MANTECH Publications

Journal of Engineering Mathematics & Statistics
Volume 6, Issue 2, July-December 2022

Maximum Likelihood Robust Regression Model for Downscaling Hydrology Data

Hitesh Bisht[1], V. R. Raman[2]

Assistant Professor[1], Student[2]

Department of Mathematics

Priyadarshini Bhagwati College of Engineering

***Corresponding Author's Email id**: ramanvinist@gmail.com[2]*

Abstract

Several studies have been undertaken on the use of time series models for the prediction and forecasting of hydrologic parameters such as precipitation, water discharge levels, rainfall runoff, and so on. The main job of data management in time series analysis is to estimate the standard parameters of the data utilising some resilient new methodologies. Currently, the study proposes estimating the parameters using the maximum likelihood technique and creating a robust regression line for future predictions and forecasts for downscaling data sets. Finally, for the best model fitting, the RMSE technique and the chi-square test were used.

Keywords: - *Robust regression models, ARIMA models, Maximum likelihood method, etc*

INTRODUCTION

Data acquired using devices on the ground, as well as on a variety of airborne and space-borne platforms, is used in remote sensing investigations. Sensor technology for remote sensing for hydrological applications covers a wide spectrum of electro-magnetic field images. Remote sensing techniques measure hydrological factors indirectly, as can electromagnetic variables acquired by remote sensing techniques.

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved

In water resource engineering challenges, mathematical statistical approaches such as the Auto Regressive and Auto Regressive Integrated Moving Average models are commonly employed for time series analysis. In some circumstances, Artificial Neural Network models may also be used to anticipate the future of hydrological problems. Relevance Vector Machine (RVM) is a kernel-based method for classification and regression that has been popular in recent years.

The output of GPR is a normal distribution. In the literature, there are several applications of GPR (Ruiz and Binefa,2012; Nearing et al.,2014, Chen et al.,2014). Huang et al. (2008) investigated the use of RVM to stock prices in share market business challenges. The Gaussian Process Regression (GPR) model is probabilistic and non-parametric. Edoardo Bertone et al. (2017) created a hybrid regressive and probabilistic model that can anticipate the storage volume of the Little Nerang dam. For seasonal stream predictions, the model applies to nonlinear threshold auto regressive models utilising the Monte-Carlo technique. SK.Khadar Babu et al. (2011) investigated auto-regressive models for rain-fall flow time series prediction and simulated the data for all recorded values. Lunchakorn Prathumratana et al. (2008) investigated the different properties of water using the correlation coefficient approach.

In the current study, Andhra Pradesh is used as the study region for discharge prediction. The primary rivers in Andhra Pradesh include the Krishna, Godavari, Palar, and Penna, which are utilised to support agriculture, industrial power generation, and residential needs. Simultaneously, further floods are being recorded on the Krishna and Godavari rivers in Andhra Pradesh. As a result, historical data from the Mantralayam gauging station on the Krishna-Thunga Bhadra river is utilised to make projections for the future.

In this work, the Maximum Likelihood Robust Regression Model (MLRRM) was utilised to predict and forecast downscaling data sets.

The Root Mean Square Error (RMSE) and correlation coefficient are used to choose the optimal model.

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved

METHODS AND MATERIALS

Gholam Abbas et.al(2015), developed a model for forecasting and prediction of seasonal precipitation time series data analysis. The model is used to predict the future values using Box-Jenkins approach. The present study proposed to apply maximum likelihood procedure (ML) for parameter estimation for robustness and make a robust regression model for future predictions.

Maximum Likelihood Method of Estimation

In statistical analysis, the major task is to estimate the parameters , that are identified in the study area. Actually, based on the data sets, we can first choose the distribution for best fit of the data, and then this is achieved by using the distribution function, in which the parameters are replaced by their standard pre-defined estimates.

