Journal of Spatial Hydrology

Volume 15 | Article 5

2019

Stochastic variate difference approach for water level discharge data sets at Mantralayam of Andhra Pradesh of India

JOSH JOURNAL OF SPATIAL HYDROLOGY
Journal of Spatial Hydrology Vol.15, No.2 Fall 2019

Stochastic variate difference approach for water level discharge data sets at Mantralayam of Andhra Pradesh of India

S. Sathish[1*], Dr. SK. Khadar Babu[2] and Dr. M. Venkata Ramanaiah[3]
[1,2] Dept. of Mathematics. SAS, VIT University, Vellore, Tamil Nadu, India,
[3]Professor Department of Statistics, Sri Venkateswara University, Tirupathi.

Email id: [1]subramanisathish88@gmail.com, [2]khadar.babu36@gmail.com,
[3]mvrstat@yahoo.com

Abstract: The stochastic variate difference method generates downscaled data in the hydrology and engineering seasonal periods. The new method is developed in the stochastic linear fitted values to fit a straight line to observed data. The paper explains in detail the standard measures such as mean, standard deviation, skewness and kurtosis of the downscaling data in hydrology. The new model - MLE by Gaussian distribution - is a method that was used to find the values to best fit the data sets. Water is one of the sensitive environmental parameters of the hydrological processes. Therefore, the study of water resources exploitation sustainable to the environment is important. Water resources development is an important part of Andhra Pradesh and poses a key issue in the management strategy. This work provides a methodological approach for prediction of the discharge water level for further development and management practices. The present paper predicts future values using stochastic variate difference method for specified fitted values by using downscaling method on daily data.

Keywords: Stochastic process, Seasonal periods, Stochastic variate difference methods, Maximum likelihood estimation.

1. Introduction

The stochastic process in the variate difference method on generated downscaling data is an estimation approach in stochastic statistical analysis. A fundamental problem of time – series analyses is the estimation or description of the relationship among the random variable values at different times. The statistics used to describe this relationship is the autocorrelation. Several estimates of the autocorrelations have been suggested in the past, though a simple and satisfactory estimate was recommended by Jenkins and Watts (1968). The observed value of F is insignificant process similar to stochastic process of good fit to a straight line. The generated downscaling data gave results to fitted values by stochastic variate difference method. The application of stochastic process, which are trend on hydrology, management science, seasonal periods and climate change can be predicted by a stochastic linear normal equation useful in real life. One of the most important differences between streams and lakes lies in the movement or velocity of water (Stendera, 2006). Streams are characterized by high flushing rates and turbulent mixing. Water movement in lakes is slower and influenced by factors such as stratification (Wetzel, 1983). A comparison of the autocorrelation and variance is provided by the case when the joint distribution of the random variables are independent normal variates (Kendall and Stuart, 1966). The aim of this paper is to show the fit to a straight line to the observed data by stochastic variate difference method at mantralayam per day data sets. Wide reading in the variate difference method was expected based on stochastic difference table.

2. Review of literature

The discussions regarding moving average modelling in discussed in the scientific literature, but the rationale for model building was perhaps best expressed by Rosenblueth and Wiener (1945). The mathematical modelling taken over the most important result in problem and predict the future value in the hydrology resources. The applied climatology in the downscaling of the quantity of water falling to earth at a specific place within a specific period of time has already been explained. To identify and quantify the periodicity in the hydrology or climatology time series, the time scale is to be considered less than a year. The Probability Concepts in Engineering Planning and Design has been illustrated by Ang and Tang (1975) and Sathish and Babu (2017). Benefits from using combined dynamical-statistical downscaling approaches lessons from a case study in the Mediterranean region. Leh et al. (2018) has reported that the modeling can be conducted with limited data too. The Probability estimates based on small normal-distribution samples is explained by Bread (1960). The Generated stream flows have been called synthetic to distinguish them from historical observations (Fiering, 1967). The markov decision process to describe stochastic demand by Lautenbacher and Stidham (1999). The activity has been called the stochastic hydrological modelling. The dynamic and temporal hydrologic modeling for the entire watershed is another alternative to predicting hydrological flow (Arnold et al, 2012; Singh, 2015; Singh and Saraswat, 2016). More detailed presentations can be found in Salas and Bartolini (1993).

Regarding the use of time series modelling and forecasting hydrological parameters such as rainfall flow/wind flow, multiple researches have been conducted (Lachtermacher and Fuller, 1994; Singh et al, 2018; Singh and Kumar, 2017). By the previous literature, scholars built the models and then simulate these parameters for analyzing the variations of climatic parameters. The stochastic frontier analysis by means of maximum likelihood and the method of moments is described by Chandra (2011).

3. Description of Study Area

3.1 Geographical location of the studied region

The region is located at Matralayam in Andhra Pradesh on Tungabhadra and covers the area between the villages of Andhra Pradesh and Telangana. The area is identified in Figure 1. Mantralayam is temple city of Andhra Pradesh at a distance of 74 km from Kurnool, 148 km from Nandyal and 253 km from Hyderabad. In Andhra Pradesh, during heavy rains approximately 11 tmc ft of water unutilized into the Bay of Bengal from prakasm barrage. Now, increasing the demand of water, then how to store and predict water levels for future generations is a vital objective in front of the government of Andhra Pradesh. The present models are very useful to predict and forecast the water discharge level of the newly and existed barrage in our study area.

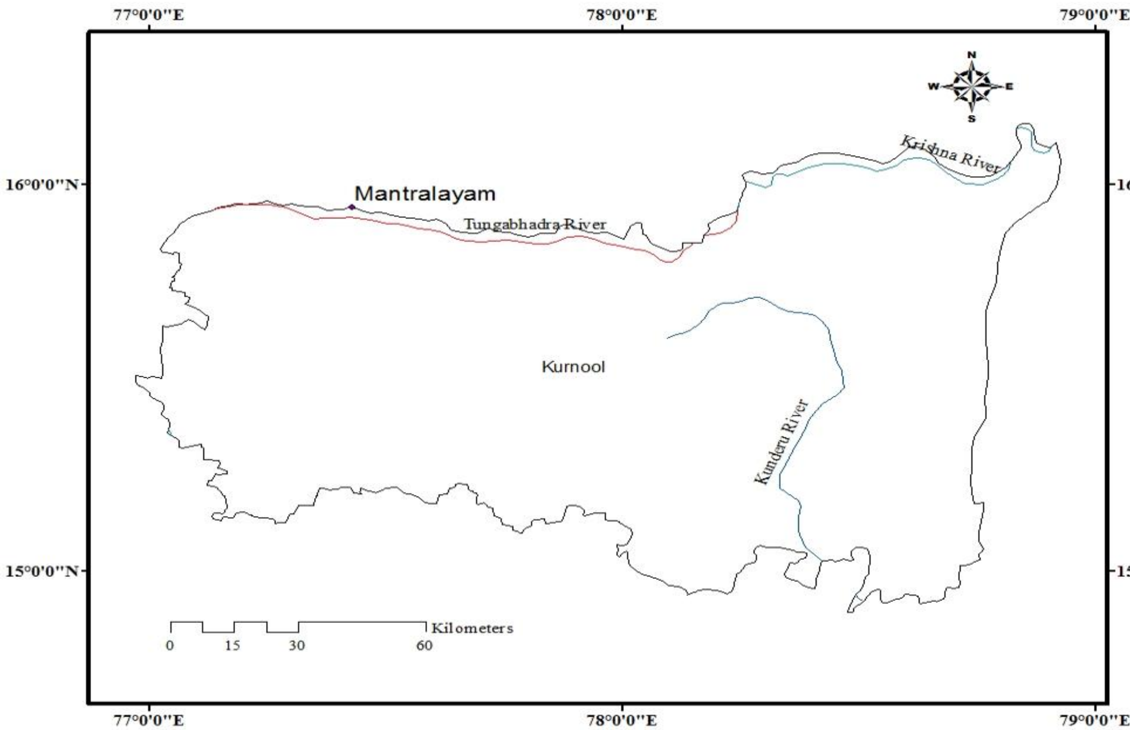


Figure 1. Location studied region at Andhra Pradesh

4. Methodology

4.1 Variate difference method:

The model $y_t = T_t + S_t = R_t + k_t$ (4.1)

We assume that R_t can be fitted to any degree of approximation by a polynomial and that R_t constitutes mainly of the trend values.

