

Open Access Article



<https://doi.org/10.55463/issn.1674-2974.49.12.20>

## Dam Environmental Impact Assessment in Java Island, Indonesia

Dony Hermawan<sup>1\*</sup>, Muhammad Bisri<sup>2</sup>, Lily Montarcih<sup>2</sup>, Evi Nurcahya<sup>2</sup>

<sup>1</sup> Doctoral Program Civil Engineering, Faculty of Engineering, University of Brawijaya, Malang, Indonesia

<sup>2</sup> Department of Water Resources, Faculty of Engineering, University of Brawijaya, Malang, Indonesia

\* Corresponding author: [dony.hermawan@pu.go.id](mailto:dony.hermawan@pu.go.id)

Received: August 14, 2022 / Revised: October 12, 2022 / Accepted: November 10, 2022 / Published: December 30, 2022

**Abstract:** Increasing demand for water requires the improvement of water infrastructure. Dams are one of the most prominent infrastructures for water supply, especially as water demand increases. Due to its size, the dam is bound to affect the local environment, including the hydrometeorological and hydro/geological condition. The impact of dam on the local scale is supposedly dependent on several variables, namely, the height of the dam, its age, and the area of the river basin. This study studies the impact of dams on the surrounding environment by investigating 12 dams located in Java and are spread all throughout Java in particularly West Java, East Java, Central Java and lastly Jogjakarta. The dam environmental impact will be analyzed using: firstly, linear regression to determine the relationship between dams and environmental impact and secondly cluster analysis to divide the researched dams into categories. Dams in the study area could be divided into three clusters based on the similarity in height, age and river basin area. Additionally, this study finds that there is no correlation between the river basin area and both (ground) water quality and local climate. However, there was a correlation between the dam height with temperature and precipitation, as well as the dam age with Dissolved Oxygen (DO), Total Dissolved Solid (TDS) and pH. Regarding groundwater, the values of DO, TDS and pH are higher in dams located inside of groundwater basins compared to the opposite.

**Keywords:** dam, water demand, dam impact.

## 印度尼西亚爪哇岛大坝环境影响评价

**摘要:** 对水的需求不断增加, 需要改善水利基础设施。水坝是最重要的供水基础设施之一, 尤其是在用水需求增加的情况下。由于其规模, 大坝必然会影响当地环境, 包括水文气象和水文/地质条件。据推测, 大坝对当地规模的影响取决于几个变量, 即大坝的高度、年龄和流域面积。本研究通过调查位于爪哇的12座水坝来研究水坝对周围环境的影响, 这些水坝遍布整个爪哇, 特别是西爪哇、东爪哇、中爪哇和日惹。大坝环境影响分析将使用: 首先, 线性回归确定大坝与环境影响之间的关系, 其次聚类分析将研究的大坝分类。根据高度、年龄和流域面积的相似性, 研究区内的大坝可分为三类。此外, 本研究发现流域面积与(地下)水质和当地气候之间没有相关性。然而, 大坝高度与温度和降水量之间存在相关性, 并且与溶解氧(做)、总溶解固体(TDS)和酸碱度值之间存在相关性。关于地下水, 位于地下水盆地

内部的水坝的做、TDS和酸碱度值高于相反的水坝。

**关键词：**大坝、需水量、大坝影响。

### 1. Introduction

Water scarcity is one of the most prominent problems for the 21<sup>st</sup> Century. Water availability is not constant throughout the year, with people lacking access to water in the drought season [1]. Environmental changes in terms of hydrological drought could be lessened by the building of mega reservoirs [2]. Additionally, Dam could also be used for hydropower generation, flood control and water storage. In other words, dams are important in supporting agriculture, energy and water resources [3].

Dams are one of the most important water structures in Indonesia. Based on the Ministry of Public Works and Public Housing [4], there are currently 205 operating dams in Indonesia compared with 201 dams in 2019. Moreover, in the period between 2020-2024, there will be additional 45 new dam projects that costs 92.9 trillion Rupiahs.

The presence of dams in certain locations means that there will be changes in the land use. The change in land use could affect rainfall intensity in certain areas [5], for example, the effect of urban areas on rainfall intensity. The most notable change caused by damming is the change of land cover to a water body, altering the hydro (geo) logical cycle in the surrounding area due to increasing the availability of water moisture [6]. Several studies [6]-[7] found that dams increase the precipitation and, consequently, increase the risk for extreme precipitation and widen the gap between rainfall extremes and normal rainfall in the Three Gorges Dam in China [7].

This research will focus on the effect of dam construction to hydro(geo)logical cycle by focusing on three parameters: water quality, groundwater condition, local climate and river. Firstly, the water quality could change in relation to its chemistry. For groundwater, in particular, the groundwater condition could improve or decrease due to the reservoir water interaction with groundwater. Dams also affect the local climate by changing the local precipitation and air temperature. Lastly, it could also affect the river regime.

