

# Co-Integration Models for Koyna and Warna Reservoirs, India

D. V. Ramana<sup>1\*</sup>, J. Pavan Kumar<sup>2</sup>, R. N. Singh<sup>1</sup>, R. K. Chadha<sup>1</sup>

<sup>1</sup>National Geophysical Research Institute (CSIR), Hyderabad, India

<sup>2</sup>Indian National Centre for Ocean Information Services (MoES), Hyderabad, India

Email: [\\*dvr@ngri.res.in](mailto:dvr@ngri.res.in), [chadha@ngri.res.in](mailto:chadha@ngri.res.in), [rn Singh@mgri.res.in](mailto:rn Singh@mgri.res.in), [jpawan33@gmail.com](mailto:jpawan33@gmail.com)

Received 30 August 2015; accepted 27 October 2015; published 30 October 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Koyna region, a seismically active region, has many time series observations such as seismicity, reservoir water levels, and many bore well water levels. One of these series is used to predict others since these parameters are interlinked. If these series were stationary, we used correlation analysis. However, it is seen that maximum of these time series are nonstationary. In this case, co-integration method is used that is extracted from econometrics and forecast is possible. We have applied this methodology to study time series of reservoir water levels of this region and we find them to be co-integrated. Therefore, forecast of water levels for one of the reservoir is done from the other as these will never drift apart too much. The outcomes demonstrate that a joint modelling of both data sets based on underlying physics resolves to be sparingly useful for understanding predictability issues in reservoir induced seismicity.

## Keywords

Co-Integration, Seismicity, Forecast, Koyna-Warna, Reservoirs, Water Levels

---

## 1. Introduction

Fluid pressure and transport are playing crucial roles in controlling the seismicity in the near subsurface and described in two ways of assessing their roles. In theoretical approach, a model of the region is developed and tested to be validated with the data provided. Further, this model is used for understanding of the process and also forecasting of future events. Such a methodology requires a comprehensive knowledge of the distribution of physical properties and forces, which is not available due to the high cost of acquiring relevant data. In the other empirical approach, the responses of the region are studied to make statistical models for forecasting the onset of

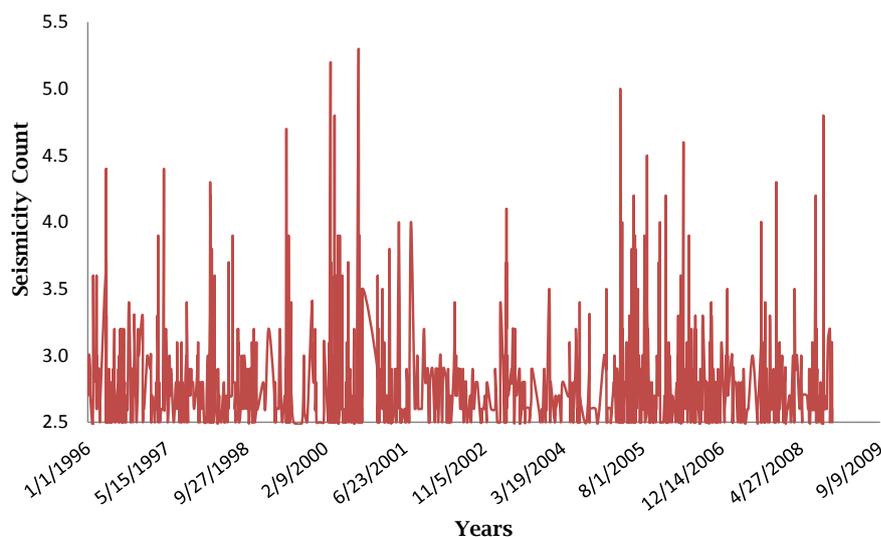
---

\*Corresponding author.

the future events in which society has immediate interest. For this approach, we have two datasets, viz., reservoir water level time series and seismicity time series and it should be fascinating to know any relationship between them. Hence we have an opportunity to find long term relationships between them about each other and this should be pattern even after short term departure.