We know that $\Delta^q T_t$ = constant this implies $\Delta^{q+i}\, T_t = 0$, i>0

Now $\Delta^m y_t = \Delta^m T_t + \Delta^m k_t$

Hence $V(\Delta^m y_t) = V(\Delta^m T_t) + V(\Delta^m k_t) + 2\mathrm{cov}(\Delta^m T_t, \Delta^m k_t) = V(\Delta^m k_t)$ for $m \geq q$

$$= V[(E-1)m\, k_t] = V\left(\sum_{r=0}^{m}(-1)^{m-r}\binom{m}{r} k_{t+r}\right)$$

$$= \sum_{r=0}^{m}\binom{m}{r}_2 \; V(k_{t+r}) = \binom{2m}{m}\sigma_e^2 \quad (4.2)$$

Assuming k_t's are uncorrelated random variables with $V(k_t) = \sigma_e^2$ $\forall$ t.

$$\text{Thus } \frac{V(\Delta^q y_t)}{\binom{2q}{q}} = \frac{V(\Delta^{q+1} y_t)}{\binom{2q+2}{q+1}} = \frac{V(\Delta^{q+2} y_t)}{\binom{2q+4}{q+2}} = \ldots\ldots = \sigma_e^2 \text{ (a constant).} \quad (4.3)$$

Therefore, one has to calculate $\frac{V(\Delta^r y_t)}{\binom{2r}{r}}$, r = 1, 2 ,…..

And find when these ratios stabilize to a constant. To ensure the equality of estimates two variances

$$\frac{V(\Delta^r y_t)}{\binom{2r}{r}} \text{ and } \frac{V(\Delta^{r'} y_t)}{\binom{2r'}{r'}}, r' > r$$

a F-test has to be done. However, such a test is based on two independents χ^2- statistic.

Therefore, we must have two independent estimates of σ_e^2.

$$\Delta^q y_j, \Delta^q y_{j+2q+3}, \Delta^q y_{j+2(2q+3)}, \ldots \quad (4.4)$$

Hence
$$\hat{V}(\Delta^q y_t) = \frac{2q+3}{(N-q)} \sum_{\substack{t=j,j+2q+3, \\ j+2(2q+3),\ldots \\ j=1,2,\ldots}} (\Delta^q y_t)^2 = V' \quad (4.5)$$

Similarly,

$$\hat{V}(\Delta^{q+1} y_t) = \frac{2q+3}{(N-q-1)} \sum_{\substack{t=j+k+1,j+k+1+(2q+3), \\ j+k+1+2(2q+3),\ldots \\ j=1,2,\ldots}} (\Delta^q y_t)^2 = V'' \quad (4.6)$$

Hence
$$Ғ = \frac{V'\binom{2q+2}{q+1} n_2}{V''\binom{2q}{q} n_2} \sim Ғ_{n1,n2,..} \quad (4.7)$$

Where n_1 is the number of observations on which V' is based and n_2 is the number of observations on which V'' is based. If the observed value of Ғ is insignificant at 100% level, then the data are consistent with the null – hypothesis and a polynomial of degree q can be fitted to the required degree of approximation.

4.1.1 Maximum likelihood estimation for discharge data using Gaussian distribution:

Maximum likelihood estimation is a method that determine values for the parameters of a model. The parameters values are found such that they maximize the likelihood that the process described by the model produced the data that were actually observed. The generation process can be adequately described by a Gaussian distribution. The Gaussian distribution has two parameters: the mean μ and the standard deviation σ. Maximum likelihood estimation is a method that we will find the values of μ and σ that result in the curve that fit the data. In 18 per day discharge data points at time and we assume that they have been generated from a process that is adequately described by a Gaussian distribution.

5. Results and Discussion

Discharge data at mantralayam – Downscaling data (1995 January)

Stochastic variate difference method is shown in Table 1.

Table 1. Discharge level data sets in variate difference method

t	Discharge (m3/s) Water level-s_t	Δ_{st}	Δ_{st}^{2}	Δ_{st}^{3}	Δ_{st}^{4}	t'	$t'st'$	$t'^2 st'$
1	54.43					-15	-816.45	12246.7
		-2.98						
2	51.45		-4.88			-14	-720.3	10084.2
		-7.86		11.93				
3	43.59		7.05		-16.11	-13	-566.67	7366.7
		-0.81		-4.18				
4	42.78		2.87		-4.93	-12	-513.36	6160.3
		2.06		-9.11				
5	44.84		-6.24		17.25	-11	-489.28	5425.6
		-4.18		8.14				
6	40.66		1.9		-6.28	-10	-406.6	4066
		-2.28		1.86				
7	38.38		3.76		-2.76	-9	-345.42	3108.7
		1.48		-0.9				
8	39.86		2.86		-7.02	-8	-318.88	2551.04
		4.34		-7.92				
9	44.2		-5.06		11.31	-7	-309.4	2165.8
		-0.72		3.39				
10	43.48		-1.67		0.692	-6	-260.88	1565.2
		-2.39		4.082				
11	41.09		2.412		-6.516	-5	-205.45	1027.2
		0.022		-2.434				
12	41.112		-0.022		7.584	-4	-164.44	657.7
		0		5.15				
13	41.112		5.128		3.064	-3	-123.33	370.008
		5.128		8.214				

14	46.24		13.342		-36.956	-2	-92.48	184.96
		18.47		-28.74				
15	64.71		-15.4		37.812	-1	-64.71	64.71
		3.07		9.07				
16	67.78		-6.33		26.02	0	0	0
		-3.26		35.09				
17	64.52		28.76		-8.77	1	64.52	64.52
		25.5		26.32				
18	90.02		55.08		-207.08	2	180.04	360.08
		80.58		-180.76				
19	170.6		-125.6		332.04	3	511.8	1535.4
		-45.1		151.28				
20	125.5		25.6		-180.97	4	502	2008
		-19.5		-29.69				
21	106		-4.09		48.37	5	530	2650
		-23.59		18.68				
22	82.41		14.59		10.2	6	494.46	2966.7
		-9		-8.48				
23	73.41		6.11		2.64	7	513.87	3597.09
		-2.89		-5.84				
24	70.52		0.27		5	8	564.16	4513.28
		-2.62		-0.84				
25	67.9		-0.57		-6.91	9	611.1	5499.9
		-3.19		-7.75				
26	64.71		-8.32		25.86	10	647.1	6471
		-11.51		18.11				
27	53.2		9.79		-13.15	11	585.2	6437.2
		-1.72		4.96				
28	51.48		-4.83		6.29	12	617.76	7413.1
		-6.55		11.25				
29	44.93		6.42			13	584.09	7593.1
		-0.13						
30	44.8					14	627.2	8780.8