### 2. Research Locations

Twelve dams were chosen for this research, with Java as the focus. The dams are spread in four big provinces in Java: West Java, Central Java, East Java and Jogjakarta (Table 1 and Fig. 1). The oldest dam researched is Cipancuh, which has been operating since 1927 at the opposite of Wonorejo, which is the youngest dam.

Table 1 Dam research location (<https://sda.pu.go.id>)

No.	Dam	Operated since	Province	Height
1	Cipancuh	1927	West Java	7 m
2	Darma	1959	West Java	37.5 m
3	Malahayu	1940	Central Java	31.3 m
4	Cacaban	1958	Central java	38 m
5	Wadas Lintang	1988	Central Java	125 m
7	Delingan	1923	Central Java	23 m
6	Sermo	1996	Jogjakarta	58.6
8	Pondok	1993	East Java	30.67
9	Selorejo	1972	East Java	49 m
10	Lahor	1977	East Java	74 m
11	Sampean Baru	1983	East Java	26 m
12	Wonorejo	2001	East Java	100 m



Fig. 1 Dam locations (Developed by the authors)

### 3. Data and Methods

#### 3.1. Framework

The research was performed by focusing on dividing four main parameters into its own framework (Fig. 1). Each parameter has its own dataset which needs to be collected, either directly in the field or through literature study. After that, the water quality data, groundwater data and local climate data will be analyzed. Finally, the resulting analysis and model will be analyzed further using the linear regression and cluster analysis.



Fig. 2 Research framework

### 3.2. Research Data

The data is divided into four main datasets: dam, surface water quality, groundwater, and local climate.

Table 2 Research data

No.	Parameter	Indicator	Data	Method	Output
1	Reservoir Water Quality	1. Physical 2. Chemical	1. Secondary Data 2. Field Campaign	Linear Regression	Regression model for water quality
2	Groundwater	1. Groundwater level 2. Groundwater quality	1. Secondary Data 2. Field Campaign	Linear Regression	Regression Model for groundwater
3	Local Climate	1. Mean precipitation 2. Temperature	1. Secondary Data 2. Field Campaign	Linear Regression	Regression Model for hydrometeorology
4	Dam Characteristics	1. Height 2. Age 3. River Basin Area	1. Secondary Data 2. Field Campaign	Cluster Analysis	Clustered Dam

### 3.3. Analysis Method

#### 3.3.1. Regression Analysis

Regression analysis is a simple statistical method used to investigate the effect of certain variables (predictor variables) on other variables (response variables) [8]. This will result in a probabilistic model, also known as the regression model. Even though it is the simplest statistical model, it is still relevant to be used even to analyze big data [8].

#### 3.4. Linear Regression

A linear regression model determines the relations between dependent and independent variables linearly [9] with the following formula:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (1)$$

where:

$Y$  - dependent/response variable;

$X_i$  - independent/predictor variable,  $i = 1, 2, \dots, k$ ;

$\beta_i$  - parameter with unknown value, was also called regression coefficient,  $i = 0, 1, \dots, k$ ;

$\epsilon$  - error model, assumed to have normal distribution with a mean value of 0 and constant variance,  $\sigma^2$ , written as  $\epsilon \sim N(0, \sigma^2)$ .

If the value of the predictor variable is 1 ( $k = 1$ ), the model is called a simple regression model [8]. A linear regression with only one independent variable could be calculated with the formula [9]:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \quad (2)$$

#### 3.5. Parameter Estimation

One of the most important factors in regression analysis is to determine the value of  $\beta_0$  and  $\beta_1$ . The values of  $\beta_0$  and  $\beta_1$  could be calculated using the least squares method to find the  $\beta_0$  and  $\beta_1$  estimator,  $b_0$  and  $b_1$  respectively [9]:

$$b_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (3)$$

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (4)$$

where:

$\sum X_i$  - sum of data from predictor variable  $X$ ;

$\sum Y_i$  - sum of the data from the response variable  $Y$ ;

$\bar{X}$  - mean value of predictor variable  $X$ ;

$\bar{Y}$  - mean value of predictor variable  $Y$ ;

$n$  - amount of data.

After calculating  $b_0$  and  $b_1$ , the relationships between predictor variable  $X$  and response variable  $Y$  could be shown by:

$$\hat{Y} = b_0 + b_1 X_1 \quad (5)$$

#### 3.6. Testing the Coefficient Significance

The coefficient significance will be tested only after the regression estimation model result from equation 3 is obtained. The coefficient value should be greater or less than zero to conclude that there is a significant relationship between the response and predictor variable [10]. Table 3 shows the significance testing results using Minitab with the height of the dam as a predictor variable ( $X$ ) and the change in rainfall intensity as the response variable ( $Y$ ).