The seismicity in the month of September is 90% correlated with a rapid increase in the water levels in the month of June/July due to the rain [1]. The cross correlation between the seismicity and water levels in the Koyna reservoir is studied and recommended that a time lag of 223 days resembles a maximum correlation coefficient [2]. Many times stationary time series methods such as correlation analysis are applied to them and false causal relationships are inferred. In stationary time series, mean and variance are constants and this is not observed in seismicity and reservoir water level data. The near subsurface is constantly changing due to variety of processes (both internal and external), making these times series as nonstationary. From the above consequences the cointegration method has been found to be a good application and this method have the scope to find the linear combination between two nonstationary time series, which transforms into stationary time series. Cointegration method has several fascinating applications in nonstationary time series. Further, we apply this methodology to find relationship between seismicity and reservoir water level time series. Error correction model based on the cointegration theory is applied to forecast the annual runoff of the Basishan station and Fengman station in the Songhua river basin [3]. Koyna region has an impounded reservoir that continues to be seismically active and seismic activities have become noticeable after the start of impoundment since 1961 [4] [5]. This area experienced a major shock in 1967 of magnitude 6.5. Since then, no other shock of such magnitude has taken place. Nevertheless the shock of lesser magnitudes has occurred magnitude 5 and above occurred in 1967, 1968, 1973, 1993, 1994, 2000, 2005 and 2009. **Figure 1** shows the seismicity time series with minimum shock magnitude as 2 based upon Gutenberg-Richter relationship ( $\text{Log}N = a + (b * M)$ , (N-number of earthquakes and M-magnitude) has been fitted to this data [6] and b values are observed to range within **0.9 to 1.2**. **Figure 2(a)** shows the Koyna water level data since the impounding of the reservoir in 1967. The water level increases after rainfall and releases subsequently. The maximum water level recorded is about 660 m and the minimum value is 610 m, it follows an annual cycle in reservoir level data. **Figure 2(b)** shows the Warna reservoir water level data from 1985 after the impounding of the reservoir. Various reasons have been given to the cause of these earthquakes. Water load of the reservoir has been single such mechanism. It has also been argued that water seepage from the reservoir through porous basement and faults leads to the rise of pore pressure in the focal region to induce the faulting. The pore pressure increase at focal region depends on the hydraulic properties.

Therefore, the seismicity in the region requires a causative relationship between water level variation and seismicity data. As both time series are nonstationary, simple correlation coefficient analysis gives false relationship. It should be desirable to use co-integration method for finding a causative relationship and this is done in the converse.



