

Forecasting Using Vector Autoregressive Models (VAR)

Applying Vector Autoregressive Model For Smart Irrigation

V Karthik Srinivasan, G Rajkumar Reddy.

Abstract— Forecasting data can give a better understanding, control and manage unexpected results better. Vector Autoregressive Models (VAR Models) have for long been used to find trends in a set of non-discreet values. In this paper, we focus on building VAR models, to determine the best fit using various tests and the results obtained when we applied VAR models to estimate the future values and trends in soil moisture. Along with the application of VAR models to predict the soil moisture, we have also additionally applied it to the temperature and mean sea level pressure forecasting, results of which are presented.

Keywords—VAR; Vector Autoregressive; VWC; Volumetric Water Content; Forecasting;

I. INTRODUCTION

Vector Autoregressive (VAR) models can be used to predict the future values with the help of past values. It could be applied to any series of values to predict the future outcome. In a project we carried out, we measured the soil moisture and estimated the Volumetric Water Content (VWC) in the soil. The future values of soil moisture were predicted by using VAR models.

The VAR model has been used applied regularly for describing the dynamic behavior of economic and financial time series and for forecasting. Forecasts from VAR models are quite flexible because they can be made conditional on the expected future paths of applied variables in the model.

II. STAGES OF VAR MODEL

Following are the steps to be followed to build VAR model.

- i. Importing and pre-processing data.
- ii. Specifying a model.
 - a. Specifying Models to set up a model using `vgxset`, involving the following
 - Specification Structures with Known Parameters.
 - Specification Structures with No Parameter.
 - Specification Structures with Selected Parameter.
 - b. Determining an Appropriate Number of Lags.
- iii. Fitting model to data, which could include,
 - a. Fitting Models to Data to use `vgxvarx` to estimate the unknown parameters in your models.
- iv. Analysing and forecasting using fitted model, including
 - a. Checking Stability to determine whether your model is stable and invertible.

- b. Forecasting with `vgxpred` to forecast directly from models.
- c. Comparing Forecasts with Forecast Period Data to compare the results of your model's forecasts to your data.

III. IMPORTING AND PRE-PROCESSING DATA

Data obtained should be converted to a matrix form, where each column should contain a time-series of data. Following which the data is imported for data processing. On Checking whether data is stationary or non-stationary, unit root test is performed to determine the presence of unit root in the data. In order to maintain consistent, reliable results, the non-stationary data needs to be transformed into stationary data.

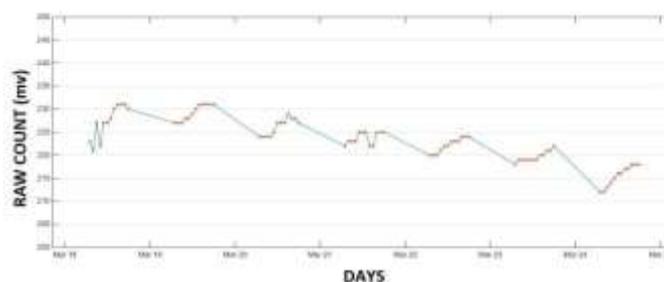


Figure 1. Sample of unprocessed data, readings of soil moisture sensor.

Figure 1. shows a sample of unprocessed data we used for this experiment. The graph depicts a series of readings of sensor data over a period of time. This data was calibrated in the soil mechanics lab to obtain the required raw data i.e. VWC or Volumetric Water Content, the data on which the experiment was conducted. Observe that the graph illustrates the unprocessed data as time series. Whatever inconsistencies or noise obtained in data should be corrected in this stage.

A. Adf Test

Augmented Dickey-Fuller test is a test for a unit root for a time series data.

$$h = adfstest(Y)$$

The values $h = 0$ indicate that tests fail, thus to reject the null hypothesis of a unit root against the trend-stationary alternative.

Non-stationary time series can be converted into stationary time series. For time series with exponential growth, pre-processing the data by taking the logarithm of the growing series is necessary. Data from multiple sources may have missing data, which has to be filled before proceeding. Taking the missing values as unchanged from the previous value is a simple way.

