$N = 0.83A^{0.2} \quad (1)$$

where N denotes the number of days after the storm hydrograph's peak that surface runoff stops and A is drainage area in km^2.

The fixed interval technique assigns baseflow to the lowest discharge during each period separated by a time interval of $2N^*$ days. N^* is the odd integer that comes closest to the determined value of N. Baseflow is assigned to the lowest discharge within a time frame equal to one-half of the interval $2N^*$ minus one day using the sliding-interval approach. The local-minimum technique compares each day to the lowest discharge in half the period minus one day $[0.5(2N^* - 1)$ days] before and after the day in consideration (Gregor 2010).

The most frequent tools for processing and evaluating the hydrograph signal are RDFs, which can be used to distinguish high-frequency signals from low-frequency signals by setting a suitable threshold (Gregor 2010). When using an RDF to separate the baseflow from the quick flow, quick flow signals with a high frequency are removed from the hydrograph, while base flow signals with a low frequency are extracted (Galli et al. 2021) (Table 2).

BFLOW-LHA consists of forward and backward passes. Following the initial forward pass, the separated baseflow is subjected to a reverse pass to eliminate any phase distortion (i.e., to reduce lag), and the process is repeated for additional separation (Ladson et al. 2013). For daily data from all catchments, a fixed value of $k = 0.925$ has been proposed (Nathan and McMahon 1990; Murphy et al. 2009).

Table 2 RDFs used for this study

Filter name	Equation
Chapman algorithm (Chapman 1999)	$q_{f(i)} = \frac{3\alpha-1}{3-\alpha} q_{f(i-1)} + \frac{2}{3-\alpha}\left(q_{(i)} - q_{(i-1)}\right)$ (2) $q_{b(i)} = q_i - q_{f(i)}$ (3)
BFLOW-LHA (Nathan and McMahon July 1990; Lyne and Hollick 1979)	$q_{f(i)} = \alpha q_{f(i-1)} + \left(q_{(i)} - q_{(i-1)}\right)\frac{1+\alpha}{2}, q_{f(i)} \geq 0$ (4) $q_{b(i)} = q_i - q_{f(i)}$
Eckhardt algorithm (Eckhardt 2005)	$q_{b(i)} = \frac{(1-BFI_{max})\alpha q_{b(i-1)} + (1-\alpha)BFI_{max} q_{(i)}}{1-\alpha BFI_{max}}$ (5)

$q_{f(i)}$ is the filtered quick flow for the ith sampling instant, $q_{f(i-1)}$ is the filtered quick flow for the previous sampling instant to i, $q_{(i)}$ is the original streamflow for the ith sampling instant, $q_{(i-1)}$ is the original streamflow for the previous sampling instant to i, $q_{b(i)}$ is calculated baseflow at day i, α is filter parameter, BFI_{max} is the maximum baseflow index (constant)

The filter's output is restricted so that the separated baseflow is not negative or greater than the original streamflow (Lott and Stewart 2016). This approach is ideal for analyzing and comparing massive data sets from multiple places. By using a BFImax Genetic-Algorithm module built into the WHAT1 (an automated version of the Eckhardt filter method) system, the BFImax values and filter parameter of the Eckhardt filter equation can be optimized (Lim and Schoenung 2010). A detailed review of each RDF method can be found in different literatures (Rammal et al. 2018).

In addition to this, calibration guidelines presented by Duncan (2019) for master recession curves were adopted by modifying the baseflow graph. Because, the methods used in this study for baseflow separation like RDF, can serve the smoothing purpose. The guidelines are; (1) the separated baseflow should show a step up during significant rain, (2) the separated baseflow should not lie much below total flow in the absence of rain, and (3) the separated baseflow should not cling tightly to the total flow in the absence of rain and should not be greater than total flow.

Evaluation metrics

In this study, we evaluate the performance of the selected methods using R^2, MAE, BIAS, NSE, and RMSE metrics. The formulas for each metric are shown in Table 3 below.