Table 3 Coefficient significance testing (Developed by the authors)

Variable	Coefficient	SE	T-Value	P-Value
Constant	18.9686	2.577	7.36	0
Dam Height	-0.1534	0.0428	-3.59	0.005

Estimator  $b_0$  dan  $b_1$  could be seen in the *Coefficient column* with the values of  $b_0 = 18.9686$  and  $b_1 = -0.1534$ .

#### 3.7. Hypotheses Testing

The testing was conducted using  $t$ -test, with the assumption that if the value of  $H_0 : \beta_1 = 0$ , there is no correlation between  $X$  and  $Y$ . However, if the value of  $H_1 : \beta_1 \neq 0$ , there is a correlation between  $X$  and  $Y$ .

The  $H_0$  testing criteria will be rejected if  $|T\text{-value}| > t_{\alpha/2}$  or  $P\text{-value} < \alpha$ , with:

$T\text{-value}$  - testing statistic;

$t_{\alpha/2}$  - obtained from distribution table t with degrees of freedom ( $n-2$ );

$\alpha$  - a rank of significance (acceptable error, determined by researcher);

P-value - probability that was obtained from the samples and supports the decision to reject  $H_0$ . The smaller the value of P-value, the stronger the evidence to reject  $H_0$ . In this research,  $H_0$  will be rejected if  $P\text{-value} < \alpha$ .

For example, based on the data from Table 3 (P-value = 0.005),  $H_0$  will be rejected. This implies that the height of the dam affects the precipitation value.

**3.8. The Model Goodness of Fit and Its Relationship**

The goodness of the model could be seen from its dependent coefficient: how much the X variable contributes to the predictions of the Y value. In other words, how much the behavior of Y variable could be explained by X in the linear regression model (Eq. 5). The determinant coefficient is calculated from  $R^2$  or  $R^2(\text{adj})$  value with the majority using the value of  $R^2(\text{adj})$ . However, the relation between X and Y variables could be seen in coefficient correlation (R).

$$\begin{array}{ccc} \underline{R} & \underline{R^2} & \underline{R^2(\text{adj})} \\ 0.7505 & 56.33\% & 51.96\% \end{array} \quad (6)$$

**3.9. Model Interpretation**

After the coefficient significance test, the regression model estimation in Equation 3 could be interpreted as follows:

- If the value of  $b_1 = 0$ , there is no correlation between X and Y variables
- If the value of  $b_1 > 0$ , there is a positive correlation between variable X and Y. If the value of X increases, so does the value of  $\hat{Y}$  in a linear relationship.
- If the value of  $b_1 < 0$ , there is a negative correlation between variable X and Y. If the value of X increases, the opposite will happen to the value of  $\hat{Y}$  in a linear relationship.

For example, the value of  $b_1$  is  $-0.1534 < 0$  in Table 3. That implies that the height of the dams has a significant (negative) impact in correlation with precipitation change. The change in one unit in dam height (in m) will decrease the precipitation change value in one unit (0.1534 mm).

**3.10. Cluster Analysis**

*3.10.1. Cluster Analysis Definition*

Cluster analysis could be grouped as Interdependent Techniques. The purpose of this analysis is to group the objects (Dam in this case) based on the similarity or differences in characteristics [11]. Objects that are grouped together ideally should have similar characteristics.

The characteristics of a good cluster are:

- a. The high homogeneity of objects within a

cluster.

- b. The high heterogeneity of one cluster with the other clusters.

*3.10.2. Cluster Analysis Process*

The process of cluster analysis:

*a) Similarity Measurement*

The first and the most important step in cluster analysis is to group objects with similar characteristics. This objective could be achieved by measuring similar the objects are. The methods for similarity measurement are:

- Measuring the correlation between objects in several variables measuring the distance between objects. One popular method is using *Euclidian distance*.

*b) Cluster Generation*

In this research, cluster generation used *the complete linkage* method, which groups objects using the farthest distance. The first step is to group two or more objects, which are the most similar followed by objects that have the second-most similarities. The process continues until it produces a “tree”, which shows the similarity hierarchy between objects. All objects will form a cluster that could be explained using Dendrogram.

**4. Results and Discussions**

**4.1. Local Climate Analysis**

*4.1.1. The Impact of Dam Height on Minimum, Mean and Maximum Temperature Analysis*

*Hypotheses for dam effect to minimum temperatures:*

$H_0$ : The height of the dams does not have a significant impact on temperature.

$H_1$ : The height of the dams has a significant impact on temperature.