IV. SPECIFYING VAR MODEL

vgxset function is used to specify VAR models. *vgxset* function is independent of the time series used.

There are two functions that help to determine the appropriate number of lag values for the model to be built:

- i. *lratiotest*: identifies the appropriate number of lags.
- ii. *aicbic* test: determines the minimal appropriate number of required lags.

lratiotest function takes inputs loglikelihood of an unrestricted model, the loglikelihood of a restricted model and the number of degrees of freedom. Obtain the loglikelihood of the model from *vgxvarx* function and degrees of freedom from *vgxcount* function.

$$reject1 = lratiotest(LLF4, LLF1, n4p - n1p)$$

If $reject1 = 1$, you reject model 1 in favour of model 4. $reject1 = 0$ indicates use fewer lags than in model 4.

aicbic function takes inputs loglikelihood of model and number of active parameters in model

$$AIC = aicbic([LLF1 LLF2 LLF3 LLF4], [n1p n2p n3p n4p])$$

- Lower values are better than the higher values.

V. VAR MODEL ESTIMATION

vgxvarx function is used to create multivariate time series specification structure. Here lagged models are fitted to the data.

To fit a lagged model to data, the data is partitioned into three sections:

- i. Pre-sample data (*Ypre*).
- ii. Estimation data (*Yest*).
- iii. Forecast data (*YF*).

$$[EstSpec, EstStdErrors, LLF, W] = \dots \\ vxvarx(VARmodell, Yest, [], Ypre);$$

- *EstSpec* structures are the fitted models.

If the highest lag value is 24 then pre-sampled data is from 1 to 24, estimated data is from 24 to 90% of complete data and forecast is remaining 10% of data.

Examine whether fitted models are stable or unstable by using the function *vgxqual*.

$$[isStable, isInvertible] = vxqual(Spec)$$

If the value of *isStable* is 1 then the fitted model *Spec* is stable.

VI. ANALYSING AND FORECASTING USING FITTED MODEL

Now that we have different fitted models, we identify which fitted model is the best fit and use the particular fitted model for predictions. Analyzing the fitted model involves:

- i. Sum of mean squared error.
- ii. R-squared test.

A. Sum of Mean Squared Error

To compare the predictions of the fitted models against the forecast data *YF*, *vgxpred* function is used for forecasting. Fitted models are used as inputs for *vgxpred* function. Fitted models are created using 90% of the available data and use *vgxpred* function to predict remaining 10% of data. Compare the predicted values with the forecast data i.e., last 10% of available data and the errors found are squared. The fitted model with least squared error i.e., the least variance is used for forecasting data.