To select the best baseflow separation method on the study area, we conduct an experiment by setting optimal parameters which minimize the error stated in Table 3 above. Parameter values for all the selected methods are calibrated for dry periods of the study area and by adopting the calibration guidelines presented by Duncan (2019). The optimal values obtained for the representative watershed (Dawa at Melka Guba) are shown in Table 4 below.

where, N is the number of days after the storm hydrograph's peak that surface runoff stops, f and α are filter parameters, and BFI_{max} is the maximum base flow index value.

Table 3 Evaluation metrics and their formulas

Metrics	Formula
R^2	$\frac{(\sum_{i=1}^{n}(b_i-b_m)(Q_i-Q_m))^2}{(\sum_{i=1}^{n}(b_i-b_m))^2\times(\sum_{i=1}^{n}(Q_i-Q_m))^2}$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert b_i - Q_i\rvert$
BIAS	$\frac{\sum_{i=1}^{n}(Q_i-b_i)}{\sum_{i=1}^{n}Q_i}$
NSE	$1-\left(\frac{\sum_{i=1}^{n}(Q_i-b_i)^2}{\sum_{i=1}^{n}(b_i-b_m)^2}\right)$
RMSE	$\left(\frac{1}{n}\sum_{i=1}^{n}(b_i - Q_i)^2\right)_{1/2}$

where, b_i is baseflow at time i, b_m is mean baseflow, Q_i is streamflow at time i, b_m is mean streamflow, and n is the number of observations.

Table 4 Parameter values used on each baseflow separation methods

Methods	Optimal parameters
Fixed interval	$N=6$
Sliding interval	$N=6$
Local minimum	$N=6$ and $f=0.75$
Chapman	Alpha, $\alpha=0.90$
BFLOW-LHA	Alpha, $\alpha=0.91$
Eckhardt	Alpha, $\alpha=0.99$ and $BFI_{max}=0.8$

Using the value of parameters shown in Table 4, we conduct a comparative analysis between the six baseflow separation methods for the dry period and the result is presented as shown in Table 5 below.

Table 5 above indicated that BFLOW-LHA of RDF outperforms the remaining methods in terms of R^2 (0.98), MAE (0.51), BIAS (0.05), NSE (0.97), and RMSE (2.61). The evaluation results indicated that BFLOW-LHA of RDF represents the physical characteristics of the watershed in a reasonable way. Average BFI of the studied methods: fixed interval (0.84), sliding interval (0.85), local minimum (0.78), Chapman algorithm (0.57), BFLOW-LHA (0.91), and Eckhardt algorithm (0.83). BFI value obtained from all methods is greater than 50% of the total streamflow, this indicates that the study area is baseflow dominant.

Figure 3 below depicts the separated baseflow from total streamflow by using all the studied methods for representative watershed and selected time frame. Whereas, Fig. 4 and Fig. 5 show the separated baseflow of the representative watershed by using BFLOW-LHA for the entire period and selected period, respectively.

After calibration and evaluation, the baseflow separation method which can best fit the characteristics of the study area is selected based on the value of statistical

Table 5 Comparative analysis of the six baseflow separation methods in terms of different evaluation metrics

Methods	Evaluation metrics				
	R^2	MAE	BIAS	NSE	RMSE
Graphical (HYSEP)	0.95	1.41	0.14	0.91	7.87
Fixed interval	0.97	1.21	0.12	0.94	5.46
Sliding interval	0.91	1.68	0.16	0.83	13.26
Local minimum	0.95	1.41	0.14	0.91	7.87
RDF					
Chapman	0.94	4.35	0.43	0.23	31.32
BFLOW-LHA	0.98	0.51	0.05	0.97	2.61
Eckhardt	0.96	0.86	0.08	0.94	5.60

