Fitting of Logistic Model between Rainfall and Ground Water Levels – A Case Study

Raju Sake [1], P. Mohammed Akhtar[2]
[1] Research Scholar, Department of Statistics, Sri Krishnadevaraya University, Anantapuramu,(A.P), India
[2] Professor, Department of Statistics, Sri Krishnadevaraya University, Anantapuramu,(A.P), India

---

***Abstract*—** *Time series analysis is done primarily for the purpose of making forecasts for future and also for the purpose of evaluating past performances. In this paper we have considered the Statistical Analysis of Rainfall and Ground Water Levels in Anantapuramu District of Andhra Pradesh. It deals with the application of Time Series model to analyze and predict Rainfall (RF) and Ground water levels (GWLs) in Anantapuramu district based on the data collected from January 2007 to December 2016. Through with Logistic model by using Hotelling's method for the purpose of analysis the district is divided into five Revenue Divisions (zones). We have estimated or forecast the Logistic model values and compared them by using the data. Further, validation of the fitted model identified the best suitable Revenue Division. i.e., least Mean Square Error (MSE) value of the Revenue Division and forecast on the Rainfall and Ground water levels of this district. We also find the relationship between rainfall and ground water levels in this district and conclusions are drawn based on the results obtained.*

***Keywords*—** *Rainfall, Ground Water Level, Logistic model – Hotelling's method, Validation of the model, forecast.*

---

1. INTRODUCTION

A Time series depicts the relationship between two variables, one of them being time. For example, the temperature (U_t) of a place on different days (t) of the week; rainfall (U_t) of a place on different days (t) of the month; ground water levels (U_t) of a place on different months (t) of the year etc. Where U_t is the value or the phenomenon (or variable) under consideration at time t. The values of t may be given yearly, monthly, weekly, daily or even hourly, usually but not always at equal intervals of time [1,2,3,4,5,6,7,11].

Logistic curve is also known as **Pearl Reed Curve.** It is the most commonly used sigmoidal function to describe population growth, growth of tumors, reaction models in chemistry. There are four types of methods used for the fitting of Logistic curve. viz.., 1. Method of Three selected Time Points. 2. Yule's Method. 3. Hotelling's Method and 4. Method of successive approximations. All these methods of fitting Logistic curve Hotelling's Method is a simple and accurate method [8].

The Logistic curve is generally used in population data of various kinds (both human and non-human). It is also used in economic series relating to industrial growth [8]. Logistic curve is applicable in so many fields like artificial neural networks, biology (especially ecology), bio-mathematics, chemistry, demography, economics, geosciences, mathematical psychology, probability, sociology, political science, linguistics and Statistics.

Water is the main source for all living things for their existence and also any developmental activity. Water resources/water facilities are measured through Rainfall and Ground water levels in any region [9, 10,12]. Identifying the importance of Ground water level a time series analysis is proposed to analyze the data relating to Anantapur district because this district has been a drought prone area since so many decades [13]. In order to forecast Ground water levels of this district, at an attempt is made to predict Rainfall and Ground water levels through Modified exponential model [14] and Gompertz model.

In this direction, Average Ground Water Level (GWL) measured in meters (m) from 194 Piezometer points spread throughout the district and Average Rainfall measured in mille meters (m.m) of the district are considered. The data on the above variables are collected from the records of Ground Water and Water Audit Department Anantapuramu on Ground Water Levels (GWLs) and the data on Rainfall is collected from the Chief planning office, Anantapuramu from 2001 Jan to 2017 Oct. Further, Rainfall data is recorded on daily basis and Ground Water Levels are recorded on monthly basis from the respective records maintained by them [13, 14].

For the present work the data collected, the data relating to January to December months from 2007 to 2016 is considered for the purpose of analysis of this paper on both the variables i.e. Ground Water Level and Rainfall. Further Anantapuramu district consisting of 63 mandals is divided into **five** Revenue Divisions for the administrative convenience and hence for the analysis these five Revenue Divisions are considered as five zones [13,14].

Similarly, Zonal wise or Revenue Division wise Piezometer Points are also provided from which GWLs are measured [13,14].

The data is collected on Average Rainfall and Average Ground Water Levels are given in the following Table-1.1 for a ready reference.