Table 4 Regression model between dam height and temperature change (Developed by the authors)

Model	Climate Variable	Independent Variable	Coefficient	T-value	p-value	R	R <sup>2</sup>
1.	Temperature change (min)	Constant Dam Height	25.202 -0.0086	131.94 -2.72	0.000* 0.022*	0.6522	42.54%
2.	Temperature change (mean)	Constant Dam Height	2.097 -0.0123	9.54 -3.38	0.000* 0.007*	0.7305	53.37%
3.	Temperature change (max)	Constant Dam Height	1.935 -0.0126	9.14 -3.58	0.000* 0.005*	0.7495	56.19%

\* Significant at  $\alpha = 0.05$

*4.1.2. Dam Height Impact on Minimum Temperature*

According to the regression analysis model 1 (Table 4), the value of t-test to regression coefficient with *p-value* = 0.022. The value of p is less than the  $\alpha$  value (0.022 compared to 0.05), inferring that the height of dams has a significant impact on the minimum temperature. The lower the height of the dams, the

higher the temperature, with the goodness of model 42.54% and a slightly stronger correlation (0.6522). The regression model for minimum temperature is:

$$\text{Min. temperature increase} = 25.202 - 0.0086 * \text{height of dam} \quad (7)$$

The increase of dam height by 1-m results in the decrease of temperature value by 0.0086 unit.

4.1.3. Dam Height Impact on Mean Temperature

According to the regression analysis model 2 (Table 4), the value of t-test to regression coefficient with *p-value* = 0.007. The value of *p* is less than the  $\alpha$  value (0.007 compared to 0.05), indicating that the height of dams has a significant impact on the mean temperature. The lower the height of the dams, the higher the mean temperature with the goodness of model 53.37% and a slightly strong correlation (0.7305). The regression model for the mean temperature is:

$$\text{Mean temperature increase} = 2.097 - 0.0123 * \text{height of dam} \quad (8)$$

The increase of dam height by 1-m results in a decrease in temperature value by 0.0123 unit.

4.1.4. Dam Height Impact on Maximum Temperature

According to the regression analysis model 3 (Table 4), the value of t-test to regression coefficient with *p-value* = 0.005. The value of *p* is less than the  $\alpha$  value (0.022 compared to 0.05), inferring that the height of dams has a significant impact on the maximum temperature. The lower the height of the dams, the higher the maximum temperature with the goodness of model 56.19% and a slightly strong correlation 0.7496. The regression model for the maximum temperature is:

$$\text{Max temperature increase} = 1.935 - 0.0126 * \text{height of dam} \quad (9)$$

The increase of dam height by 1 m results in a

decrease in temperature value by 0.0126 unit.

4.1.5. The Analysis of Dam Height Impact on Precipitation

Hypotheses:

$H_0$ : The height of the dams does not have a significant impact on precipitation.

$H_1$ : The height of dams has a significant impact on precipitation.

$$\text{Precipitation decreases} = 18.9686 - 0.1534 * \text{height of dam} \quad (10)$$

Table 5 Coefficient significance tests (Developed by the authors)

Variable	Coefficient	SE	T-Value	P-Value
Constant	18.9686	2.577	7.36	0.000
Height of Dam	-0.1534	0.0428	-3.59	0.005

According to the regression analysis above, the value of the regression is -0.1534 with *p-value* = 0.005. The value of *p* is less than the  $\alpha$  value (0.005 compared to 0.05), inferring that the height of dams has a significant impact on precipitation. The lower the height of dams, the higher the precipitation, with the goodness of model 56.33% and a slightly strong correlation of 0.7505.

4.1.6. The Analysis of River Basin Area Impact on Local Climate (Temperature and Precipitation)

Hypotheses:

$H_0$ : The river basin area does not have noticeable impact on local climate change\*

$H_1$ : The river basin area has a significant impact on local climate change\*

According to the regression analysis in Table 6, all models supported the analysis result that the total area of the river basin does not have a significant impact on the local climate.

Table 6 Four regression models for the correlation between river basin area and local climate change (Developed by the authors)

Model	Climate Variable	Independent Variable	Coefficient	T-value	p-value	Decision
1.	Temperature Increment (min)	Constant	1.5272	5.39	0.0003	Insignificant
		River Basin Area	0.0007	0.57	0.5832	
2.	Temperature Increment (mean)	Constant	1.4609	7.31	0.0000	Insignificant
		River Basin Area	-0.00007	-0.09	0.9317	
3.	Temperature Increment (max)	Constant	1.39416	7.38	0.0000	Insignificant
		River Basin Area	-0.0009	-1.04	0.3242	
4.	Precipitation Decrement	Constant	10.4055	4.32	0.0015	Insignificant
		River Basin Area	0.0036	0.35	0.7343	

4.2. Reservoir Water Quality Analysis

4.2.1. The Analysis of Dam Age Impact on Reservoir Water Quality

Hypotheses:

$H_0$ : The age of dams does not have a significant impact on water quality.

$H_1$ : The age of dams has a significant impact on water quality.