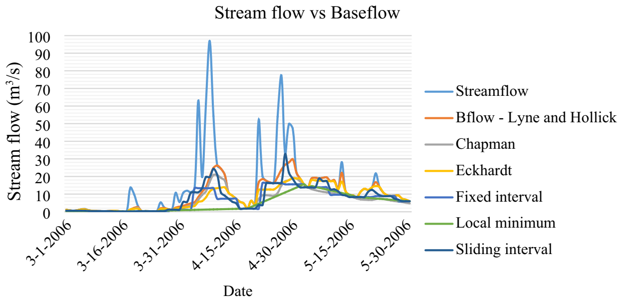

Fig. 3 A plot of streamflow vs baseflow using six baseflow separation methods

analysis. The selected method (BFLOW-LHA) is then integrated or coupled with a deep learning algorithm to get more accurate results and to adopt for the rest of watersheds within the study area.

Experiment

In this section, we investigate the performance of the integrated model which is calibrated to a basin or gage-specific characteristics. To evaluate the performance of the model, the baseflow of the remaining three watersheds (Weib, Genale Halwen, and Genale Chenemasa) is separated by using the integrated model. In addition, we compared the proposed model's results to that of BFLOW-LHA over the three watersheds. All experiments are carried out in a Windows 10 environment on a machine equipped with a core i7 processor and 16 GB of RAM. The number of epochs was used as a predictor in experiments. The details, settings, and evaluation methods used in the experiment are listed below.

Deep learning training parameters

We used a grid search to find the best value of the model's parameter when tuning hyper parameters of the deep learning model on the dataset. Selecting the best model parameter results in the best model performance. We used LSTM to predict the baseflow point from each streamflow point

Fig. 4 Baseflow separated by BFLOW-Lyne and Hollick for Dawa Melka Guba watershed

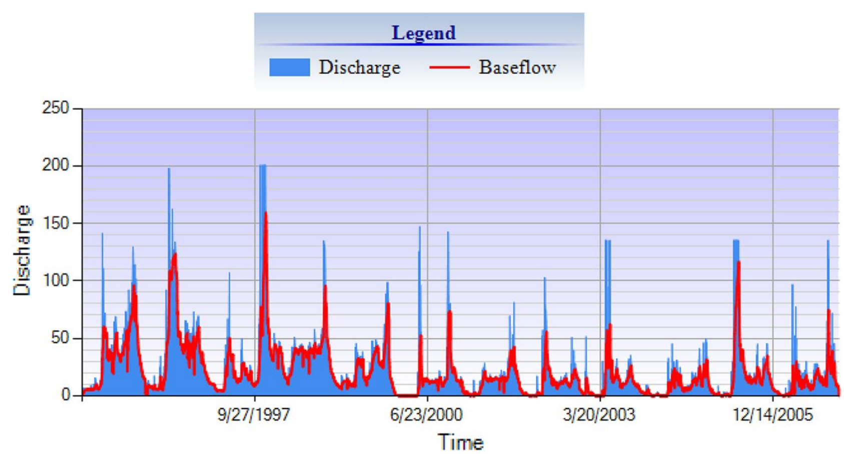

Fig. 5 Sample separated baseflow for year 2006

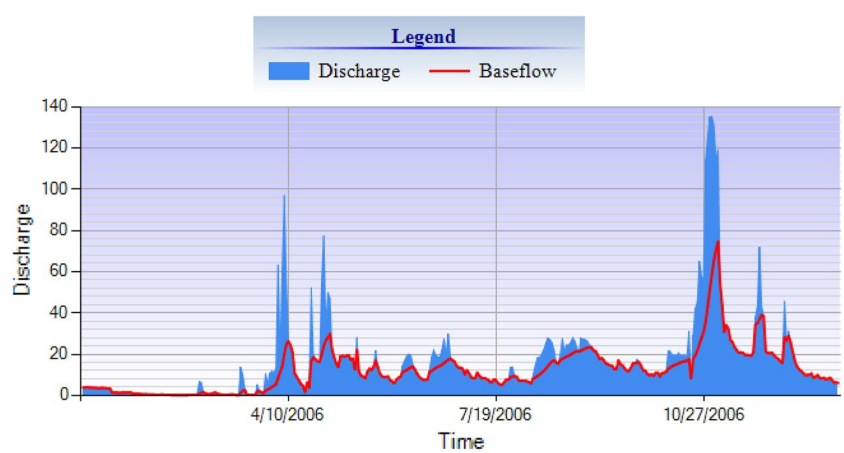

Table 6 Deep learning hyper parameters

Hyperparameters	Values
Number of neurons	500
Dense layer	128
Dropout	0.5
Activation	Softmax
Optimizer	Adam
Epoch	50

