Volume 70, Issue 1, 2026
Journal of Scientific Research
of
The Banaras Hindu University

Statistical Modelling of Groundwater Table using Linear Regression in the semi-arid Sivakasi Region, Tamil Nadu

Vishnupriya S*[1], Gowtham B[2]

*[1] Department of Mathematics, Sri Kaliswari College, Sivakasi, TamilNadu, India, vishnuagilesh03@gmail.com
[2]Department of Civil Engineering, National Engineering College, Kovilpatti,gowthambalu92@gmail.com

Abstract: **The present study investigates the influence of climatic factors on groundwater level (GWL) variations in the Sivakasi region of Tamil Nadu, India, during 2015–2025. Groundwater resources in this area are under severe stress due to extensive agricultural and industrial use. A linear regression model was applied to examine the effects of rainfall and temperature on GWL using data obtained from the Central Ground Water Board. The analysis revealed rainfall as the dominant factor influencing groundwater fluctuations, showing a negative association with GWL, while temperature exhibited a weak positive relationship. The model demonstrated good statistical reliability, confirming that rainfall variation plays a predominant role in variation of groundwater table in the Sivakasi region.**

Index Terms: **Groundwater table, rainfall, regression, temperature, water conservation.**

I. INTRODUCTION

Groundwater is an essential natural resource in India and more than half the population depends on it for daily water needs. And also, it is the largest consumer of groundwater, accounting for about 25% of global extraction. The rural population depends 85% and urban population 50% on the Ground water resources for drinking. The high dependence on the Ground water resources lead to the depletion of ground water table. A study by NASA shows that India's groundwater is declining by an average of 25 cm each year and is particularly severe in states like Rajasthan, Punjab, Haryana, and Tamil Nadu. The Central Ground Water Board (CGWB) mentions that around 1/3rd of country's groundwater assessment units are facing overuse, underscoring the urgent need for water conservation policies. Groundwater levels vary due to factors such as rainfall, soil type, temperature and land use, and making prediction important for creating effective strategies to ensure the long-term availability of this vital resource (Mohammad Ghadir zamani,2022).

Climatic factors, such as rainfall and temperature, primarily drives groundwater recharge and loss. The main source of recharge is Rainfall whereas temperature effect is also accounted as it affects the evapotranspiration rate. This can bring in a significant impact on net water availability (Yan et al.2018). However, the relationship is also complicated by the characteristics of soil, land use, and human activities like groundwater extraction. Simple predictive models, such as multiple linear regression, can offer useful initial insights into these complex relationships. They can help policymakers and water managers to make informed decisions (Dangar et al.2021). This study aims to create a predictive multiple linear regression model to examine how rainfall and temperature influence groundwater levels in the Sivakasi region. It uses an eleven-year dataset to analyze these relationships. The goal is to provide a foundational understanding of the variations in groundwater levels in this area.

II. MATERIALS AND METHODS

A. *Study Area and Data Collection*

Sivakasi taluk, popularly called as Little Japan, is located in the Virudhunagar District of Tamil Nadu, southern India (Meer et al.2024). It is located between 9.45° N and 77.82° E with an elevation of 101m above Mean Sea Level (MSL) (Balaji et al.2013). It is spread around 579.25 km² with 52 villages. According to the 2011 Census, the taluk had a population of 230,505, with a nearly equal gender ratio. It is one of the important revenue division among the taluks of Virudhunagar district (Karunanidhi, D et al.2022). The main source of surface water source is Arjuna river, which is a tributary of the Vaippar

* Corresponding Author

DOI: 10.37398/JSR.2026.700104

basin. It is a flat terrain area with abundant black cotton soil and red soil. The annual rainfall is between 778 mm and 812 mm on an average. It is below the state average of 950 mm and the climate is mostly hot and semi-arid with temperature soaring to 40̊ C during summers. The rainfall mostly occurs during the North-East monsoon from September to December. Agriculture mostly depends on rain. The cropping season is Rabi, which is between August and January. As the area is more arid and drier, dry crops such as maize, cotton, and various pulses (red gram, black gram, Bengal gram, cowpea), along with millets, chilies, and onions are cultivated. And also, horticultural crops like banana, guava, and mango are also grown which relies on Groundwater source for irrigation. However, Over-extraction of water from the ground shows that the aquifer system is critically stressed (Bharani Baanu et al.2024).

The Groundwater data for this study is obtained from the website of Central Ground Water Board (CGWB), which regularly monitors and updates groundwater levels, and water quality across India. The rainfall data are obtained from the Indian Meterological department. These data cover the period from 2015 to 2025 and are given in the Table I.

