

## Original Research Article

# Rainfall versus- Water Baalance Components using R Software: A Case Study on a Micro-watershed in the Mid-lands of Kerala

### ABSTRACT

The conversion of rainfall to other water balance components is a complex hydrologic process influenced by various storm and drainage characteristics. In the ungauged watershed, measuring the water balance components for planning the water management activities takes much work. Thus, it is essential to understand their relationship in any watershed, which engineers can use for planning various water resources projects. In this study, R software was used to determine these relationships. The Pearson's correlation coefficient ( $r$ ) between rainfall - runoff was 0.83, rainfall - soil moisture was 0.80, rainfall - depth to GW table was 0.72, and rainfall - evapotranspiration was -0.64. The relationship between rainfall vs. runoff, rainfall vs. soil moisture, rainfall vs. depth to GW table, and rainfall vs. evapotranspiration was a polynomial equation with degrees 4, 3, 5, and 3, respectively. The study concludes that by measuring the rainfall, it is possible to estimate most of the other water balance components using the regression estimates.

*Keywords: R software, rainfall, relationship, ungauged watershed, water balance*

### 1. INTRODUCTION

Rainfall and water balance components are closely interconnected and play a vital role in the overall hydrological cycle of an area. Water balance refers to the equilibrium between the inputs, outputs, and storage of water within a defined region, typically measured over a specific period. Rainfall is one of the most significant input factors in the water balance equation. It is the source of water that replenishes water bodies and contributes to various hydrological processes (Biddoccu *et al.*, 2013; Krishnan *et al.*, 2018; Zheng *et al.*, 2021).

Runoff is the portion of precipitation that flows over the land surface and eventually reaches rivers, streams, and other water bodies. The relationship between rainfall and runoff is directly linked. When rainfall intensity exceeds the soil's infiltration capacity or the land's ability to retain water, excess water becomes runoff. Runoff is an essential factor in assessing flooding risks and water availability.

Evapotranspiration is the combined process of water vapour being released into the atmosphere through evaporation from soil and water surfaces, as well as transpiration from plants. The interactive relation between rainfall and evapotranspiration is crucial for determining water availability and the water deficit in a region. If rainfall exceeds evapotranspiration, the surplus water can contribute to runoff and groundwater recharge. Conversely, if evapotranspiration exceeds rainfall, it can lead to a water deficit. Rainfall

**Comment [EV1]:** State the main objective of the study clearly

**Comment [EV2]:** Rephrase sentence to make it clear

**Comment [EV3]:** Mention where you go the data

**Comment [EV4]:** Add citations because these are borrowed ideas so you need to acknowledge source

affects soil moisture content, influencing various water balance components. Adequate soil moisture is necessary for plant growth and maintaining the water balance within the ecosystem. Excessive rainfall can saturate the soil, increasing runoff and reducing infiltration.

**Comment [EV5]:** Add citations because these are borrowed ideas so you need to acknowledge source

Groundwater recharge occurs when water from precipitation percolates through the soil and reaches the groundwater table. The relation between rainfall and groundwater recharge is evident in regions with permeable soils and suitable hydrogeological conditions. Higher rainfall usually leads to increased groundwater recharge, essential for sustaining groundwater levels. The change in storage of water bodies, such as lakes, reservoirs, and wetlands, is influenced by the difference between inflows (primarily from rainfall and runoff) and outflows (evaporation, groundwater discharge, and human withdrawals). Rainfall contributes directly to the inflow component of storage change.

**Comment [EV6]:** Add citations because these are borrowed ideas

Understanding the relationships between rainfall and water balance components is crucial for water resources management, flood prediction, agricultural planning, and ecosystem sustainability. Changes in rainfall patterns due to climate change can significantly impact these relationships and a region's overall water balance (Verma *et al.*, 2020a, 2020b; Lian *et al.*, 2020; Kumar *et al.*, 2021).

The ungauged watershed needs measurement devices to collect hydrological data, such as rainfall, runoff, soil moisture, and groundwater table. Thus/However, there are limitations for hydrologists and environmental scientists to study such watersheds to understand the movement of water through a landscape and manage water resources effectively. However, by carefully considering available information and modelling techniques, it is possible to gain insights into the water dynamics of such areas and make informed decisions regarding water resources management (Abdulla and Al-Shurafat, 2020). The R software is a versatile computerised/computerized statistical package that supports different regression techniques using the R programming language. The integrated features in R software can be used to write, execute, and visualise/visualize regression models efficiently. This study focuses on developing relationships between rainfall and other water balance components of a micro watershed.