automatically. Table 6 shows the deep learning hyper parameters used in this study.

Result and discussion

Using the analytical methods' default filtering parameters or constants does not give findings that are well correlated with basin characteristics. The results of the HYSEP and RDF approaches will be greatly enhanced once they have been calibrated by taking into account the physical characteristics of the basin. The un-calibrated analytical methods' statistical match to physical properties is poor due to a lack of calibration to basin and gage-specific characteristics. When compared to calibrated values, the results of this investigation show that un-calibrated analytical base flow separation approaches do not produce highly significant baseflow values. Once calibrated, the analytical methods can separate the baseflow component of the streamflow reasonably.

The methods' prediction power is greatly improved when the best fit baseflow separation method is combined with deep learning. After the methods' have been calibrated to basin or gage-specific characteristics and integrated with deep learning, they can be applied to other watersheds with similar characteristics without the need for additional calibration. The hybrid model is easy to use and does not take up much time. In this section, the results of the hybrid baseflow separation model is presented for the three testing watersheds in the study area and the results are compared to baseflow separated by BFLOW-LHA. Figures 6, 7, 8 depicts the streamflow and separated baseflow by proposed model for Genale Chenemasa, Genale Halwen, and Weib watersheds of the study area, respectively.

Figures 9, 10, 11 show the separated baseflow using the BFLOW-LHA and the proposed model fitted to daily streamflow data from the three test watersheds. The separated baseflow curve in the proposed model steps up quickly during significant rain, then smoothly recedes until the next significant rain. The baseflow curve between events closely matches the expected baseflow behavior at each testing watershed, providing a more accurate estimate of expected baseflow behavior during events than BFLOW-LHA. The separated baseflow meets the standard baseflow