Table-1.1
Average Rainfall and Average Ground water levels data from 2007 to 2016

Year	Zone-I		Zone-II		Zone-III		Zone-IV		Zone-V	
	RF (in mm)	GWL (m)	RF (in mm)	GWL (m)	RF (in mm)	GWL (m)	RF (in mm)	GWL (m)	RF (in mm)	GWL (m)
2007	65.60	10.57	58.20	22.58	67.20	14.23	52.00	14.97	60.50	17.03
2008	53.90	9.96	77.90	20.73	65.20	9.27	61.30	10.88	62.70	9.09
2009	45.40	12.17	50.60	17.53	46.30	11.08	57.10	9.58	38.70	10.24
2010	53.90	12.74	71.50	15.02	70.80	12.03	64.60	8.58	56.30	11.79
2011	39.50	12.69	42.30	15.20	48.90	11.48	31.80	8.93	36.60	12.84
2012	43.20	14.98	43.40	20.49	45.30	16.08	40.50	13.76	41.90	13.22
2013	35.00	15.94	52.30	23.03	47.10	18.69	34.80	16.98	38.10	14.30
2014	31.10	15.87	30.30	23.40	27.10	21.16	37.10	18.92	22.80	16.30
2015	44.10	14.90	62.60	26.88	66.30	25.80	46.00	19.26	54.30	17.66
2016	33.50	15.57	33.40	27.27	32.30	15.35	25.70	19.51	30.10	16.15

2. STATISTICAL ANALYSIS

Some of the Preliminary Statistical analysis is done for the data provided in the above table -1.1, such as yearly averages of Rainfall and Ground water levels are calculated and Karl-Pearson's Correlation Coefficient (r) is calculated between Average Rainfall(X) and Average Ground water levels (Y) Zonal wise [13,14].

By studying the Correlation Coefficients we can observed that all the Correlation Coefficients are negative, that is the relation between Rainfall and Ground Water levels is negative, that is, if Rainfall is increases the Ground water level decreases, it is true, because the depth of the water level will decrease.

To forecast **Rainfall** and **Ground Water Levels** through **Logistic model by using Hotelling's method** for different zones, we consider

The Logistic Model $y_t = \frac{k}{1+e^{a+bt}} \quad , b < 0$ … (2.1)

Where, y_t is the value of time series at time t 'k', 'a' and 'b' are parameters of this curve.

$\frac{dy_t}{d_t} = -by_t(1 - \frac{y_t}{k})$ (Properties of Logistic Curve) … (2.2)

$\Rightarrow \frac{1}{y_t} * \frac{dy_t}{d_t} = -b(1 - \frac{y_t}{k})$ …..(2.3)

If the interval is not too large as an approximation of

$\frac{1}{y_t} * \frac{dy_t}{d_t}$ We can take $\frac{1}{y_t} * \frac{\Delta y_t}{\Delta_t}$ ..….(2.4)

Therefore $\frac{1}{y_t} * \frac{\Delta y_t}{\Delta_t} = -b + \frac{b}{k} * y_t$ …… (2.5)

Let $u = \frac{1}{y_t} * \frac{\Delta y_t}{\Delta_t}, A = -b,\ B = \frac{b}{k}$ …… (2.6)

Then $u = A + By_t$ …...(2.7)

We can estimate A and B through them b and k by least square method.

To fit the above **Logistic model by using Hotelling's method** and to estimate the values of the parameters 'a', 'b' and 'k' by solving the related normal equations and following trend curve is fitted for the data given in table 1.1 and fitted model is given below.

Table-2.1
The fitted Logistic model for Average RF and Average GWLs

Zones	RF Fitted Logistic Curve	GWL Fitted Logistic Curve
Zone-I	$\hat{y}_t = \frac{41}{1+e^{(2.26)+(-0.41)*t}}$	$\hat{y}_t = \frac{(15.50)}{1+e^{(1.71)+(-0.31)*t}}$
Zone-II	$\hat{y}_t = \frac{56}{1+e^{(9.24)+(-1.68)*t}}$	$\hat{y}_t = \frac{24}{1+e^{(1.32)+(-0.24)*t}}$
Zone-III	$\hat{y}_t = \frac{(55.33)}{1+e^{(9.13)+(-1.66)*t}}$	$\hat{y}_t = \frac{(17.50)}{1+e^{(1.93)+(-0.35)*t}}$
Zone-IV	$\hat{y}_t = \frac{44}{1+e^{(2.42)+(-0.44)*t}}$	$\hat{y}_t = \frac{19}{1+e^{(1.05)+(-0.19)*t}}$
Zone-V	$\hat{y}_t = \frac{47}{1+e^{(7.76)+(-1.41)*t}}$	$\hat{y}_t = \frac{14}{1+e^{(3.08)+(-0.56)*t}}$