Observed values are $\Delta^p s_i$, $\Delta^p s_{i+2p+3}$, $\Delta^p s_{i+2(2p+3)}$

S (x , t) : Predictor V (x', t) : Predictand

Normal equation on

Stochastic difference (S, Ƒ, Δ)

i ,j = 1 p , q = 1

A Stochastic process denoted by {X (e,t) ; eϵS, tϵT} , where T is time St' = a + bt' , The stochastic variate difference method by the formula p , q =1 and i , j =1

Let Δ_{s1} , Δ_{s6} , Δ_{s11} , Δ_{s16} , Δ_{s21} , Δ_{s26}

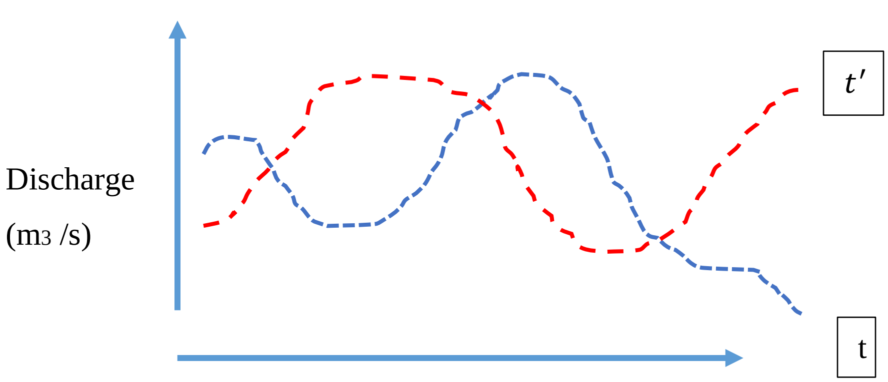
$$S = V' = (\Delta^P S_t) = \frac{(2p+3)}{(\text{N}-\text{P})} \sum_{t=i,\, i+2p+3, i+2(2p+3),\ i=1,2,...}$$

Similarly we consider Δ^2S3 , Δ^2S8 , Δ^2S13 , Δ^2S18 , Δ^2S23

$$S = V'' = (\Delta^{P+1} S_t) = \frac{(2p+3)}{(\text{N}-\text{P}-1)} \sum_{t=i+k+1,\ i+k+1+(2q+3),\ i+k+1+2(2p+3),\ i=1,2,...}$$

A set of data can be observed

Example:



data t and t' are for different downscaling data on per day

$$\sigma_1^2 = \frac{V'}{2} = \frac{123.047}{2} = 61.52 \quad , \quad \sigma_2^2 = \frac{V''}{6} = \frac{2865.82}{6} = 477.63, (S, Ƒ, \Delta) = Ƒ = \frac{61.52 \times 6}{6 \times 477.63} = \frac{369.14}{2865.78} = 0.1288$$

$$Ƒ_{6,6\,;\,0.95} = 4.28 \qquad Ƒ_{6,6\,;\,0.99} = 8.47$$

The observed value of Ƒ is insignificant in that p = 1 and fit a straight line to the observed data changing t to t' = t-16 by the Stochastic process on St' = a + bt'. The normal equations are

$$\sum st' = N_{a0} + a_1 \sum t'$$
$$\sum t'st' = a_0 \sum t' + a_1 \sum t'^2$$
, 1855.71 = 30 a_0 , $a_0 = \frac{1855.71}{30} = 61.85$ and $a_1 = \frac{1697.55}{2253} = 0.7534$

Fitted value St' = a + bt' is shown in Table 2 and generated downscaling data is shown in Figure 2.

Table 2. Fitted values for Discharge level data sets

t	St	T	St	t	St	t	St	t	St
1	50.54	7	55.06	13	59.58	19	64.11	25	68.63
2	51.30	8	55.82	14	60.34	20	64.86	26	69.38
3	52.05	9	56.57	15	61.09	21	65.61	27	70.13
4	52.80	10	57.32	16	61.85	22	66.37	28	70.89
5	53.56	11	58.08	17	62.60	23	67.12	29	71.64
6	54.31	12	58.83	18	63.35	24	67.87	30	72.39

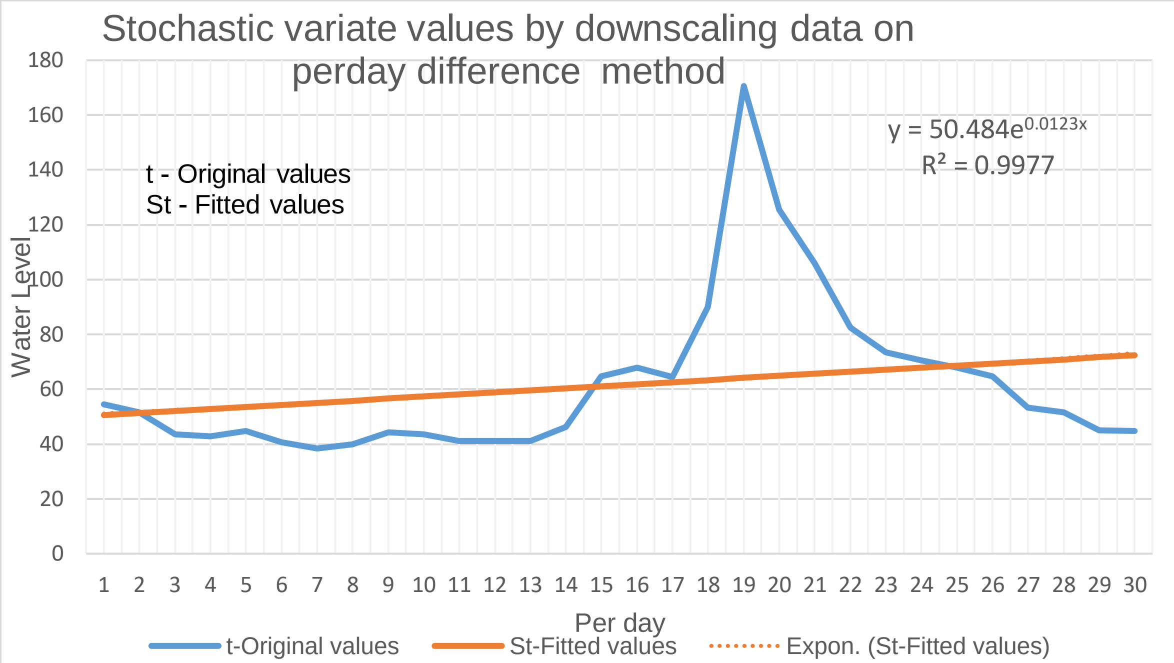


Figure 2. Generated downscaling data at Mantralayam (1995-January)

In the above discharge level, we conclude mean = 61.85 and standard deviation = 29

Calculate the maximum likelihood estimates of the parameter values of the Gaussian distribution μ and σ January-1995:

$$P(x;\ \mu, \sigma) = \prod_{x=1}^{30} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$P(30\text{Perdays};\ \mu, \sigma) = \prod_{x=1}^{30} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(54.43-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(51.45-\mu)^2}{2\sigma^2}\right) \times \ldots$$

$$\ldots\ldots\times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(44.8-\mu)^2}{2\sigma^2}\right)$$

