



River Discharge Forecasting Using Land Cover Dynamics and Hydroclimatological Data with an Adaptive Statistical Approach: A Case Study of the Citarum-Majalaya Catchment, Indonesia

Dadang Ruhiat^{1,2}, Indratmo Soekarno^{1*}, Hadi Kardhana¹, Rusmawan Suwarman³

¹ Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Bandung 40132, Indonesia

² Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Bale Bandung, Bandung 40132, Indonesia

³ Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung 40132, Indonesia

Corresponding Author Email: indratmo@itb.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijdne.210403>

ABSTRACT

Received: 13 February 2026

Revised: 18 April 2026

Accepted: 25 April 2026

Available online: 30 April 2026

Keywords:

Multivariable Spline Regression, Hydrologic Engineering Center-Hydrologic Modeling System, land cover, discharge forecasting, nonparametric statistics, hydroclimatological data

River discharge prediction modeling is crucial for water resource planning and environmental management. Physically-based methods are often constrained by data availability, while statistical approaches can be applied despite hydrological data typically being nonlinear, non-stationary, and not normally distributed, which complicates the use of parametric methods. This study develops an adaptive discharge forecasting model using land cover dynamics and historical hydroclimatic data from the Citarum-Majalaya Catchment, selected due to rapid land cover changes and the strategic national importance of the Citarum River. Two approaches are employed: Multivariable Spline Regression (MSR) and the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) hydrological model. The MSR model demonstrates high flexibility and good accuracy with an MAPE of 16.74% and RMSE of 1.97, while HEC-HMS produces an MAPE of 9.4% and RMSE of 1.04. Both models show strong performance, but HEC-HMS is quantitatively superior. The MSR model, requiring simpler data, effectively replicates discharge patterns and can predict monthly discharge up to one year ahead. Parameter significance tests indicate that all variables simultaneously have a significant impact on discharge variation, with residential land cover being the most dominant. These findings emphasize the importance of land cover management, particularly controlling residential areas, to maintain the hydrological stability of the catchment.

1. INTRODUCTION

River discharge is influenced by the interaction of various hydroclimatic variables, such as rainfall and evapotranspiration, as well as biophysical factors, particularly land cover. Rainfall plays a key role as the primary source of river flow through surface runoff and base flow, while evapotranspiration regulates the distribution of water on the land surface and other surfaces. Meanwhile, land cover dynamics, such as the conversion of forests to agricultural land or urbanization, can alter infiltration properties and water runoff, as well as affect base flow capacity. Previous studies have shown that land cover changes significantly contribute to river flow alterations and hydrological responses in the watershed [1].

River discharge modeling is a crucial aspect of water resource infrastructure planning and environmental management, including the design of dams and reservoirs, flood management, and drainage system maintenance, as river flow serves as a key indicator in the operation of hydrological

structures and water risk mitigation [2, 3]. The complexity of hydrological systems, driven by the interaction of rainfall, evapotranspiration, and various types of land cover, often results in data characteristics that are non-linear, non-stationary, and non-normally distributed. These conditions complicate the application of conventional parametric methods, which require assumptions about data distribution [4]. As an alternative, nonparametric statistical approaches offer greater flexibility in handling complex and unstructured data. Physical-based methods, such as HEC-HMS, although powerful in simulating hydrological processes in detail, often require highly complete and adequate input data, as well as assumptions that may not always reflect the dynamic field conditions. Furthermore, physical models necessitate a deep understanding of the underlying physical processes governing river flow, which, in many cases, is difficult to obtain, especially in areas with limited hydrological data.

Several studies have employed non-parametric methods in river discharge modeling. The Non-Parametric Bayesian Networks (NPBN) method has been used to model the

relationship between rainfall, discharge, and other hydrometeorological variables for estimating daily maximum discharge [5]; The nonparametric quantile regression method has been applied to model rivers that are difficult to measure directly, with a focus on estimating river discharge based on river width derived from remote sensing data [6]; Nonparametric functional data methods are used for predicting monthly average discharge and analyzing flood discharge [7]; Nonparametric quantile regression methods are used to predict the discharge levels of critical flow in discharge quantiles [8]; Nonparametric smoothing in transfer function models is used to forecast river discharge as a non-linear function of hydrometeorological inputs, serving as a comparison to classical ARIMA methods [9]; Nonparametric time series models are employed for river discharge forecasting and river system response analysis [10]; K-Nearest Neighbors Regression (KNNR), Artificial Neural Networks (ANN), and Adaptive Neuro-Fuzzy Inference System (ANFIS) models are used for monthly discharge prediction, and to evaluate their relative performance in capturing discharge variability [11]; An ANN was also employed to predict flood inundation depth in the Majalaya Watershed [12]; The Mann-Kendall method is used for trend analysis of variations and the relationship between river discharge and rainfall [13]; The Random Forest model is used for daily streamflow prediction, and it was found that Random Forest performs exceptionally well compared to several other methods for short-term forecasting [14]; and Gaussian Process Regression (GPR) is capable of providing predictions along with uncertainties, but it is highly dependent on kernel selection and requires significant computational costs [15]. The main advantage of several of these methods lies in their ability to provide accurate predictions, even when the data do not meet the assumption of normality.

However, although various non-parametric approaches have been developed, most studies remain focused on improving predictive accuracy without explicitly accounting for changes in the relationships between hydroclimatic variables and land use dynamics over time. Numerous non-parametric methods have been applied in river discharge modeling; however, they still exhibit several limitations, including sensitivity to parameter selection, high data requirements, potential long-term instability, and relatively high computational cost. In addition, most of these approaches have not adequately represented the complex interactions between climatic factors and anthropogenic activities, particularly land use change, within a flexible and adaptive framework. The main challenge in river discharge modeling lies in developing an approach that can simultaneously handle non-stationary hydroclimatic behavior and dynamically changing land use conditions, while maintaining predictive accuracy, interpretability, and practical applicability in water resources and environmental management.

In line with the development of non-parametric statistical approaches, Multivariable Spline Regression (MSR) is an alternative method that has been widely applied in various fields. This method offers high flexibility in capturing data patterns and has demonstrated good predictive performance [16, 17]. However, the application of MSR in hydrological modeling remains relatively limited. Therefore, this study aims to develop and evaluate an MSR model for river discharge forecasting by considering the interactions among land cover, rainfall, and evapotranspiration. Specifically, MSR

is selected due to its ability to model nonlinear relationships that may vary across different ranges of data, making it more adaptable to non-stationary hydrological data characteristics compared to several conventional approaches.

The novelty of this study lies in the application of MSR to explicitly model the combined effects of land-use dynamics, rainfall, and evapotranspiration within a non-stationary hydrological system. To the best of the authors' knowledge, this approach has received limited attention in previous river discharge prediction studies. This study is expected to provide a scientific basis for the development of more sustainable and adaptive watershed management strategies in response to environmental changes. Furthermore, the findings of this research are anticipated to support water resource management and the planning of hydrological infrastructure, such as weirs, reservoirs, drainage systems, and water-related disaster risk mitigation.

2. STUDY AREA, DATA, AND METHODS

2.1 Description of study area

The study location is situated in the catchment area of the Citarum-Majalaya River discharge monitoring station, which is part of the upper Citarum watershed. Administratively, this area is located in Bandung Regency and covers six sub-districts. The majority of the catchment area is in Kertasari Sub-district (33.5%), followed by Pacet Sub-district (30.8%), Ibun Sub-district (26.6%), with smaller portions in Majalaya, Pangalengan, and Paseh Sub-districts. Geographically, the catchment area is located between $7^{\circ}3'1.96''$ - $7^{\circ}14'33.1''$ S and $107^{\circ}37'50.7''$ - $107^{\circ}48'40.8''$ E [18]. The selection of the study location is based on the dynamic land cover changes in the area, which occur at a faster rate compared to other regions, potentially increasing annual flood discharge. Additionally, the location was chosen due to the strategic role of the Citarum River as a national river. The study location is shown in Figure 1.

The study area exhibits a broad elevation range, from 664.3 m a.s.l. at the stream-gauging station to approximately 2,595 m a.s.l. in the mountainous zone, with a reported river gradient of about 26.9%. Its physiography is dominated by gently sloping plains in the southern sector, and becomes steeper in the hilly and mountainous zones. Climatically, mean rainfall is approximately 5.55 mm/day, with peak precipitation typically occurring in January (around 289 mm/month). Extreme rainfall events contribute to frequent flooding (≥ 6 events per year); for example, a daily rainfall of 87.84 mm on 22 February 2018 increased discharge to 155 m³/s and generated inundation over roughly 5 km². In terms of river morphology, sedimentation rates are relatively high, averaging 1.9 cm/month, with mean deposit thickness reaching approximately 40 cm, which is associated with land-cover dynamics and high rainfall intensity [19]. Under low-flow conditions, the Majalaya Watershed water balance indicates pronounced interannual rainfall variability ($\sim 1,492$ – $3,834$ mm/year, mean $\sim 2,317$ mm/year), with surface runoff increasing during wet years, whereas baseflow remains comparatively small (~ 0.15 – 0.26 m/year) yet plays a critical role in sustaining dry-season streamflow [20].