Therefore $t = 1,2,3,\ldots..$ substitute it then required estimated Logistic curve values for RF and GWL

3. VALIDATION OF THE FITTED MODEL

Validation of the fitted model is necessary to check the suitability of the model for the given data and which is done by considering X = Years and Y = Average RF or Average GWL given in table-1.1 and estimated the Average RF (Y) or Average GWL (Y) denoted by $\hat{y}$. The estimated Average RF and Average GWL are given in the following tables.

Table-3.1
Estimated Average RF $\hat{y}$ for Logistic Curve.

Year	Zone-I		Zone-II		Zone-III		Zone-IV		Zone-V	
	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates
2007	65.60	05.57	58.20	00.03	67.20	00.03	52.00	05.34	60.50	00.08
2008	53.90	07.85	77.90	00.16	65.20	00.17	61.30	07.77	62.70	00.33
2009	45.40	10.79	50.60	00.83	46.30	00.86	57.10	10.99	38.70	01.34
2010	53.90	14.34	71.50	04.17	70.80	04.24	64.60	14.99	56.30	05.04
2011	39.50	18.36	42.30	16.89	48.90	16.80	31.80	19.59	36.60	15.49
2012	43.20	22.54	43.40	39.11	45.30	38.53	40.50	24.41	41.90	31.40
2013	35.00	26.57	52.30	51.83	47.10	51.09	34.80	29.01	38.10	41.92
2014	31.10	30.13	30.30	55.17	27.10	54.47	37.10	33.01	22.80	45.65
2015	44.10	33.08	62.60	55.84	66.30	55.16	46.00	36.23	54.30	46.66
2016	33.50	35.38	33.40	55.97	32.30	55.30	25.70	38.66	30.10	46.92

Table-3.2
Estimated Average GWL $\hat{y}$ for Logistic Curve.

Year	Zone-I		Zone-II		Zone-III		Zone-IV		Zone-V	
	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates
2007	10.57	03.07	22.58	06.08	14.23	02.99	14.97	05.65	17.03	01.04
2008	09.96	03.90	20.73	07.24	09.27	03.96	10.88	06.43	09.09	01.73
2009	12.17	04.87	17.53	08.50	11.08	05.13	09.58	07.26	10.24	02.77
2010	12.74	05.96	15.02	09.86	12.03	06.48	08.58	08.13	11.79	04.22
2011	12.69	07.13	15.20	11.28	11.48	07.96	08.93	09.03	12.84	06.03
2012	14.98	08.33	20.49	12.72	16.08	09.49	13.76	09.93	13.22	07.97
2013	15.94	09.50	23.03	14.14	18.69	10.98	16.98	10.82	14.30	09.78
2014	15.87	10.59	23.40	15.50	21.16	12.33	18.92	11.69	16.30	11.23
2015	14.90	11.57	26.88	16.76	25.80	13.51	19.26	12.53	17.66	12.27
2016	15.57	12.41	27.27	17.92	15.35	14.49	19.51	13.31	16.15	12.96

In the above tables-3.1 and 3.2 for the validation of the model Mean Square Errors (MSE's) are calculated zone wise by considering

$$\textbf{MSE} = \sum (y - \hat{y})^2 \qquad \text{.....(3.1)}$$

Where y represents actual or observed values given in table-1.1 and $\hat{y}$ is the estimated values through fitted Logistic model is given in tables-3.1 and 3.2 using fitted Logistic model respectively. MSE's were calculated and are given in the following Table-3.3.

Table-3.3
MSE's for Average RF- Logistic Model.