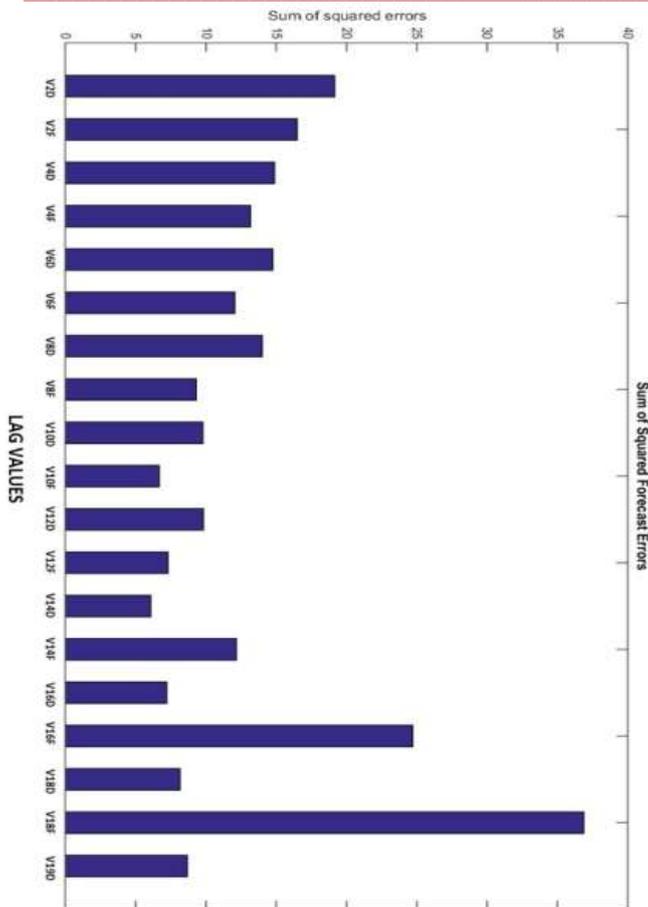


Figure 2. Sum of mean squared errors of various lag values

Figure 2. shows the bar graph representation of sum of squared errors. It is used to compare the predictions of the twenty-four models (FYi) against the forecast data YF, and use the vxpred function to get forecast data (YF). Take the difference between FYi and YF and square the result gives the sum of squared errors. The result with least value is preferred over other values.

From the figure we can see that V14D has the least sum of squared error. Therefore, we use corresponding fitted model for predictions.

B. R-squared Test

R-squared test is used to calculate how closely the predicted values and estimated values are. It gives r-squared values for each time series individually

r-squared value ranges from 0 to 1. This essentially means the r-squared is just an indicator of the deviation between the estimated and the predicted value and not an exact or elaborate measurement. When there is greater deviation between estimated and predicted value, i.e. the r-squared value being closer to 0 than 1, it is advisable to discard the model which is being tested.

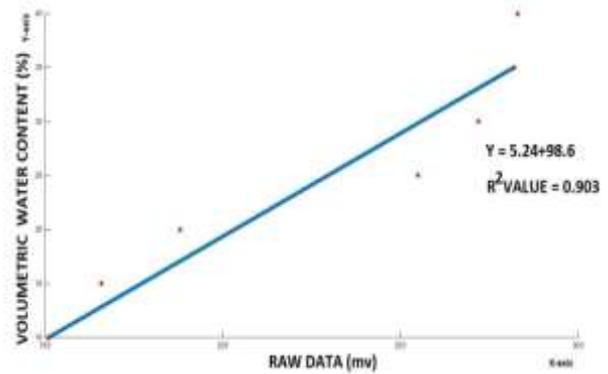


Figure 3. Time Series Analysis Graph

Figure 3. shows the relevance between the tested value and obtained value using an R-squared test. The figure gives a correspondence between the raw data obtained from the sensor and the volumetric water content (VWC). Also, note the r-squared value mentioned in the figure gives an account of how close the predicted and estimated values are. The closer the value is to 1, the greater is the correspondence of both the values on a scale of 0 to 1. The results of the r-squared values and the difference between the estimated and predicted values is discussed with an illustration in next section, section VII.

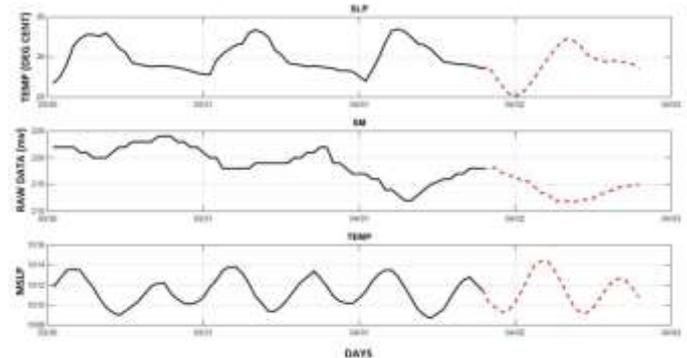


Figure 4. Time Series Analysis Graph

Figure 4. shows the predictions of sea level pressure, soil moisture and temperature respectively. Using one-week log file one-day data was predicted. Dotted curves indicates predictions. As, from the graph we can understand that time series are uniform. Therefore, we can directly use the time series to predict data. If the time series is in exponential growth, then take logarithmic of that particular time series and use the data for predictions. Use ADF-test to determine whether time series is stationary or non-stationary i.e. presence of unit root.

VII. PREDICTED MODEL VERSUS ACTUAL DATA

The r-squared values, apart from helping to choose the right VAR model, also help us to understand how a particular model varies in comparison to actual data which is obtained. Greater the amount of lag values used, the more accurate and closer

will be the predicted value, thus giving a foresight on how the data being monitored, behave in the future.

could be watered automatically without human involvement, in an appropriate setup.