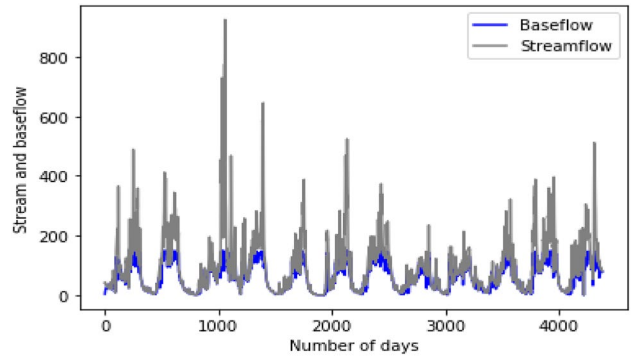

Fig. 6 Baseflow separation using the proposed model on Genale Chenemasa

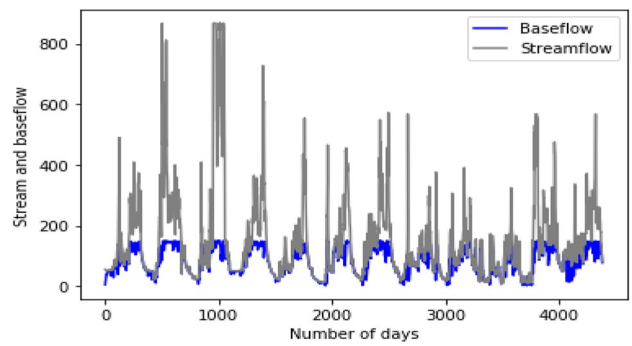

Fig. 7 Baseflow separation using the proposed model on Genale at Halwen

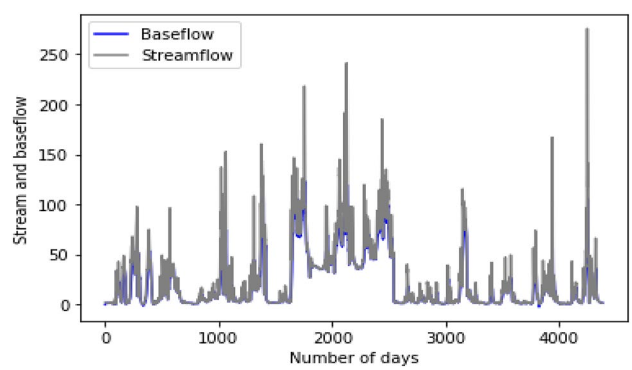

Fig. 8 Baseflow separation using the proposed model on Weib at Sof Umer

characteristics in all of the testing watersheds (i.e., low flow prior to an event is typically all baseflow, the baseflow peak falls after the total hydrograph peak, and baseflow rejoins the total hydrograph as quick flow ceases) (Duncan 2019).

The current study found a good linear correlation of BFI values with other studies that use CMB to calibrate four RDF methods in the Sudano-Sahelian Watershed, Burkina Faso (Kouanda et al. 2018), and Stable Water Isotopes

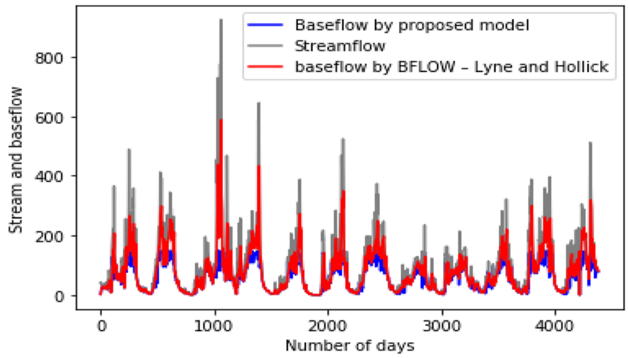

Fig. 9 Comparison between the proposed model with BFLOW-LHA on Genale Chenemasa

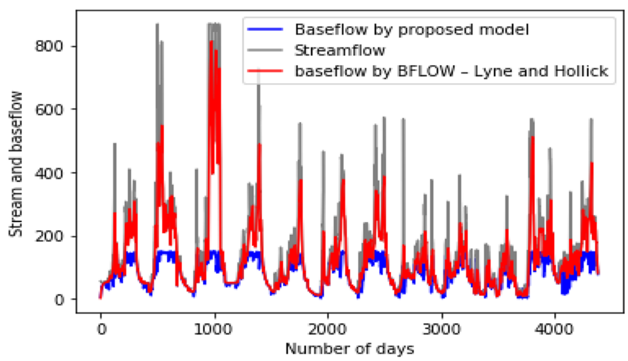

Fig. 10 Comparison between the proposed model with BFLOW-LHA on Genale Halwen

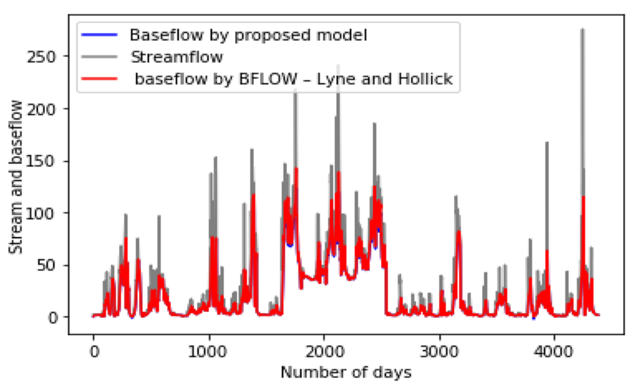

Fig. 11 Comparison between the proposed model with BFLOW-LHA on Weib at Sof Umer

(SWIs) for verification of baseflow separations using RDF methods in the Grand River basin in southern Ontario, Canada (Stadnyk et al. 2015). This study's average recession constant value correlates with the value obtained by Duncan (2019) for daily flow data (i.e., 0.91–0.98), and the baseflow curves obtained match the results of various studies (Aboelnour et al. 2020; Lott and Stewart 2016; Stewart et al. 2007; Eckhardt 2005; Mazvimavi et al. 2004; Ladson et al. 2013).