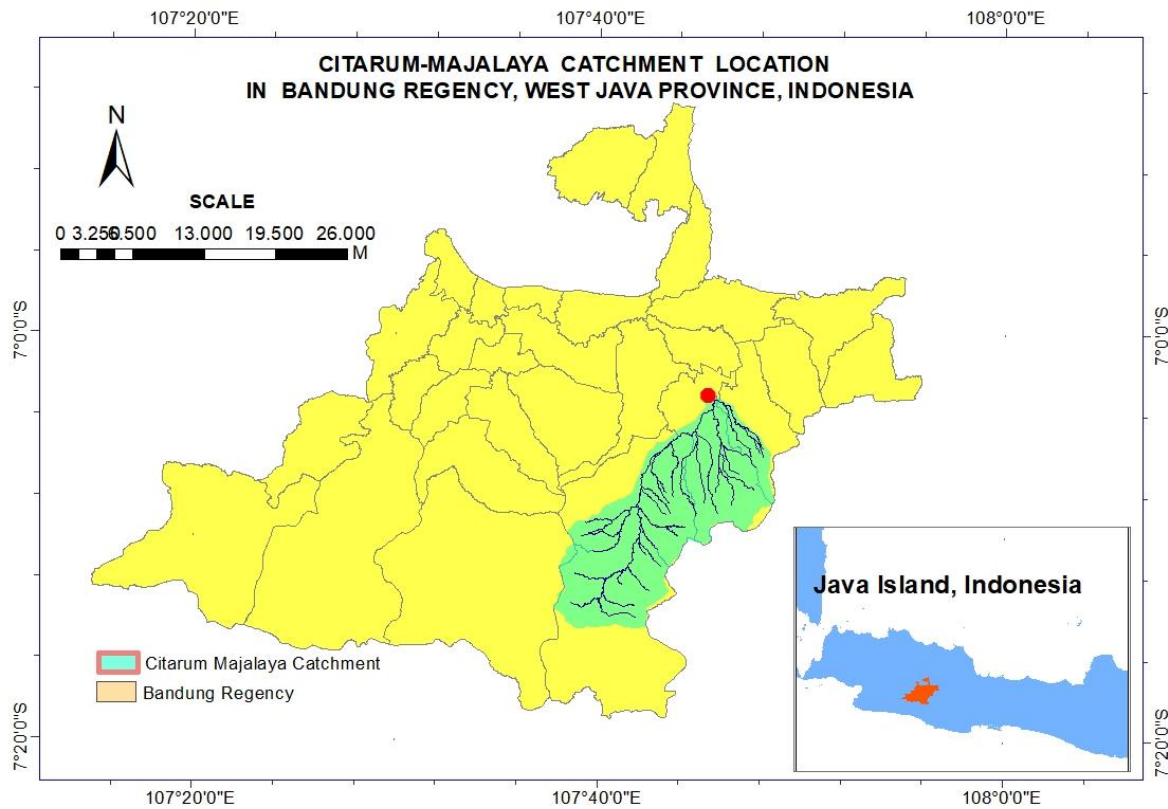


Figure 1. The Citarum-Majalaya Catchment in Bandung District, Indonesia

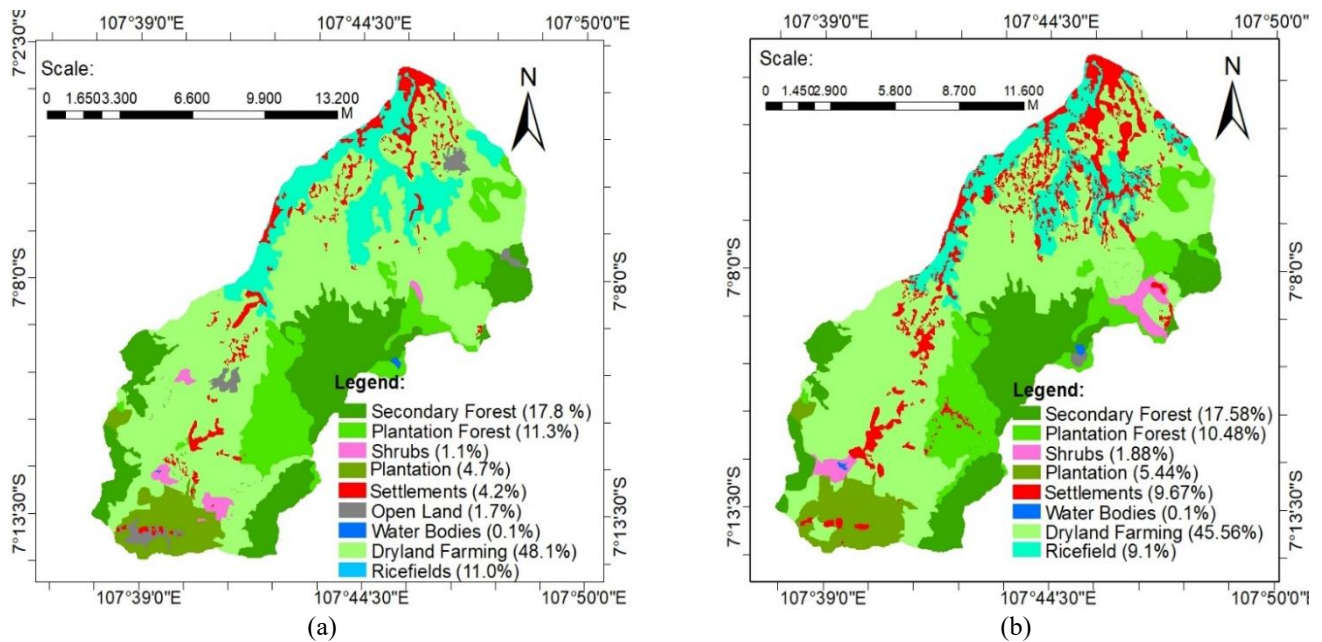


Figure 2. Land cover of the Citarum-Majalaya Catchment: (a) 2000; (b) 2021 [18]

2.2 Data

Several studies have involved land cover dynamics as a key variable in hydrological modeling, providing a solid foundation for the inclusion of land cover variables in this research. Land cover is also positioned as a key factor, as changes in land cover over recent years have been shown to alter baseflow values in the watershed [21]. In the Gajahwong watershed, the use of the HEC-HMS model shows that land cover is the main control of hydrological response, where land

cover maps are used to determine Curve Number (CN) values. As a result, the conversion of agricultural land to built-up areas is reflected in an increase in CN and a decrease in infiltration capacity [22]. The use of land cover maps as input for the SWAT model, along with rainfall and discharge data, indicates that the conversion of green land to developed land increases surface runoff, total river flow, and monthly peak discharge, while decreasing baseflow [23]. The change in land cover from permeable to impermeable surfaces has also been reported to increase CN values, reduce infiltration, and

increase surface runoff, which ultimately leads to higher flood peak discharge [24]. Several other studies in different locations also emphasize that the conversion of vegetated land to built-up areas or land with lower infiltration capacity increases CN values, enhances and accelerates surface runoff, which impacts the rise in peak discharge and flood runoff volume [25-29]. Thus, land cover variables play a key role as a main factor in hydrological modeling, particularly in influencing hydrological responses such as baseflow and surface runoff, which in turn determine the river discharge magnitude.

The data used in this study include land cover data, rainfall data, and climatological data (air temperature, humidity, wind speed, and solar radiation duration). The land cover data in SHP file format was obtained from the Ministry of Environment and Forestry for the period from 2000 to 2021. This data is used to analyze the dynamics of land cover

changes in the Citarum-Majalaya watershed over three-year intervals. Land cover data correction was performed using Geographic Information System (GIS) with ArcGIS 10.7.1 software to ensure the data's alignment with the actual conditions [18]. The land cover in the study area is classified according to SNI 7645-2010 on land cover classification [30], which includes categories such as secondary forest, plantation forest, shrubs, plantation, settlements, open land, water bodies, dryland farming, and ricefields, as illustrated in Figure 2. For modeling purposes, the land cover classification is simplified by grouping secondary forest and plantation forest into the forest category, and combining dryland farming, plantation, and ricefields into the agriculture category. Thus, the land cover classification used consists of six categories: forest, agriculture, settlements, shrubs, open land, and water bodies, as presented in Table 1.

Table 1. Land cover of the Citarum-Majalaya Catchment 2000-2021

Years	Land Cover (km ²)					
	Forest	Agriculture	Settlements	Shrubs	Open Land	Water Bodies
2000	59.6	130.4	8.6	2.3	3.5	0.1
2003	59.6	130.9	9.5	2.6	1.7	0.1
2006	59.4	129.6	10.8	2.9	1.7	0.1
2009	58.5	128.6	12.6	2.9	1.8	0.1
2012	58.4	128.5	13.5	1.9	2.1	0.2
2015	58.4	127.7	15.8	0.6	1.8	0.2
2018	57.6	126	18.7	0.9	1.2	0.2
2021	57.4	122.9	19.8	3.8	0.3	0.2

Table 2. Rainfall data availability

Rainfall Station	Longitude	Latitude	Data Availability	Influence Area
Cipaku-Paseh	107.677	-7.192	2002-2021	41.40%
Cibeureum	107.764	-7.057	2002-2021	43.10%
Situ Cisanti	107.661	-7.208	2014-2021	15.50%

Source: Hydrology and Water Environment Center, Directorate General of Water Resources, Ministry of Public Works

Daily rainfall data were obtained from the Citarum River Basin Organization of the Ministry of Public Works, located in Bandung City. The data were collected from three rainfall stations: Cibeureum Station, Situ Cisanti Station, and Cipaku-Paseh Station, which are situated around and within the Citarum-Majalaya catchment area. The availability of rainfall data varies: data from Cibeureum Station is available for the period from 2002 to 2021 (20 years), data from Situ Cisanti Station is available for the period from 2014 to 2021 (8 years), and data from Cipaku-Paseh Station is available for the period from 2001 to 2021 (21 years). For analysis purposes, daily rainfall data from the three stations were integrated into a spatial (areal) rainfall dataset. However, prior to conversion into areal rainfall, the quality of the daily rainfall data from each station was first verified. Outlier detection was conducted using the boxplot method. Identified outliers were not directly removed; instead, they were validated through input error checks and spatial consistency analysis with nearby stations to distinguish between technical errors and extreme meteorological events. Valid extreme rainfall events were retained, whereas erroneous or corrupted data were corrected using the normal ratio imputation method. This quality control procedure ensures data reliability and continuity, thereby maintaining consistent input quality for hydrological modeling.

The daily areal rainfall for the Citarum-Majalaya Catchment was calculated using the Thiessen polygon method.

The dominant influence area of the rainfall stations is Cibeureum Station (43.1%), followed by Cipaku-Paseh Station (41.4%), and Situ Cisanti Station at approximately 15.5%, as illustrated in Figure 3(a). The daily rainfall data of the catchment area (Figure 3(b)) was then converted into monthly average rainfall (Figure 3(c)). The availability of daily rainfall data for each station is presented in Table 2.

One of the parameters involved in the modeling is evapotranspiration. The evapotranspiration value in a given area is influenced by several factors, including geographic location (altitude, latitude, and longitude) and climatic conditions, which are represented by climatological data encompassing air temperature (°C), humidity (%), solar radiation duration (hours), and wind speed (km/day). Climatological data were obtained from the Bandung Geophysical Station (BMKG Bandung) for the period from 1980 to 2017. To ensure the quality, continuity, and accuracy of evapotranspiration calculations, missing data were handled using the daily climatological mean method. This approach is appropriate for datasets with long observation periods. An overview of the climatic conditions at the study location during this period, in the form of monthly average data, is presented in Figure 4.

Figure 4 also illustrates climate change trends from 2000 to 2017. During this period, the mean air temperature exhibited an increasing trend, whereas the mean air humidity, mean wind speed, and mean duration of sunshine showed a

decreasing trend. The minimum mean air temperature was recorded at 21.63 °C in September 2013, while the maximum mean temperature was recorded at 25.34 °C in April 2014. The minimum mean wind speed was 8.9 km/day in June 2009, and the maximum was 173.5 km/day in October 2006. The minimum mean air humidity was 59.9% in September 2013, whereas the maximum was 88.0% in April 2007. The minimum mean sunshine duration was 1.4 hours in September 2007, while the maximum reached 7.3 hours in June 2015. Furthermore, stationarity testing using the Augmented

Dickey-Fuller test (ADF), along with seasonality analysis based on the Autocorrelation Function (ACF) plot and spectral regression methods, indicates that the monthly mean temperature and humidity data are stationary and exhibit a seasonal pattern with a period of 6. In contrast, the monthly mean wind speed data are non-stationary and show a seasonal pattern with a period of 216, while the monthly mean sunshine duration data are stationary and exhibit a seasonal pattern with a period of 12.