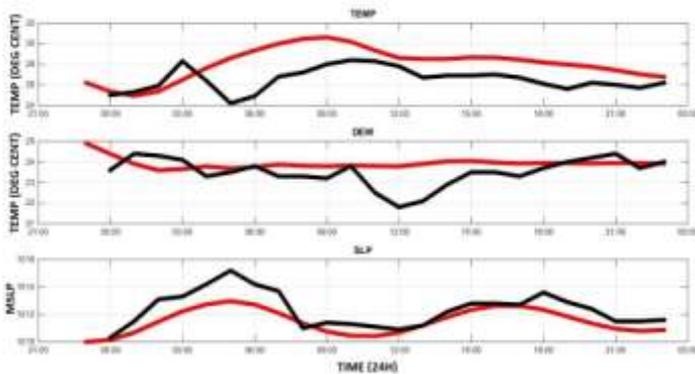


Figure 5. Graph depicting predicted model and actual data.

Figure 5. shows the graphs of predictions and estimations of temperature, dew and sea level pressure respectively, red colored graph indicates the predictions and black colored graph indicates estimations. Main purpose of the graph is to show the variation between the estimation data and predicted data. Similarity between the prediction and estimation are calculated using Rsquare test values. Rsquare value for temperature, dew and sea level pressure are 0.776, 0.546, and 0.801 respectively. This graph shows the predictions and estimations of temperature, dew and sea level pressure are almost similar.

VIII. CONCLUSION

We have in this work, outlined the various steps which are to be done to analyze and build a VAR Model which could be applied to forecast a particular indiscrete time series data with a reasonable amount of accuracy. When applied, we observed a good accuracy with the soil moisture level as well as the sea level pressure. At times the accuracy may be greater, and at times it may be lesser. To obtain a greater correspondence between the actual values and predicted values, large data set and greater lag values are required to be used. When the right model is selected, and which exhibits a closer resemblance to the actual data, it can be used to forecast values of data for which the model is obtained.

In our work, we built a VAR model for the soil moisture, atmospheric temperature, humidity (dew), and the mean sea level pressure all of which when forecasted can be used to build a model of smart irrigation using sensors which do the active monitoring and forecasting along could have a greater control over the irrigation system wherein, for example the plants

IX. ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Revathi Venkataraman, Professor, Department of Computer Science and Engineering, SRM University, who guided us through this project. We would also like to thank K.Sornalakshmi, Research Assistant, Department of Computer Science and Engineering, SRM University, for providing us with thorough inputs and providing her timely assistance and Dr. M.Pushpalatha, Professor, Department of Computer Science and Engineering, SRM University, for giving her inputs and encouragement over this experiment.

X. REFERENCES

- [1] Adi Nugroho, Khabib Mustofa, Sri Hartati, Subanar, "Vector Autoregression (Var) Model for Rainfall Forecast and Isohyet Mapping in Semarang – Central Java – Indonesia" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 11, 2014.
- [2] Jian Jin, Yao Chen, "VAR-Based Research on Energy Consumption in China" Vol 3, No. 18, 2009.
- [3] Jose Alberto, Mauriccio. (1999) "An algorithm for the exact likelihood of a stationary vector auto regression moving average". Journal of Time Series Analysis, vol. 23, no. 4, ISSN 0143-9782/02/04, pp. 473-486.
- [4] Lutkepohl, H. (1999). "Vector Autoregressions," unpublished manuscript, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- [5] Lutkepohl, Helmut. (2005). "New introduction to multiple time series analysis". Berlin: Springer Science+Business Media, Inc.
- [6] Research and Development sites of agricultural commodity – Department of Agriculture. Retrieved July 20, 2013 from http://bbsdlp.litbang.deptan.go.id/tamp_komoditas.php
- [7] Watson, M. (1994). "Vector Autoregressions and Cointegration," in Handbook of Econometrics, Volume IV. R.F. Engle and D. McFadden, Elsevier Science Ltd., Amsterdam.