**Figure 1.** Time series of earthquake counts from 1996-2008.

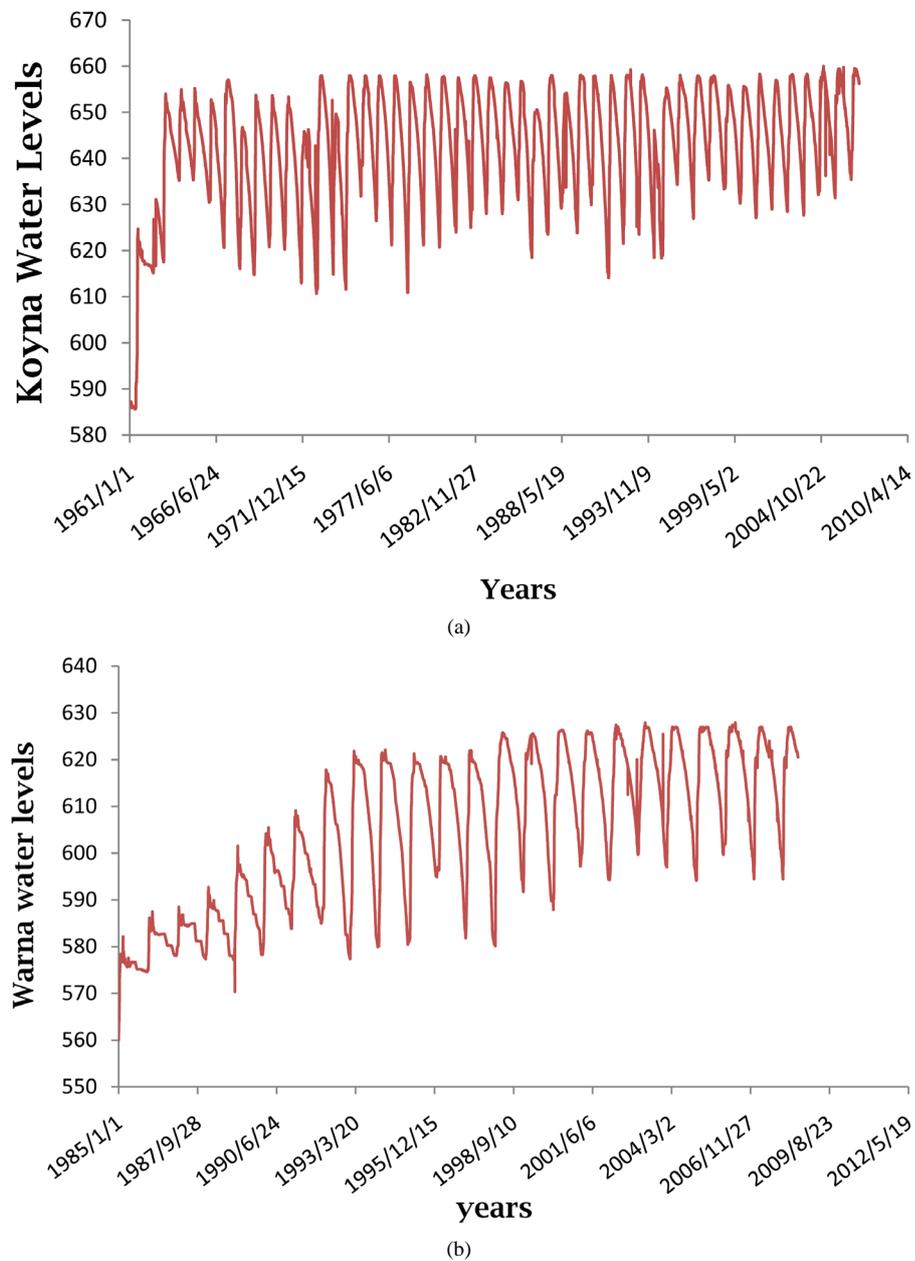


Figure 2. (a) Koyna Reservoir water levels from 1961-2008; (b) Warna Reservoir water level from 1985-2008.

## 2. Methodology

Non-stationary time series contain trends. The trends can be deterministic or stochastic. For instance a random walk model has a stochastic trend. At first we need to determine the presence of stochastic trend in the time series. A test for the presence of stochastic trend in the time series is widely called the Augmented Dickey Fuller (ADF) test. In this test the time series is represented as  $\Delta y_t$ , called autoregressive model of order  $p$ , denoted as  $AR(p)$ .

$$\Delta y_t (\cong y_t - y_{t-1}) = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t \tag{1}$$

$\alpha$ ,  $\beta$ ,  $\delta_i$ , and  $\gamma$  are constants. “ $\varepsilon_t$ ” Is the random error term and “ $p$ ” indicates number of lags during this test,

hypothesis are tested for, the coefficient of lag term to be at zero. If this null, then hypothesis cannot be rejected and the time series is nonstationary. First we need to find the value of lag order “p”. This is done with the help of Akaike Information Criterion (AIC) or Bayesian information criterion (BIC). After determining the order, Equation (1) fits with the time series using ordinary least squares (OLS) method to get constants:  $\alpha, \beta, \gamma, \delta_i$  and variance of  $\varepsilon_i$ . From mean and standard error of constant,  $\gamma$  the value of the **tau** statistics defined by  $t = \hat{\gamma} / SE(\gamma)$  is calculated. If this tau statistic is more than a critical value, then the hypothesis cannot be rejected and the time series has stochastic trend.

After determining the nonstationary nature of the time series, further we proceed to establish a relation between these two time series, the seismicity time series (reservoir water level changes),  $y_t$ , with water level time series,  $x_t$ , as

$$Y_t = a_0 + a_1 X_t + u_t \tag{2}$$

Here  $u_t$  is stationary time series. The values of  $a_i$ 's are obtained by Ordinary Least Squares (OLS) theory and times series of residual  $u_t$  is obtained using the coefficients derived by Ordinary Least Squares (OLS) method. This residual series fits with the following model:

$$\Delta u_t = \alpha_1 + \beta_1 t + \gamma_1 u_{t-1} + \sum_{i=1}^p \varphi_i \Delta u_{t-i} + \varepsilon_t \tag{3}$$

To test the null hypothesis the coefficient  $\gamma_1$  must be zero, ADF test is performed. If the relevant “**tau**” statistics are greater than the critical value and the presence of a unit root cannot be rejected then the both times series are not cointegrated otherwise they are cointegrated. This method is called the **Engle-Granger** two step methods. There are more advanced methods for multivariate problems. We will be using freely available software, **EViews**, for implementing the cointegration mechanism.