Type of the Model	Zone-I	Zone-II	Zone-III	Zone-IV	Zone-V
Logistic	9557.75	18275.55	17730.11	10351.53	12997.83

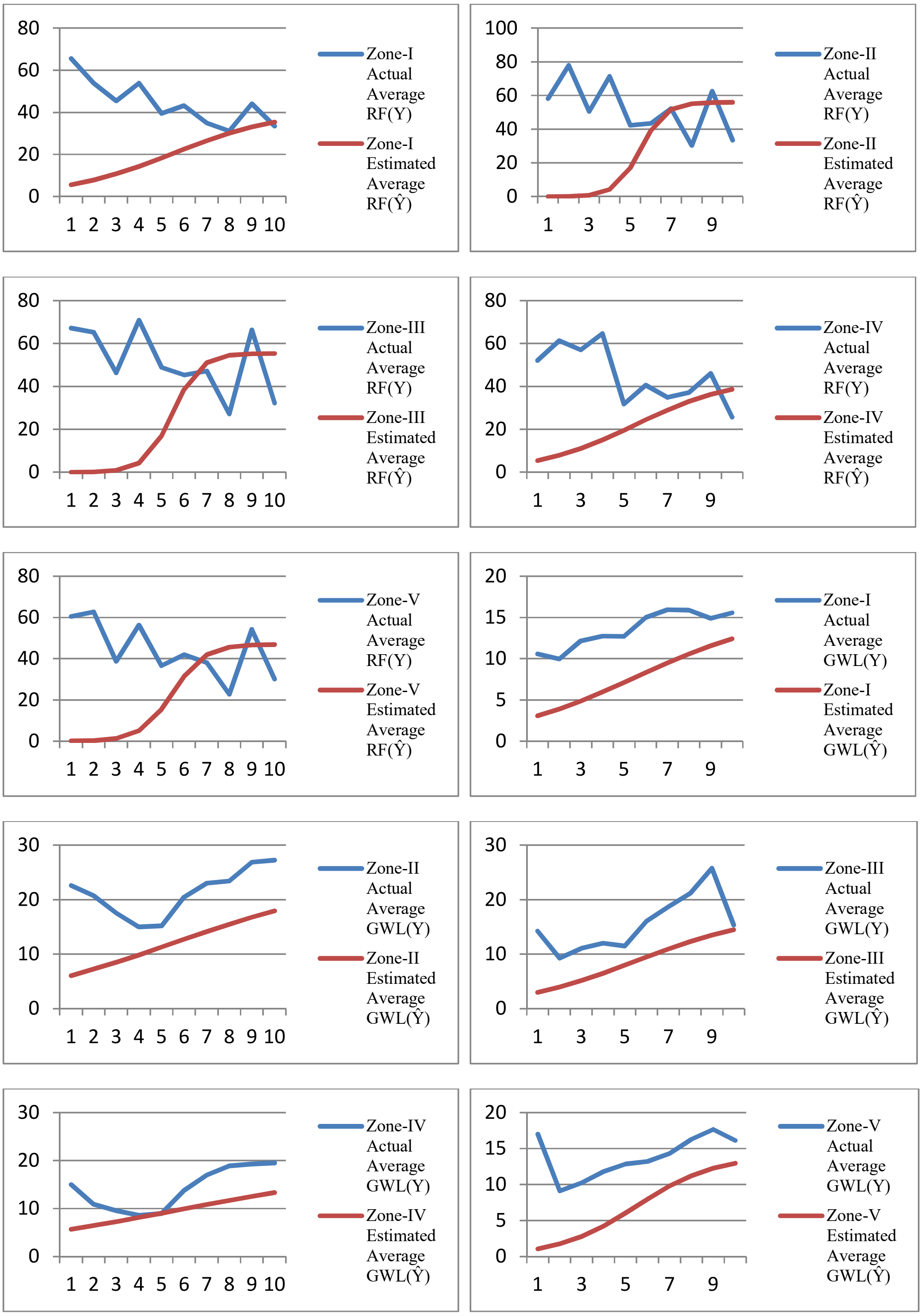
Table-3.4
MSE's for Average GWL – Logistic Model.

Type of the Model	Zone-I	Zone-II	Zone-III	Zone-IV	Zone-V
Logistic	357.79	969.41	565.75	300.87	582.25

By Comparing MSE's for RF and GWLs through Logistic model under consideration, for RF of zone-I is least and also comparing to GWLs for zone-IV is least, Logistic model is the most suitable model because MSEs for zone-I in RF and Zone –IV in GWL is least. In RF next to zone-I, zone-IV has least MSE. Thus in GWL next to zone-IV for zone-I is least MSE. Further, the behaviors of RF and GWL through this model i.e. logistic model in different Zones are represented in the following Figure-3.1. Similar conclusions can be drawn from the following graphs also.