Taking logs of the original expression is given us

$$\text{'In}\,(P(x;\ \mu, \sigma)) = \sum_{x=1}^{30} \text{In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(-54.43-\mu)^2}{2\sigma^2} + \text{'In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(51.45-\mu)^2}{2\sigma^2} + \ldots\ldots$$

$$\ldots\ldots + \text{'In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(44.8-\mu)^2}{2\sigma^2}$$

We can simplify above expression using the laws of logarithms to obtain

$\text{ʹln}\,(P(x;\ \mu\,,\sigma)) = -30\,\text{ʹln}\,(\sigma) - \frac{30}{2}\,\text{ʹln}\,(2\pi) - \frac{1}{2\sigma^2}[(54.43-\mu)^2 + (51.45-\mu)^2 + \ldots\ldots\ldots$

$\ldots\ldots + (44.8-\mu)^2]$

This expression can be differentiated to find the maximum. We can find the MLE of the mean μ . To do this we take the partial derivative of the function with respect to μ , giving

$$\frac{\partial\text{ʹln}(P(x;\ \mu\,,\sigma))}{\partial\mu} = \frac{1}{\sigma^2}\,(54.43 + 51.45 + \ldots + 44.8 - 30\mu)$$

Finally, setting the left-hand side of the equation to zero and rearranging for μ gives

$$\mu = \frac{54.43+51.45+\cdots+44.8}{30} = 61.85$$

Similarly, we have to get σ

$$\frac{\partial\text{ʹln}(P(x;\ \mu\,,\sigma))}{\partial\ \sigma} = \frac{1}{2\ \sigma}\,(54.43 + 51.45 + \ldots + 44.8 - \mu)$$

$$2\ \sigma = [-0.48+0.29+\ldots+0.33\text{—Fitted values }(1914.18)]$$

$$\sigma = 29.2,\ \mu = 61.85,\ \sigma = 29.2$$

Comparison of Stochastic variate difference method and MLE method:

Results:

(i) Stochastic variate difference method

Mean = 61.85

Standard deviation = 29

(ii) Maximum likelihood estimation

Mean = 61.85

Standard deviation = 29.2

Discharge data at Mantralayam – Downscaling data (2016 January)

Stochastic variate difference method is shown in Table 3.

Table 3. Discharge level data sets in variate difference method

t	Discharge (m3/s) Water level-s_t	Δ_{st}	Δ_{st}^2	Δ_{st}^3	Δ_{st}^4	$t^{'}$	$t^{'}st^{'}$	$t^{'2}st^{'}$
1	16.82					-15	-252.3	3784.5
		-1.02						
2	15.80		0.35			-14	-221.2	3096.8
		-0.67		-2.71				
3	15.13		-2.36		42.52	-13	-196.69	2556.97
		-3.03		39.81				
4	12.10		37.45		-112.3	-12	-145.2	1742.4
		34.42		-72.49				
5	46.52		-35.04		106.77	-11	-511.72	5628.92
		-0.62		34.28				
6	45.90		-0.76		-33.81	-10	-459	4590
		-1.38		0.47				
7	44.57		-0.29		12.54	-9	-401.13	3610.17
		-1.67		13.01				
8	42.90		12.72		-38.13	-8	-343.2	2745.6
		11.05		-25.12				
9	53.95		-12.4		44.96	-7	-377.65	2643.55
		-1.35		19.84				
10	52.60		7.44		-34.68	-6	-315.6	1893.6
		6.09		-14.84				
11	48.69		-7.4		22.43	-5	-243.45	1217.25
		-1.31		7.59				
12	47.38		0.19		-7.14	-4	-189.52	758.08
		-1.12		0.45				
13	46.26		0.64		2.29	-3	-138.78	416.34
		-0.48		2.74				
14	45.78		3.38		-17.9	-2	-92.56	183.12

		2.9		-15.16				
15	48.68		-11.78		38.54	-1	-48.68	48.68
		-8.88		23.38				
16	39.80		11.6		-43.36	0	0	0
		2.72		-19.98				
17	42.52		-8.38		20.51	1	42.52	42.52
		-5.66		0.53				
18	36.86		-7.85		15.37	2	73.72	147.44
		-13.51		15.9				
19	23.35		8.05		-21.25	3	70.05	210.15
		-5.46		-5.35				
20	17.89		2.7		2.92	4	71.56	286.24
		-2.76		-2.43				
21	15.13		0.27		2.41	5	75.65	378.25
		-2.49		-0.02				
22	12.64		0.25		0.98	6	75.84	455.04
		-2.24		0.96				
23	10.40		1.21		-1.14	7	72.8	509.6
		-1.03		-0.18				
24	9.37		1.03		1.84	8	74.96	599.68
		0		1.66				
25	9.37		2.69		-7.04	9	84.33	758.97
		2.69		-5.38				
26	12.06		-2.69		8.65	10	120.6	1206
		0		3.27				
27	12.06		0.58		-3.83	11	132.66	1459.26
		0.58		-0.56				
28	12.64		0.02		0.55	12	151.68	1820.16
		0.6		-0.01				
29	13.24		0.01			13	172.12	2237.56
		0.61						
30	13.85					14	193.9	2714.6

$\sigma_1^2 = \frac{V'}{2} = \frac{3.14817}{2} = 1.574$, $\sigma_2^2 = \frac{V''}{6} = \frac{291.70}{6} = 48.61$, (S, Ƒ, Δ) = Ƒ = $\frac{1.574 \times 6}{6 \times 48.61} = \frac{9.444}{291.66}$ = 0.0323

$Ƒ_{6,6\,;\,0.95} = 4.28$ $Ƒ_{6,6\,;\,0.99} = 8.47$

The observed value of Ƒ is insignificant in that p = 1 and fit a straight line to the observed data changing t to t' = t-16 by the Stochastic process on St' = a + bt'

The normal equations are

$$\sum st' = N_{a0} + a_1 \sum t'$$
$$\sum t' st' = a_0 \sum t' + a_1 \sum t'^2$$

$864.26 = 30\ a_0$, $a_0 = \frac{864.26}{30} = 28.80$ and $a_1 = \frac{2553.09}{2255} = 1.132$

Fitted value St' = a + bt' is shown in Table 4 and the generated downscaled data as graph shown in Figure 3.

Table 4. Fitted values for Discharge level data sets

t	St	T	St	t	St	t	St	t	St
1	11.92	7	18.96	13	25.79	19	31.19	25	37.98
2	12.95	8	19.74	14	26.98	20	32.32	26	38.12
3	14.95	9	20.98	15	27.86	21	33.46	27	39.25
4	16.52	10	22.43	16	28.87	22	34.98	28	40.38
5	16.80	11	23.84	17	29.93	23	35.82	29	41.51
6	17.68	12	24.98	18	30.96	24	36 .85	30	42.64

