Integrated Streamflow Forecasting System: A Step Towards Smart Flood Management

Irfan Ardiansah[1*], Alfonsus Mario Agung[2], Chay Asdak[2], Nurpilihan Bafdal[2], Roni Kastaman[1], Selly Harnesa Putri[1] and Desy Nurliasari Suparno[1]
[1]Department of Agro-Industrial Technology, Faculty of Agro-Industrial Technology, Universitas Padjadjaran, Indonesia
[2]Department of Agriculture Engineering and Biosystem, Faculty of Agro-Industrial Technology, Universitas Padjadjaran, Indonesia
e-mail: irfan@unpad.ac.id
[*]Coresponding author

Keywords: streamflow forecasting, regression model, precipitation data, water resource management, flood mitigation

Received: May 26, 2023

This study aims to help users manage water resources and prevent flooding by creating an online monthly streamflow forecasting system. We have integrated a regression model into the system, using historical information on rainfall and streamflow selectivity from a number of monitoring stations in the Upper Cimanuk sub-basin. Users can access the online system to input and view rainfall and streamflow data and enumerate monthly streamflow rate projections. To verify the system's forecast accuracy, we compared it with manual calculations employing the velocity-area method and field observations. The system provides reasonably accurate forecasts, as indicated by the system's high coefficient of determination (R^2) value of 0.91. Nevertheless, the differences between predictions and measurements suggest there is scope to improve the accuracy of the system by including additional variables and more comprehensive data. Future enhancements may include additional validation using a wider range of field data, as well as the inclusion of precipitation intensity, duration, catchment shape and size. The developed monthly streamflow forecasting system is a valuable tool for analyzing and forecasting streamflow rates, providing a basis for informed decision making in water resource management and flood disaster mitigation.

Povzetek: V članku je predstavljen spletni sistem za napovedovanje mesečnega pretoka rek (opozorilo za poplave) v Cimanuku z visoko natančnostjo (R^2 = 0.91).

1 Introduction

[1] defines flooding as the inundation of an area that occurs when water overflows beyond its drainage capacity, causing physical, social, and economic losses. Flooding happens when the river or canal overflows on either its right or left side because the channel capacity cannot manage the streamflow. Flooding occurs due to overflowing to the left or right side of the river/canal because the capacity of the river channel is not sufficient for the streamflow [2]. The complexity of flooding in a basin involves numerous main elements, which function both as natural physical objects or targets, and as subjects or actors utilized by humans. The elements interact and mutually influence each other, leading to the hydrological condition of the basin [3].

The Watershed Management Board of Cimanuk – Citanduy recorded flood and landslide report data, noting more than one flood event in the Upper Cimanuk Sub-basin. Flooding struck Sindangsari Village, Garut Regency on March 28th, 2014, with runoff reaching 165 cm. In 2015, floods took place between March 15 and 16 with a streamflow of 384 m^3/s. A flash flood hit Sukakarya Village, Tarogong Kidul Subdistrict, Garut Regency, on September 20, 2016, with a runoff of 50 - 200 cm, claiming three lives [4]. In terms of administration, two governmental regions divide the Upper Cimanuk sub-basin: Garut Regency, which covers the majority with 20 sub-districts, and Sumedang Regency, which covers the remaining 12 sub-basins with an area of 156,020 hectares [5].

Streamflow forecasting is used to try to anticipate flooding. Basin rivers take longer to observe precipitation observations than streamflow observations [6]. Studying the relationship between the two variables is important. We can express the relationship of hydrological variables in mathematical formulas, which can then be used for hydrological analysis, such as forecasting, extension, repair, and data filling [7].

In linear regression analysis, we examine the relationship between two or more variables. We use a linear regression model to establish the relationship between dependent and independent variables. Linear regression comes in two types: simple and multiple, depending on the number of independent variables. Multiple linear regression involves more than one independent variable, while simple linear regression involves only one [8]. In this study, we use simple linear regression, treating precipitation as the independent variable.

Considering the rise in internet users in Indonesia and the recurring floods in the Upper Cimanuk sub-basin, we plan to develop a monthly online streamflow forecasting system. This system will leverage widespread internet access to share crucial information about potential floods, thereby minimizing their adverse effects on local communities. According to the Indonesian Internet Service Providers Association survey, Indonesia has 204.7 million active internet users, making up 74.23% of the total population. Java Island hosts the highest number of internet users, comprising around 43.92% of Indonesian users [9].

The relationship between precipitation and river flow is unidirectional in both the dry and wet seasons; as precipitation increases, so does river flow, and vice versa. We plan to build an online monthly streamflow forecasting system, considering the growing number of Internet users and the flooding in the Upper Cimanuk subbasin in Indonesia. We will implement a linear regression model to forecast streamflow, using precipitation as the independent variable and streamflow as the dependent variable. The system will analyze both variables to predict monthly streamflow. Given the substantial number of internet users in Indonesia, we believe this online system will be accessible to everyone who needs it and will help reduce flooding.

The purpose of the study is to develop a web-based monthly streamflow forecasting system for the Upper Cimanuk sub-basin that users can access online to analyze and forecast streamflow. The system calculates streamflow forecasts using rainfall and streamflow information and compares the precision of its forecasts to manual calculations. The purpose of this study is to provide a tool for analyzing and predicting streamflow to serve as a foundation for decision-making in water resource management and flood disaster prevention. This research includes only the Upper Cimanuk sub-basin and focuses on developing and evaluating an online application that offers users access to precipitation and streamflow data and allows them to forecast streamflow based on the provided rainfall input.

The findings of this study will grant convenient access to information on streamflow and precipitation for users in the Upper Cimanuk sub-watershed, benefiting planning and growth activities in the area.

2 Related works

In recent years, there has been some interest in the field of water resource management and flood disaster mitigation. By integrating advanced modeling techniques and data-driven approaches, these systems have paved the way for flood management strategies. Streamflow forecasting stands as water resource management and flood disaster mitigation tool. An array of methodologies has been cultivated for streamflow forecasting, spanning statistical methods, hydrological models, and machine learning techniques. This chapter offers an overview of researches within the integrated streamflow forecasting systems domain.

The exploration into streamflow prediction has led to a series of investigations. Regression models have been a prominent focus, revealing their potential to forecast streamflow dynamics. [10] study investigates the marginal advantage of a different methods using initial hydrologic conditions (IHC), focusing on seasonal water supply forecasts (WSF) with case studies on five watersheds located in the US Pacific Northwest region. The researchers found that climate information can increase the reliability of forecasts from IHC, but strict control over sample size must be observed to avoid overtrained forecast solutions.

The rise of online platforms has significantly impacted water resource management. [11] study introduces a serious gaming framework to assist stakeholders in the decision-making process for water resources scheming and disaster mitigation. The framework includes a Multi-Hazard Tournament (MHT) and a web-based decision support tool. The framework was evaluated in a case study and found to be effective in increasing collective understanding and awareness of water-related hazards and mitigation strategies.

The validation of forecasting systems plays a crucial role in establishing their credibility. [12] introduced a machine learning model that can be used to predict drought events in the eastern Mediterranean. The bagging algorithm was the most accurate in the training stage, but the bagging and random forest algorithms were more dynamic in drought capturing. The results of the research can help decision-makers with drought mitigation plans.

Spatial and temporal factors emerge as pivotal determinants of forecast accuracy. [13] developed a distributed hydro-meteorological forecasting approach to provide information at unexplored sites. The system was validated with respect to actual road inundations and the results are promising. The system could be used to identify areas at risk and adopt appropriate safety and rescue measures.