### The Akaike Information Criterion

(AIC) is an approach of selecting a model from a set of models. The chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. It’s based on information theory, but a heuristic way to think about it is as a criterion that seeks a model that has a good fit to the truth but few parameters. It is defined as:

$$AIC = -2(\ln(\text{likelihood})) + 2 * K \tag{4}$$

where likelihood is the probability of the data given a model and  $K$  is the number of free parameters in the model. AIC scores are often shown as  $\Delta AIC$  scores, or difference between the best model (smallest AIC) and each model (so the best model has a  $\Delta AIC$  of zero).

### 3. Results and Discussions:

The reservoir water levels in Koyna and Warna are taken from years 1996-2008 to establish the Cointegration theory. The lag parameters using Akaike information criteria for individual reservoir water levels as well as a combination of both the reservoirs are obtained and shown in **Table 1** and **Table 2**. The trend of AIC value changes in both the reservoirs observed at lag 2.

Thus the fitted model is AR (2). Now we have to test the stationary nature of all the data sets. To achieve the above, we need to apply the ADF test for both time series with the order p. and the results are shown in **Table 3**.

From this table (the tau statistics) is recorded. The variables of X and Y can be made ordinary least square regression as the following form:

$$Y_t = 1.012132 * X_t - 40.17355 \tag{5}$$

**Table 1.** Lagging order for both Koyna-Warna reservoir water levels.

Lag Selection Criterion	Values	Lag Order
Akaike information criterion	15.94125	1
Akaike information criterion	15.27168	2
Akaike information criterion	15.34040	3

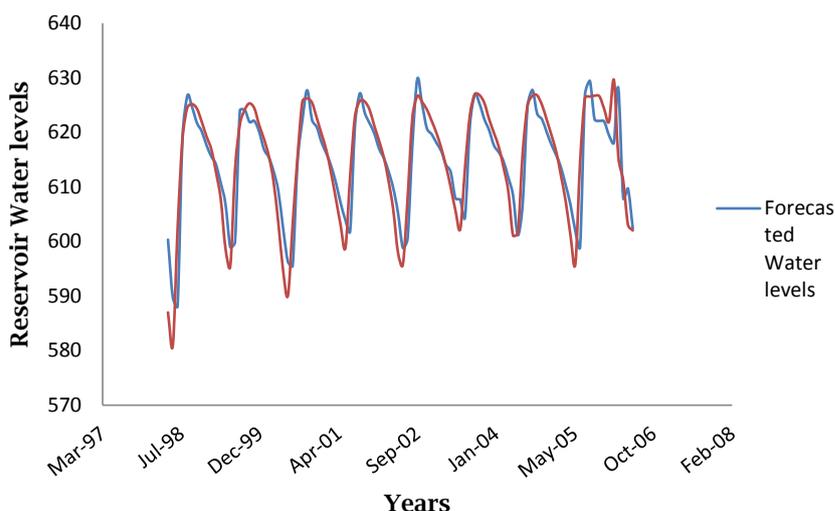
where  $Y_t$  and  $X_t$  are the Warna and Koyna reservoir water levels respectively. From the residual equation the unit root test is done for the residuals and it is established that both the variables of  $X$  and  $Y$  are cointegrated. The vector error correction model (VECM) is estimated according to the formula Equation (1) as follows:

$$\Delta Y_t = -0.0987 * Y_{t-1} + 1.7877 * X_{t-1} - 2416.5535 + 0.3475 * \Delta Y_{t-1} - 0.3258 * \Delta Y_{t-2} + 0.3445 * \Delta X_{t-1} + 0.3879 * \Delta X_{t-2} + 0.0356 \tag{6}$$

The error correction model according to the estimation of formula Equation (2) as follows:

$$Y_t = 1.17701 * Y_{t-1} - 0.4894 * Y_{t-2} + 0.1327 * X_{t-1} - 0.1865 * X_{t-2} + 221.1733 \tag{7}$$