Fig. 1. TamilNadu District Maps

Fig. 2. Virudhunagar District Map with Sivakasi

Table I. Hydrological data

Observ ations	Year	Rainfall (mm)	Temperat ure ^{0}C	GWL (m)
1	2015	696	36	5.48
2	2016	680	36	5.48
3	2017	720	38	5.27
4	2018	745	36	5.07
5	2019	750	37	5.07
6	2020	660	40	5.92
7	2021	810	34	2.76
8	2022	800	37	2.95
9	2023	820	37	2.9
10	2024	650	39	6.59
11	2025	620	40	7.85

*Courtesy central ground water board, India & Metrological department

Table II. Regression statistics summary

Multiple R	0.965483196
R Square	0.932157801
Adjusted R Square	0.915197251
Standard Error	0.466481107
Observations	11

Table III. *p* and *t-stat* Value based on Regression analysis

Parameters	**Average Rainfall in mm**	**Temperature in Celsius**
Standard Error	0.002950059	0.108940562
t-stat	−7.362419564	0.45391642
p-value	7.89904E-05	0.661949989

Table IV. ANOVA Results

	df	**SS**	**MS**	**F**	**Significance F**
Regression	2	23.919253	11.959626	54.96035315	2.11836E-05
Residual	8	1.7408369	0.2176046		
Total	10	25.660090			

B. Statistical Analysis

A linear regression model was developed to predict the Groundwater Level (GWL). GWL is taken as a dependent variable and the Rainfall and Temperature are taken as independent variables. The model equation for the regression can be written in a linear equation and as follows:

$$Y = mX + C\text{-----------(1)}$$

Where Y is the output (Ground water Table), X is the dependent variables like rainfall, temperature and humidity and C is the intercept. In this study, the effect of rainfall and temperature on ground water table is studied and analysed using standard statistical software Excel. The model's performance was evaluated using regression statistics, including the coefficient of multiple correlation (Multiple R), the coefficient of determination (R^2), and the Adjusted R^2 (MacDonald et al.2012). The overall model significance was determined using the Analysis of Variance (ANOVA) *F-statistic*. The significance of individual predictors was assessed using their *p-values.*

III. Results and Discussion

A. Variation of Groundwater level with respect to rainfall

The variation of groundwater level with rainfall is plotted in the Figure 1. The plot of observed and the predicted values from linear regression analysis is also shown. The observed GWL value is having an inverse relationship with rainfall, meaning that increase in rain over a year has increased the ground water level. It can also be inferred that the aquifer gets recharged effectively during rainy periods. The predicted GWL values are in alignment with the observed data and it implies that the data set is good without major deviations (Maréchal et al.2006). The equation obtained with the regression model for predicting the GWL is given by:

GWL (m) = −0.0226 × Rainfall (mm) + 21.389-------- (2)

The value of −0.0226 indicates that increase in rainfall of 1mm will reduce the groundwater depth by ~0.0226 m, which means that the ground table is raised in this context. The model also has a value of $R^2 = 0.9986$, suggesting that the variance is over 99% for the observed GWL.

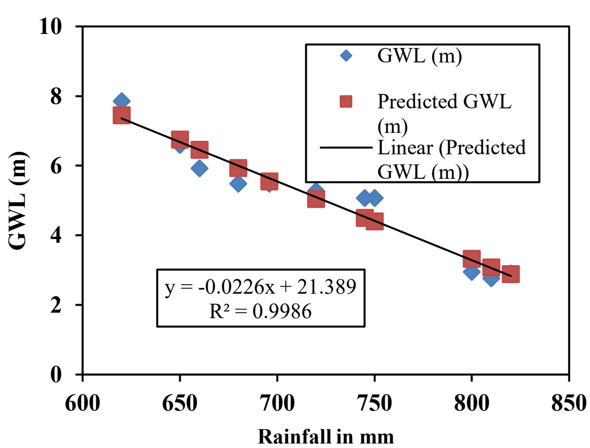

Fig. 3. Variation of GWL with rainfall

B. Residual analysis for the variation of Ground water level

The Figure 2 shows the residual analysis for the GWL variation with rainfall. The deviations has no clear pattern and randomly spread around zero. And it can be seen that the data is linearly distributed. As the distribution was symmetrical, it can be validated that the errors are normally distributed. The abnormalities in certain rainfall ranges may be due to local variances in hydrogeology, abstraction, and evapotranspiration. The regression model provides an effective way to predict groundwater level (GWL) from rainfall data. Diagnostic tests confirm that it is statistically valid. This method gives a useful tool for hydrological forecasting in semi-arid areas. Future enhancements could incorporate other climate and human factors for broader application (Scanlon.2006). The Table II, III and IV show the regression statistics and ANOVA results. From the table, it can be inferred that it supports the alternative hypothesis, indicating that the groundwater level depends on rainfall and that a linear relationship exists. The intercept value is also greater than zero, which supports the alternative hypothesis.