Table 7 Regression model for the correlation between the age of the dam and reservoir water quality (Developed by the authors)

Model	Water Quality	Independent Variable	Coefficient	T-value	p-value	R	R <sup>2</sup>
1(a).	DO	Constant	0.1777	0.49	0.6371	0.9298	86.44%
		Age of Dam	0.0461	7.99	0.0000*		
1(b).	DO	Age of Dam	0.0487	22.93	0.0000*	0.9897	97.95%
2.	PH	Constant	0.2783	2.29	0.0446*	0.8976	80.57%

		Age of Dam	0.0123	6.44	0.0001*		
3(a).	TDS	Constant	24.3911	1.51	0.1626	0.6847	46.89%
		Age of Dam	0.7597	2.97	0.0140*		
3(b)	TDS	Age of Dam	1.1160	10.83	0.0000*	0.9562	91.44%

\* Significant at  $\alpha = 0.05$

4.2.2. Age of Dam Impact on Dissolved Oxygen (DO)

According to the regression analysis model 1(b) (Table 7), the value of t-test to regression coefficient with *p-value* = 0.000. The value of p is less than the  $\alpha$  value (0.000 compared to 0.05), inferring that the age of dams has a significant impact on Dissolved Oxygen. The older the dams are, the more the value of DO with the goodness of model value of 97.95% and a compelling correlation of 0.9897. The regression model for the dam impact on DO is:

$$DO\ increase = 0.0487 * age\ of\ dam \quad (11)$$

The increase of dam age in 1 year will result in the increase of DO value by 0.0487 unit.

4.2.3. Age of Dam Impact on pH

According to regression analysis model 2 (Table 7), the value of t-test to regression coefficient with *p-value* = 0.0001. The value of p is less than the  $\alpha$  value (0.0001 compared to 0.05), inferring that the age of dams has a significant impact on pH. The older the dams are, the more the value of pH with the goodness of model value of 80.57% and a strong correlation of 0.8976. The regression model for dam impact on pH is:

$$pH\ increase = 0.2783 + 0.0123 * age\ of\ dam \quad (12)$$

The increase of dam age in one year will result in

the increase of DO value by 0.0123 unit.

4.2.4. Age of Dam Impact on TDS

According to the regression analysis model 3(b) (Table 7), the value of t-test to regression coefficient with *p-value* = 0.014. The value of p is less than the  $\alpha$  value (0.0001 compared to 0.05), inferring that the age of dams has a significant impact on TDS.

The older the dams are, the more the value of TDS with the goodness of model value of 46.89% and a above slightly strong correlation of 0.6847. The regression model for dam impact on TDS is:

$$TDS\ increase = 1.116 * age\ of\ dam \quad (13)$$

The increase of dam age in one year will result in the increased TDS value by 1.116 unit.

4.3. The Analysis of River Basin Area Impact on Reservoir Water Quality

*Hypothesis:*

$H_0$ : River basin areas do not have a significant impact on reservoir water quality\*

$H_1$ : River basin areas have a significant impact on reservoir water quality\*

Table 8 Regression model for river basin area impact on reservoir water quality (Developed by the authors)

Model	Water Quality	Independent Variable	Coefficient	T-value	p-value	Decision
1.	DO	Constant	2.6041	5.71	0.0002	Insignificant
		River Basin Area	0.0019	0.98	0.3517	
2.	PH	Constant	0.887	7.49	0.0000	Insignificant
		River Basin Area	0.0008	1.57	0.147	
3.	TDS	Constant	61.5957	6.17	0.0001	Insignificant
		River Basin Area	0.0516	1.19	0.2612	

According to the regression analysis in Table 8, all models supported the analysis result that the total area of the river basin does not have a significant impact on reservoir water quality.

4.3.1. The Analysis of River Basin Area and Dam Age Impact on Reservoir Water Quality

According to the regression analysis in Table 9, model 1 and model 2 show that there is a significant impact of the age of dam on DO and pH. However, model 3 shows that there is no significant impact of either the river basin area or the age of the dam on TDS.

Table 9 Regression model for river basin area and age of dam impact on reservoir water quality (Developed by the authors)

Model	Water Quality	Independent Variable	Coefficient	T-value	p-value	Decision
1.	DO	Constant	0.1769	0.46	0.6566	Insignificant
		Age of Dam	0.0464	7.19	0.0001*	Significant
		River Basin Area	-0.0001	-0.13	0.9003	Insignificant
2.	pH	Constant	0.2805	2.34	0.0438	Insignificant
		Age of Dam	0.0116	5.77	0.0003*	Significant
		River Basin Area	0.0003	1.12	0.2904	Insignificant
3.	TDS	Constant	0.2837	1.24	0.2465	Insignificant
		Age of Dam	-0.0012	-0.31	0.7605	Insignificant
		River Basin Area	0.0006	1.22	0.2527	Insignificant

\* Significant at  $\alpha = 0.05$

### 4.4. Groundwater Quality Analysis

#### 4.4.1. The Analysis of the Location Impact (in Groundwater Basin or GB) on Groundwater Quality

$H_0$ : Location does not have a significant impact on

groundwater quality.

$H_1$ : Location has a significant impact on groundwater quality.