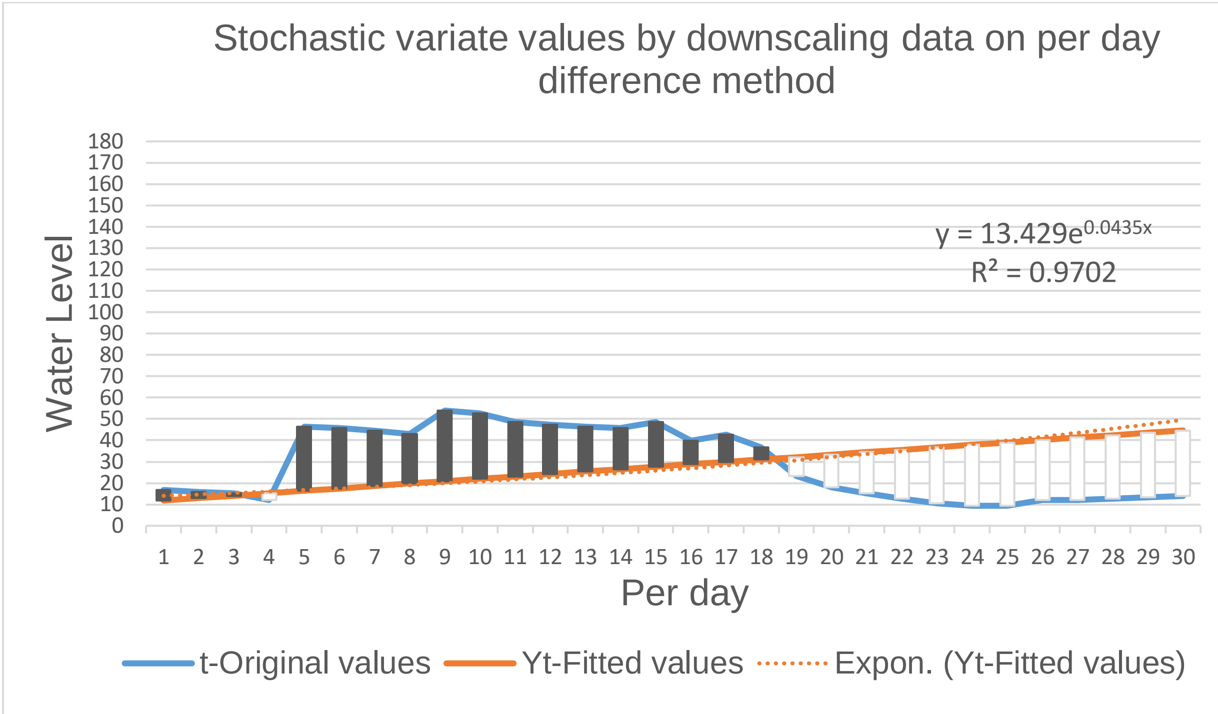


Figure 3. Generated downscaling data at Mantralayam (2016 January)

In the above discharge level, we conclude that Mean = 28.80 and Standard deviation = 16.13

Calculate the maximum likelihood estimates of the parameter values of the Gaussian distribution μ and σ for January 2016:

$$P(x;\ \mu,\sigma) = \prod_{x=1}^{30} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$P(30\text{Perdays};\ \mu,\sigma) = \prod_{x=1}^{30} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(16.82-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(15.80-\mu)^2}{2\sigma^2}\right) \times \ldots$$

$$\ldots\ldots \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(13.85-\mu)^2}{2\sigma^2}\right)$$

Taking logs of the original expression is given us

$$\text{In}(P(x;\ \mu,\sigma)) = \sum_{x=1}^{30} \text{In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(16.82-\mu)^2}{2\sigma^2} + \text{In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(15.80-\mu)^2}{2\sigma^2} + \ldots\ldots$$

$$\ldots\ldots + \text{In}\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(13.85-\mu)^2}{2\sigma^2}$$

We can simplify above expression using the laws of logarithms to obtain

$$\text{In } (P(x; \ \mu, \sigma)) = -30 \text{ In } (\sigma) - \frac{30}{2} \text{ In } (2\pi) - \frac{1}{2\sigma^2}[(16.82 - \mu)^2 + (15.80 - \mu)^2 + \ldots\ldots\ldots \ldots\ldots + (13.85 - \mu)^2]$$

This expression can be differentiated to find the maximum. We can find the MLE of the mean μ . To do this we take the partial derivative of the function with respect to μ , giving

$$\frac{\partial \text{In}(P(x; \ \mu, \sigma))}{\partial \mu} = \frac{1}{\sigma^2}(16.82 + 15.80 + \ldots.. + 13.85 - 30\mu)$$

Finally, setting the left-hand side of the equation to zero and rearranging for μ gives

$$\mu = \frac{16.82 + 15.80 + \cdots + 13.85}{30} = 28.80$$

Similarly, we have to get σ

$$\frac{\partial \text{In}(P(x; \ \mu, \sigma))}{\partial \sigma} = \frac{1}{2\sigma}(16.82 + 15.80 + \ldots.. + 13.85 - \mu)$$

$$2\sigma = [16.82 + 15.80 + \ldots + 13.85 \text{—Fitted values } (832)]$$

$$\sigma = 16.13$$

μ = 28.80, σ = 16.13

Comparison of Stochastic variate difference method and MLE method

Results:

(i) Stochastic variate difference method

Mean = 28.80

Standard deviation = 16.44

(ii) Maximum likelihood estimation

Mean = 28.80

Standard deviation = 16.13

6. Measure of quality of Fit

With respect to the coefficient of determination, $R_1^2 = R_2^2$=0.999 it is nearly equal to 1.0, it notes that if the fit is perfect, all residuals are zero, but if SSE is only slightly smaller than SST.

The coefficient of determination suggests that the model fit to the data explains 99.9% of the variability observed in the response.

Comparison of past and current observed discharge level of downscaled daily precipitation generated for years 1995 and 2016 at Mantralayam location in AP

Table 5. Coefficient of determination values

Years: Jan 1995 to 2016 Location: Mantralayam State: Andrapradesh	TSS	SSE	R^2
Data-1 Year-1995	3217933.7	SSE = 149.3284	R^2 = 0.9995
Data-2 Year-2016	678646.4	SSE = 27.248	R^2 = 0.9995

7. Stochastic Forecasting process

The discharge level values changes through time according to probabilistic methods is called stochastic process in variables. An observed time series is considered to be one realization of a stochastic process. A single observation of a random variable is one possible value assume. In the development here, a stochastic forecasting process is a sequence of a random variables $\{X(t)\}$ ordered by a discrete time variable t = 1, 2, 3,….. Stochastic forecasting must generally be determined from a single time series or realization. First, one generally assumes that the process is stationary. This means that the probability distribution of the process is not changing over time. If a process is strictly stationary, the joint distribution of the random variables X (t_1) ,…., X (t_n) is identical to the joint distribution of X (t_{1+1}) ,…., X (t_{n+t}) for any t. The joint distributions depend only on the differences t_i - t_j between the times of occurrence of the events. For a stationary stochastic process, one can write the mean and variance are

$$\mu_X = E[X(t)] \quad (5.1)$$

and

$$\sigma^2 = \text{var}[X(t)] \tag{5.2}$$

Both independent of time 't'. The autocorrelations, the correlation of 'X' with itself, are given by

$$\rho_X(\kappa) = \frac{\text{cov}[X(t), X(t+\kappa)]}{\sigma_X^2} \tag{5.3}$$

For any positive integer time log κ . These are the statistics most often used to describe stationary stochastic process. When one has available only a single time series, it is necessary to estimate the values. If μ_X , σ_X^2 and $\rho_X(\kappa)$ from the values of the random variable that one has observed, the mean and variance are generally estimated essentially as

The sample estimate of the mean $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ (5.4)