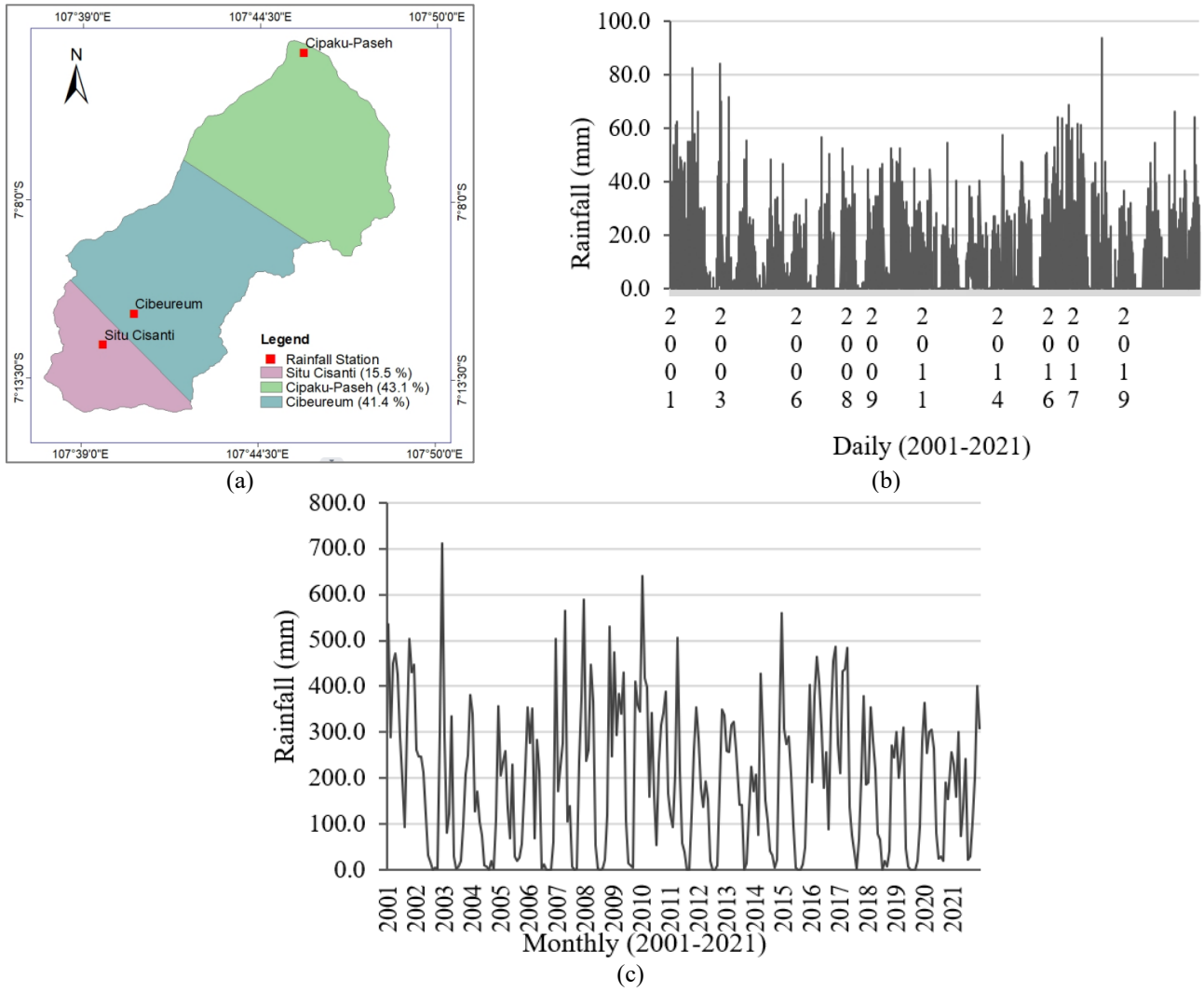
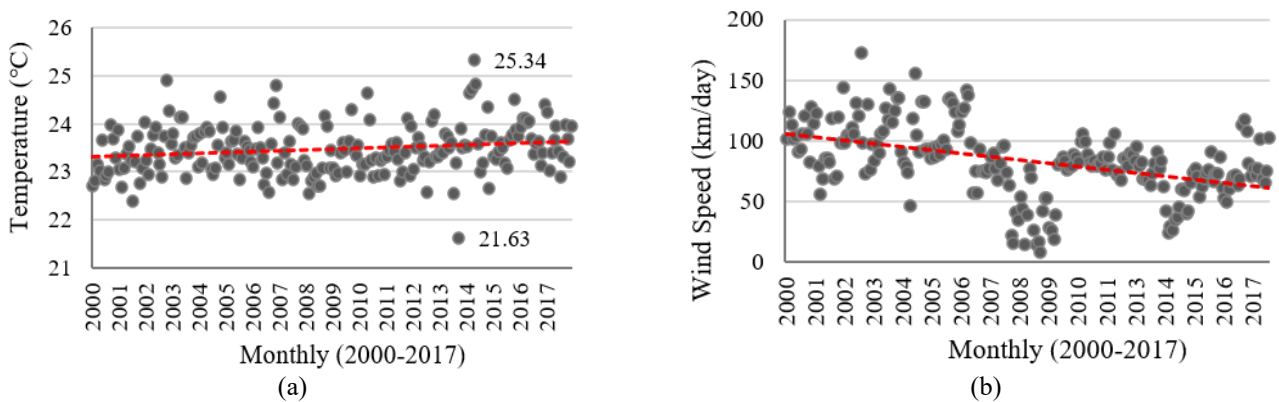


Figure 3. (a) The Thiessen polygon, (b) the daily rainfall, (c) the monthly rainfall



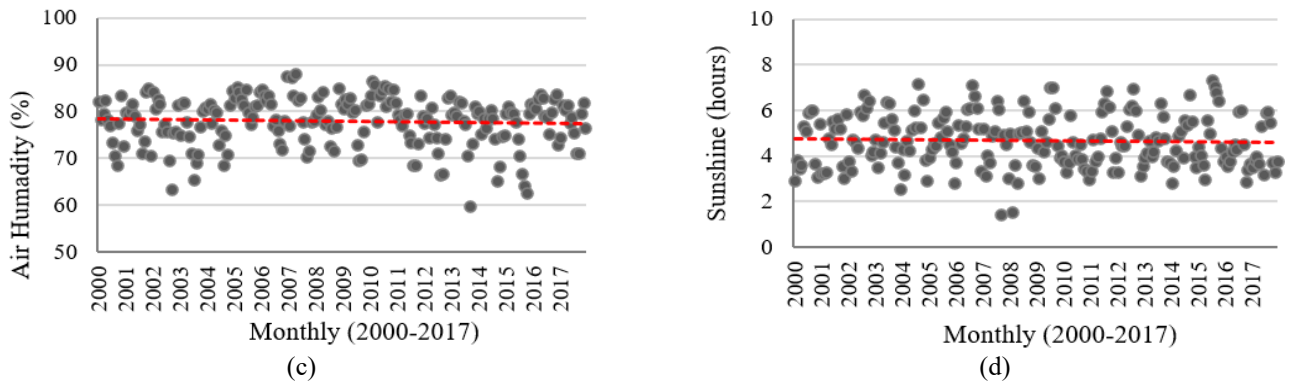


Figure 4. Climatological data (Monthly average): (a) temperature (°C), (b) wind speed (km/day), (c) air humidity (%), and (d) sunshine (hours)

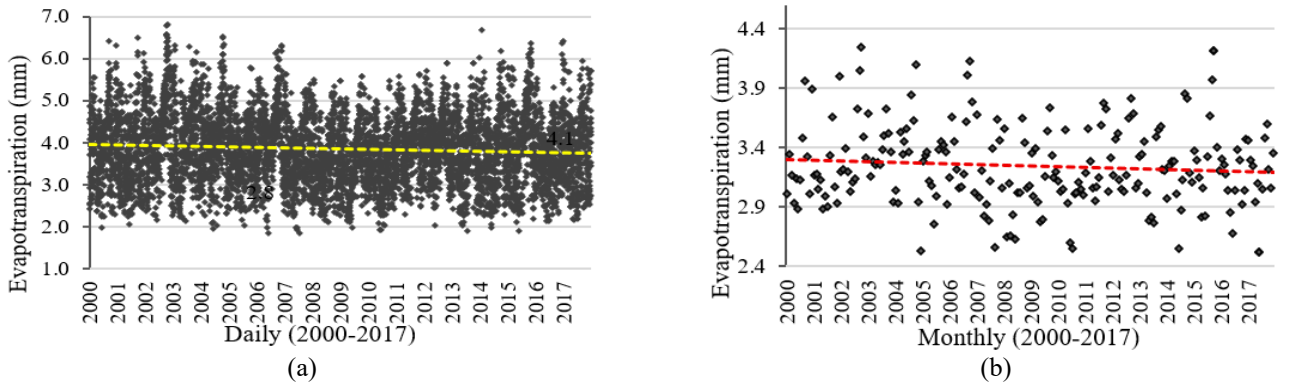


Figure 5. Evapotranspiration (mm/day): (a) daily (2000-2017); (b) monthly average (2000-2017)

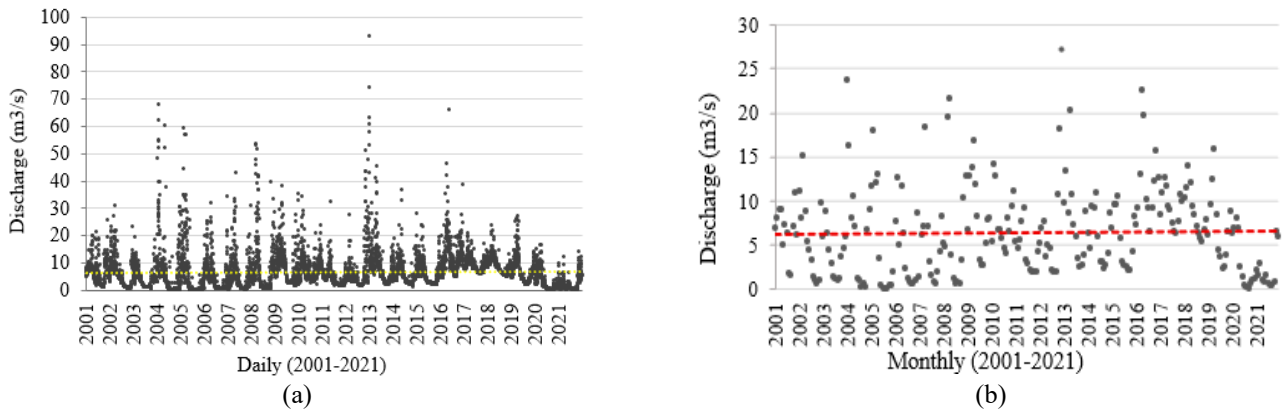


Figure 6. Discharge of Citarum-Majalaya (m^3/det): (a) daily (2000-2021); (b) monthly average (2001-2021)

The daily and monthly evapotranspiration values were calculated using the Penman and Penman-Monteith methods for modeling purposes, integrating meteorological variables such as air temperature, humidity, solar radiation duration, and wind speed. Monthly evapotranspiration calculations were conducted using the CROPWAT 8.0 software, which applies the FAO 56 approach for estimating evapotranspiration (ET_o). The results show that the average monthly evapotranspiration values over an 18-year period (2000-2017) ranged from 2.51 to 4.21 mm/day, while the daily evapotranspiration values ranged from 1.8 to 6.8 mm. A visualization of the evapotranspiration values is presented in Figure 5.