Table 10 Regression model for location (in GB) impact on groundwater quality (Developed by the authors)

Model	Groundwater Quality	Independent Variable	Coefficient	T-value	p-value	R	R <sup>2</sup>
1(a).	GWL	Constant	0.2429	1.18	0.2671	0.8137	66.21%
		GB	1.4171	4.43	0.0013*		
1(b).	GWL	GB	1.6600	6.67	0.0000*	0.8955	80.19%
2(a).	TDS	Constant	4.2857	1.62	0.1368	0.8994	80.90%
		GB	26.7143	6.51	0.0001*		
2(b).	TDS	GB	31.0000	9.23	0.0000*	0.9411	88.57%
3(a).	pH	Constant	0.1143	1.55	0.1522	0.7769	60.36%
		GB	0.4457	3.90	0.0030*		
3(b).	pH	GB	0.5600	6.50	0.0000*	0.8767	76.86%

\* Significant at  $\alpha = 0.05$

#### 4.4.2. The Location Impact on Groundwater Level (GWL)

According to the regression analysis model 1 (Table 10), the value of t-test to regression coefficient with  $p$ -value = 0.0013. The value of  $p$  is less than the  $\alpha$  value (0.0001 compared to 0.05), inferring that locations have a significant impact on the groundwater level. The goodness of the model value is 66.21% with a strong correlation of 0.81737. The regression models for the impact on GWL are:

$$\begin{aligned} \text{Location (in GB)} &= 0.2429 + 1.4171 \\ \text{Location (outside GB)} &= 0.2429 \end{aligned} \tag{14}$$

#### 4.4.3. The Location Impact on Groundwater TDS

According to regression analysis model 2 (Table 10), the value of t-test to regression coefficient with  $p$ -value = 0.0001. The value of  $p$  is less than the  $\alpha$  value (0.0001 compared to 0.05), inferring that location has a significant impact on TDS. The goodness of the model value is 80.90% with a strong correlation of 0.8994. The regression model for the impact on TDS are:

$$\begin{aligned} \text{Location (in GB) on TDS} &= 4.2857 + 26.7143 \\ \text{Location (outside GB) on TDS} &= 4.2857 \end{aligned} \tag{15}$$

#### 4.4.4. The Location Impact on Groundwater pH

According to the regression analysis model 3 (Table 10), the value of t-test to regression coefficient with  $p$ -value = 0.003. The value of  $p$  is less than the  $\alpha$  value (0.0001 compared to 0.05), inferring that location has a significant impact on pH. The goodness of the model value is 60.36% with a strong correlation of 0.7769. The regression model for the impact on pH are:

$$\begin{aligned} \text{Location (in GB) on pH} &= 0.1143 + 0.4457 \\ \text{Location (outside GB) on pH} &= 0.1143 \end{aligned} \tag{16}$$

### 4.5. Special Analysis for Dams Located inside Groundwater Basin

- a. Groundwater-Level values increase in average to 1.66 m (33.2%) from the previous value.
- b. The TDS values increase on average to 31 ppm (62%) from the previous TDS value.

- c. pH values increase on average by 0.56 (8%) of the previous value.

The correlation between reservoir water TDS and Groundwater pH is shown by the  $r$  value, which is only 0.1475 (feeble). From statistics perspective, the correlation is irrelevant, supported by  $p$ -value of 0.647, which is larger than the  $\alpha$  value of 0.05.

### 4.6. Cluster Analysis

The importance analysis using AHP needs supporting opinion from experts to determine the priority for the variables: reservoir water quality (A), groundwater quality (B) and climate (C).

#### 4.6.1. Dam Grouping Using Cluster Analysis

- a) Two clusters (using complete-linkage method)  
Similarity level: 66.3456%.

Table 11 Cluster centroids for two clusters' analysis (Developed by the authors)

Variable	Cluster 1	Cluster 2	Grand centroid
Dam height (m)	31.5	54.3809	52.474
Dam age (year)	95.0	55.1818	58.500
River Basin Area (km <sup>2</sup> )	699.0	89.3818	140.183

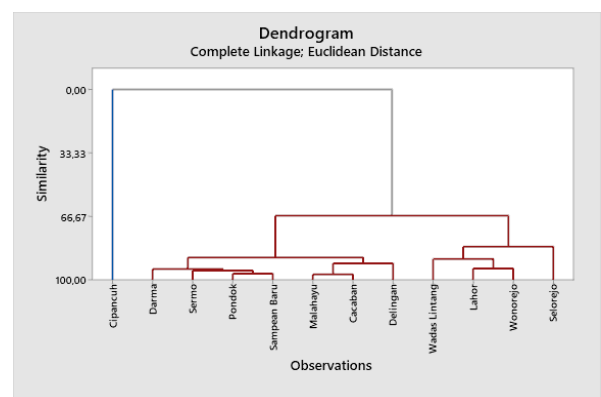


Fig. 3 Cluster result for two clusters analysis (Developed by the authors)

- b) Three clusters (using complete linkage method)  
Similarity level: 82.5436%.