The sample estimate of the variance $\hat{\sigma}_X^2 = S_X^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})2$ (5.5)

$$\bar{\mu}_X = \bar{X} = \frac{1}{\mathrm{T}} \sum_{i=1}^{T} (X_t) \tag{5.6}$$

$$S_X^2 = \frac{1}{\mathrm{T}} \sum_{i=1}^{T} (X_t - \bar{X})2 \tag{5.7}$$

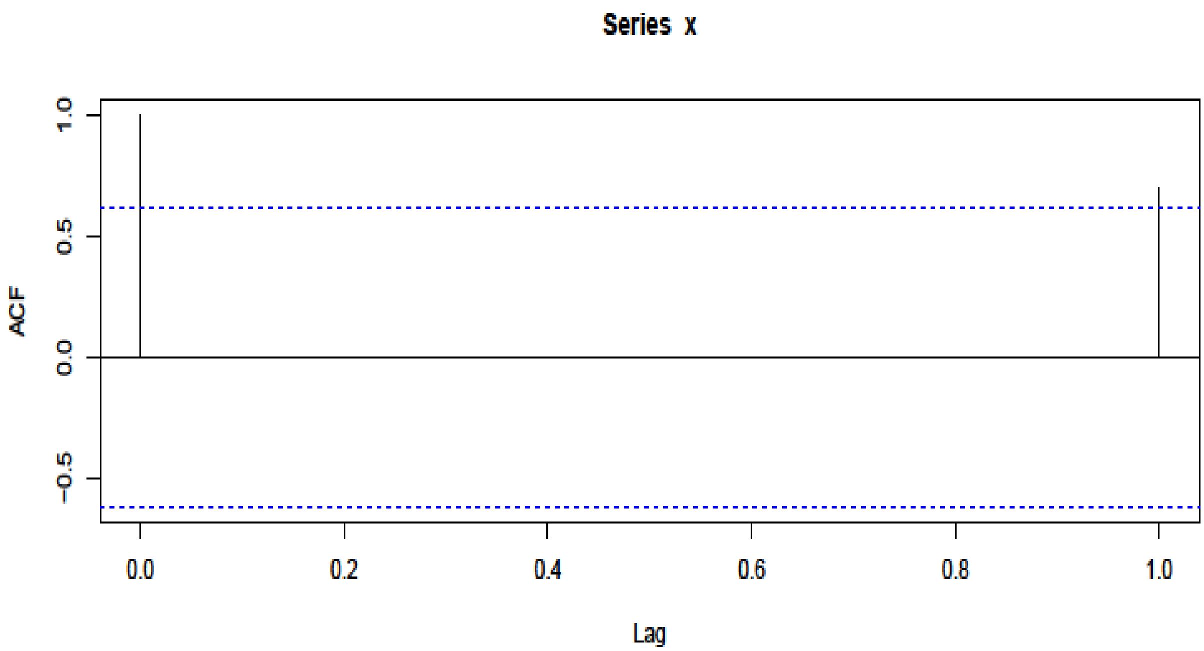
While the autocorrelation $\rho_X(\kappa)$ for any time log κ can be estimated as (Jenkins and Watts, 1968)

$$\bar{\rho}_X(\kappa) = \Upsilon_\kappa = \frac{\frac{1}{\mathrm{T}} \sum_{i=1}^{T-k} (X_{t+k} - \bar{X})(X_t - \bar{X})}{\frac{1}{\mathrm{T}} \sum_{i=1}^{T} (X_t - \bar{X})2} \tag{5.8}$$

R-Code:

```
x<-54.43:44.8
> ### function to find autocorrelation when lag=1;
> my.auto<-function(x) {+ n<-length(x) + denom<-(x-mean(x)) %*% (x-mean(x))/n
+ num<-(x [-n]-mean(x)) %*% (x [-1]-mean(x))/n+ result<-num/denom+ return (result)}
> my.auto(x)
    [, 1]
```

[1,] 0.7



Discharge level series-1995 (January– Per day)

Random series:

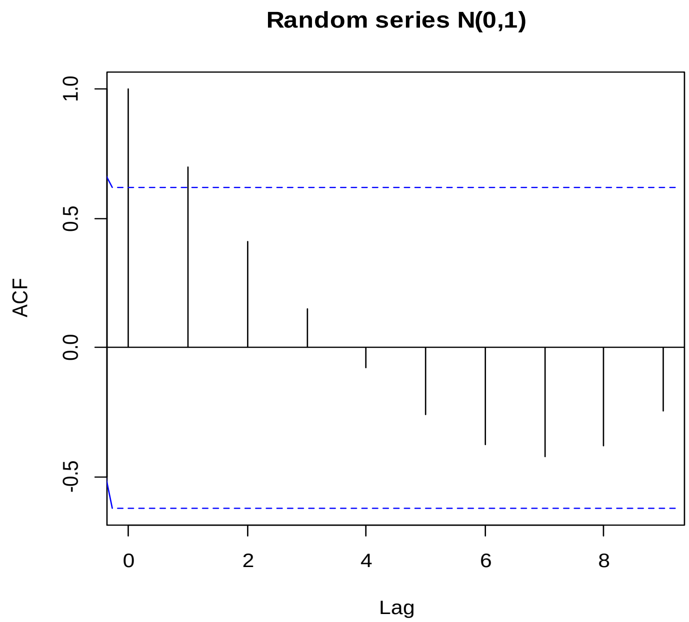
>### autocorrelation function for

> ### random series

> set. Seed (1995);

> x<-rnorm (30days);

> acf (x, lag=30, main="Random series N(0,1)");

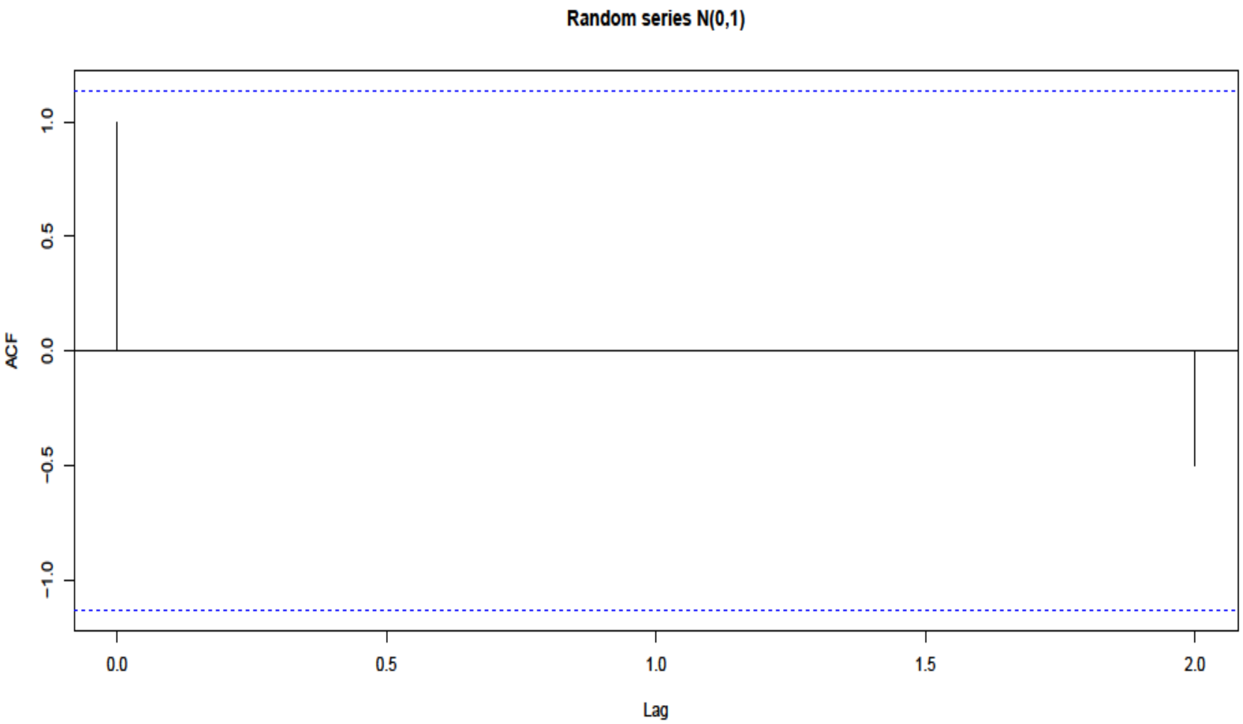


Discharge level Random series-1995 (January- Per day)