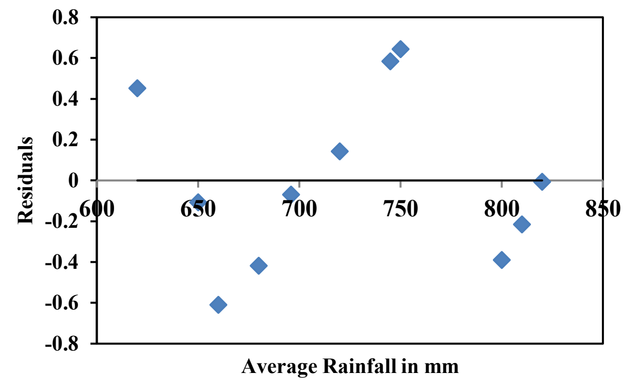

Fig. 4. Residuals for Rainfall

C. Variation of Groundwater level with respect to temperature

The variation of ground water level with the temperature is shown in Figure 3. The plot of observed and the predicted values from linear regression analysis is also shown .The regression equation arrived from the model is as follows:

GWL (m) = 0.5956 × Temperature (°C) - 17.167----- (3)

The slope value of 0.5956 shows that as temperature increases GWL depth is lowering. This may be attributed to the greater evaporation rates. These rates will lessen the groundwater availability because of increased water demand and delays in recharge processes (Putu Doddy Heka Ardana et al.2022). However, the model's value of $R^2 = 0.5069$, indicates that temperature dependency on ground water level variation is only about 51%. From this, it can be understood that the ground water variation is not only attributed with the temperature but also with the other factors such as rainfall, aquifer characteristics, and land use. It can also be validated with the difference between the observed and predicted values (Montgomery, 2021).

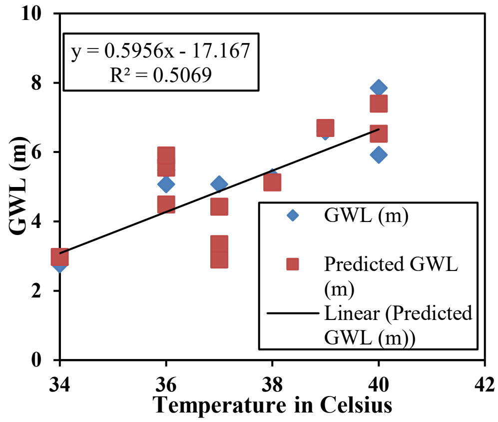

Fig. 5. Variation of GWL with temperature

D. Residual analysis for the variation of Ground water level with respect to temperature

The plot in Figure 4 shows the residual analysis of the model. The residuals are scattered above and below the zero line, within ±0.7 m. The lack of a clear pattern implies that it is linearly distributed. The uniform variance across the temperature range indicates that there occurred symmetrical distribution and it can be accounted that the errors are normally distributed (Poursaeid et al.2022). In the case of modelling, the regression model shows a statistically significant linear connection between temperature and GWL, but it is not as strong as that of rainfall-based models. It can be understood that the temperature can be taken as a secondary predictor, along with rainfall. The temperature is not going to play a predominant role for predicting groundwater in semi-arid areas (Klemen Kenda et al.2020).

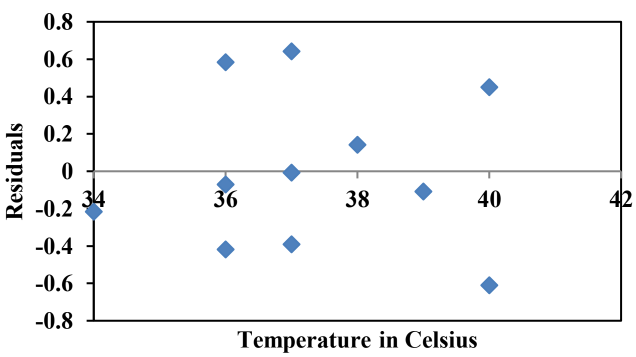

Fig. 6. Residuals for Temperature

E. Standard Residuals

On regression analysis it is found that the standard residuals of the data collected are less than or equal to 3 and it varied between only −1.46 and +1.54. It is shown in Table V. This means there are no outliers within the data and the errors are normally distributed terms (Gowtham B,2025). The residuals varied symmetrically around zero, indicating that the model predictions are unbiased and error variance is consistent. The findings also show that the fitted regression model is statistically strong for predictions (Anantha Natarajan et al.2022). The lack of influential deviations improves the reliability and increases the predictive validity of the proposed model (Kutner, 2005).