The river discharge data used for modeling were obtained from the Citarum-Majalaya Hydrological Observation Station (PDA) and sourced from the Citarum River Basin Organization (BBWS). For modeling purposes, the discharge data is presented in daily and monthly forms. The maximum

daily discharge reached 92.8 m^3/second , while the maximum monthly average discharge was recorded at 27.2 m^3/second , with the minimum value approaching zero. A visualization of the average monthly river discharge in the Citarum-Majalaya can be seen in Figure 6.

2.3 Methods

This modeling is to identify and analyze the relationship between rainfall, evapotranspiration, and various types of land cover as predictor variables, with river discharge as the response variable. This relationship is quantitatively represented through a regression model, enabling the prediction of future river discharge based on the available input variables. The model development was carried out by incorporating and utilizing several analytical approaches, including GIS for the analysis of land cover change dynamics,

the Penman and Penman-Monteith methods for calculating evapotranspiration values, and the nonparametric MSR method for building a river discharge prediction model. For comparison, the river discharge model was also developed using the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) software. To determine the statistical significance of each input variable in this study, a significance level (α) of 0.05 was employed. Simultaneous significance testing was conducted using the F-test, while partial significance testing was performed using the t-test. In addition, the analysis was complemented by the evaluation of standard error (SE) and p-values for each variable.

Figure 7 presents a brief flowchart of the modeling process, outlining the main stages, including the determination of required input and output data, selection of the most optimal model, model validation, comparison of model results, and drawing conclusions. The base land cover data in shapefile format for the period from 2000 to 2021 was corrected using

ArcGIS 10.7.1 software [18]. Evapotranspiration values were calculated using climatological data inputs (humidity, air temperature, solar radiation, and wind speed) with the Penman-Monteith method for monthly averages using CROPWAT software [31, 32] and for daily values using the Penman method. The monthly average evapotranspiration was used as input for the MSR model, while the daily evapotranspiration was used as input for the HEC-HMS model. Subsequently, the river discharge modeling process was carried out through statistical computation using the R programming language. Additionally, the HEC-HMS software with the Soil Conservation Service-Curve Number (SCS-CN) approach was also used to predict river discharge [33] and the results were compared with those obtained from the statistical approach. The SCS-CN approach was applied based on the consideration that this method is suitable for small watersheds ($\leq 250 \text{ km}^2$) [34].

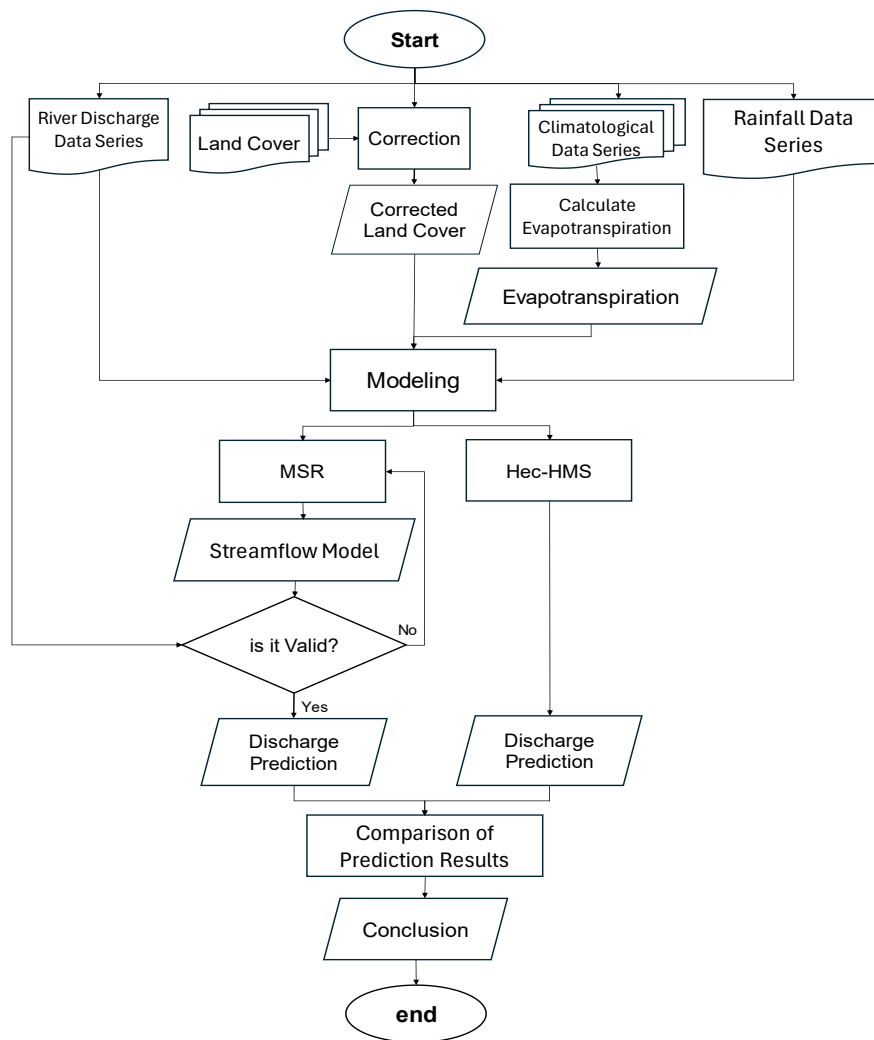


Figure 7. Modeling process flowchart

For the MSR model, the input data were monthly, so the predicted discharges were also on a monthly basis. In contrast, the HEC-HMS model was run using daily input data to produce daily discharges for the same year. These daily discharges were then aggregated into monthly values by summing all daily discharges within each month. In this way, the outputs of both models were aligned on the same temporal scale (monthly), allowing for a fair comparison. All error

metrics (RMSE, MAPE, R^2) were calculated on the monthly discharges to ensure standardization and consistency in the comparison between the MSR and HEC-HMS models.

The prediction period (forecast horizon) in this hydrological modeling is divided based on its objectives. In the review chapter on "Streamflow forecasting" [35], it is stated that river flow forecasting is categorized according to the time horizon: short-term (for flood forecasting), medium-term (for reservoir

operation), and long-term (for water resource planning). Several other literatures generally divide the prediction period into three categories: 1) Short-term, daily to two-week predictions for flood warning and intensive rainfall management [36]; 2) Medium-term, weekly to monthly predictions (1–3 months) for reservoir operation, irrigation, and river management on a seasonal cycle [37]; 3) Long-term, monthly to annual predictions for water resource planning, watershed management, climate change mitigation, and water allocation policies [38]. The time horizon categories are also used in several previous river discharge models, including study [37], which discusses predictions with a lead time of one month or more; study [39], which demonstrates the capability of ML/ANN/LSTM models in river flow prediction with long lead times; and study [40], which classifies monthly predictions into short-term, intermediate-term, and long-term categories. The river discharge modeling in this study aims to predict and project the average monthly discharge for the next 12 months; thus, based on the aforementioned prediction time horizon criteria, it falls under the long-term category.

2.3.1 Multivariable Spline Regression discharge model

Advances in nonparametric statistical approaches have introduced several methods for nonparametric regression estimation, including Fourier series, spline, and kernel methods [16]. In particular, spline regression demonstrates advantages in handling data patterns characterized by abrupt changes, whether increasing or decreasing [16]. Spline estimators tend to adapt to data variability, thereby producing models that closely reflect the underlying data structure [41]. One specific form of spline regression, namely MSR, has been applied in various modeling domains, except in the context of hydrological modeling.

The river discharge modeling in this study uses the MSR method. A spline is a piecewise polynomial function that is continuous, allowing it to adapt more effectively to sharp upward or downward data patterns with the help of knot points, producing a relatively smooth curve [16]. Modeling with the MSR method involves more than one predictor variable, with the general equation as stated in Eq. (1) [42].

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

where, f is the regression curve whose form is unknown. If f exhibits an additive property and can be approximated by a spline, the regression can be expressed in the form of Eq. (2).

$$y_i = \sum_{j=1}^p f(x_{ji}) + \varepsilon_i, i = 1, 2, \dots, n \quad (2)$$

where,

$$f(x_{ji}) = \beta_{j0} + \sum_{l=1}^{m_j-1} \beta_{jl} x_{ji}^l + \sum_{k=1}^{r_j} \beta_{j(m_j+k)} (x_{ji} - K_{jk})_+^{m_j-1} + \varepsilon_i \quad (3)$$

$$i = 1, 2, \dots, n$$

Thus, Eq. (1) can be written in the form of Eq. (4) as follows:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) = \beta_{00} + \sum_{j=1}^p \left(\sum_{l=1}^{m_j-1} \beta_{jl} x_{ji}^l + \sum_{k=1}^{r_j} \beta_{j(m_j-1+k)} (x_{ji} - K_{jk})_+^{m_j-1} \right) + \varepsilon_i \quad (4)$$

with $\beta_{00} = \sum_{j=1}^p \beta_{j0}$

$$(x_{ji} - K_{jk})_+^{m_j-1} = \begin{cases} (x_{ji} - K_{jk})^{m_j-1}, & x_{ji} \geq K_{jk} \\ 0, & x_{ji} < K_{jk} \end{cases} \quad (5)$$

The modeling of the Citarum-Majalaya river discharge was carried out with several spline order alternatives to obtain the most optimal model, namely order 1 (linear), order 2 (quadratic), and order 3 (cubic). The predictor variables include evapotranspiration, rainfall, and several types of land cover, which include forest land, agriculture, settlement areas (built-up land), shrubland, and open land. The estimation of the regression coefficients for the MSR model was performed through statistical computation using R. The success of the modeling is highly influenced by the selection of the optimal knot points for each predictor variable in relation to river discharge as the response variable. The optimal knot points will yield a model with a higher level of accuracy. The approach used to determine the optimal knot points is based on the Mean Squared Error (MSE) and Generalized Cross Validation (GCV), which are expressed in Eqs. (6) and (7) respectively [16].

$$MSE = n^{-1} \sum_{i=1}^n (Y_i - f(x_i))^2 \quad (6)$$

where, n represents the number of observation data; Y_i denotes the response variable at the i -th observation; x_i represents the predictor variable at the i -th observation; and $f(x_i)$ represents the regression function at the i -th observation with the knot point.

$$GCV = \frac{n^{-1} \sum_{i=1}^n (Y_i - f(x_i))^2}{(n^{-1} Tr(I - G))^2}, n^{-1} Tr(I - G) < 0 \quad (7)$$

where, $Tr(I - G)$ represents the trace of the matrix $(I - G)$; I is the identity matrix; $G = W(W^T W)^{-1} W^T$; and W is the column identity matrix.