The incorporation of additional variables has emerged as a promising avenue. [14] built a method for disaggregating daily rainfall observations into hourly rainfall. The method was applied in Singapore and was found to produce intensity–duration–frequency curves with significantly improved accuracy.

One contemporary example of a forecast system model is presented by [15]. By employing a cross-disciplinary collaboration between life scientists and expert users of Earth system models will greatly enhance the likelihood of developing robust evidence to address climate change challenges. This will make caveats more explicit and place decisions regarding potential tradeoffs in the hands of the user.

Another notable example of how the use of big data and machine learning technologies has the potential to impact many facets of environmental and water management comes from the following research by [16]. They found that big data and machine learning have the potential and benefits to enable data-driven research in environmental and water management, provide an overview of key concepts and approaches in big data and machine learning, and discuss key issues and challenges.

Table 1 provides a summarized overview of key findings from various literature reviews concerning streamflow forecasting systems, encompassing those utilizing regression models and online platforms. The compilation underscores that while these state-of-the-art methodologies offer valuable insights, they often exhibit limitations in terms of accuracy and suitability for specific hydrological contexts.

The significance of our work becomes apparent within the context of streamflow forecasting, owing to the following attributes:

- Integrated Approach: Our study adopts an integrated methodology, amalgamating multiple techniques to enhance accuracy and resilience.
- Online Accessibility: Implementation through an online platform expands accessibility to a diverse user base.
- Real-world Validation: Evaluation using historical data from a practical watershed affirms the practicality of our approach.

The outcomes of this endeavor suggest that our proposed system holds potential to furnish precise and trustworthy streamflow forecasts for an array of applications. Its implications extend to refining water resource management, mitigating flood disasters, and safeguarding lives and assets against inundation risks.

3 Research methodology

In this study we used (1) 10 years of precipitation data as the source of the database, collected from the Watershed Management Office West Java Region, (2) 10 years of streamflow data from the Watershed Management Office West Java Region to form the basis for streamflow forecasting, (3) a 1:50,000 scale map of the Upper Cimanuk Sub-basin, obtained from the Cimanuk -

Table 1: Summary of: Related Works

Author	Results	Advantages	Disadvantages
[10]	Investigated seasonal water supply forecasts in five US Pacific Northwest watersheds. Found that climate information enhances forecast skill but cautioned against over-trained solutions.	Climate info improves seasonal forecasts Insights into handling sample size limitations	Limited to specific regions Risk of over-training
[11]	Introduced a serious gaming framework for water resource planning and hazard mitigation. Evaluated effectiveness in a case study, enhancing awareness and understanding of hazards and mitigation strategies.	Engages stakeholders through serious gaming Enhances collective awareness	Specific to decision-making contexts, May require technological infrastructure
[12]	Introduced a machine learning model for drought prediction in the eastern Mediterranean. Bagging and random forest algorithms were dynamic in drought capturing.	Machine learning for dynamic drought prediction Insights for mitigation planning	Focus on drought prediction, Algorithm complexity
[13]	Developed a distributed hydro-meteorological forecasting approach to identify ungauged sites at risk of road inundation. Promising results for safety and rescue measures.	Provides info for ungauged sites Identifies at-risk areas for inundation	Specific to road inundation, May require data infrastructure
[14]	Built a method to disaggregate daily rainfall into hourly observations. Applied in Singapore, producing improved accuracy in intensity–duration–frequency curves.	Improved accuracy in rainfall curves Valuable for hydrological modeling	Focus on rainfall disaggregation, Regional applicability
[15]	Emphasized cross-disciplinary collaboration for robust climate change evidence. Advocated for informed decision-making by users through explicit caveats and tradeoff considerations.	Integrates Earth system models and expert users, Enhances user decision-making	Reliant on interdisciplinary collaboration, Potential for complex communication
[16]	Highlighted big data and machine learning's potential in environmental and water management research. Discussed concepts, approaches, benefits, and challenges.	Enabling data-driven research, Overview of key concepts and approaches	General overview, Lack of specific applications

Citanduy Watershed Management Board to create Thiessen polygons, and (4) Flood and Landslide Reports from the Cimanuk - Citanduy Watershed Management Board to determine the minimum streamflow required during flood events.

The deliberate selection of the Upper Cimanuk sub-basin as the focal point of our study is a result of careful deliberation guided by multifaceted considerations. Acknowledging the sub-basin's proclivity towards flood occurrences due to its heterogeneous topography, land use patterns, and hydrological attributes, our deliberate choice is anchored in the intention to confront the tangible challenges entwined with flood management and optimization of water resources. This regional spotlight provides an impeccable crucible for evaluating the efficacy of our integrated streamflow forecasting system amidst the complexities and dynamism of the environment.

The Upper Cimanuk sub-basin offers a manifold of advantages stemming from its hydrological diversity, rendering it an invaluable paradigm for comprehending the dynamics of our system's performance across an array of terrains and land utilization. Simultaneously, the region's susceptibility to floods accentuates the criticality of precise streamflow forecasts as a bedrock for potent flood disaster mitigation strategies.

Our methodology involved the comprehensive development and evaluation of a monthly streamflow forecasting system tailored specifically for the Upper Cimanuk sub-basin. At the heart of our Integrated Streamflow Forecasting System lies the intricate framework of the regression model. Our methodology seamlessly integrates a straightforward linear regression model, meticulously combining historical rainfall and streamflow data gathered from a diverse array of monitoring stations scattered throughout the expansive Upper Cimanuk sub-basin. This particular model is thoughtfully crafted to illuminate the complex interplay between rainfall patterns and the resulting streamflow rates unique to this region.

To establish the relationship between rainfall and flow data, we incorporated a simple linear regression model into our system design process. This model helps to understand and describe the relationship between rainfall and river flow dynamics. To make things easy and accessible for users, we've developed this application using web-based technology. This lets you connect to the system from anywhere by simply using the internet.

To ensure the system is functioning accurately, we execute manual computations employing the area velocity technique as a standard for precision in forecasting the river's flow. The velocity-area approach is a customary means of approximating the velocity of the river's current by gauging the dimensions of the river's cross-section and determining the speed of the water at various locations. We subdivided the length of the river into ten segments and gauged the swiftness of the water at particular depths (at 20% of the complete depth and at 80% of the entire depth) [17]. This technique depends on the principles of fluid dynamics to provide us with an estimate of the speed at which the fluid is traveling.

The velocity-area approach is a frequently employed and comparatively straightforward method to employ. Nevertheless, its precision can be affected by numerous factors such as disparities in flow velocity at distinct segments of the river, turbulence, and errors committed by individuals. To guarantee the tool functions correctly in measuring river flow, we perform a form of testing known as black box testing. This approach aids us in verifying that every facet of the tool operates precisely and in accordance with user requirements [18]. We employ black box testing to assess: user registration pages, sign-in pages, user-submitted flow forecast pages, and flow prediction pages.

Application of the approach entails statistical analysis, software development, and system assessment through juxtaposition with manual computations, field inspection tasks, and black box testing. The objective is to furnish a user-friendly monthly river flow prediction system for planning and development endeavors in the Cimanuk Hulu sub-basin.

3.1 Data collection

In this study, we utilized a quantitative research method. By obtaining secondary data from the Cimanuk - Citanduy Watershed Management Board, the Upper Cimanuk Sub Basin map, and a Flood and Landslide Report, which includes 10 years of rainfall data and water flow data.

3.2 Determination of simple linear regression equation

The coefficient of determination of a simple linear regression equation is used to study the relationship between two variables, the independent variable (X) and the dependent variable (y). This allows researchers to develop a linear mathematical model to predict the value of the dependent variable, Y, based on the value of the independent variable, X [8], [19].