**Comment [EV7]:** Explain what an ungauged watershed is?

**Comment [EV8]:** Mention if any attempts have been done to solve the problem and why it failed (if attempts exist) .

**Comment [EV9]:** Can you advance reasons why there are limitations by citing the source of information if any?

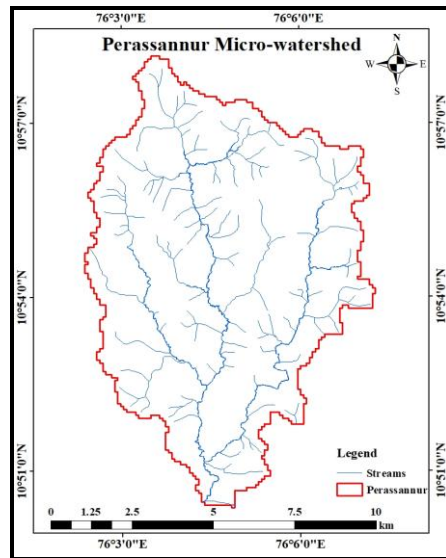
## 2. MATERIALS AND METHODS

### 2.1 Study Area

The Perassannur micro-watershed is located in the Malappuram district of Kerala, between 10°50' and 10°58' N latitude and 76°2' and 76°7' E longitude, as shown in (Fig. 1)-4. The area and perimeter of the water basin were are 79.66 km<sup>2</sup> and 56.60 km, respectively. The average annual rainfall in this area was is 2582.16 mm. The and the average temperature in this region was 27 is 27.38°C, with an annual mean yearly humidity of 77%. Also, the area has an the annual mean annual wind velocity was of 2.16 km.hr<sup>-1</sup>. The climate of the region's climate is humid tropical, and it experiences high humidity during the monsoon months from June to October. The main streams play an essential role in the socio-economic development of society by providing water for domestic agriculture and other activities.

**Comment [EV10]:** Add your main objective of the study

**Comment [EV11]:** Cite the sources of these information if available



**Fig.1. Perassannur Micro-watershed with Drainage Network**

**Comment [EV12]:** Cite the source of the map if you are not the author

## 2.2 Data Acquisition

The study watershed was ungauged, and hence, the water balance components were determined by installing various instruments for measuring rainfall, runoff, soil moisture storage, and groundwater storage for 17 months from August 2021 to December 2022. There were four rain gauges for recording rainfall, a gauging station for recording runoff, soil moisture sensors for recording soil moisture, and a water table sensor for recording groundwater table.

**Comment [EV13]:** Rather give the name of the watershed

## 2.3 R Software

R software is an Integrated Development Environment (IDE) for the R programming language, which is widely used for statistical computing, data analysis, and data visualisation/visualization. It provides a user-friendly interface that helps researchers, data scientists, and analysts work efficiently with R. It is an efficient statistical tool for conducting regression analysis in the R programming language. It provides an interactive and flexible environment for each step of the regression analysis process, from data import to model evaluation. Regression analysis can be efficiently performed by combining R's statistical capabilities with R software's user-friendly interface.

**Comment [EV14]:** Acknowledge idea by citing the source of information if any

## 2.4 Steps Followed for Regression Analysis in R Software

The following are the steps followed for performing regression analysis:

1. **Loading of Data:** Import your data into R Software. You can read data from various file formats such as CSV, Excel, or databases. The read.csv() function is commonly used to read CSV files.