The river discharge modeling of the Citarum-Majalaya using the MSR method involves several predictor variables, including evapotranspiration (X_1), rainfall (X_2), and land cover area, which consists of categories such as forest (X_3), agriculture (X_4), settlement (X_5), shrubland (X_6), and bare land (X_7), with river discharge serving as the response variable (Y).

The modeling process begins with organizing the input variables into a matrix using Excel or Notepad, followed by creating scatter plots to visualize the relationships between predictor variables and the response variable (river discharge). The optimal knot location is then determined based on the minimum values of MSE and GCV. The dataset is divided into a training subset covering 17 years (2001–2017) and a

validation subset for one year (2018). Model parameters are subsequently estimated, and the predicted discharge values and residuals are calculated. Model performance is evaluated using MSE (mean square error), generalized cross-validation (GCV), and MAPE, with the best model selected based on a combination of MAPE, MSE, RMSE, and R². Model validation is conducted using MSE and MAPE, and the parameters are tested for statistical significance and standardized coefficients. All computations and analyses are performed using the statistical software *R*, allowing the MSR procedure to be applied systematically and reproducibly.

The significance test of the model parameters is conducted both simultaneously and partially. The simultaneous test uses the F-statistic and aims to evaluate the significance of all model parameters simultaneously in explaining the variation in river discharge. Meanwhile, the partial test is conducted using the t-statistic to assess the significance of each parameter individually with respect to the response variable [43]. To assess the contribution of each predictor variable to the variation in river discharge, this can also be done based on the standardized coefficient values. These coefficients are the result of converting the model's coefficient values by multiplying them by the ratio of the standard deviation of each predictor variable to the standard deviation of the response variable (river discharge).

The general formulation of the F-test statistic is expressed in Eq. (8), with the following testing hypothesis:

H0: $\beta_1 = \beta_2 = \dots = \beta_{q+m} = 0$, ($q + m$) represents the number of regression parameters, excluding β_0 .

H1: at least one $\beta_k \neq 0$ where $k = 1, 2, \dots, q + m$

$$F_{calculated} = \frac{MS_{reg}}{MSE} \quad (8)$$

where $Var(\hat{\beta}_k) = diag[(X'X)^{-1}]_k$, and the parameter is considered significant if $|t_{calculated}| > t_{\frac{\alpha}{2}, (n-(q+m)-1)}$.

2.3.2 Hydrologic Engineering Center-Hydrologic Modeling System discharge model

The HEC-HMS is a software commonly used for rainfall-runoff simulation and demonstrates the application of the SCS-CN method in a watershed [44]. Modeling with HEC-HMS consists of four main components: Basin Model, Meteorological Model, Control Specifications, and Terrain Data. The delineation of the Citarum-Majalaya Catchment in the Basin Model component is derived from the DEMNAS, which was downloaded from the official website of the Geospatial Information Agency. The time series data required includes daily rainfall data and daily river discharge data for the specified period. The rainfall data serves as input for the Meteorological Model. The simulation period, time step, and warm-up settings for calibration/validation purposes are configured through the Control Specifications.

The HEC-HMS configuration utilizes several methods and various key parameters: SCS Curve Number for the loss component, including initial abstraction (mm) and CN values; SCS Unit Hydrograph for rainfall-runoff transformation, including lag time parameter (minutes); Recession for baseflow, including initial discharge (m³/s), recession constant, and threshold flow (m³/s); Muskingum-Cunge for routing, including channel geometry (length, slope, width, side slope), Manning's roughness coefficient, and flow index (m³/s); and Percolation for subsurface exchange (rate,

m³/s/1,000 m²) representing loss/gain.

2.3.3 The goodness and validity of the model

The goodness and validity of the model can be evaluated using several statistical criteria, including Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and R-squared (R²). MAPE measures the average percentage of absolute error between the predicted and actual values, with smaller MAPE values indicating better model performance. MSE calculates the average squared difference between the predicted and actual values; smaller MSE values indicate better model performance. RMSE, the square root of MSE, measures the standard deviation of the prediction errors; smaller values indicate a model with better performance [45]. NSE is used to assess the fit between the predicted and observed data, with a model considered excellent when the NSE value approaches 1, and poor when it approaches $-\infty$ [46]. Meanwhile, R² (R-squared) measures the extent to which the model can explain the variation in the actual data. The R² value close to 1 indicates that the model fits the actual data very well. The values of these statistical criteria are calculated using Eqs. (9) through (12).

$$MAPE = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{Y_i} \right)^2} \quad (10)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

$$R = \frac{n \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n y_i \sum_{i=1}^n \hat{y}_i}{\sqrt{(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)(n \sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2)}} \quad (12)$$

where, y_i = actual value; \hat{y}_i = model value, and n = number of data.

3. RESULTS AND DISCUSSION

3.1 Optimal knot point

The selection of optimal knots in MSR directly governs the model's predictive accuracy, as knots determine the flexibility of the regression curve in capturing complex variable interactions. As illustrated in Figure 8, each optimal knot marks a critical transition threshold where the behavioral relationship between the predictors (e.g., evapotranspiration and residential area) and river discharge shifts significantly. Visually, Figure 8 demonstrates the representative optimal knot plots for Order-1, 2, and 3 using evapotranspiration and residential land as examples, which are mathematically identified by minimizing both MSE and GCV criteria. A similar optimization mechanism was systematically applied to determine the optimal knots for all other predictors, with the comprehensive knot configuration detailed in Table 3.

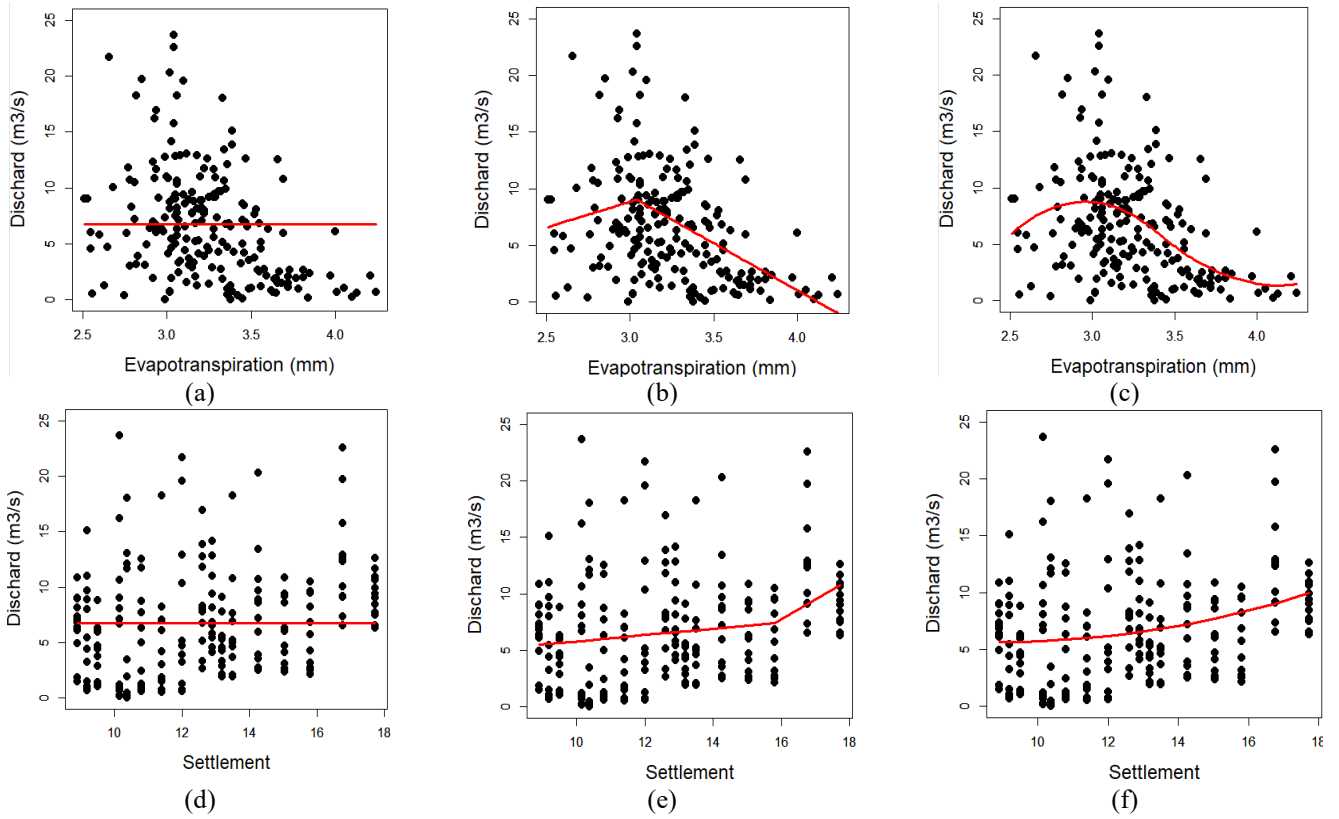


Figure 8. Representative optimal knot plots for Order-1, 2, and 3

Note: Optimal knots for evapotranspiration are shown in (a) Order-1, (b) Order-2, (c) Order-3; and for settlement in (d) Order-1, (e) Order-2, (f) Order-3