We are using rainfall as the independent variable (X), while flow is representing the dependent variable (Y). We develop the model to investigate the relationship between rainfall and runoff and predict runoff [20].

The steps to determine the simple linear regression equation include:

1. Obtaining precipitation (X) and streamflow (y) data from relevant sources, such as the West Java Provincial Watershed Management Office.
2. Calculating mean values for the independent variable (X) and dependent variable (y).
3. Measuring the correlation coefficient (R) between the independent variable (X) and dependent variable (y) to determine the strength of the linear relationship between the two variables.

$$R = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left\{n\sum X^2 - \left(\sum X\right)^2\right\}\left\{n\sum Y^2 - \left(\sum Y\right)^2\right\}}}$$

We calculate regression coefficients (a and b) using a predetermined formula. Coefficient a is a constant, while coefficient b is the slope of the regression line [19].

$$a = R\left(\frac{\sigma_y}{\sigma_x}\right)$$

$$b = R\left(\frac{\sigma_x}{\sigma_y}\right)$$

We can construct a simple linear regression equation with the calculated regression coefficients (a and b) in the form of:

$$y = a + bX$$

By evaluating the fit of the regression model, we can calculate the coefficient of determination (R^2). The value of R^2 ranges from 0 to 1, where a higher value indicates a better model in explaining the variation in the data. After forming the simple linear regression equation, we validate and interpret the resulting model. We will use the formulated regression model to forecast streamflow based on observed precipitation and aid in water resources planning and management in the Upper Cimanuk sub-basin region.

3.3 Determining the thiessen polygon value

We use the Polygon Thiessen technique in spatial analysis to estimate variables in a region based on values observed at specific measurement points [21]. In this research, we employ Thiessen polygons to ascertain the scope of the Upper Cimanuk sub-basin by considering the location of weather stations in the region.

We split the Cimanuk Hulu sub-basin into sections based on the position of the weather stations using ArcGIS software. This process entails developing Thiessen polygons, where each weather station point becomes the nucleus of the polygon. Every Thiessen polygon represents a region nearest to its core weather station.

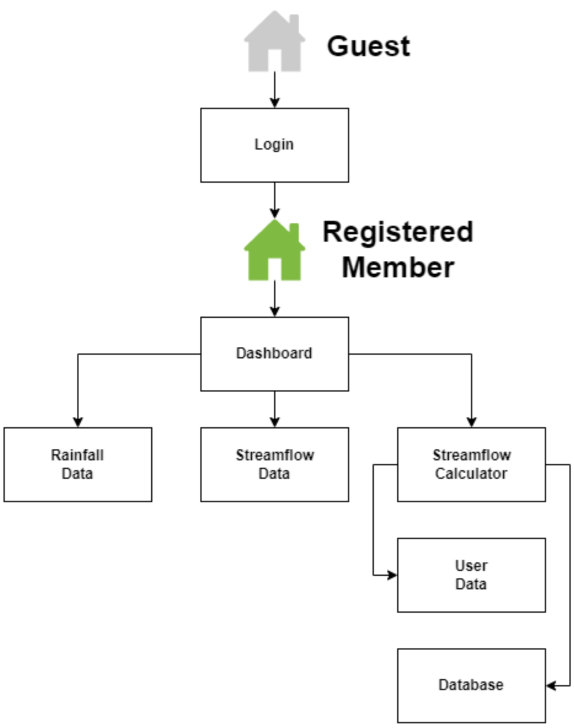


Figure 1: Site map of streamflow forecast system

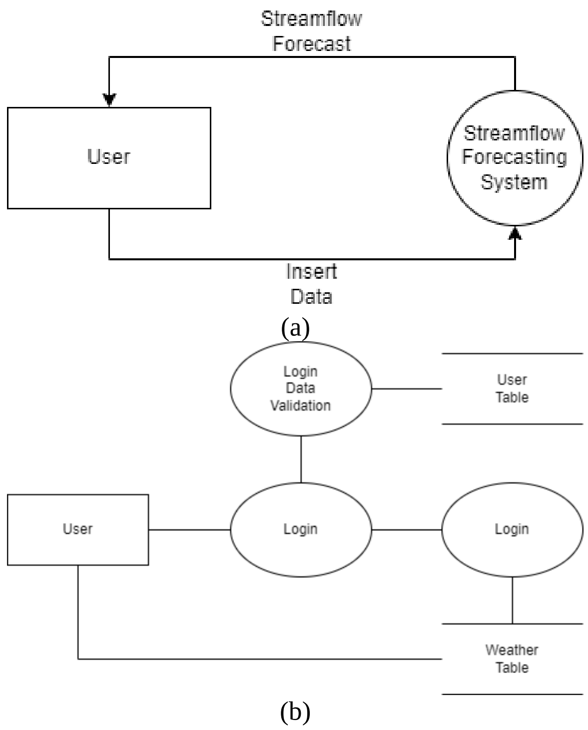


Figure 2: (a) Data context diagram (b) Data flow diagram of streamflow forecast system

After Thiessen polygons are generated, we calculate each polygon area to determine the monthly precipitation at the closest weather station to the streamflow station [13]. As a result, past data from weather stations can be used to estimate precipitation and flow at different points in the Upper Cimanuk catchment area.

The Thiessen polygon method is critical for regional water resource management, strategy development and decision making, as it provides accurate precipitation and runoff information on the sub-basin distribution [21], [22].

3.4 Overview of system design

The main system architecture objective is to enable people in the Upper Cimanuk sub-basin to evaluate rainfall and streamflow data for flood forecasting and management purposes. Users must either log in (if they already have an account) or register (if they do not) to access the database. By visiting the database page, users can search for data and observe how it has changed over time. They can also calculate streamflow by selecting that option from the streamflow calculation menu, which aids in planning and determining flood management and water resource administration in the region.

Users can access the streamflow calculation menu page without logging in. This page offers a tool that calculates streamflow based on user-provided inputs, such as precipitation and station information, and generates estimated streamflow to assist in planning and decision-making. Figure 1 displays the system's site map.

Therefore, the system provides easy and quick access to relevant precipitation and streamflow information and a

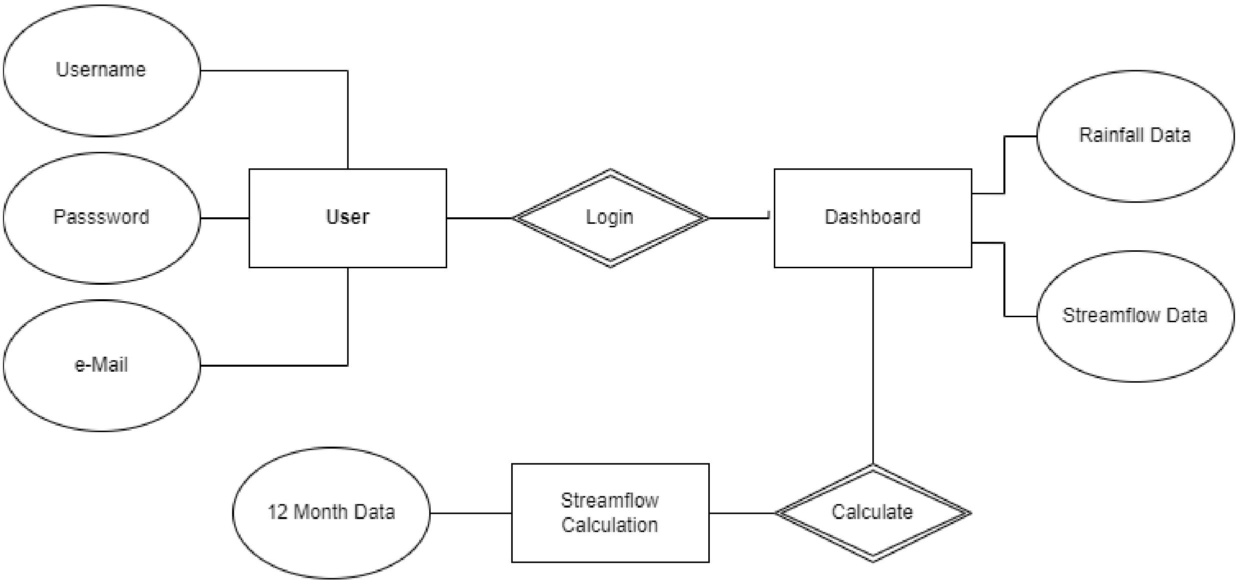


Figure 4: Entity relationship diagram of streamflow forecast system

useful streamflow calculation tool for users like students, lecturers, researchers, and practitioners in the field of water resources management.