2. Exploring of Data: Exploring and understanding your data is essential before performing regression analysis. R Software provides various tools for data visualisation and summary statistics to help you understand the distribution and relationships among variables.
3. Preprocessing of Data: Clean and preprocess your data as needed. This might involve handling missing values, transforming variables, or dealing with outliers.
4. Choosing Regression Model: Decide on the type of regression model you want. Common types include:
  - a. Linear Regression: Modeling a linear relationship between the dependent and independent variables.
  - b. Multiple Regression: Extending linear regression to include multiple independent variables.
  - c. Logistic Regression: Used for binary or categorical dependent variables.
  - d. Polynomial Regression: Modeling non-linear relationships using polynomial terms.
5. Fitting Regression Model: Use R's regression functions to fit the chosen model to your data. For linear regression, the `lm()` function is typically used. For logistic regression, you might use functions like `glm()`.
6. Assessing the Fitted Model: Evaluate the goodness-of-fit of your regression model using various statistics, such as R-squared for linear regression, deviance for logistic regression, and more.
7. Interpretations of Results: Interpret the regression model coefficients to understand the relationships between the independent and dependent variables. The statistical significance of coefficients can be assessed using p-values.
8. Checking of Assumptions: Ensure that the assumptions of regression are met. These include assumptions about linearity, independence of errors, constant variance, and normality of residuals.
9. Making Predictions: Use the fitted regression model to predict new or unseen data.
10. Visualization of Results: R Software offers various packages for creating plots and visualisations to help you visualise the relationships between variables, the model fit, and the residuals.
11. Reporting and Documentation: R Markdown in R Software can create a document that includes your code, explanations, visualisations, and results in a single file.

**Comment [EV15]:** It is better to report this in a block using paragraphs where necessary rather than using points. But this can still be left out and provided upon request

## 2.5 Development of Relationships between Rainfall and Water Balance Components

The following are the statistical parameters for determining the relations between various water balance components.

### 2.5.1 Pearson's Correlation Coefficient

Pearson's correlation measures the strength of the linear relationship between two variables. It has a value between -1 and +1, with a value of -1 meaning a total negative linear correlation, 0 being no correlation, and +1 indicating an absolute positive correlation. The Pearson's correlation coefficient is given by Eq. 1.

$$r = \frac{\sum_{i=1}^n [(O_i - \bar{O}) * (P_i - \bar{P})]}{\sqrt{\left[ \sum_{i=1}^n (O_i - \bar{O})^2 * \sum_{i=1}^n (P_i - \bar{P})^2 \right]}} \dots(\text{Eq.1})$$

Where,

i is the counter for individual observed and predicted values

n is the number of observations

O<sub>i</sub> is the observed values

$\bar{O}$  is the mean of observed values

P<sub>i</sub> is the predicted values

$\bar{P}$  is the mean of predicted values

R software determined the Pearson's correlation coefficient between all the measured water balance components for monthly and weekly time scales.

**Comment [EV16]:** Mention why this statistical test was used. For instance which variables in your study were correlated?

### 2.5.2 Regression Model

Regression analysis is a powerful technique for uncovering the associations between variables observed in data, but it cannot easily indicate causation. In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

This study developed a regression model after testing the relationship for linear, polynomial, exponential, logarithmic, and power functions in the R-Software software. The model was selected based on the coefficient of determination ( $R^2$ ), shown in Eq. 2, and the p-value obtained from R software.

$$R^2 = \left\{ \frac{\sum_{i=1}^n [(O_i - \bar{O}) * (P_i - \bar{P})]}{\sqrt{\left[ \sum_{i=1}^n (O_i - \bar{O})^2 * \sum_{i=1}^n (P_i - \bar{P})^2 \right]}} \right\}^2 \dots(\text{Eq.2})$$

The polynomial regression analysis appeared better and represents a non-linear relationship between rainfall and other water balance components, meaning a curved best-fit line. The regression plot was made with a 95% confidence interval, and the mathematical equation was determined.

### 2.5.3 Residual vs. Fitted Plot

A residual plot has the residual values on the vertical axis and the independent variable on the horizontal axis. A "residuals versus fitted plot" is frequently created when conducting a residual analysis. It is a scatter plot of residuals on the y-axis and fitted values

(estimated responses) on the x-axis. The plot is used to detect non-linearity, unequal error variances, and outliers. The residuals are calculated from Eq. 3.

$$e_i = y_i - \bar{y}_i \quad \dots \text{(Eq. 3)}$$

Where,

$i$  is the counter for individual observed and predicted values

$e_i$  is the residual

$y_i$  is the predicted value

$\bar{y}_i$  is the observed value

**Comment [EV17]:** Mention why you carried out residual analysis

#### 2.5.4 Normal Q-Q (Quantile-Quantile) Plot

The Normal Q-Q plot is handy in graphically analysing and comparing two probability distributions by plotting their quantiles against each other. If the two distributions are exactly equal, then the points on the Q-Q plot will perfectly lie in a straight line.