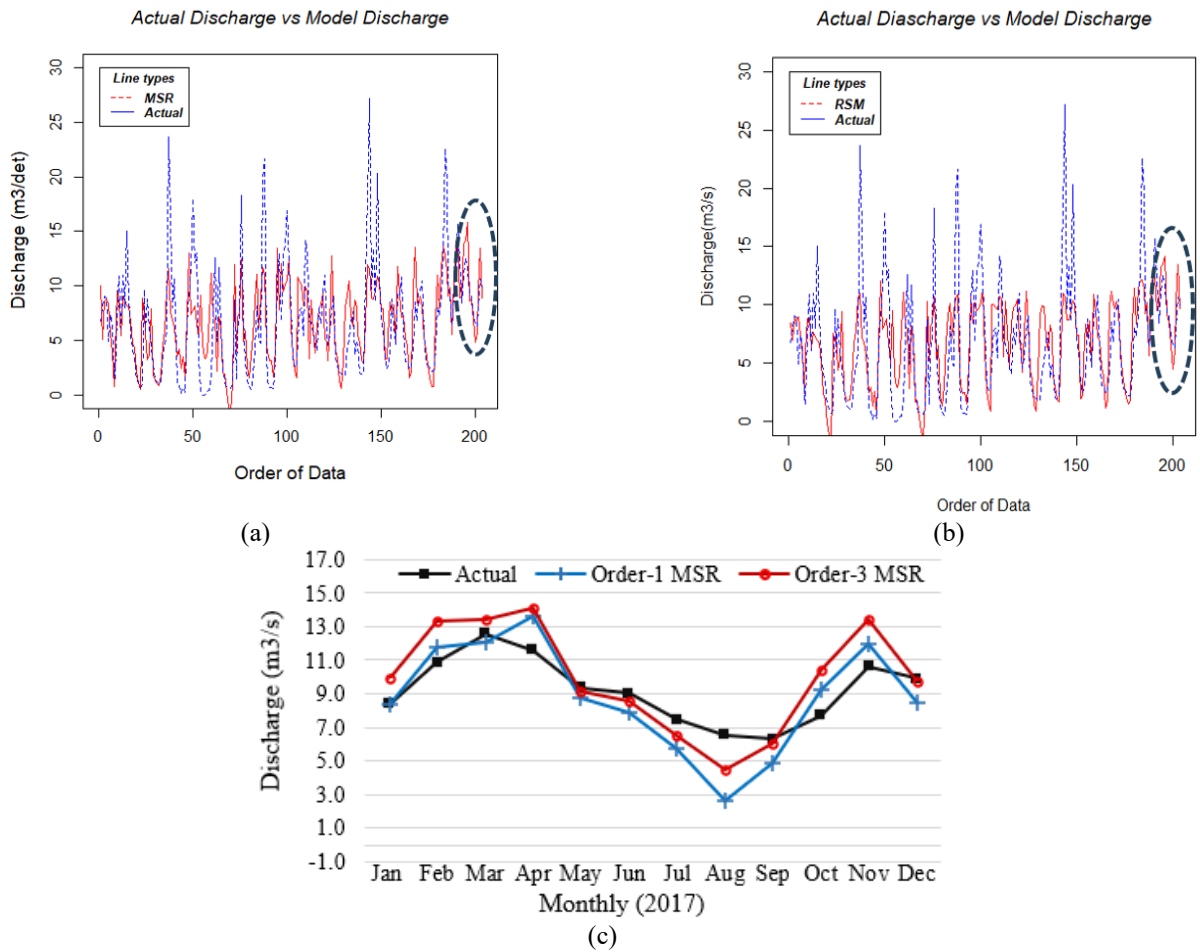


Figure 9. Plot of actual and Multivariable Spline Regression (MSR) model discharge: (a) order-1 model (2001-2017); (b) order-3 model (2001-2017); (c) monthly (2017)

Table 3. Summary of optimal knot for spline order

Variable	Order-1			Order-2			Order-3		
	Knot	GCV	MSE	Knot	GCV	MSE	Knot	GCV	MSE
X1(Evapotranspiration)	3.17	26.07	25.81	3.04	21.76	21.13	3.40	21.93	21.08
X2 (Rainfall)	536.64	26.07	25.81	338.84	15.69	15.23	116.22	17.75	15.14
X3 (Forest)	58.43	26.07	25.81	58.43	24.62	23.90	58.13	24.67	23.72
X4 (Agriculture)	130.73	26.07	25.81	127.70	24.72	24.00	126.57	24.85	24.12
X5 (Settlement)	9.20	26.07	25.81	15.80	24.80	24.08	17.73	24.98	24.26
X6 (Shrubs)	2.50	26.07	25.81	0.70	25.00	24.27	0.70	24.87	23.91
X7 (Open land)	2.30	26.07	25.81	1.70	25.47	24.73	1.80	26.09	25.08

Table 4. Goodness-of-fit criterion value

MSR	MAPE (%)	MSE	RMSE	R ²
Order 1	16.74	3.89	1.97	0.85
Order 2	34.28	11.09	3.33	0.81
Order 3	15.61	2.96	1.72	0.81

The best MSR model is determined based on the model goodness criteria values, including MAPE, MSE, RMSE, and the R² (R-squared) between the actual data and predictions. Table 4 indicates that the third-order model yields the lowest MAPE (15.61%), although its difference from the first-order model (16.74%) is relatively marginal. Considering the principle of parsimony, the first-order model was selected as it provides the best balance among predictive accuracy, model simplicity, and ease of interpretation. A visual comparison of these two models is presented in Figure 9.

3.2 Multivariable Spline Regression discharge model

The first-order MSR model shows an R² of 0.85. This indicates that approximately 85% of the river discharge variation at the Citarum-Majalaya PDA can be explained by the relationship between river discharge and the other variables used in the model. These variables include evapotranspiration (X₁), rainfall (X₂), and land cover dynamics, which consist of forest (X₃), agricultural land (X₄), built-up areas or settlements (X₅), shrubland (X₆), and open land (X₇). Meanwhile, the remaining 15% of the variation is influenced by other factors not included in the model, such as local hydrological characteristics, soil conditions, and dynamic anthropogenic factors.

Based on the coefficient values and optimal spline knot points presented in Table 3, and referring to Eq. (3), the formulation of the first-order MSR model for the Citarum-Majalaya river discharge is expressed in Eq. (13). A comparison between the actual discharge data and the model results is visualized in Figure 9.

$$\hat{Y} = -2456.05 - 2.13X_1 - 1.21(X_1 - k_1)_+ + 0.02X_2 - 0.06(X_2 - k_2)_+ + 12.17X_3 - 8.19(X_3 - k_3)_+ + 44.27X_4 - 35.71(X_4 - k_4)_+ - 455.6X_5 + 489.6(X_5 - k_5)_+ + 27.21X_6 + 11.54(X_6 - k_6)_+ + 29.82X_7 - 248.34(X_7 - k_7)_+ \quad (13)$$

where, X_i = i-th variable; k_i = the optimal knot point of the i-th variable, and

$$(X_i - k_i)_+ = \begin{cases} 0, & \text{if } X_i \leq k_i \\ (X_i - k_i), & \text{for other values of } X_i \end{cases}$$

Figure 9 illustrates a comparison between the MSR-predicted discharge and the observed discharge for the year

2017. Overall, the model adequately captures the discharge fluctuations and trends, as evidenced by the high correlation between the predicted and observed data. Nevertheless, the model does not fully reproduce extreme values observed in the dataset. The figure also indicates that during the dry season, particularly in August and September, the modeled discharge tends to be lower than the observed discharge. This discrepancy is presumed to be influenced by return flow contributions originating from domestic and industrial water use in the Majalaya area and its surroundings, which are not explicitly represented in the model configuration.

Figure 9 presents the comparison between the predicted monthly river discharge using the MSR model and the observed discharge for the year 2017. Overall, the model captures the general trends and fluctuations of the observed discharge, as supported by the high correlation between the predicted and observed values. However, the model does not fully reproduce extreme values, particularly during the dry season in August and September, when predicted flows are consistently lower than observed flows. This discrepancy is likely influenced by local return flows originating from domestic and industrial water use in Majalaya and surrounding areas, which are not explicitly represented in the model configuration. Recognizing this limitation highlights the importance of considering anthropogenic contributions when interpreting low-flow conditions and suggests potential areas for model refinement in future studies.

3.3 Parameter significance testing and model performance evaluation

The simultaneous test for the MSR model parameters for the Citarum-Majalaya river discharge at a significance level (α) = 0.05 resulted in an $F_{calculated} = 14,37$ and $F_{table} = 0.395$. These results indicate that all independent variables simultaneously have a significant effect on the variation in Citarum-Majalaya river discharge. Meanwhile, the results of the partial significance test using the t-test showed that rainfall had a t-statistic value of 9.84 and a p-value of 0.00, indicating that it was the most significant variable influencing river discharge variation. This was followed by residential land use, which yielded a t-statistic value of -2.08 and a p-value of 0.04. The remaining variables did not provide sufficient statistical evidence to indicate a significant effect on river discharge. The SE values, t-statistics, and p-values for each variable are presented in Table 5.

Furthermore, to assess the contribution of each independent variable to the variation in river discharge, the model coefficients were converted into standardized coefficients. Standardized coefficients were obtained by multiplying the model coefficient values by the ratio between the standard deviation of the independent variables and the standard deviation of the response variable (river discharge). Based on

the magnitude of the standardized coefficients, it is found that settlement land cover, with the largest standardized coefficient value (248.09), has the strongest influence on the variation in river discharge, followed in order by agricultural land (10.7), shrubland (4.44), open land (1.9), and forest (1.37). The dominance of the effect of settlement land cover is due to the relatively rapid rate of land use change, while the influence of forest land is relatively small because its area tends to remain stable or show no significant changes during the observation period.

Table 5. Partial parameter significance test results

Variable	SE	$\hat{\beta}$	t	p	Notes
X1(ET)	1.97	-2.13	-1.079	0.28	not sig.
X2 (Rainfall)	0.00	0.02	9.837	0.00	sig..
X3 (Forest)	27.01	12.17	0.451	0.65	not sig.
X4 (Agriculture)	30.01	44.26	1.475	0.14	not sig.
X5 (Settlement)	219.16	455.6	-2.079	0.04	sig.
X6 (Shrubs)	25.36	27.21	1.073	0.28	not sig.
X7 (Open land)	26.32	29.82	1.133	0.26	not sig.

The model performance test results, with a MAPE value of 20.6%, an RMSE of 2.55, and an R-squared (R^2) coefficient of 61.7%, indicate that the model is capable of representing the relationship between each predictor variable (land cover, rainfall, and evapotranspiration) and the variation in river discharge as the response variable. The model demonstrates predictive ability that can be categorized as good in depicting

the relationship between hydrometeorological factors and river discharge.

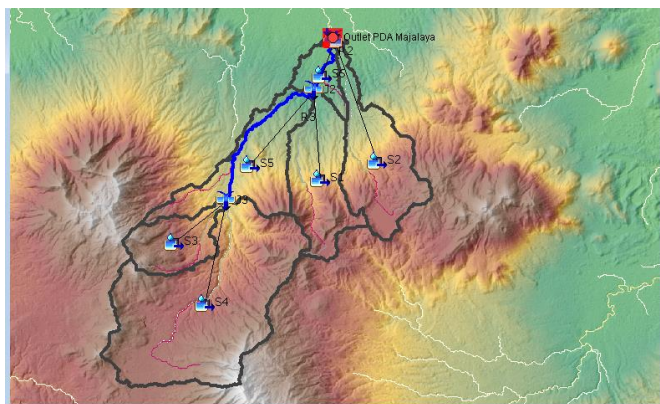
3.4 Modeling with Hydrologic Engineering Center-Hydrologic Modeling System

In comparison with the statistical approach, hydrological modeling using HEC-HMS is conducted by utilizing time series data for discharge, rainfall, and evapotranspiration in a daily format (Figures 6(a) and (b)). The influence of land cover factors on the catchment is reflected through the CN values. The catchment boundary delineation is based on DEMNAS data downloaded from the official website of the Geospatial Information Agency (BIG). The delineation process results in seven sub-catchments and three river reaches, as shown in Figure 10(a). Parameter estimation for each component of the HEC-HMS model is determined through calibration, as presented in Table 6. Meanwhile, the daily and monthly river discharge results from the HEC-HMS model are displayed in Figures 10(b) and (c).