Equation (7) is used to calculate the water level in the Warna reservoir at a given time  $t$ . The water levels are calculated from the years 1998 to 2006 and plotted in **Figure 3**. From **Figure 3** it is very clear that the model values are well matched with the observed values (original values). Further it is used to calculate the water levels in the Warna reservoir at different times since Jan 2005 to Dec. 2006 and the values are given in **Table 4**. The results show a maximum relative error of the forecasting values is 25%. The numerical results also show that the forecasting for a month after the period of observation the error is merely 3% but if the forecasting is for a few months before, then the error is more. These results conclude that our model can forecast the values in the tolerance limit are up to six points (months) only.



**Figure 3.** Forecast of water levels in the Warna reservoir (red line is the observed water level, blue line is water level forecast).

**Table 2.** Calculated lag order ‘p’ for both Koyna-Warna reservoir water levels.

Lag (p)	1	2	3	4	5
Koyna	6.359976	5.941964	5.940718	5.90818	5.86736
Warna	6.660829	6.199189	6.220249	6.1891	6.19288

**Optimal Lag**

**Table 3.** The unit root test results of Koyna and Warna reservoirs.

Variables	Tau Statistics	DW numerical	Critical number	Test type
Warna	-8.251595	2.049397	-4.030157	(c, t, 1)
Koyna	-4.840979	1.91902	-4.030157	(c, t, 1)
Del_Warna	-8.305417	2.045654	-3.481623	(c, 0, 1)
Del_Koyna	-27.64629	2.114081	-3.481217	(c, 0, 1)

**Table 4.** Calculated values of reservoir water levels.

Date	Original Water Levels	Forecasted water levels	Error %
12/04	622.33	620.03	0.37
01/05	619.42	617.68	0.28
02/05	616.26	615.72	0.09
03/05	612.14	613.61	-0.24
04/05	607.29	610.29	-0.5
05/05	601.13	606.74	-1.0
06/05	595.9	601.88	-1.0
07/05	614.6	599.04	2.6
08/05	626.35	626.86	-0.04
09/05	626.57	629.46	0.16
10/05	626.73	622.41	0.69
11/05	626.62	622.11	0.72
12/05	624.31	621.99	0.37
01/06	621.98	619.46	0.41
02/06	629.63	618.13	2.45
03/06	615.52	628.12	-2.04

## 4. Conclusion

The error correction model (ECM) based on the co-integration theory to do analysis and discuss the cointegration methodology of daily water level changes in the Koyna and Warna reservoirs of the seismically active region is Established. The non-stationary time series of the model are cointegrated and the regression analysis can be used to forecast the time sequences.

## Acknowledgements

The authors are grateful to the Director, National Geophysical Research Institute, Hyderabad for his kind permission to publish this research work.

## References

- [1] Gupta, H.K. (2002) A Review of Recent Studies of Triggered Earthquakes by Artificial Water Reservoirs with Special Emphasis on Earthquakes in Koyna, India. *Earth Science Review*, **58**, 279-310. [http://dx.doi.org/10.1016/S0012-8252\(02\)00063-6](http://dx.doi.org/10.1016/S0012-8252(02)00063-6)
- [2] Telesca, L. (2010) Analysis of the Cross Correlation between Seismicity and Water Level in the Koyna Area of India. *Bulletin of the Seismological Society of America*, **100**, 2317-2321. <http://dx.doi.org/10.1785/0120090392>
- [3] Li-ping, Song, X.Y., Sheng, T.Y. and Peng, T. (2009) The Annual Runoff Forecasting Research Based on the Theory of Cointegration and Error Correction Model.
- [4] Gupta, H.K. and Rastogi, B.K. (1976) Dams and Earthquakes. Elsevier Publishers, Amsterdam, 299 p.
- [5] Gupta, H.K. (1983) Induced Seismicity Hazard Mitigation through Water Level Manipulation: A Suggestion. *BSSA*, **73**, 679-682.
- [6] Singh, C., Bhattacharya, P.M. and Chadha, R.K. (2008) Seismicity in the Koyna-Warna Reservoir Site in Western India: Fractal and b-Value Mapping. *Bulletin of the Seismological Society of America*, **98**, 476-482. <http://dx.doi.org/10.1785/0120070165>