The Normal Q-Q plot allows the selection of the points that do not fall close to the reference line. The points on the Normal Q-Q plot indicate the univariate normality of the dataset. If the data is normally distributed, the points will fall on the 45-degree reference line. On the other hand, the points will deviate from the reference line if the data is non-normally distributed.

**Comment [EV18]:** Mention why you used the Q-Q plot in your study

### 3. RESULTS AND DISCUSSION

#### 3.1 Water Balance Components

The results of monthly water balance components for 17 months from August 2021 to December 2022 for the Perassannur watershed are shown in Table 1.

**Table 1. Water Balance Components of the Perassannur Watershed**

Month	Rainfall (mm)	Runoff (mm)	Soil Moisture (%)	Height of Ground Water Table from MSL (m)	Evapotranspiration (mm)
Aug-21	356.63	220.47	29.98	44.99	65.36
Sep-21	309.62	233.27	29.78	44.98	71.90
Oct-21	529.08	412.22	31.21	45.63	69.95
Nov-21	279.18	230.00	30.52	45.18	58.51
Dec-21	12.05	129.18	25.74	44.58	85.32
Jan-22	0.00	70.28	19.95	44.39	102.55
Feb-22	0.00	25.27	17.63	44.26	101.99
Mar-22	35.96	4.70	19.36	43.88	127.17
Apr-22	68.67	1.57	22.84	43.48	113.74
May-22	342.89	77.29	26.90	44.18	97.23
Jun-22	392.19	141.76	28.55	44.81	76.11
Jul-22	502.62	391.95	29.85	45.95	60.38

**Comment [EV19]:** Use the values in the table and discuss possible trends that are of importance. For instance say something about the values in the table

Aug-22	351.32	352.78	28.31	45.11	54.68
Sep-22	180.54	182.20	27.25	44.93	83.57
Oct-22	114.69	108.94	26.12	44.91	82.19
Nov-22	59.58	60.83	26.81	44.74	69.73
Dec-22	135.40	77.04	26.93	44.72	72.40

### 3.2 Relationships between Rainfall and Water Balance Components

#### 3.2.1 Pearson's Correlation Coefficient (r)

Fig. 2 shows the Pearson's correlation coefficient (r) between the monthly water balance components for the Perassannur watershed. It was observed that 'r' between rainfall and runoff was 0.83, rainfall and soil moisture was 0.80, rainfall and depth to GW table was 0.72, and rainfall and evapotranspiration was -0.64. Pearson's correlation coefficient was also observed as 0.76 between runoff and soil moisture, 0.89 between runoff and depth to GW table, and -0.76 between runoff and ET. The 'r' value between soil moisture and depth to the GW table was 0.76, and between soil moisture and ET was -0.85. Also, the 'r' value between depth to GW table and ET was -0.86. This indicates that evapotranspiration negatively correlates with other water balance components. The values of 'r' between monthly rainfall vs. runoff and rainfall vs. soil moisture indicate an excellent correlation; rainfall vs. depth to GW table indicates good. In contrast, rainfall vs. ET shows a weak correlation.

**Comment [EV20]:** What does it mean when r is very close to one

**Comment [EV21]:** Is this the right term to use? Use statistical language

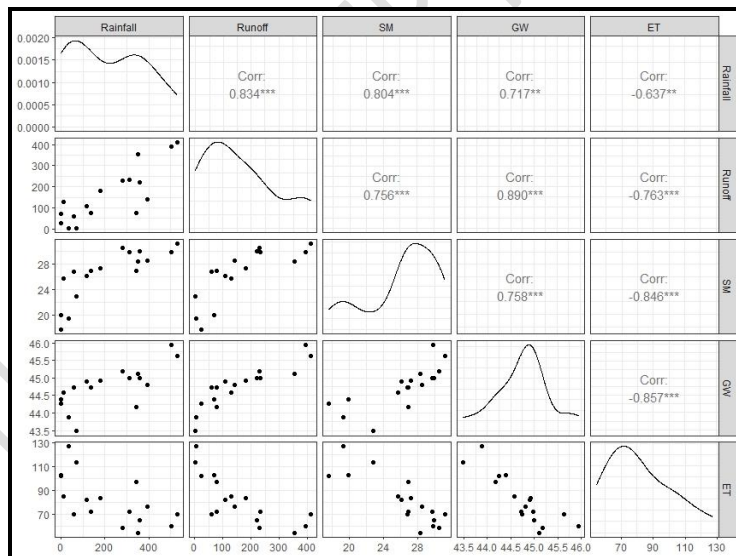


Fig. 2. Pearson's Correlation Coefficient for Perassannur Watershed

#### 3.2.2 Regression Model

The polynomial non-linear regression model was developed between monthly rainfall vs. runoff, rainfall vs. soil moisture, rainfall vs. depth to GW table, and rainfall vs. ET with a 95 % confidence interval for the Perassannur watershed as shown in Fig. 3, Fig. 4, Fig. 5, and Fig. 6, respectively.