4 Proposed system

The system forecasts streamflow, particularly in the Upper Cimanuk sub-basin. Users can access the database only after logging in. As for streamflow forecasting, users do not need to log in. The user's experience with the system starts with the Data Context Diagram (DCD), shown in Figure 2a. The DCD visually represents the framework and primary components of the monthly streamflow forecasting system. It shows the interaction between users and the system, as well as the underlying processes that occur when the system processes user input [23]. The DCD helps describe the flow of information and interactions between the user and the system, and provides an overview of how the monthly streamflow forecasting system works.

Figure 2b shows the Data Flow Diagram (DFD) of the system development. The DFD models the flow of data through a system or process, as well as how the data is processed and stored. The DFD is more detailed than the DCD in describing the system flow process. In the monthly streamflow forecasting system, users must log in to access the database. The system matches the username and password entered by the user with the registered data in the user database. A successful login provides users with streamflow and precipitation data according to the desired annual data of the selected station. The DFD helps acknowledge how data flows through the system and how the system produces the desired output. It also makes it easy to point out and enhance the processes that occur in the monthly streamflow forecasting system [24].

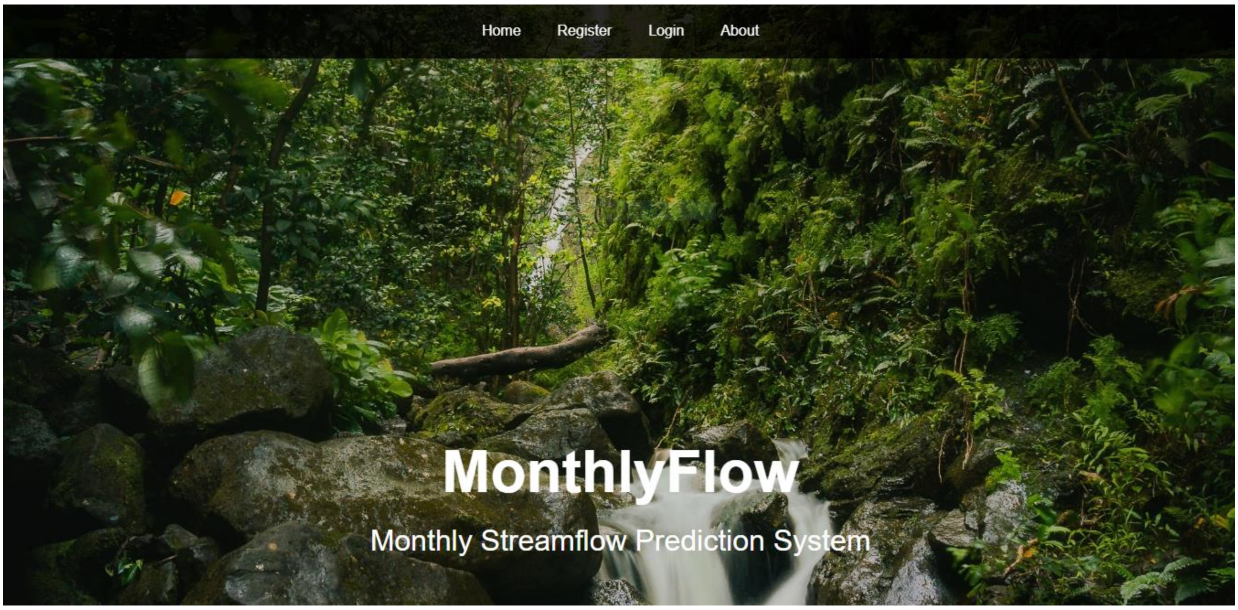


Figure 3: Home page of online streamflow forecast system

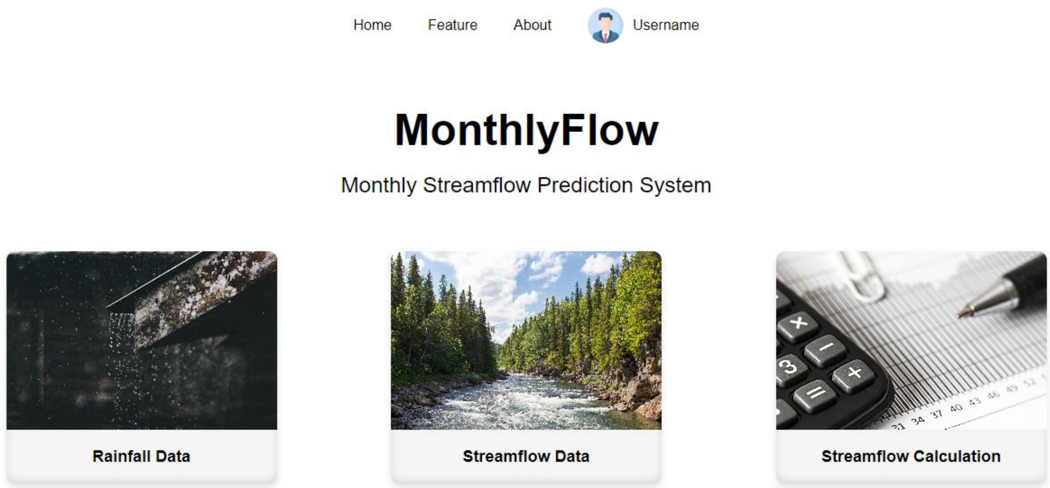


Figure 5: Dashboard of online streamflow forecast system as seen by registered user

The Entity Relationship Diagram (ERD) defines the data structure and system entities relationships. The ERD helps identify the main entities and their interaction with each other [25]. Figure 4 illustrates the relationship between user, precipitation, and streamflow entities; its interpretation is as follows (1) After successful login, users have access to precipitation data and streamflow data, (2) Precipitation and streamflow data are used in the calculation process of streamflow forecasting. The ERD model makes it easier to explain the data structure and relationships between entities in the monthly streamflow forecasting system. With ERD, system developers can design and optimize the system's data structure more efficiently.

5 Results and discussion

An online web application is the monthly streamflow forecasting system. Users can use Internet-connected desktops, laptops, tablets, and smartphones to access the system. Users can access the system anytime and anywhere as needed. When first accessing the system, users will see a home view that contains general information about the monthly streamflow forecasting system. Figure 3 illustrates the home view. To access the features in the system, users must log in by entering their registered username and password. If the user does not have an account, they can register through the join menu available on the login page. The upper center of the web application will show log on username signifying that they have successfully entered the system and can use its features as shown in Figure 5.