Table 12 Cluster centroids for three clusters' analysis (Developed by the authors)

Variable	Cluster 1	Cluster 2	Cluster 3	Grand centroid
Dam height (m)	31.5	36.1700	86.250	52.474
Dam age (year)	95.0	60.4286	46.000	58.500
River Basin Area (km <sup>2</sup> )	699.0	37.8429	179.575	140.183

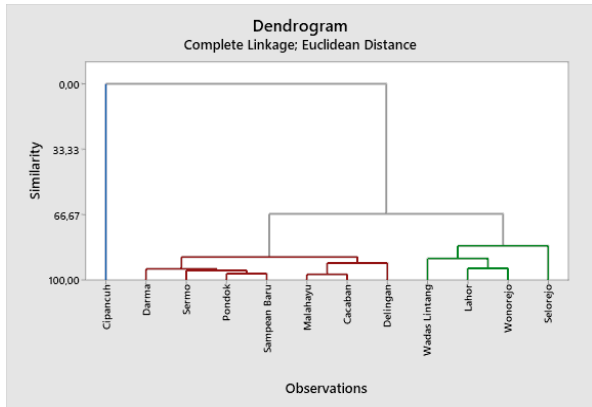


Fig. 4 Cluster result for three clusters' analysis (Developed by the authors)

Table 13 Cluster result for two clusters' analysis (Developed by the authors)

No.	Cluster	Number of Dams	The Name of Dam	Characteristics
1	Cluster 1	1	Cipancuh	The height of dam is insignificant. The dam is very old. The highest total river basin area
2	Cluster 2	11	Darma, Sermo, Pondok, Sampean Baru, Malahayu, Cacaban, Delingan, Wadas Lintang, Lahor, Wonorejo, Selorejo	No significant difference in height. The three highest dams are Wadas Lintang, Lahor and Wonorejo. 80% of dams are less than 75 years old, except Malahayu and Delingan. 72% have less than 100 km <sup>2</sup> river basin area, except Wadas Lintang, Lahor, and Wonorejo

Table 14 Cluster result for three clusters' analysis (Developed by the authors)

No.	Cluster	Numbers of Dam	The Name of Dam	Characteristics
1	Cluster 1	1	Cipancuh	Has a below average height. Very old. Highest river basin area
2	Cluster 2	7	Darma, Sermo, Pondok, Sampean Baru, Malahayu, Cacaban, Delingan	Average height. Not so old. Small river basin area
3	Cluster 3	4	Wadas Lintang, Lahor, Wonorejo, Selorejo	Above average

3	Lahor, Wonorejo, Selorejo	height Young Average river basin area
---	---------------------------	---------------------------------------

### 5. Conclusions

Dams have been used as a supporting infrastructure to supply water in Indonesia. The dam dimensions often mean that dams will have an impact on the local hydrometeorological and hydro (geo) logical system. Twelve dams were investigated in this research with variable characteristics with the aim to investigate their effect on local climate and water quality. From the results in the section above, it was found that linear regression could be used to determine the dam impact on the surrounding area. Furthermore, cluster analysis is beneficial for grouping dams to observe the parameter similarities in each cluster.

Regarding climate, the dam height has a significant impact on the temperature change in the opposite value. The taller the dam, the cooler the temperature. A similar relationship could be seen for precipitation, in which the taller the dam is, the less the precipitation value will be. However, the research revealed no correlation between the river basin area and local climate (temperature and precipitation).

The older the dam is, the higher are the values of DO, pH and TDS in the dam. However, the impact is not significant on TDS. Additionally, there are no correlation between the river basin and water quality.

The research showed a correlation between dams located in groundwater basins with groundwater quality (DO, pH and TDS). The values are higher for dams located in groundwater basins compared to dams outside groundwater basin. Overall, there are correlations between dam infrastructure and climate and water quality in the research area.

The investigated Dams could be clustered into three or two clusters with 82.5 and 66.3 similar level, respectively. Cipancuh is the only cluster that appeared in both two and three clusters using the complete linkage method because of its notable difference in height, age and river basin area with other dams.

### Acknowledgements

The authors would like to offer their gratitude to the University of Brawijaya for their support.

### References

[1] GUO Y, HUANG Q, HUANG S, *et al.* Elucidating the effects of mega reservoir on watershed drought tolerance based on a drought propagation analytical method. *Journal of Hydrology*, 2021, 598. <https://doi.org/10.1016/j.jhydrol.2020.125738>

[2] ZHAO Y, LIU S. & SHI H. Impacts of dams and reservoirs on local climate change: a global perspective. *Environmental Research Letters* 2021, 16(10): 104043.