Field measurement data is required to evaluate the performance of the proposed model. However, because those data (chemical and radioactive tracers) are unavailable, the separated baseflow is compared to theoretical baseflow curve behaviors. Furthermore, if the proposed model's parameters are not optimized for basin or catchment characteristics, it may result in poor performance in some catchments.

Conclusion

An attempt was made in this study to introduce a new baseflow prediction model by combining an analytical method with deep learning. The adjustment of parameters in the existing baseflow separation methods varied from watershed to watershed. Instead of adjusting parameters for different watersheds each time, train the model once and use it to predict baseflow wherever it is needed. The dataset used to test and evaluate the proposed model's performance was obtained from MoWIE. The model was trained using streamflow and baseflow separated from the Dawa Melka Guba watershed using the BFLOW-LHA and evaluated on baseflow prediction for the basin's three watersheds: Weib Sof Umer, Genale Halwen, and Genale Chenemasa. The hybrid model presented here aims to preserve as much physical applicability as feasible in the separated components of streamflow by calibrating parameters to gage-specific or basin characteristics. The experimental results on the three watersheds show that the proposed model outperforms the BFLOW-LHA. The model is straightforward and intuitively satisfying in its description of baseflow in the study area. The proposed model can be used as an efficient model for baseflow prediction. As a result, the proposed hybrid baseflow prediction model is suitable for use in a variety of applications where baseflow separation is critical, such as overland flow determination, water resources planning and management, and construction of hydraulic structures. However, the suggested model's performance will be tested on larger datasets and compared to chemical and radioactive tracers (where data is available) in future research.

Acknowledgements This study was performed by three academic staff at Jimma Institute of Technology, Jimma University, Ethiopia. The authors would like to thank the institute for its assistance with various resources, as well as MoWIE of Ethiopia for providing the dataset for our experiments. The authors would like to thank Jimma University for its support during the research work.

Funding This study received no outside funding.

Data availability All the relevant data are uploaded on GitHub and accessible via the following URL: https://github.com/wondiye8/baseflow

Declarations

Conflict of interest The authors declare that they do not have any conflicts of interest about this work.

References

Aboelnour M, Gitau MW, Engel BA (2020) A comparison of streamflow and baseflow responses to land-use change and the variation in climate parameters using SWAT. Water 191(12):1–29

Chapman D (1999) Access Regulation Risk Management And Framework Design. Econ Press 18(4):68–79

Chapman T (1999) A comparison of algorithms for stream flow recession and baseflow separation. Hydrol Process 13(5):701–714

Duncan HP (2019) Baseflow separation—a practical approach. J Hydrol 575(5):308–313

Eckhardt K (2005) How to construct recursive digital filters for baseflow separation. Hydrol Process 19(2):507–515

Gregor M (2010) BFI+ 3.0 user's manual. HydroOffice Software Package

Galli T, Chiclana F, Siewe F (2021) Quality properties of execution tracing, an empirical. Appl Syst Innov 20:35

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735

Huyck AAO, Pauwels VRN, Verhoest NEC (2005) A base flow separation algorithm based on the linearized Boussinesq equation for complex hillslopes. Water Resour Res 41:W08415

Kouanda B, Coulibaly P, Niang D, Fowe T, Karambiri H, Paturel J-E (2018) Analysis of the performance of base flow separation methods using chemistry and statistics in Sudano–Sahelian watershed, Burkina Faso. Hydrol Curr Res 9(2):1–11