The monthly rainfall and runoff relationship for the Perassannur watershed is given by Eq. 4, with  $R^2$  of 0.78 and a p-value of 0.0006, having an F-statistic of 10.64 on 4 and 12 degrees of freedom.

$$121.55x^4 + 34.01x^3 + 85.13x^2 + 434.60x + 159.99 = 0 \quad \dots (\text{Eq. 4})$$

The monthly relationship between rainfall and soil moisture for the Perassannur watershed is given by Eq. 5, with  $R^2$  of 0.77 and a p-value of 0.0001, having an F-statistic of 14.99 on 3 and 13 degrees of freedom.

$$3.49x^3 - 4.72x^2 + 13.14x + 26.34 = 0 \quad \dots (\text{Eq. 5})$$

The monthly relationship between rainfall and depth to GW table for the Perassannur watershed is given by Eq. 6, with  $R^2$  of 0.69 and p-value of 0.01, having an F-statistic of 4.84 on 5 and 11 degrees of freedom.

$$-0.69x^5 + 0.52x^4 + 0.40x^3 + 0.32x^2 + 1.73x + 44.75 = 0 \quad \dots (\text{Eq. 6})$$

The relation between monthly rainfall and evapotranspiration for the Perassannur watershed is given by Eq. 7, with  $R^2$  of 0.45 and a p-value of 0.04, having an F-statistic of 3.62 on 3, and 13 degrees of freedom.

$$-4.93x^3 + 17.57x^2 - 52.06x + 81.93 = 0 \quad \dots (\text{Eq. 7})$$

**Comment [EV22]:** Don't present figures without describing what is happening in the figures. I suggest you match each description and equation with its picture

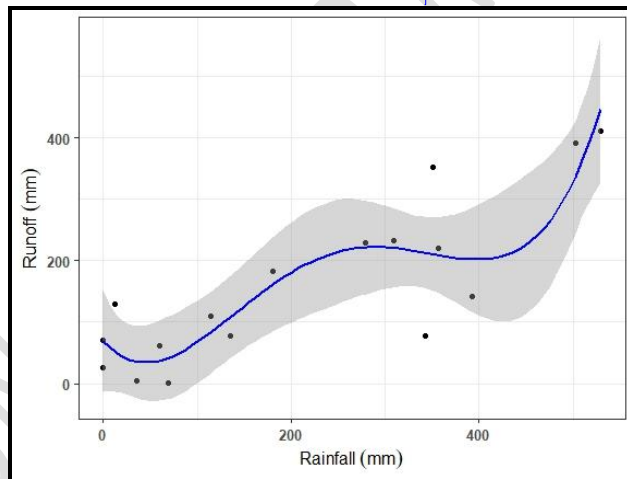
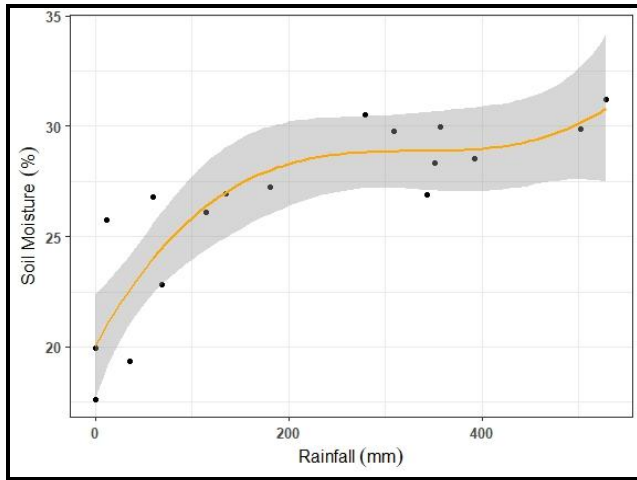
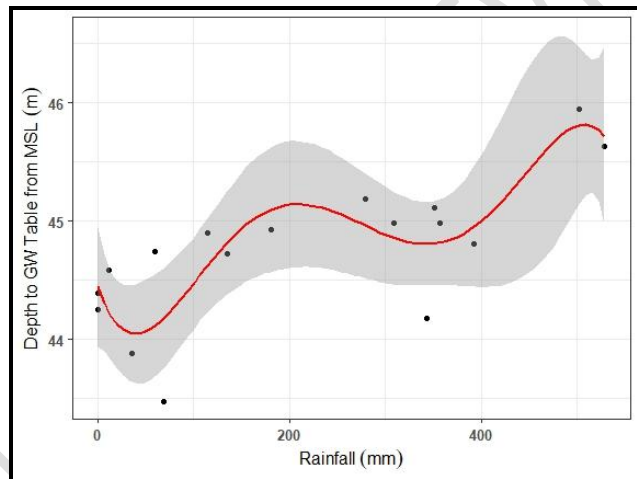