Users can access precipitation and streamflow database menus through the monthly streamflow forecasting system after logging in successfully as shown in Figure 6. Users will see the feature menu located in the upper center corner of the system page. Users will have access to precipitation data, streamflow data, and streamflow calculations. This menu provides access to precipitation data based on weather stations in the Upper Cimanuk sub-basin and the selected year. By selecting from the menu in the middle, users can also access the streamflow database. Streamflow menu has the same presentation as the precipitation database, including the choice of annual streamflow station data.

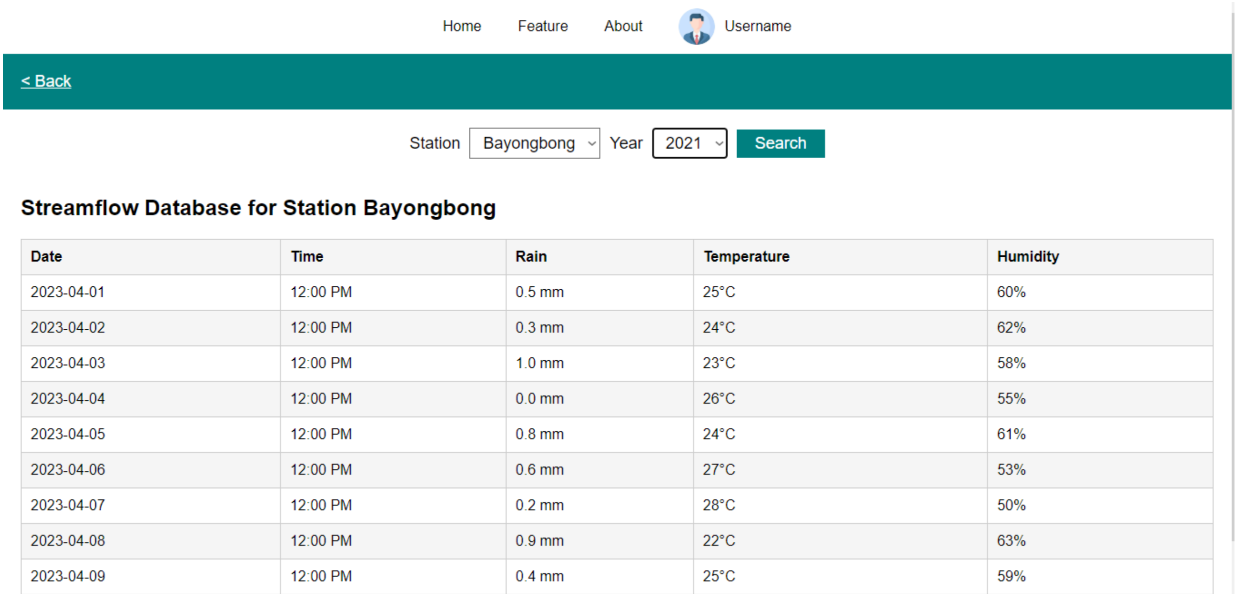
When users search for the desired data, they will get a tabular perspective of precipitation data from the annual elected weather station data as shown in Figure 7. In addition, users will also see a small map showing the area of the weather station. With this online monthly streamflow forecasting system, users can easily access information on precipitation and streamflow in the Upper Cimanuk sub-basin for analysis, planning, and development purposes in various related fields.

5.1 System evaluation

We assess the monthly streamflow forecasting system by comparing its calculation results with manual calculations. The results obtained are in accordance with the manual method, indicating that the system works accurately. The system conducts the calculation process from user input. The user inputs precipitation and streamflow data twelve times, representing monthly data, and then compares the data with manual calculations. After the user inputs the data, the system displays the final calculation table used to determine the regression equation, correlation coefficient (R), and coefficient of determination (R^2). The system provides a limit of two decimal digits to facilitate data reading by the user.

Comparison results show that the R, R^2, and regression equation values obtained from both methods are quite close, indicating that the system is accurate. The system displays streamflow forecasting based on previous calculations by substituting precipitation values into the regression equation. However, the system has the disadvantage of not being able to provide recommendations for users in preventing flood hazards. The system can only display monthly streamflow forecasting based on precipitation values that will occur.

s



Date	Time	Rain	Temperature	Humidity
2023-04-01	12:00 PM	0.5 mm	25°C	60%
2023-04-02	12:00 PM	0.3 mm	24°C	62%
2023-04-03	12:00 PM	1.0 mm	23°C	58%
2023-04-04	12:00 PM	0.0 mm	26°C	55%
2023-04-05	12:00 PM	0.8 mm	24°C	61%
2023-04-06	12:00 PM	0.6 mm	27°C	53%
2023-04-07	12:00 PM	0.2 mm	28°C	50%
2023-04-08	12:00 PM	0.9 mm	22°C	63%
2023-04-09	12:00 PM	0.4 mm	25°C	59%

Figure 6: Rainfall database of online streamflow forecast system as seen by registered user

The system calculates the streamflow forecasting result for each unique streamflow station. We perform ground check operations to verify that the predicted streamflow values are consistent with the measured values in the field. These ground check activities serve to verify the accuracy of the data and ensure consistency between the data analysis results and the real-world conditions in the field.

We used the current meter to gauge the speed of the river's flow while measuring the streamflow using the velocity-region technique. As part of the ground check procedures, we take streamflow field measurements to utilize the velocity-region technique. For this study, we estimated the average velocity of the river by taking readings at two depths (0.2d and 0.8d) from the total depth. We determined the surface region of each segment by looking at whether the segment was triangular or trapezium shaped. Using the below formula, we determined the streamflow at each individual segment:

$$Q = A * V$$
(streamflow = surface region * flow velocity)

and then summed the value of each segment to obtain the total streamflow. By comparing the forecasted streamflow values of the system with the measured values, we can gain insights into the validation of our streamflow forecasting results. Table 2 presents a comprehensive comparison between field-measured streamflow values and the forecasted streamflow values produced by our system. The results affirm that our system offers relatively accurate forecasting outcomes, yet disparities exist between the forecasted and observed results.

Table 2: Comparison of measurement results at several weather station

Location	Q_{ground} (m^3/s)	$Q_{forecast}$ (m^3/s)
Bayongbong	2.46	3.37
Bojongloa	3.01	4.49
Cibatu	4.14	11.96
Leuwidaun	5.29	19.79
Leuwigoong	4.85	12.26

Analyzing the data in Table 2, we observe distinct variations among different weather stations. For instance, at the Bayongbong streamflow station, the observed streamflow (Q) value is recorded as 2.46 m3/s, while the corresponding estimated Q value is 3.37 m3/s, indicating a slight difference between the two. Conversely, at the Leuwidaun streamflow station, the observed Q value is 5.29 m3/s, significantly deviating from the estimated Q value of 19.79 m3/s, marking the largest discrepancy in this comparison.

The observed discrepancies can be attributed to a variety of factors. One significant factor is the absence of certain data points used in our study. Additionally, other variables that were not included in the regression process for streamflow forecasting, such as precipitation intensity, duration of precipitation time, and basin shape and size, play a role in these differences. These unaccounted variables can influence the accuracy of our calculations, leading to discrepancies between forecasted and measured streamflow values.

This underscores the importance of considering a broader spectrum of variables that impact streamflow. By incorporating these additional variables into our forecasting model, we can enhance the accuracy of our streamflow predictions. As part of further research, we recommend the collection of supplementary data and a comprehensive analysis to refine the accuracy of our streamflow forecasting system. Incorporating these relevant variables will bring our forecasting results closer to real-world field conditions, ultimately providing more precise and valuable insights for a range of analytical, planning, and developmental applications in related domains.