[3] CHEN J, SHI H, SIVAKUMAR B. & PEART MR. Population, water, food, energy and dams. *Renewable and*

*Sustainable Energy Reviews* 2016, 56: 18-28.

[4] Ministry of Public Works and Public Housing. 2021. <https://data.pu.go.id/dataset/bendungan>

[5] LIU J, & NIYOGI D. Identification of linkages between urban heat Island magnitude and urban rainfall modification by use of causal discovery algorithms. *Urban Climate*, 2020, 33: 100659. <https://doi.org/10.1016/j.uclim.2020.100659>

[6] ZHU G, SANG L, ZHANG Z, *et al.* Impact of landscape dams on river water cycle in urban and peri-urban areas in the Shiyang River Basin: Evidence obtained from hydrogen and oxygen isotopes. *Journal of Hydrology*, 2021, 602: 126779. <https://doi.org/10.1016/j.jhydrol.2021.126779>

[7] LÜ M, JIANG Y, CHEN X. & CHEN J-L. Spatiotemporal Variations of Extreme Precipitation under a Changing Climate in the Three Gorges Reservoir Area (TGRA). *Atmosphere* 2018, 9(1): 24. <https://doi.org/10.3390/atmos9010024>

[8] MILLER JD. *Statistics for data Science: Leverage the Power of Statistics for Data Analysis, Classification, Regression, Machine Learning and Neural Network*. Mumbai: Packt Publishing, 2017.

[9] NIRWANA MB, & WULANDARI D. Comparison of Simple and Segmented Linear Regression Models on the Effect of Sea Depth toward the Sea Temperature. *Enthusiasti: International Journal of Applied Statistics and Data Science*, 2021, 1(2), 68-75. <https://doi.org/10.20885/enthusiastic.vol1.iss2.art3>

[10] PANT A, & RAJPUT RS. Linear Regression Analysis Using R for Research and Development. In *Writing Qualitative Research Paper of International Standard* (pp.180-195), 2019. [https://www.researchgate.net/publication/336981868\\_Linear\\_Regression\\_Analysis\\_Using\\_R\\_for\\_Research\\_and\\_Development](https://www.researchgate.net/publication/336981868_Linear_Regression_Analysis_Using_R_for_Research_and_Development)

[11] JOHNSTONE E C, OWENS D C, LAWRIE S M, *et al.* (ed.). *Companion to Psychiatric Studies*. Churchill Livingstone; 8th edition, 2010

大气 2018, 9(1): 24. <https://doi.org/10.3390/atmos9010024>

[8] MILLER J D. 数据科学统计：利用统计的力量进行数据分析、分类、回归、机器学习和神经网络。孟买：包出版社，2017。

[9] NIRWANA M B 和 WULANDARI D. 简单和分段线性回归模型对海深对海温影响的比较。热情的：国际应用统计和数据科学杂志，1(2)，68-75。 <https://doi.org/10.20885/enthusiastic.vol1.iss2.art3>

[10] PANT A, 和 RAJPUT R S. 使用R进行研究和开发的线性回归分析。在撰写国际标准的定性研究论文（第 180-195 页），2019 [https://www.researchgate.net/publication/336981868\\_Linear\\_Regression\\_Analysis\\_Using\\_R\\_for\\_Research\\_and\\_Development](https://www.researchgate.net/publication/336981868_Linear_Regression_Analysis_Using_R_for_Research_and_Development)

[11] JOHNSTONE E C, OWENS D C, LAWRIE S M 等。（编辑）。精神病学研究的伴侣。丘吉尔·利文斯通；第八版，2010

## 参考文献:

[1] GUO Y, HUANG Q, HUANG S, 等。基于干旱传播分析方法阐明大型水库对流域耐旱性的影响。水文学杂志，2021，598。 <https://doi.org/10.1016/j.jhydrol.2020.125738>

[2] ZHAO Y, LIU S. 和 SHI H. 大坝和水库对当地气候变化的影响：全球视角。环境研究快报 2021, 16(10): 104043。

[3] CHEN J, SHI H, SIVAKUMAR B. 和 PEART M R. 人口、水、食物、能源和水坝。可再生和可持续能源评论 2016, 56: 18-28。

[4] 公共工程和公共住房部。2021. <https://data.pu.go.id/dataset/bendungan>

[5] LIU J, 和 NIYOGI D. 通过使用因果发现算法识别城市热岛强度与城市降雨变化之间的联系。城市气候，2020年，33：100659。 <https://doi.org/10.1016/j.uclim.2020.100659>

[6] ZHU G, SANG L, ZHANG Z, 等。景观大坝对石羊河流域城市和城郊地区河流水循环的影响：来自氢氧同位素的证据。水文学杂志，2021，602：126779。 <https://doi.org/10.1016/j.jhydrol.2021.126779>

[7] LÜ M, JIANG Y, CHEN X. 和 CHEN J-L. 三峡库区（TGRA）气候变化下极端降水的时空变化。