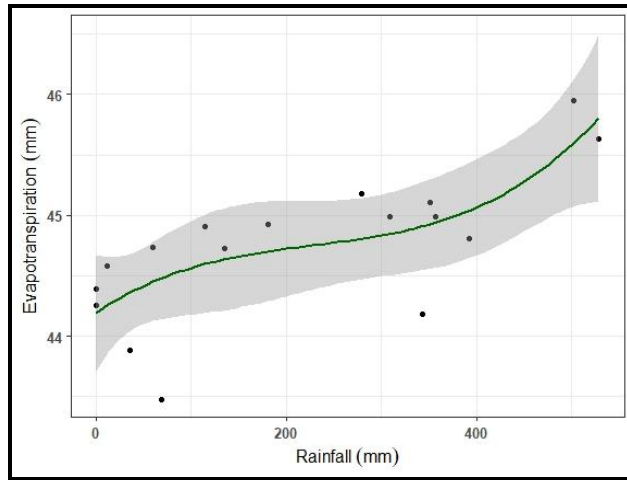
Fig. 3. Relation between Monthly Rainfall and Runoff



**Fig. 4. Relation between Monthly Rainfall and Soil Moisture**



**Fig. 5. Relation between Monthly Rainfall and Depth to GW Table**

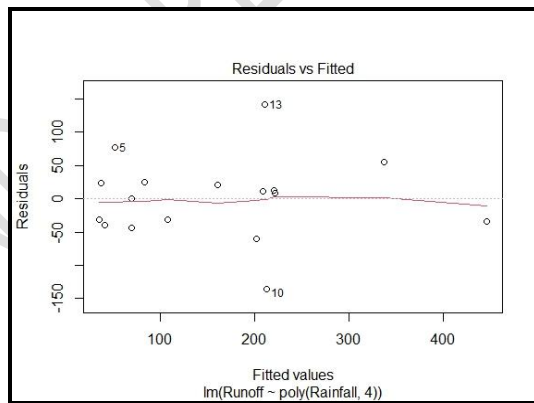


**Fig. 6. Relation between Monthly Rainfall and Evapotranspiration**

**3.2.3 Residual vs. Fitted Plot**

Fig. 7, Fig. 8, Fig. 9 and Fig. 10 show the residual vs. fitted plot between rainfall vs. runoff, rainfall vs. soil moisture, rainfall vs. depth to GW table, and rainfall vs. evapotranspiration for the Perassannur watershed. The x-axis indicates the water balance component's monthly value, and the y-axis indicates the residuals. All these figures show that the fitted line is not a straight line passing through zero residual, indicating non-linearity between the plotted variables.