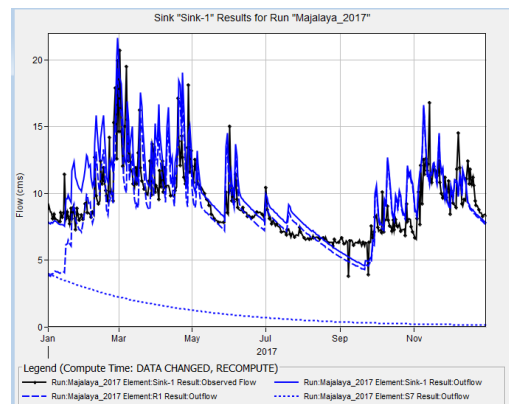
The HEC-HMS model was iteratively calibrated by adjusting Manning’s roughness and Muskingum-Cunge routing parameters through a trial-and-error approach. The CN was derived from a composite analysis of land cover and HSG, while the initial abstraction (Ia) was fine-tuned to optimize model performance. The calibrated model achieved a Nash–Sutcliffe Efficiency (NSE) of 0.46, demonstrating a satisfactory ability to reproduce observed streamflow patterns in the catchment.

Table 6. Calibration parameters of the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) model

Methods	Parameters	Values
SCS Curve Number	Initial Abstraction (mm)	4.6
	Curve Number (CN)	79.5
SCS Unit Hydrograph	Lag time (minute) pada 7 sub-basin (S):	492.1 (S1); 579.8 (S2); 354.7(S3); 557.4 (S4); 548.7 (S5); 315.9 (S6); 34.4 (S7)
Recession	Initial discharge (m ³ /s)	4
	Recession constant	0.99
Muskingum-Cunge	Threshold flow (m ³ /s)	9
	River geometry:	
	• Long (m): Reach-1; Reach-2, and Reach-3	11162.71; 4442.86; and 127.28
	• Slope (mm): Reach-1; Reach-2, and Reach-3	0.0271; 0.0105; and 0.00176
Manning Values	• wide:	7
	• Side slope:	1
Percolation	Manning Values	0.035
	Index flow (m ³ /s)	6.43
	rate (m ³ /s/1000 m ²)	8.43



(a)



(b)

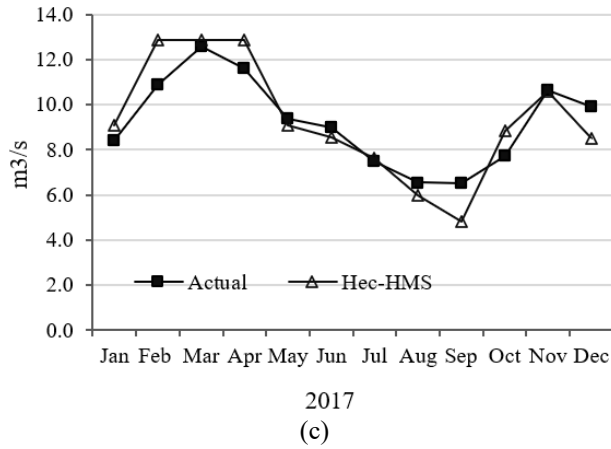


Figure 10. (a) Catchment delineation; (b) daily HEC-HMS model calibration (2017); (c) monthly HEC-HMS model calibration (2017)

HEC: Hydrologic Engineering Center-Hydrologic Modeling System

As part of the calibration phase, the rainfall loss component was modeled using the SCS-CN scheme, resulting in an initial abstraction (Ia) of 4.6 mm and a CN of 79.5, reflecting a relatively small initial interception/infiltration capacity under the studied land cover conditions. The rainfall-runoff transformation was performed using the SCS Unit Hydrograph with lag times varying across sub-catchments, ranging from 34.4 to 579.8 minutes (average ± 411.9 minutes), indicating the heterogeneity of morphometric characteristics and slopes across the catchment units. The baseflow component was calibrated using a recession model, providing an initial discharge of 4 m³/s, a recession constant of $k = 0.99$, indicating a slow baseflow, and a threshold flow of 9 m³/s. River flow routing was represented by the Muskingum-Cunge scheme with calibrated channel geometry: reach lengths of 11.16 km, 4.44 km, and 0.13 km; slopes of 0.0271, 0.0105, and 0.00176; an effective width of 7 m; bank slope of 1H:1V; and Manning's roughness coefficient of 0.035. Meanwhile, the flow index was recorded at 6.43 m³/s, and the percolation rate was 8.43 m³/s/1000 m². Overall, the combination of these parameters resulted in a good reproduction of the hydrograph at the daily and monthly aggregate scales (see Figure 10(b) and (c)), with a summary of the values presented in Table 5. Further evaluation was conducted during the validation phase using 2018 rainfall data.

Figure 10(b) and (c) show that the pattern and magnitude of river discharge produced by the HEC-HMS model are in good agreement with the actual observed data. However, during the dry season, particularly in August, the model's forecasted discharge values are lower than the observed data. This discrepancy is suspected to be caused by the contribution of return flow from domestic and industrial water usage in the Majalaya region and its surroundings.

3.5 Validation and comparison of the model

The validation of the MSR and HEC-HMS models was conducted using independent data that were not involved in the model development phase, thus reflecting the models' ability to generalize to new data. Performance was evaluated using two main metrics: MAPE and RMSE. Based on MAPE, performance categories are defined as follows: excellent ($\leq 10\%$); good ($>10 - \leq 30\%$); fair ($>30 - \leq 50\%$); and poor ($>50\%$). Meanwhile, the interpretation of RMSE is monotonically decreasing, where smaller values (closer to

zero) indicate better performance, as they represent the magnitude of the average error in discharge units. The validation results obtained using this evaluation framework directly reflect the model's performance when applied to data that was never "seen" during the modeling process.

The validation results show that the HEC-HMS model has a MAPE of 11.8% and an RMSE of 1.17, while the MSR model has a MAPE of 17.5% and an RMSE of 2.20. Both models fall into the "good" performance category (based on $MAPE \leq 30\%$), but the HEC-HMS model performs better quantitatively due to its smaller error values. Visually, both models replicate the pattern and trend of observed discharge well, and they exhibit similar discharge patterns and trends. More noticeable deviations from the discharge values of both models are observed in March, when the simulation pattern diverges from the actual data.

In terms of data requirements, the MSR is simpler, relying mainly on a few secondary time series (rainfall, river discharge, evapotranspiration, and land cover dynamics), whereas HEC-HMS requires a larger set of physical data and parameters. Therefore, although HEC-HMS demonstrates better performance than MSR, the MSR remains a viable alternative, offering an efficient option for monthly discharge forecasts up to one year ahead. A comparison of the forecast curves from both models against actual data is presented in Figure 11.

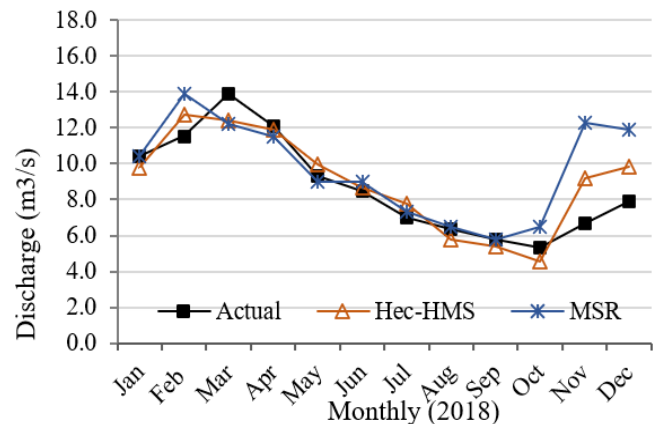


Figure 11. Plot of actual discharge and model predicted discharge

4. CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

The MSR river discharge model, developed using land cover, precipitation, and evapotranspiration data, demonstrated strong performance during validation and effectively predicted both the magnitude and patterns of monthly river discharge up to 12 months in advance. Significance analysis indicated that these three variables collectively influenced discharge variability, while standardized coefficient analysis revealed that the conversion of land to residential areas was the dominant factor increasing surface runoff, whereas changes in forest cover contributed minimally. Comparison with the HEC-HMS model showed that although HEC-HMS outperformed MSR quantitatively, MSR remained reliable and more efficient, as it requires only secondary time series data and captures hydrological dynamics driven by interactions between climatic factors and land use. Scientifically, these findings confirm that land use change is a primary control on river discharge dynamics and highlight the necessity of integrating both biophysical and land-use aspects in hydrological modeling. Practically, the results of this study have implications for water resource management, including water allocation planning, reservoir and dam operation, and the mitigation of hydrological risks such as floods and droughts. Monthly streamflow predictions can support timely and effective decision-making, thereby enhancing the practical relevance of the study.

4.2 Recommendations

Future research should aim to refine the model by incorporating more detailed vegetation cover variables and expanding the spatial and temporal scope of the data to enhance the accuracy and stability of river discharge forecasting in response to climate variation. Additionally, this method should be tested on catchments with different scales and land cover complexities to assess the model's reliability consistency and its sensitivity to the diversity of biophysical characteristics.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the the Faculty of Civil and Environmental Engineering (FTSL) at Institut Teknologi Bandung (ITB for their support and approval in conducting this research. Special thanks are also extended to the Ministry of Environment and Forestry of Indonesia for providing land cover data, and to the Citarum River Basin Organization and the Hydrology and Water Environment Institute for supplying river discharge and rainfall data, which greatly contributed to the smooth execution of this.

REFERENCES

[1] Wang, H. (2019). Investigating the relationship between hydrological variation, land use/cover change and climate change at regional and local scales under future scenarios. University of Connecticut. Doctoral Dissertations.

<https://digitalcommons.lib.uconn.edu/dissertations/2198>.

[2] Sharma, P., Machiwal, D. (2021). Advances in Streamflow Forecasting - From Traditional to Modern Approaches. Elsevier. <https://doi.org/10.1016/C2019-0-02163-2>

[3] Yifru, B.A., Lim, K.J., Lee, S. (2024). Enhancing streamflow prediction physically consistently using process-based modeling and domain knowledge: A review. *Sustainability*, 16(4): 1376. <https://doi.org/10.3390/su16041376>

[4] Yaseen, Z.M. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Engineering and Sciences*, 4(3): 65-103. <https://doi.org/10.51526/kbes.2023.4.3.65-103>

[5] Ragno, E., Hrachowitz, M., Morales-Nápoles, O. (2022). Applying non-parametric Bayesian networks to estimate maximum daily river discharge: Potential and challenges. *Hydrology and Earth System Sciences*, 26(6): 1695-1711. <https://doi.org/10.5194/hess-26-1695-2022>

[6] Elmi, O., Tourian, M.J., Bárdossy, A., Sneeuw, N. (2021). Spaceborne river discharge from a nonparametric stochastic quantile mapping function. *Water Resources Research*, 57(12): e2021WR030277. <https://doi.org/10.1029/2021WR030277>

[7] Quintela-del-Río, A., Francisco-Fernández, M. (2018). River flow modelling using nonparametric functional data analysis. *Journal of Flood Risk Management*, 11: S902-S915. <https://doi.org/10.1111/jfr3.12282>

[8] Kiarie, F., Mwita, P.N. (2016). Estimation of critical streamflow discharge level using nonparametric quantile regression model. <http://ir.mksu.ac.ke/handle/123456780/4746>.