Fig-3.1

Behavior of RF and GWL actual and Logistic Curve Forecasts in Zone –I, II, III, IV and V



Note: In the above graphs x-axis represents years in the last decade i.e. from 2007 to 2016.

On y-axis RF measured in Mille Meters or Average GWLs measured in Meters.

4. FURTHER STATISTICAL ANALYSIS

Now we proceed to analyze the given estimates in tables-3.1 and 3.2 using ANOVA two- way classification by considering rows as different years and columns as different zones and the following Null Hypothesis are formed and tested.

H_{01} : There is no significant difference between different years of Average RF in Anantapuramu District.
H_{02} : There is no significant difference between Average RF of different zones in Anantapuramu District.
H_{03} : There is no significant difference between different years of Average Ground Water Levels in Anantapuramu District.
H_{04} : There is no significant difference between Average Ground Water Levels of different zones in Anantapuramu District.

Table-4.1
ANOVA Two-way Table for RF

Source of variation	**d.f**	**S.S**	**M.S.S**	**F-cal**
Rows (years)	9	16039.17	1782.13	31.31645
Columns (Zones)	4	457.0626	114.2656	2.007931
Error	36	2048.657	56.90714	
Total	49	18544.89		

By comparing F-calculated value of Rows (Years) with F-critical value at 5 % level of significance we reject the H_{01} i.e. There is a significant difference between different years of Average RF in Anantapuramu District. Similarly by comparing F-calculated value of Columns (Zones) with F-critical value at 5 % level of significance we accept the H_{02} i.e. There is no significant difference between different zones of Average RF in Anantapuramu District.

Table-4.2
ANOVA Two-way Table for GWL

Source of variation	**d.f**	**S.S**	**M.S.S**	**F-cal**
Rows (years)	9	616.1695	68.46328	118.8048
Columns (Zones)	4	149.0495	37.26237	64.66165
Error	36	20.74561	0.576267	
Total	49	785.9646		

By comparing F-calculated value of Rows (Years) with F-critical value at 5 % level of significance we reject the H_{01} i.e. There is a significant difference between different years of Average GWL in Anantapuramu District. Similarly by comparing F-calculated value of Columns (Zones) with F-critical value at 5 % level of significance we reject the H_{02} i.e. There is a significant difference between different zones of Average GWL in Anantapuramu District.

Since F-cal value related to rows(years) in RF and rows(years) in GWL is high so there is a necessity for Critical Difference (C.D) Test for sub-grouping various years using the following formula[11][12].

C.D. = $\sqrt{2 \times Error\ M.S.S/m}$ × $t_{0.01}$ for error d.f. in tables -4.1 and 4.2 ......(4.1)

Where m represents number of replicates in each zone and as well as year.

5. CRITICAL DIFFERENCE (C.D) TEST: Average RF for Years

Table-5.1
Year wise Aggregate Average RF for Logistic Curve estimates

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Average	2.21	3.256	4.962	8.556	17.426	31.198	40.084	43.686	45.394	46.446
Ranking	I	II	III	IV	V	VI	VII	VIII	IX	X

Table 5.2
If we can arranged Ascending Order

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Average	2.21	3.256	4.962	8.556	17.426	31.198	40.084	43.686	45.394	46.446

S.E $= \sqrt{2 \times Error\ M.S.S/m}$
= 4.77

1% l.o.f C.D = 2.58×4.77
= 12.31

2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

Above notation indicates that 2007, 2008, 2009, 2010 years Average RF come under one category and 2010, 2011 years Average RF and 2012, 2013 years Average RF and also 2013, 2014, 2015, 2016 come under another category because there is no Significant Difference in average RF. These years are ranked based on their respective Average RF.