R-code:

Random series:

```
>### autocorrelation function for
> ### random series
> set. Seed (2016);
> x<-rnorm (30days);
> acf (x, lag=30, main="Random series N(0,1)");
```



Discharge level-2016 (January- Per day)

>x=c(54.43,51.45,43.59,42.78,44.84,40.66,38.38,39.86,44.2,43.48,41.09,41.112,41.112,46.24,64.71,

67.78,64.52,90.02,170.6,125.5,106,82.41,73.41,70.52,67.9,64.71,53.2,51.48,44.93,44.8)

>y=c(50.54,51.30,52.05,52.80,53.56,54.31,55.06,55.82,56.57,57.32,58.08,58.83,59.58,60.34,61.09,61.85

,62.60,63.35,64.11,64.86,65.61,66.37,67.12,67.87,68.63,69.38,70.13,70.89,71.64,72.39)

> fit=lm(y~x)

> fit

Call:

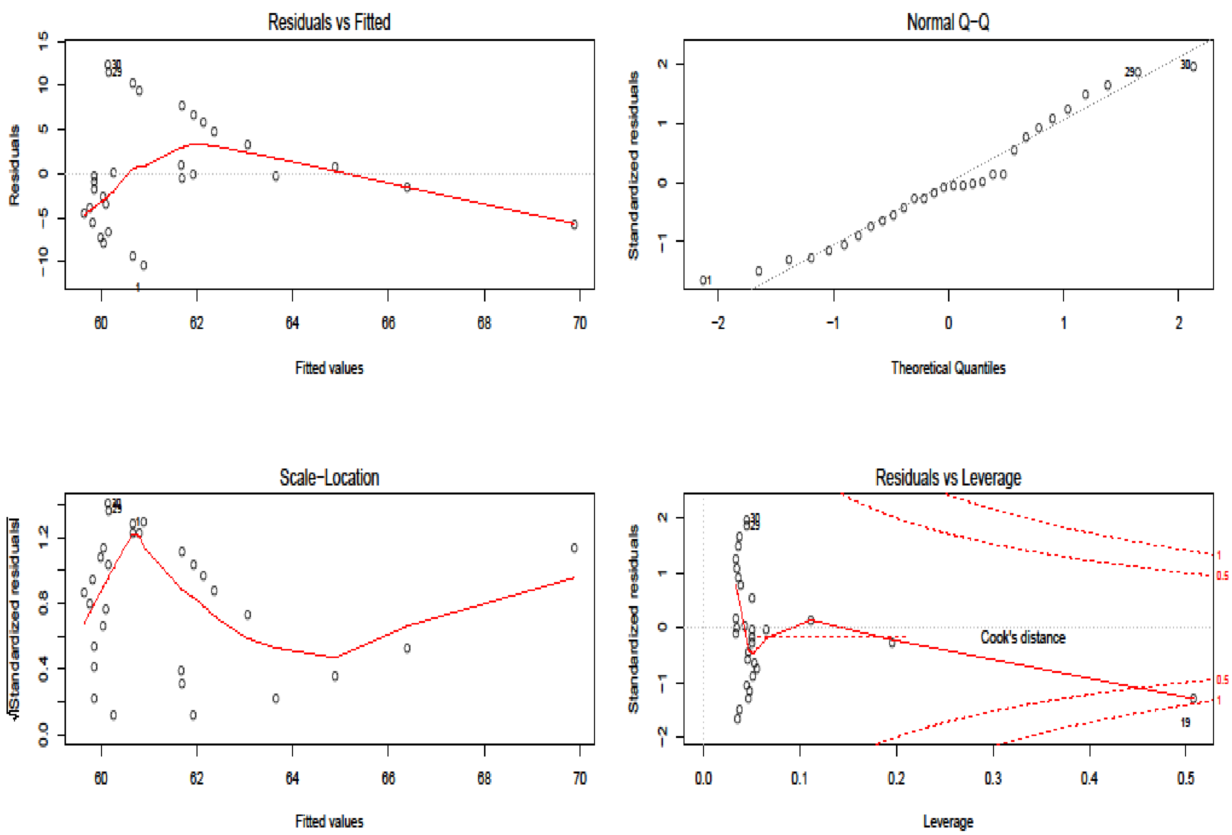
lm(formula = y ~ x)

Coefficients:

(Intercept) x

56.68012 0.07741

```
> summary (fit)
Call:
lm(formula = y ~ x)
Residuals:
    Min       1Q   Median       3Q      Max
-10.3534   -4.4297 -0.4488   4.3957 12.2420
Coefficients:
            Estimate Std. Error t value Pr (>|t|)
(Intercept) 56.68012    2.74204   20.671   <2e-16 ***
      x          0.07741    0.04018   1.926   0.0642 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' * ' 1
Residual standard error: 6.343 on 28 degrees of freedom
Multiple R-squared:  0.117,     Adjusted R-squared:  0.0855
F-statistic: 3.711 on 1 and 28 DF,   p-value: 0.06425
> Par (mfrow=c (2,2));
> Plot (fit)
> Par (mfrow=c (1,1));
```