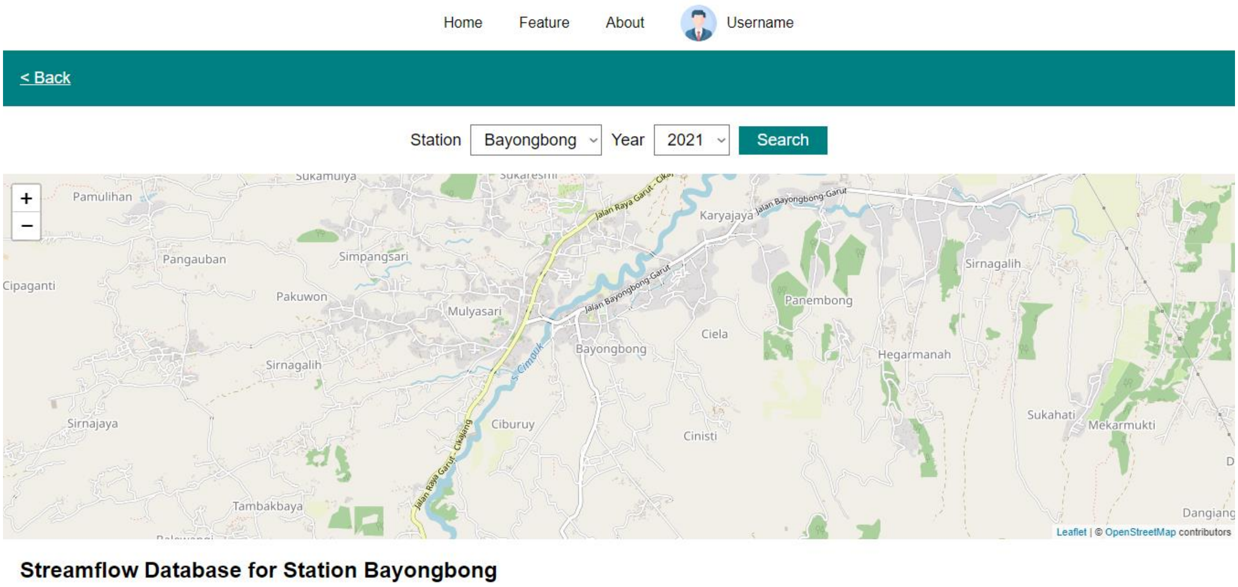


Figure 7: Weather station location and streamflow database as seen by registered user

From Figure 8, a high value of 0.91 for R^2 shows that the forecasting results in this study have high accuracy and are close to the original data value, because an R^2 value close to one indicates that the regression model used can explain the data variation. Based on the comparison between the value of the system forecasting calculation and field measurements, we conclude that the system produces good forecasts. Although differences exist between the forecasted results and the field measurements, the system can produce streamflow estimates that are useful in the context of this study.

The limitations and challenges of the system include the absence of data used in the study and other variables not included in the regression process of streamflow forecasting, such as precipitation intensity, duration of precipitation time, basin shape and size, among others. Researchers can collect additional data and conduct a more comprehensive analysis to enhance the accuracy of the streamflow forecasting system, as well as incorporate additional relevant variables in the forecasting model. Additionally, regular updates of the system with new data can help improve its accuracy over time. Integrating the current system with other tools or systems is possible to provide additional, more comprehensive information and recommendations for flood hazard prevention and water resource management.

5.2 Black box testing

Black box testing on this monthly streamflow forecasting system aims to check each part of the system's functions to see if they run well or not. This test includes:

- User register page: If a user leaves fields blank, the system displays an error message, limits character filling in the username field, and checks username availability,
- Login page: If a user leaves fields blank or the username and password don't match, the system displays an error message. The forgotten password and unregistered functions also work well,
- Flow-rate forecasting page from the user: The calculate flow rate button works well. The system displays a warning if there are empty fields or input other than numbers. If everything checks out, the system previews a monthly forecast of streamflow.
- Flow rate forecasting page from the database: All buttons work properly, including the streamflow station select box and the beginning and ending year select boxes. The final year select box adjusts to the selected initial year select box, facilitating user selection and preventing errors.

With this black box test, the monthly streamflow forecasting system successfully ensures that all functions run properly and in accordance with user needs.

5.3 Discussion

In the realm of integrated streamflow forecasting systems, our study embarks on an expansive discourse, intricately juxtaposing our freshly conceived forecasting system with the landscape outlined in Table 2. This illuminating comparative analysis serves as a prism, elucidating profound insights into the diverse echelons of forecast accuracy achieved through various methodologies. Most notably, it unveils the distinctive advantages harnessed from the harmonious amalgamation of a regression model with historical rainfall and streamflow data.

Our crafted monthly streamflow forecasting system, a testament to our study's rigor, proudly bears the emblem of precision, conspicuously manifested by the substantial R^2 value of 0.91 underscored in Figure 8. This robust correlation serves as an unequivocal testament to the prowess of our regression model, proficiently dissecting and explicating a significant fraction of the data variance. The upshot is an assemblage of forecasts that harmoniously converge with authentic streamflow values. Though discernible disparities linger between our

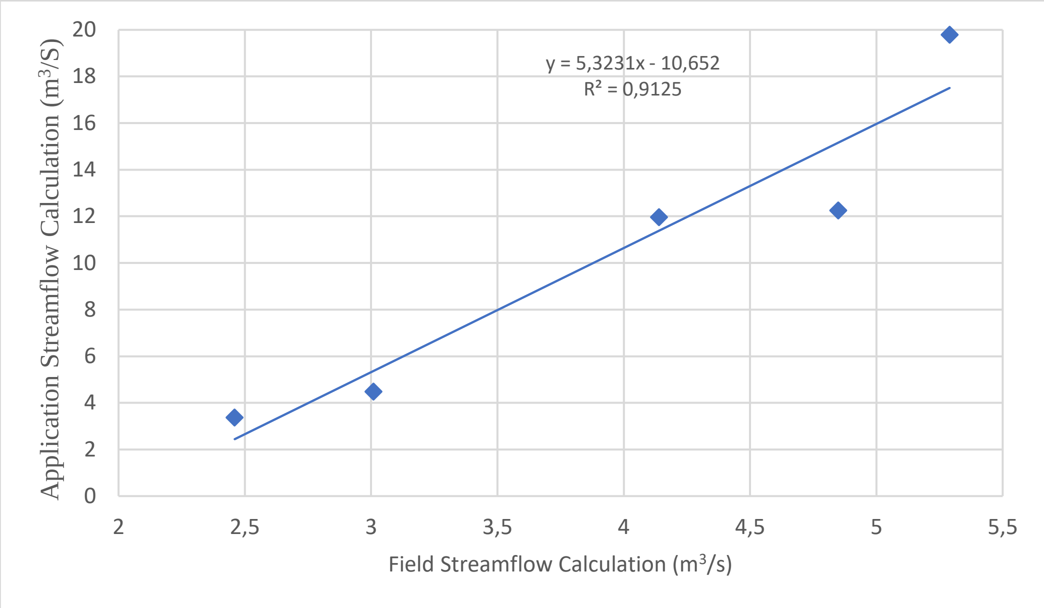


Figure 8: Comparison of measurement results between field measurements and system measurements

foreseen outcomes and the tangible field observations, the prognostications furnished by our system remain germane within the contextual contours of our study's purview.

In juxtaposition to antecedent inquiries, our study discerns its stronghold through a nuanced interplay of advantages. Within the annals of streamflow prediction, the focus on regression models, as evidenced by the endeavors of [10], has unfurled a promising tapestry of potential. Noteworthy is their aptitude in illuminating the dynamics of streamflow, as proven by their ability to forecast with precision. Our comparative analysis with this lineage underscores a pivotal divergence, attributed to the infusion of a regression model with historical rainfall and streamflow data. This synergy of variables serves as our study's fulcrum, endowing our system with an elevated plane of accuracy and efficacy.