At Present study, for parameters choosing a Gaussian distribution and apply the maximum likelihood estimations was adopted for parameter estimation. We can apply the method of moments, the least squares method etc,. for parameter estimation.

It involves that it gives the maximum probability priority occurrence of the observations in a problem. Let the probability density function (Continuous case) f(x) and its parameters α1, α2, α3,…, αk are the parameters of the distribution.

Then the Likelihood function with k parameters as

$$L(\alpha 1, \alpha 2, \ldots, \alpha k) = \prod_{i=1}^{n} f(x_i, \alpha 1, \alpha 2, .., \alpha k) \quad (1)$$

The parameters of α1, α2,…, αk that maximises the partial differentiation of the likelihood function with respect to the parameters and equating to zero as in Eq.2. The resulting to get set of equations solved simultaneously, then we the estimates of the parameters α1, α2,…, αk.

$$\frac{\partial L(\alpha 1,\alpha 2,\ldots,\alpha k)}{\partial \alpha i} = 0, \quad i = 1.2\ldots,k \quad (2)$$

In many cases for easier calculations we can apply natural logarithms of the likelihood function by using

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved

$$\frac{\partial lnL(\alpha 1,\alpha 2,\ldots,\alpha k)}{\partial \alpha i} = 0, \quad i = 1.2\ldots,k \tag{3}$$

Using the above procedure , we can estimate the parameters by adopting the suitable probability distribution. The present article, an attempt was to assume Gaussian probability distribution perfectly fitted for the discharge levels of our proposed data sets.

Maximum Likelihood Robust Regression Approach

The relationship between two variables gives in form of regression and the word regression means stepping backward to the average. It denotes the linear regression. But, the present study generates naturally in a single dimensioned variable flow time series.

Let the observed values $X_1, X_2, \ldots, X_t, ,$ is used to fit the following regression model

$$(X_t - \mu) = \beta_1(X_{t-1} - \mu) + \varepsilon_t \tag{4}$$

Where

The population parameters calculated by using ML method of estimation is as follows:

The parameter Mean μ and the standard deviation σ are the ML estimates of the data sets.

The ε_t follows the normal distribution with mean zero and Variance one. And it is the standard normal error variate.

After fitting the Robust Regression , is named as Maximum Likelihood Robust Regression Model for prediction of the time series data analysis. For validation of the model , we can obtain the correlation coefficient and Root Mean Square Error(RMSE) Approach

The correlation coefficient $r = \frac{\sum_{t=1}^{n-1}(X_i - AVG)(X_{t+1} - AVG)}{\sum_{t=1}^{n-1}(X_i - avg)^2}$ (5)

$$RMSE = SQRT\left(\frac{\sum_{i=1}^{n}(O_i - E_i)^2}{n}\right) \tag{6}$$

Correlation coefficients and RMSE values are the performance measures are attempted.

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved

RESULTS AND DISCUSSIONS

ML estimates of the parameters are mean $\mu = 24.42554$, the standard deviation $\sigma = 7.451964$.

Skewness measure $\beta 1 = 0.130413$.