CRITICAL DIFFERENCE (C.D) TEST: Average GWL for Years

Table-5.3
Year wise Aggregate Average Ground Water Levels for Logistic Curve estimates

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Average	3.766	4.652	5.706	6.93	8.286	9.688	11.044	12.268	13.328	14.218
Ranking	I	II	III	IV	V	VI	VII	VIII	IX	X

Table 5.4
If we can arranged Ascending Order

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Average	3.766	4.652	5.706	6.93	8.286	9.688	11.044	12.268	13.328	14.218

S.E $= \sqrt{2 \times Error\ M.S.S/m}$
= 0.48

1% l.o.f C.D = 2.58×0.48
= 1.24

2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

Above notation indicates that 2007, 2008 one category, similarly 2008, 2009, and 2009, 2010, and 2013, 2014, and 2014, 2015 and finally 2015, 2016 Average GWLs comes under another category because there is no Significant Difference in average ground water levels. These years are ranked based on their respective average GWLs.

ACKNOWLEDGEMENTS

The authors express heartfelt thanks to the persons responsible for giving the necessary data on Rainfall and Ground Water Levels working in Ground Water and Water Audit Department and Chief Planning Office Anantapuramu. Further the authors are profusely thankful to Mr. C. Pothulaiah, Asst. Hydro geologist, Ground Water and Water Audit Department and Mr. Madhav Reddy, Scientific Assistant PBO Anantapur Meteorological Dept. and Mr. Muralimohan Reddy, Assitant Statistical Officer, Chief Planning Office, Anantapuramu and also Mr. N.V. Subbarao, Mr. Ramakrishna and finally Mr. Muralikrishna from Jaipur in Rajasthan (State) for his timely suggestions and useful discussions. Further I expression sincere thanks to Sri Krishnadevaraya University authorities for providing me the necessary facilities in the Department.

REFERENCES

[1]. A.H. Nury, M. Koch and M.J.B. Alam, "*Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh*", Proceedings of 4th International Conference on Environmental Aspects of Bangladesh, Fukoka, Japan, 2013.

[2]. D. Machiwal and M.K. Jha, "*Time Series Analysis of Hydrologic Data For Water Resources Planning and Management*": A Review. J. Hydrol, Hydromech, 54(3), pp.237-257, 2006.

[3]. D.C Montgomery, L.A Johnson," *Forecasting and time series analysis"*. McGraw-Hill; 1976.

[4]. G.E.P Box, G.M. & G.C Reinsel, "*Time series analysis and forecasting and control"*, 3rd ed., Englewood Cliffs, N.J. Prentice Hall, 1994.

[5]. G.E.P Box, G.M Jenkins, *"Time Series Analysis, Forecasting and Control"*, Holden-Day: San Francisco, 1976.

[6]. G.E.P Box, G.M Jenkins, G.C Reinsel, *"Time series analysis: Forecasting and Control"*. John Wiley and Sons; 2008.

[7]. Otnes RK, Enochson L. *"Applied Time Series Analsis"*, Vol. 1. New York: Wiley; 1978

[8]. P.N. Arora, Sumeet Arora, S. Arora, *"Comprehensive Statistical Methods"*. S.Chand, 2012.

[9]. S. Soltani, R. Modarres and S.S. Eslamian, " *The use of time series modeling for the determination of rainfall climates of Iran*". International Journal of Climatology, 27, pp.819-829, 2007.

[10]. S.A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf and S.F. Vahdat, " *Modelling Of Weather Parameters (Temperature, Rainfall And Humidity) Using Stochastic Methods*. Internat. Conference on Environment and Industrial Innovation, IPCBEE, Singapore, pp.282-285, 2011.

[11]. S.C. Gupta, V.K. Kapoor, *"Fundamentals of Applied Statistics"*. Sultan Chand & Sons; 2003.

[12]. S.M. Ali, "Time *Series Analysis of Baghdad Rainfall Using ARIMA method* "Iraqi Journal of Science,54(4):1136-1142,2013.

[13]. S. Raju, P. Mohammed Akhtar, *"Time Series Analysis on Rainfall and Ground Water Levels Data – A Case Study"*, International Journal of Scientific Research in Mathematics and Statistical Sciences (IJSRMSS) Vol.6, Issue.1, pp.76-85, February (2019).

[14]. S. Raju, P. Mohammed Akhtar, *"Fitting of modified exponential model between rainfall and ground water levels: A case study"*, International Journal of Statistics and Applied Mathematics 2019; 4(4), pp. 01-06.