Ladson A, Brown RR, Neal B, Nathan R (2013) A standard approach to baseflow separation using the Lyne and Hollick filter. Aust J Water Resour 17:25–34

Lim SR, Schoenung JM (2010) Toxicity potentials from waste cellular phones, and a waste management policy integrating consumer, corporate, and government responsibilities. Waste Manage 30(8–9):1653–1660

Liu Z, Liu S, Ye J, Sheng F, You K, Xiong X, Lai G (2019) Application of a digital filter method to separate baseflow in the small watershed of Pengchongjian in Southern China. Forests 10(12):1065

Lott DA, Stewart MT (2016) Base flow separation: a comparison of analytical and mass balance methods. J Hydrol 535:525–533

Lu W, Li J, Li Y, Sun A, Wang J (2020) A CNN-LSTM-based model to forecast stock prices. Artif Intell Smart Syst Simul 2020:10

Lyne V, Hollick M (1979) Stochastic time-variable rainfall-runoff modelling. In: Institute of engineers Australia national conference. Institute of Engineers Australia, Barton, Vol 79, No 10, pp 89–93

Mazvimavi D, Meijerink AMJ, Stein A (2004) Prediction of base flows from basin characteristics: a case study from Zimbabwe/Prévision de débits de base à partir de caractéristiques du bassin: une étude de cas au Zimbabwe. Hydrol Sci J 49(4):715

Murphy DM, Solomon S, Portmann RW, Rosenlof KH, Forster PM, Wong T (2009) An observationally based energy balance for the Earth since 1950. J Geophs Res 114(17):16

Murphy S, Ouellon T, Ballard J-M, OuellonT CID (2011) Tritium–helium groundwater age used to constrain a groundwater flow model of a valley-fill aquifer contaminated with trichloroethylene (Quebec, Canada). Hydrogeol J 19(1):195–207

Murray G (2010) Framing globalization and work: a research agenda. J Indus Relat 52(1):1–25

Nandini GS, Kumar APS, Chidananda K (2021) Dropout technique for image classification based on extreme learning machine. Global Transit Proc 2(1):111–116

Nathan RJ, McMahon TA (1990) Evaluation of automated techniques for base flow and recession analyses. Water Resour Res 26(7):1465–1473

Pielke RA Sr, Adegoke J, BeltraáN-Przekurat A, Hiemstra CA, Lin J, Nair US, Niyogi D, Nobis TE (2007) An overview of regional land-use and land-cover impacts on rainfall. Tellus B: Chem Phys Meteorol 59(3):587–601

Rammal M, Archambeau P, Erpicum S, Orban P, Brouyère S, Pirotton M, Dewals B (2018) An operational implementation of recursive digital filter for base flow separation. Water Resour Res 54(10):8528–8540

Roy S, Mistri B (2013) Estimation of peak flood discharge for an ungauged river: a case study of the Kunur river, West Bengal. Geogr J 2013:11

Sivapalan M (2018) From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science. Hydrol Earth Syst Sci 22(3):1665–1693

Sloto R, Crouse M (1996) HYSEP: a computer program for streamflow hydrograph separation and analysis. Water-Resources Investigations Report, US Geological Survey

Spongberg ME (2000) Spectral analysis of base flow separation with digital filters. Water Resour Res 36(3):745–752

Stadnyk TA, Asce M, Gibson JJ, Longstaffe FJ (2015) Basin-scale assessment of operational base flow separation methods. J Hydrol Eng 20(5):04014074

Stewart MT, Cimino J, Ross M (2007) Calibration of base flow separation methods with streamflow conductivity. Ground Water 45(1):17–27

Tallaksen LM (1995) A review of baseflow recession analysis. J Hydrol 165(1–4):349–370

Young AR, Gustard A, Bullock A, Sekulin AE, Croker KM (2000) A river network based hydrological model for predicting natural and influenced flow statistics at ungauged sites. Sci Total Environ 251–252(12):293–304