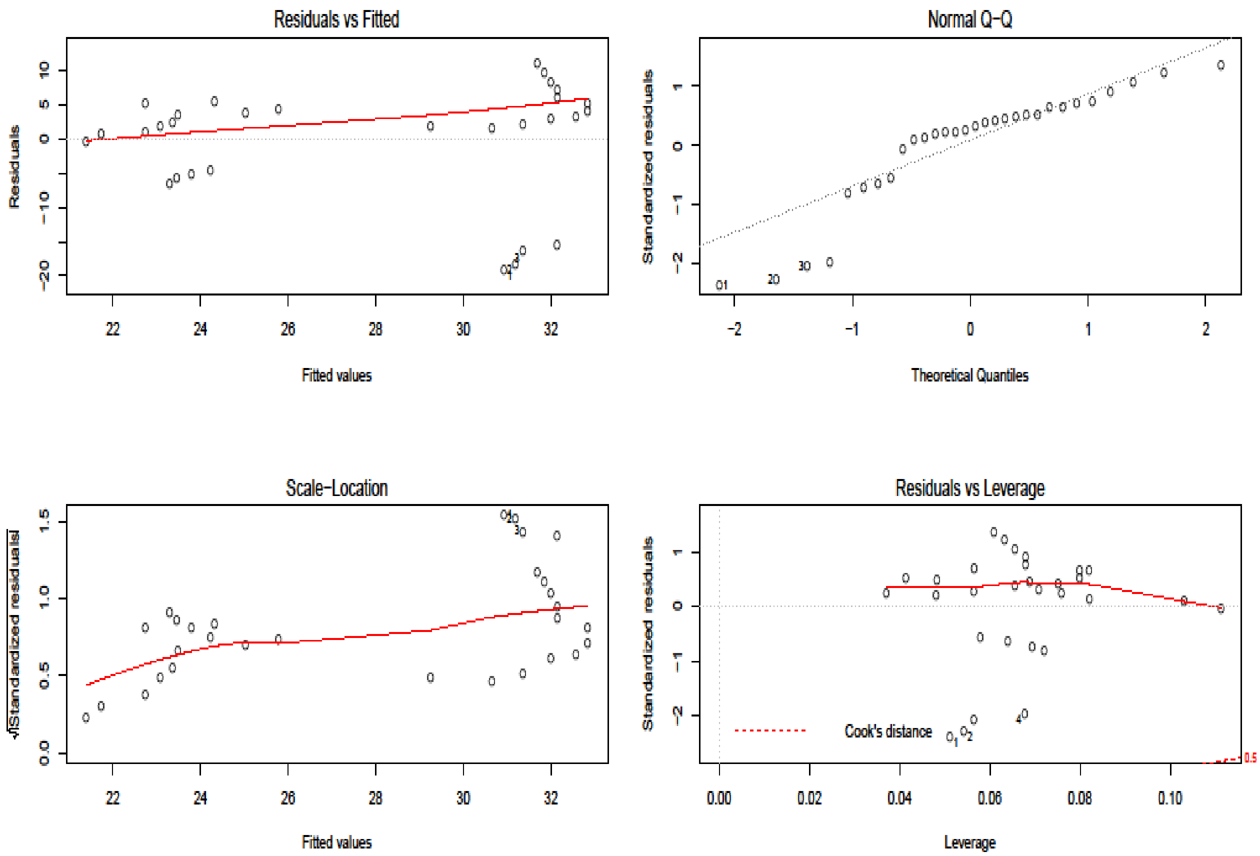


Original-Fitted values (1995-January)

```
>x=c(16.82,15.80,15.13,12.10,46.52,45.90,44.57,42.90,53.95,52.60,48.69,47.38,46.26,45.78,48.
68,
    39.80,42.52,36.86,23.35,17.89,15.13,12.64,10.40,9.37,9.37,12.06,12.06,12.64,13.24,13.85)
>y=c(11.92,12.95,14.95,16.52,16.80,17.68,18.61,19.74,20.98,22.43,23.84,24.98,25.79,26.98,27.
86,28.87
    ,29.93,30.06,31.19,32.32,33.46,34.98,35.82,36.85,37.98,38.12,39.25,40.38,41.51,42.64)
> fit=lm(y~x)
> fit
Call:
lm(formula = y ~ x)
Coefficients:
```

```
(Intercept)          x
   35.2460      -0.2569

> Summary (fit)
Call:
Lm (formula = y ~ x)
Residuals:
   Min      1Q   Median     3Q     Max
-19.006   -3.467   2.263   4.909   10.951
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.24597    3.02309   11.659 2.93e-12 ***
 x          -0.25685    0.09114  -2.818  0.00876 **
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' * ' 1
Residual standard error: 8.208 on 28 degrees of freedom
Multiple R-squared:  0.221,     Adjusted R-squared:  0.1932
F-statistic: 7.943 on 1 and 28 DF, p-value: 0.008758
> Par (mfrow=c(2,2));
> Plot (fit)
> Par (mfrow=c(1,1));
```



Original-Fitted values (2016-January)

8. Conclusions

The stochastic variate difference approach is best fit for water level discharge data time series, and it is suitable for forecasting and prediction of the future time series values. In this case, coefficient of determination (squared R) value is very close to one. We conclude that the model is best fit for the data. The proposed new model for parameter estimation, adopted MLE for Gaussian distribution. The present proposed model is to predict future values using stochastic variate difference approach. Finally, we conclude that the new procedure is comparatively approachable by past year 1995-January and present year 2016-January on generated downscaled data sets and are specified stochastic variate difference model. It would be good practice to obtain a better set of measurements and combine two years water level discharge data sets of information in order to compare and especially improve the future estimation method.

9. References

Ang, A.H.S. & Tang, W. H. (1975). Probability Concepts in Engineering Planning and Design, Vol. I, Basic Principles, Wiley, New York, 28 (3), pp-226.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R., Griensven, A., Van Liew, M., Kannan, N., & Jha, M. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491-1508

Bread, L.R (1960). The Probability estimates based on small normal-distribution samples. *Journal of Geophysical Research*, *65*(7), pp. 2143–8.

Chandra, S., Jayadeva, & Mehera, A. (2011). The Numerical Optimization with Application, Narosa Publishing House, Reprint 2011.

Fiering, M. B. (1967). Streamflow synthesis. *Cambridge Harvard University Press, 1967. 139 P*.

Jenkins, G. M., & Watts, D. G. (1968). Spectral Analysis and its Applications. Holden-Day, San Francisco.

Kendall, M., & Stuart, A. (1966). ch. 49, "spectrum theory". *The advanced theory of statistics*, *3*, 454-471.

Lachtermacher, G., & Fuller, J. D. (1994). Backpropagation in hydrological time series forecasting. In *Stochastic and statistical methods in hydrology and environmental engineering* (pp. 229-242). Springer, Dordrecht.

Lautenbacher, C. J., & Stidham Jr, S. (1999). The underlying Markov decision process in the single-leg airline yield-management problem. *Transportation Science*, *33*(2), 136-146.

Leh, M. D., Sharpley, A. N., Singh, G., & Matlock, M. D. (2018). Assessing the impact of the MRBI program in a data limited Arkansas watershed using the SWAT model. *Agricultural water management*, *202*, 202-219. doi: https://doi.org/10.1016/j.agwat.2018.02.012

Rosenblueth, A., & Wiener, N. (1945). The role of models in science. *Philosophy of science*, *12*(4), 316-321.

Salas, J., & Bartolini, P. (1993). Modelling of stream flow processes at different time scales. Water Resources. Vol. 29, No. 8, pp. 2573–87.

Singh, G. (2015). Evaluation of Conservation Practices through Simulation Modeling and Tool Development (Doctoral dissertation, University of Arkansas). Available at https://search.proquest.com/openview/727bcc071e9a422c4f9553ace87db78b/1?pqorigsite=gscholar&cbl=18750&diss=y

Singh, G., & Kumar, E. (2017). Input data scale impacts on modeling output results: A review. *Journal of Spatial Hydrology*, *13*(1).

Singh, G., & Saraswat, D. (2016). Development and evaluation of targeted marginal land mapping approach in SWAT model for simulating water quality impacts of selected second generation biofeedstock. *Environmental modelling & software*, *81*, 26-39.

Singh, G., Saraswat, D., & Sharpley, A. (2018). A sensitivity analysis of impacts of conservation practices on water quality in L'Anguille River Watershed, Arkansas. *Water*, *10*(4), 443. doi: https://doi.org/10.3390/w10040443

Stendera, S., & Johnson, R. K. (2006). Multiscale drivers of water chemistry of boreal lakes and streams. *Environmental Management*, *38*(5), 760-770.

Wetzel, R. G. (1983). Attached algal-substrata interactions: fact or myth, and when and how?. In *Periphyton of freshwater ecosystems* (pp. 207-215). Springer, Dordrecht.