**Comment [EV23]:** For each figure describe the trends, you don't present a figure without discussing it



**Fig. 7. The Residual vs. Fitted Plot between Rainfall and Runoff**

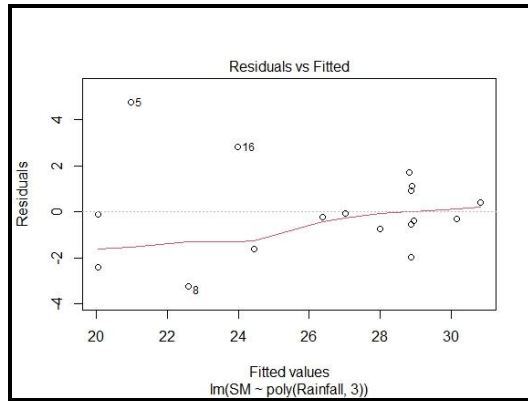


Fig. 8. The Residual vs. Fitted Plot between Rainfall and Soil Moisture

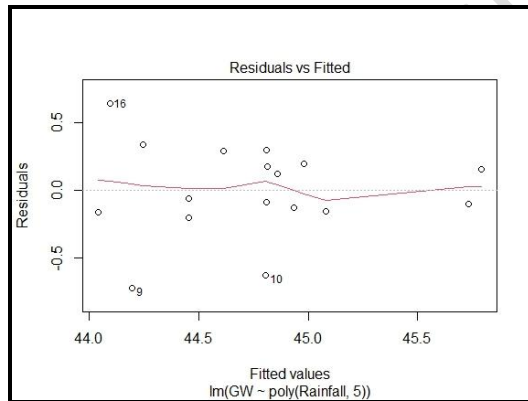


Fig. 9. The Residual vs. Fitted Plot between Rainfall and Depth to GW Table

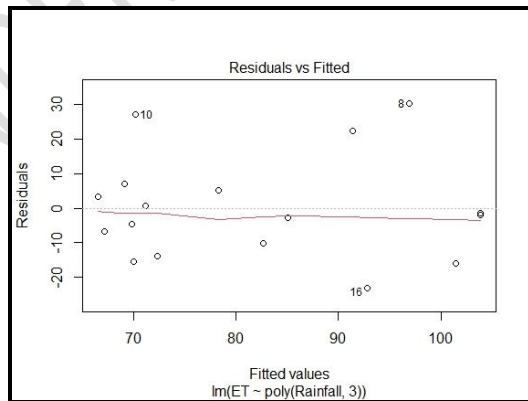
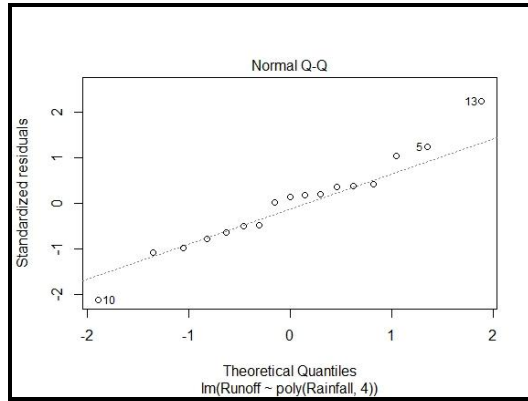


Fig. 10. The Residual vs. Fitted Plot between Rainfall and Evapotranspiration

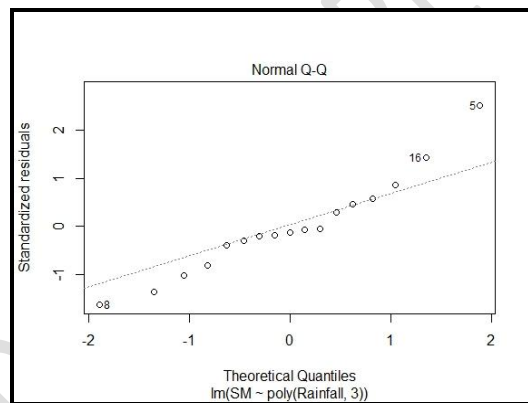
**3.2.4 Normal Q-Q Plot**

Fig. 11, 12, 13 and 14 show the Normal Q-Q plot between rainfall and runoff, rainfall and soil moisture, rainfall and depth to GW table, rainfall and evapotranspiration for the Perassannur watershed. The x-axis indicates the theoretical quantities of the water balance component, and the y-axis shows the standardised residuals.