[9] Liu, J.M. (2009). Nonlinear forecasting using nonparametric transfer function models. *Wseas Transactions On Business and Economics*, 6(5). <https://scholars.georgiasouthern.edu/ws/portalfiles/portal/49818914/Jun%20Liu%20Nonlinear%20Forecasting.pdf>.

[10] Wong, H., Ip, W.C., Zhang, R., Xia, J. (2007). Non-parametric time series models for hydrological forecasting. *Journal of Hydrology*, 332(3-4): 337-347. <https://doi.org/10.1016/j.jhydrol.2006.07.013>

[11] Khazaee Poul, A., Shourian, M., Ebrahimi, H. (2019). A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resources Management*, 33(8): 2907-2923. <https://doi.org/10.1007/s11269-019-02273-0>

[12] Burnama, N.S., Rohmat, F.I.W., Farid, M., Kuntoro, A.A., Kardhana, H., Rohmat, F.I.W., Wijayasari, W. (2023). The utilization of satellite data and machine learning for predicting the inundation height in the Majalaya watershed. *Water*, 15(17): 3026. <https://doi.org/10.3390/w15173026>

[13] Uddin, M.E., Akter, S.A., Uddin, M.J., Diganta, M.T.M. (2017). Trend analysis variations and relation between discharge and rainfall: A study on Kushiyara River. *Journal of Environmental Science and Natural Resources*, 10(2): 121-132. <https://doi.org/10.3329/jesnr.v10i2.39025>

[14] Papacharalampous, G.A., Tyrallis, H. (2018). Evaluation of random forests and Prophet for daily streamflow forecasting. *Advances in Geosciences*, 45: 201-208.

- <https://doi.org/10.5194/adgeo-45-201-2018>
- [15] Sun, A.Y., Wang, D., Xu, X. (2014). Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology*, 511: 72-81. <https://doi.org/10.1016/j.jhydrol.2014.01.023>
- [16] Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing*. CRC Press. <https://doi.org/10.1201/9781482273144>
- [17] Afa, I.B., Suparti, S., Rahmawati, R. (2018). Perbandingan metode regresi linier multivariabel dan regresi spline multivariabel dalam pemodelan indeks harga saham gabungan. *Media Statistika*, 11(2): 147-158. <https://doi.org/10.14710/medstat.11.2.147-158>
- [18] Ruhiat, D., Soekarno, I., Kardhana, H., Suwarman, R. (2025). Projected Curve Number (CN) changes and surface runoff response to corrected land cover dynamics in the Citarum-Majalaya catchment. *International Journal of Design & Nature and Ecodynamics*, 20(4): 761-775. <https://doi.org/10.18280/ijdne.200407>
- [19] Nugroho, J., Qolbi, S.A.N., Adiprayoga, M.F., Wulandari, S., Soekarno, I., Kuntoro, A.A., Kusuma, M.S.B., Wijayasari, W., Rohmat, F.I.W. (2025). Post-normalization sedimentation in the Citarum Majalaya River: Patterns, impacts, and mitigation strategies. *Results in Engineering*, 27: 106943. <https://doi.org/10.1016/j.rineng.2025.106943>
- [20] Kardhana, H., Lihawa, A.W.I., Rohmat, F.I.W., Wulandari, S., Harjupa, W., Adiprawita, W., Kardena, E., Kusuma, M.S.B. (2024). Water balance analysis in the majalaya watershed: Two-step calibration and application of the SWAT+ model for low-flow conditions. *Water*, 16(23): 3498. <https://doi.org/10.3390/w16233498>
- [21] Yusuf, A., Kusumastuti, D.I., Wahono, E.P. (2021). The effect of land cover on the base flow index of the Seputih Way Watershed, Lampung Province. *Siklus: Jurnal Teknik Sipil*, 7(2): 146-159. <https://doi.org/10.31849/siklus.v7i2.7323>
- [22] Koneti, S., Sunkara, S.L., Roy, P.S. (2018). Hydrological modeling with respect to impact of land-use and land-cover change on the runoff dynamics in Godavari River Basin using the HEC-HMS model. *ISPRS International Journal of Geo-Information*, 7(6): 206. <https://doi.org/10.3390/ijgi7060206>
- [23] Sajikumar, N., Remya, R.S. (2015). Impact of land cover and land use change on runoff characteristics. *Journal of Environmental Management*, 161: 460-468. <https://doi.org/10.1016/j.jenvman.2014.12.041>
- [24] Shaw, S.B., Marrs, J., Bhattarai, N., Quackenbush, L. (2014). Longitudinal study of the impacts of land cover change on hydrologic response in four mesoscale watersheds in New York State, USA. *Journal of Hydrology*, 519: 12-22. <https://doi.org/10.1016/j.jhydrol.2014.06.055>
- [25] Farid, M., Pratama, M.I., Kuntoro, A.A., Adityawan, M.B., Rohmat, F.I.W., Moe, I.R. (2021). Pengaruh perubahan tutupan lahan terhadap debit banjir di daerah aliran Sungai Ciliwung Hulu. *Jurnal Teknik Sipil*, 28(3): 309. <https://doi.org/10.5614/jts.2021.28.3.8>
- [26] Wisanggeni, D.H., Sitorus, J.E., Putra, K.H.P., Adityawan, M.B. (2024). The effect of land cover change on flood discharge in the Central Government Core Area (KIPP) of the capital city of the Archipelago: Indonesia. *Jurnal Teknik Sipil*, 9(2): 327-338. <https://doi.org/10.29244/jstil.9.2.327-338>
- [27] Seyam, M.M.H., Haque, M.R., Rahman, M.M. (2023). Identifying the land use land cover (LULC) changes using remote sensing and GIS approach: A case study at Bhaluka in Mymensingh, Bangladesh. *Case Studies in Chemical and Environmental Engineering*, 7: 100293. <https://doi.org/10.1016/j.cscee.2022.100293>
- [28] Handyastono, B., Alghoul, M.A., Rizki, A., Djambek, N.P., Kusuma, M.S.B., Kuntoro, A.A., Harlan, D., Nugroho, E.O., Munthe, H.M., Hazmi, M., Wisanggeni, D.H., Rizqy, M.A. (2025). Flood hazard assessment in Pemaluan Village due to land use change in IKN (Ibu Kota Nusantara) as the new capital city of Indonesia. *Case Studies in Chemical and Environmental Engineering*, 11: 101211. <https://doi.org/10.1016/j.cscee.2025.101211>
- [29] Suprapti, S., Kusuma, M.S.B., Kardhana, H., Cahyono, M. (2024). An assessment of potential infiltration areas to support groundwater supply system in Jagakarsa, South Jakarta, based on Multi-Criteria Decision-Making (MCDM) analysis. *Case Studies in Chemical and Environmental Engineering*, 10: 100799. <https://doi.org/10.1016/j.cscee.2024.100799>
- [30] Download SNI 7645:2010 tentang Klasifikasi Penutup Lahan. <https://www.indonesia-geospasial.com/2020/07/download-sni-76452010-klasifikasi.html>
- [31] Pawar, A., Lohith, H.G., Chiles Kumar, J.U., Bhargavkumar, P.G., Sindhu, D. (2021). Estimation of Evapotranspiration using CROPWAT 8.0 Model. *International Journal of Engineering Research & Technology*, 9(15): 50-54.
- [32] Singh, K., Singh, S.K., Bikramaditya, B. (2025). Implementation of FAO-CROPWAT 8.0 model for estimation of crop water requirement and planning of irrigation for Angul district, Orissa, India. *International Journal of Research in Agronomy*, 8(5): 41-46. <https://doi.org/10.33545/2618060X.2025.v8.i5Sa.2891>
- [33] US Army Corps of Engineers. (2000). *HEC-HMS Technical Reference Manual*.
- [34] Uwizeyimana, D., Mureithi, S.M., Mvuyekure, S.M., Karuku, G., Kironchi, G. (2019). Modelling surface runoff using the soil conservation service-curve number method in a drought prone agro-ecological zone in Rwanda. *International Soil and Water Conservation Research*, 7(1): 9-17. <https://doi.org/10.1016/j.iswcr.2018.12.001>
- [35] Sharma, P., Machiwal, D. (2021). Streamflow forecasting: Overview of advances in data-driven techniques. *Advances in Streamflow Forecasting*, 2021: 1-50. <https://doi.org/10.1016/B978-0-12-820673-7.00013-5>
- [36] Granata, F., Di Nunno, F. (2024). Forecasting short-and medium-term streamflow using stacked ensemble models and different meta-learners. *Stochastic Environmental Research and Risk Assessment*, 38(9): 3481-3499. <https://doi.org/10.1007/s00477-024-02760-w>
- [37] Chu, H., Wei, J., Jiang, Y. (2021). Middle-and long-term streamflow forecasting and uncertainty analysis using Lasso-DBN-Bootstrap model. *Water Resources Management*, 35(8): 2617-2632. <https://doi.org/10.1007/s11269-021-02854-y>
- [38] Widayastuti, M.T., Taufik, M. (2019). Long-term monthly discharge prediction for Cimanuk Watershed. *Agromet*, 33(2): 96-104.

- <https://doi.org/10.29244/j.agromet.33.2.96-104>
- [39] Cheng, M., Fang, F., Kinouchi, T., Navon, I.M., Pain, C.C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590: 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- [40] Syed, Z., Mahmood, P., Haider, S., Ahmad, S., Jadoon, K. Z., Farooq, R., Syed, S., Ahmad, K. (2023). Short–long-term streamflow forecasting using a coupled wavelet transform–artificial neural network (WT–ANN) model at the Gilgit River Basin, Pakistan. *Journal of Hydroinformatics*, 25(3): 881-894. <https://doi.org/10.2166/hydro.2023.161>.
- [41] Härdle, W. (1990). *Applied Nonparametric Regression* (No. 19). Cambridge University Press. <https://doi.org/10.1017/CCOL0521382483>
- [42] Budiantara, I.N. (2005). Model keluarga spline polinomial truncated dalam regresi semiparametrik. *Berkala Ilmiah MIPA*, 15(3): 241598.
- [43] Putri, T., Dwidayati, N.K. (2018). *Pemodelan regresi nonparametrik spline multivariabel*. Semarang: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
- [44] Das, R. (2014). Study of rainfall-runoff response of a catchment using SCS curve number combined with Muskingum routing techniques. *Structural and Civil Engineering*, 3(3): 81-91.
- [45] Naabil, E., Lamptey, B.L., Arnault, J., Olufayo, A., Kunstmann, H. (2017). Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *Journal of Hydrology: Regional Studies*, 12: 196-209. <https://doi.org/10.1016/j.ejrh.2017.05.010>
- [46] Lian, H., Yen, H., Huang, J.C., Feng, Q., Qin, L., Bashir, M.A., Wu, S., Zhu, A.X., Luo, J., Di, H., Lei, Q., Liu, H. (2020). CN-China: Revised runoff curve number by using rainfall-runoff events data in China. *Water Research*, 177: 115767. <https://doi.org/10.1016/j.watres.2020.115767>