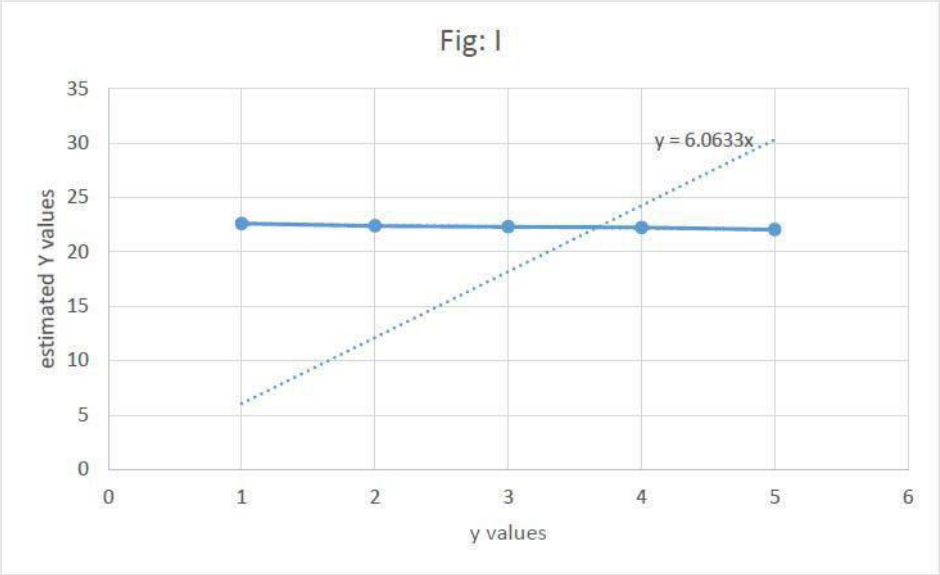
The proposed Model $X_t = \mu + \beta_1 X_t\text{-}1 - \beta 1 \mu$

$X_t = 24.42554 + 0.130413 X_t\text{-}1 - (0.130413)(24.4255)$

$X_0 = 10.514$ then $X_1 = 22.8295$, $X_2 = 24.214717$, $X_3 = 24.39835$, $X_4 = 24.42230$,

$X_5 = 24.425425$ and so on.

Fig: I



The correlation coefficient, r = 0.6743, which means that there is a positive relationship between the variables of Y and estimate Y, we can agree the method is suitable for future predictions.

Coefficient of determination is also the data contains 60% accuracy of the models ,here the data depends on the discharge levels are rare some times, this is a reason to get 60% acceptance.

Finally, which is also feasible for prediction and forecasting of the downscaling data sets by conforming through the chi-square test for goodness of the model at 95% confidence.

Root mean square error approach concludes that it is good for forecasting and predictions.

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved

CONCLUSIONS

The Pearson correlation coefficient and the coefficient of determination are the validity measures used in this research for the analysis of time series. There is a positive link between the observed and expected variables. The MLRRM is excellently suited for future prediction and forecasting of downscaling data. According to the findings of the Chi-square testing technique, the suggested model is suitable for downscaling data predictions.

REFERENCES

1. Chen J ,Yu J and Zhang Y(2014),Multivariate video analysis and Gaussian Process Regression model based soft sensor for online estimation and prediction of nickel pellet size distributions, Computers and chemical Engineering 64:13-23.
2. Edoardo Bertone,Kelvin o'Halloran,Rodney A.Stewart, Guilherme F.de oliveira(2017), Medium-term storage volume prediction for optimum reservoir management : A hybrid data-driven approach, Journal of cleaner production, Vol 154,No15,pp353-365.
3. SK.Khadar Babu, K.Karthikeyan, M.V.Ramanaiah and D.Ramanah (2011),Prediction of rain-fall flow time series using auto- regressive models, Pelagia research library, Advances in applied science research , 2(2):128-133.
4. Lunchakorn Prathumratana, Suthipong Sthiannopkao, Kyoung Woong Kim(2008), The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River, Environment International,vol.34,pp 860-866.
5. Huang SC and Wu TK (2008), Combining Wavelet-based feature extractions with relevance vector machines for stock index forecasting. Expert Systems vol25:133-149.
6. Ruiz and Benefa(2012),Modelling facial expressions dynamics with Gaussian process regression. Frontiers in Artificial intelligence and Applications.vol 248: 91-100.
7. Nearing GS,Gupta HV, and Crow WT(2014),Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity. Journal of Hydrology, vol 507(12):163-173.
8. S.K.Khadar Babu, M.Sunitha and M.V. Ramanaiah (2017), Mathematical Modelling of RMSE Approach on Agricultural and Financial Data Sets, International journal of Pure and Applied Bio-sciences, Vol 5(6),pp942-947.

Page 60-65 © MANTECH PUBLICATIONS 2022. All Rights Reserved