**Comment [EV24]:** For each table describe the trends



**Fig. 11. The Normal Q-Q Plot between Rainfall and Runoff**



**Fig. 12. The Normal Q-Q Plot between Rainfall and Soil Moisture**

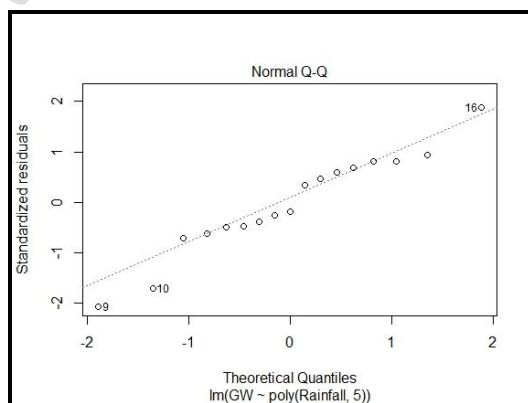


Fig. 13. The Normal Q-Q Plot between Rainfall and Depth to GW Table

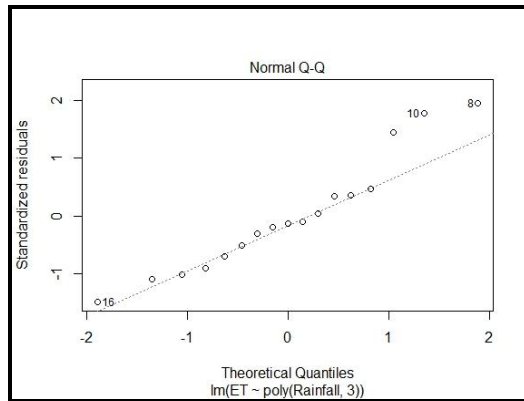


Fig. 14. The Normal Q-Q Plot between Rainfall and Evapotranspiration

#### 4. CONCLUSION

The study ~~could~~ was to determine the water balance components of a micro watershed by directly measuring each hydrological process in the watershed, which otherwise was an ungauged basin. Relationships between rainfall and other water balance components were also developed with the help of R software. It was found that good correlations existed between rainfall and other water balance components, with the 'r' value ranging between 0.70 and 0.83. The maximum correlation was seen between rainfall and runoff, followed by rainfall and soil moisture. The importance of the study lies in the fact that by measuring the rainfall, it is possible to estimate most of the other water balance components using the regression estimates. The professionals and planners can ~~utilize~~ utilize these relationships to conserve and develop water resources.

#### REFERENCES

- Abdulla F, Al-Shurafat AW. Rainfall-Runoff Modeling for Semi-arid and Trans-boundary Yarmouk River Basin. Proc Manuf. 2020;44:180-188.
- Biddoccu M, Ferraris S, Cavallo E, Opsi F, Previati M, Canone D. Hillslope Vineyard Rainfall-Runoff Measurements in Relation to Soil Infiltration and Water Content. Proc Environ Sci. 2013;19:351-360.
- Krishnan G, Kumar KS, Rejani R. Application of SWAT Model for Estimating Runoff of Nethravathi River Basin using Sequential Uncertainty Fitting Technique. Int J Sci Res. 2018;9(3):893-902.
- Kumar A, Kanga S, Taloor AJ, Singh SK, Durin B. Surface Runoff Estimation of Sind River Basin using Integrated SCS-CN and GIS Techniques. Hydro Res. 2021;4:61-74.

**Comment [EV25]:** Where is your discussion. You need to advance arguments to backup your key results or explain why you had the results you had, or give us the meaning of your results. How are your results related to the environment. Lastly you need to compare your results with what those of other authors that is if they are similar or dissimilar.

**Formatted:** Left

**Comment [EV26]:** In your conclusion you conclude based on your results.

**Comment [EV27]:** Use statistical language, what do you mean by good. I will ask again when P is closer to one what does it mean

**Comment [EV28]:** What message can your results pass through? That is what we want to see

**Comment [EV29]:** How?

**Comment [EV30]:** Cross check references to be sure they are properly written

- Lian H, Yen H, Huang C, Feng Q, Qin L, Bashir MA, et al. CN-China: Revised Runoff Curve Number by using Rainfall Runoff Events Data in China. *Water Res.* 2020;177:1-10.
- Verma S, Mishra SK, Verma RK. Improved Runoff Curve Numbers for a Large Number of Watersheds of the USA. *Hydrol Sci.* 2020a;65(16):2658-2668.
- Verma S, Singh PK, Mishra SK, Singh VP, Singh V, Singh A. Activation Soil Moisture Accounting (ASMA) for Runoff Estimation using Soil Conservation Service Curve Number (SCS-CN) Method. *J Hydrol.* 2020b;589:1-26.
- Zheng H, Nie X, Liu Z, Mo M, Song Y. Identifying Optimal Ridge Practices under Different Rainfall Types on Runoff and Soil Loss from Sloping Farmland in a Humid Subtropical Region of Southern China. *Agric Water Manag.* 2021;255:1-11.

UNDER PEER REVIEW