Table V: Standard Residuals

Observations	Predicted GWL(m)	Standard Residuals
1	5.55045	−0.16886
2	5.89796	−1.00175
3	5.12808	0.34014
4	4.48619	1.399233
5	4.42704	1.540995
6	6.53015	−1.46239
7	2.97552	−0.51655
8	3.34106	−0.93729
9	2.90667	−0.016
10	6.69790	−0.25861
11	7.39893	1.081075

CONCLUSIONS

The regression analysis showed a strong relationship existed between the predictors such as rainfall and temperature and the Ground water level, with an R^2 value of 0.93. This means that the variance in the outcome of the model is over 93%. The ANOVA results backed up the significance of the model, with $F = 54.96$ and $p < 0.001$. This supports the regression equation. Among the predictors, rainfall is known to play a predominant role in the variation in GWL depth whereas the temperature remains a secondary predictor. From the above observations and analysis, it can be concluded that the model is statistically robust and gives useful insights to the policy makers to plan on the water conservation strategies. As temperature and rainfall are natural factors beyond control, mitigating groundwater depletion in the Sivakasi region requires local water management measures. The authors suggest promoting rainwater harvesting, constructing recharge structures such as percolation ponds and check dams, and adopting water-efficient irrigation practices to enhance groundwater recharge and ensure sustainable water availability.

REFERENCES

Bharani Baanu et al., Climate change and farmers' perception in Sivakasi taluk, India: a nexus and a suggestion for sustainable water reuse, Journal of Water and Climate Change Vol 15(3), 883,2024.

Balaji, S et al., Investigation of the causes for the decline of house sparrow, *Passer domesticus* in Sivakasi Taluk, Virudhunagar District, Tamil Nadu, India. J. Pure Appl. Zool., 1(2): 160-166, 2013.

CGWB, District groundwater Brochure, Virudhunagar district, Tamilnadu Ministry of Water Resources, Government of India, 2008. (https://cgwb.gov.in/)

Dangar, Swarup et al., Causes and implications of groundwater depletion in India: A review. Journal of Hydrology 596,126103, 2021.

Gowtham B., Experimental Investigation and statistical Regression analysis for the reduction of COD from paper and pulp industry waste water using Advanced Oxidation Process. Research Journal of Chemistry and Environment; Vol. 29(5), 2025.

Karunanidhi, D et al., Perchlorate Contamination in Groundwater and Associated Health Risks from Fireworks Manufacturing Area (Sivakasi region) of South India. Expo Health 14, 359–373, 2022.

Klemen Kenda et al., Usage of statistical modeling techniques in surface and groundwater level prediction, Journal of Water Supply: Research and Technology—AQUA, Volume 69(3), 248-265, 2020.

Kutner, M. H et al., Applied Linear Statistical Models. McGraw-Hill/Irwin, 2005.

MacDonald, A. M et al., Quantitative maps of groundwater resources in Africa. Environmental Research Letters, 7(2), 024009, 2012.

Maréchal, J. C., Dewandel, B., Ahmed, S., Galeazzi, L., & Zaidi, F. K, Combined estimation of specific yield and natural recharge in a semi-arid groundwater basin with irrigated agriculture. Journal of Hydrology, 329(1–2), 281–293, 2006.

Meer, M. S. et al., A baseline study on the distribution of fluoride in drinking water and its health risk assessment in Industrial areas of Sivakasi, India. Journal of Applied and Natural Science, 16(1), 1 – 11, 2024.

Mohammad Ghadir zamani et al., Groundwater management in arid and semi-arid regions. Arabian Journal of Geosciences, 15:362, 2022.

Montgomery et al., Introduction to Linear Regression Analysis. Wiley, 2021.

Poursaeid M et al., A Comparative Study of Artificial Intelligence Models and a Statistical Method for Groundwater Level Prediction. Water Resour Manage 36, 1499–1519. 2022.

Putu Doddy Heka Ardana et al., Groundwater Level Forecasting Using Multiple Linear Regression and Artificial Neural Network Approaches. Civil Engineering and Architecture, 10(3),784 – 799, 2022.

Scanlon, B. R et al., Global synthesis of groundwater recharge in semiarid and arid regions. Hydrological Processes, 20(15), 3335–3370, 2006.

V Anantha Natarajan et al., Analysis of Groundwater Level Fluctuations and its Association with Rainfall Using Statistical Methods, Journal of Algebraic Statistics, Volume 13(3), 2022, 1895-1904, 2022

Yan et al., Understanding groundwater table using a statistical model. Water Science and Engineering 11(1), 1-7, 2018.