Moreover, the rise of online platforms, as propounded by [11], has ushered an epoch of transformation in water resource management. Our innovative solution further amplifies this transformative spirit by seamlessly integrating an online platform. This dynamic interplay empowers stakeholders, offering not only forecasting insights but also interactive tools for decision-making. By enhancing collective awareness and understanding of water-related hazards and mitigation strategies, our system underscores a pivotal shift in the approach to resource management, surpassing mere prediction and ushering in the era of informed, preemptive action.

Validation, a cornerstone of credibility, as emphasized by [12], forms a fundamental tenet of our approach. We champion this by grounding our system in meticulous validation processes, ensuring its robustness and reliability. Furthering the thread of augmentation, spatial and temporal factors, as underscored by [13], etch their influence onto our methodology. The systematic inclusion of these factors enriches the predictive prowess of our system, enabling us to identify at-risk areas and extend timely safety measures, thereby fostering a paradigm of proactive risk mitigation.

Incorporating additional variables, as exemplified by [14], kindles a beacon of promise. Our study inherently heeds this call by embracing historical rainfall and streamflow data as pivotal components in our predictive framework. This infusion enhances our accuracy, aligning our forecasts more closely with real-world dynamics. Moreover, the fusion of cross-disciplinary collaboration and the insights of expert users, as illuminated by [15], epitomizes our journey. We seamlessly weave life sciences and Earth system models, fusing empirical evidence with expert intuition. By doing so, we endow decision-makers with a holistic and nuanced understanding, empowering them to navigate complex climate change challenges with clarity and astuteness.

Yet another facet, resonating with [16], unearths the transformative potential of big data and machine learning technologies. Our system, founded upon these very principles, encapsulates the transformative spirit of data-driven environmental management. In doing so, we bridge the gap between theoretical concepts and practical application, rendering our solution not merely a scholarly endeavor but a dynamic instrument to revolutionize the very landscape of water resource management.

In synthesis, our study's discourse within the domain of integrated streamflow forecasting unfolds as a testament to the fusion of innovation, validation, and cross-disciplinary insights. By synergizing the potency of regression models with historical data, augmenting our platform with dynamic online tools, and embracing the nuances of spatial, temporal, and additional variables, our approach stands as a beacon of advancement. It is an embodiment of the metamorphosis in water resource management, forging an informed and empowered future, where foresight marries action to harmoniously navigate the dynamic tapestry of our hydrological world.

The novelty of our integrated streamflow forecasting system, which transcends its technical intricacies, lies in its profound impact within the domains of flood management and water resource optimization. A hallmark of our solution is the seamless integration of an online platform, endowing users with real-time interactive capabilities for data input and visualization. This aspect significantly amplifies the pragmatic utility and versatility of our forecasting system.

Central to the efficacy of the online platform is its ability to provide facile access to monthly streamflow and precipitation data. This resource-rich accessibility empowers decision-makers to make well-informed choices concerning water resource development within the Upper Cimanuk sub-basin. Furthermore, our platform serves as a conduit for timely initiation of proactive flood disaster mitigation strategies, including early warning systems and evacuation plans, all rooted in accurate and promptly delivered information.

Beyond its applications in flood management, our forecasting system bears relevance to diverse sectors encompassing agriculture, irrigation, and hydropower. By delivering precise streamflow predictions, it optimizes irrigation scheduling, enabling farmers to maximize crop yields while conserving water resources.

The harmonization of an online platform with our regression model engenders a synergy, positioning our integrated streamflow forecasting system as an instrument in the realm of smart flood management. The intuitive user interface augments the accessibility and usability of our findings, rendering our approach highly pertinent in the context of contemporary water resource management practices.

6 Conclusion

The monthly streamflow forecasting system generates proportionate precise forecasting results that resemble values of original data. This is evident from the 100% coefficient of determination (R^2) value, which indicates that the regression model can adequately explain the differences in the data. Nevertheless, there is potential to develop and improve the accuracy and precision of forecasting. There may be discrepancies between predicted and measured results due to lack of data and the exclusion of other variables from the streamflow regression prediction process. These include the intensity and duration of precipitation, also the shape and size of the catchment area.

To improve the quality of the system's predictions, we recommend adding more variables affecting flow and collecting more data. Extensive field data validation can also optimize the performance of the forecasting system. The developed monthly flow forecasting system can be considered as a valuable tool in the decision making process through flow analysis and forecasting in water resource management and flood prevention.

We highlight several limitations and challenges associated with the developed system. Firstly, the system accuracy is affected by the exclusion of other variables, such as rainfall intensity, rainfall duration, basin shape and size, etc. Next, the system only displays monthly streamflow forecasts based on past rainfall, and does not provide recommendations or next steps to prevent flood hazards.

To address these limitations and challenges, we propose multiple methods for improving the performance of the system. Initially, researchers can collect additional data, conduct a more thorough analysis, and incorporate additional relevant variables into the forecasting model. This can enhance the accuracy and precision of the forecasting. Second, improved system capabilities can provide users with recommendations for the next steps to prevent flood hazards. For instance, the system can recommend flood prevention measures based on the anticipated streamflow values. Thirdly, we are conducting field data validation to optimize system performance. This will help confirm the forecasting accuracy and provide valuable selective data for examination, strategy development, and the enhancement of related fields.

7 Implications and future work

Our integrated streamflow forecasting system holds significant potential for various real-world water resource management scenarios. By focusing on the Upper Cimanuk sub-basin, a region characterized by heterogeneous topography and hydrological attributes prone to flooding, our system directly confronts the challenges inherent in flood management and water resource optimization. This region serves as a testing ground for evaluating the system's efficacy within the complexities of its environment, making it an ideal crucible for assessing performance across diverse terrains and land use patterns. Moreover, the sub-basin's susceptibility to floods accentuates the critical role of precise streamflow forecasts in underpinning effective flood disaster mitigation strategies.

The intricacies of our methodology revolve around the development and evaluation of a bespoke monthly streamflow forecasting system tailored to the Upper Cimanuk sub-basin. Central to this system is a sophisticated regression model that seamlessly blends historical rainfall and streamflow data gleaned from an extensive network of monitoring stations across the sub-basin. This model artfully elucidates the intricate interplay between regional rainfall patterns and the resulting streamflow dynamics.

Our methodology deliberately incorporates a straightforward linear regression model to establish a coherent link between rainfall and streamflow data. This model serves to illuminate the complex relationships underpinning these variables. At the same time, we ensure seamless access and engagement through our user-centric web application. This platform allows users to interact with our forecasting system from any internet-connected device, making forecasting easy and convenient to use.

7.1 Challenges and limitations

It is important to recognise the potential challenges of real-world implementation, although our system offers promising results. Missing data points and unaccounted for variables such as rainfall intensity, rainfall duration, or basin shape and size can affect the accuracy of the calculations. These factors contribute to the difference between the predicted value and the manual measurement value. In addition, although widely used, the velocity-area method is difficult to detect the dynamics of flow variations when extreme events or irregular river shapes occur.

Other important considerations include scalability, performance and data availability. Forecast accuracy and real-time data acquisition become critical as systems scale to larger areas. To guarantee forecast accuracy, it is important to have consistent and reliable data sources.

7.2 Areas of improvement

We suggest several alternatives to improve system accuracy. Incorporating additional variables such as rainfall intensity, duration and watershed characteristics, can improve forecast accuracy and minimize gaps between predictions and observations. In addition, exploring the integration of machine learning and artificial intelligence techniques could be a promising technology for enhancing predictive capabilities. The system can provide more dynamic and accurate predictions, especially for extreme events, by training machine learning models on large data sets.

Reference

[1] M. Tanoue, R. Taguchi, S. Nakata, S. Watanabe, S. Fujimori, and Y. Hirabayashi, "Estimation of Direct and Indirect Economic Losses Caused by a Flood With Long-Lasting Inundation: Application to the 2011 Thailand Flood," *Water Resour. Res.*, vol. 56, no. 5, May 2020, doi: 10.1029/2019WR026092.

[2] Z. Qin, M. Storozum, H. Liu, X. Zhang, and T. R. Kidder, "Investigating environmental changes as the driving force of agricultural intensification in the lower reaches of the Yellow River: A case study at the Sanyangzhuang site," *Quat. Int.*, vol. 521, pp. 25–34, Jun. 2019, doi: 10.1016/j.quaint.2019.06.033.

[3] T. Mai *et al.*, "Defining flood risk management strategies: A systems approach," *Int. J. Disaster Risk Reduct.*, vol. 47, p. 101550, Aug. 2020, doi: 10.1016/j.ijdrr.2020.101550.

[4] E. Savitri and I. Pramono, "ANALISIS BANJIR CIMANUK HULU 2016," *J. Penelit. pengelolaan Drh. Aliran Sungai*, vol. 1, no. 2, pp. 97–110, Oct. 2017, doi: 10.20886/jppdas.2017.1.2.97-110.

[5] M. Ardiansyah, R. A. Nugraha, L. O. S. Iman, and S. D. Djatmiko, "Impact of Land Use and Climate Changes on Flood Inundation Areas in the Lower Cimanuk Watershed, West Java Province," *J. Ilmu Tanah dan Lingkung.*, vol. 23, no. 2, pp. 53–60, Dec. 2021, doi: 10.29244/jitl.23.2.53-60.

[6] F. A. Hirpa *et al.*, "Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data," *J. Hydrol.*, vol. 566, pp. 595–606, Nov. 2018, doi: 10.1016/j.jhydrol.2018.09.052.

[7] Y. Gao, C. Merz, G. Lischeid, and M. Schneider, "A review on missing hydrological data processing," *Environ. Earth Sci.*, vol. 77, no. 2, p. 47, Jan. 2018, doi: 10.1007/s12665-018-7228-6.

[8] P. C. Austin and E. W. Steyerberg, "The number of subjects per variable required in linear regression analyses," *J. Clin. Epidemiol.*, vol. 68, no. 6, pp. 627–636, Jun. 2015, doi: 10.1016/j.jclinepi.2014.12.014.

[9] Reza Pahlevi, "APJII: Penetrasi Internet Indonesia Capai 77,02% pada 2022," *katadata*, 2022. https://databoks.katadata.co.id/datapublish/2022/06/10/apjii-penetrasi-internet-indonesia-capai-7702-pada-2022

[10] P. A. Mendoza *et al.*, "An intercomparison of approaches for improving operational seasonal streamflow forecasts," *Hydrol. Earth Syst. Sci.*, vol. 21, no. 7, pp. 3915–3935, Jul. 2017, doi: 10.5194/hess-21-3915-2017.

[11] A. Teague, Y. Sermet, I. Demir, and M. Muste, "A collaborative serious game for water resources planning and hazard mitigation," *Int. J. Disaster Risk Reduct.*, vol. 53, p. 101977, 2021, doi: https://doi.org/10.1016/j.ijdrr.2020.101977.

[12] S. Mohammed *et al.*, "A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean," *Comput. Electron. Agric.*, vol. 197, p. 106925, 2022, doi: https://doi.org/10.1016/j.compag.2022.106925.

[13] J.-P. Naulin, O. Payrastre, and E. Gaume, "Spatially distributed flood forecasting in flash flood prone areas: Application to road network supervision in Southern France," *J. Hydrol.*, vol. 486, pp. 88–99, 2013, doi: https://doi.org/10.1016/j.jhydrol.2013.01.044.

[14] Y. Sun, D. Wendi, D. E. Kim, and S.-Y. Liong, "Deriving intensity–duration–frequency (IDF) curves using downscaled in situ rainfall assimilated with remote sensing data," *Geosci. Lett.*, vol. 6, no. 1, p. 17, 2019, doi: 10.1186/s40562-019-0147-x.

[15] D. S. Schoeman *et al.*, "Demystifying global climate models for use in the life sciences," *Trends Ecol. Evol.*, vol. 38, no. 9, pp. 843–858, Sep. 2023, doi: 10.1016/j.tree.2023.04.005.

[16] A. Y. Sun and B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," *Environ. Res. Lett.*, vol. 14, no. 7, p. 73001, 2019, doi: 10.1088/1748-9326/ab1b7d.

[17] S. Otuagoma, E. Ogujor, and P. Kuale, "Comparative Measurement of Stream Flow in the Ethiope River for Small Hydropower Development," *Niger. J. Technol.*, vol. 34, no. 1,

p. 184, Dec. 2014, doi: 10.4314/njt.v34i1.23.

[18] L. Yuge and T. Badarch, “Research on Contemporary Software Development Life Cycle Models,” *Am. J. Comput. Sci. Technol.*, vol. 6, no. 1, pp. 1–9, Mar. 2023, doi: 10.11648/j.ajcst.20230601.11.

[19] M. Jena and S. Dehuri, “DecisionTree for Classification and Regression: A State-of-the Art Review,” *Informatica*, vol. 44, no. 4, Dec. 2020, doi: 10.31449/inf.v44i4.3023.

[20] I. Ardiansah, N. Bafdal, E. Suryadi, and A. Bono, “Design of micro-climate data monitoring system for tropical greenhouse based on arduino UNO and raspberry pi,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 757, no. 1, 2021, doi: 10.1088/1755-1315/757/1/012017.

[21] A.-A. Hussein, V. Govindu, and A. G. M. Nigusse, “Evaluation of groundwater potential using geospatial techniques,” *Appl. Water Sci.*, vol. 7, no. 5, pp. 2447–2461, Sep. 2017, doi: 10.1007/s13201-016-0433-0.

[22] G. B. Adane, B. A. Hirpa, C.-H. Lim, and W.-K. Lee, “Evaluation and Comparison of Satellite-Derived Estimates of Rainfall in the Diverse Climate and Terrain of Central and Northeastern Ethiopia,” *Remote Sens.*, vol. 13, no. 7, p. 1275, Mar. 2021, doi: 10.3390/rs13071275.

[23] Y. Wu and P. Xie, “Exploration of Enterprise Audit Information Management System Model Based on Data Flow Diagram,” in *2021 International Wireless Communications and Mobile Computing (IWCMC)*, IEEE, Jun. 2021, pp. 1997–2001. doi: 10.1109/IWCMC51323.2021.9498870.

[24] R. Arifin, G. N. Syaifuddiin, D. Desriyanti, Z. U. Rosyidin, and G. A. Buntoro, “A Highly Accurate Internet-Based Fake Information Detection Tool for Indonesian Twitter,” *Informatica*, vol. 46, no. 9, Jan. 2023, doi: 10.31449/inf.v46i9.4416.

[25] T. A. Vakaliuk, O. V. Korotun, and S. O. Semerikov, “The selection of cloud services for ER-diagrams construction in IT specialists databases teaching,” *CTE Work. Proc.*, vol. 8, pp. 384–397, Mar. 2021, doi: 10.55056/cte.295.