Journal of Geomatics Vol. 13, No. 2, October 2019

Forecasting and visualization of NDVI series using statistical methods through Web-GIS

Ujjwal K. Gupta*[,1], Vidit Shah[2] and Markand P. Oza[1]
[1]Space Applications Centre (ISRO), Ahmedabad, 380015
[2]Ahmedabad University, Ahmedabad, 380009
*Email: ujjwal_gupta@sac.isro.gov.in

(Received: Jan 02 2019; in final form: Oct 10, 2019)

Abstract: Use of remote sensing derived information in agriculture sector plays an important role for decision makers to know status of vegetation on larger spatial scale. Researchers have developed various indices for this purpose. Normalised Difference Vegetation Index (NDVI) is one such indices that measures vegetation vigour of crop. Agricultural community has shown its uses in various applications viz. vegetation health monitoring, crop growth assessment, crop yield estimation etc. NDVI forecasting helps to make an educated guess, based on its temporal behaviour in past years, on likely vegetation condition ahead of time and thus supports decision makers to formulate mitigation strategy. In this paper, Moderate resolution Imaging Spectroradiometer (MODIS) satellite data is used for calculating NDVI and then, NDVI forecasting is performed based on 2002-2016 NDVI time series data. This paper uses two statistical approaches, (i) Conditional mean and variance based statistical approach and (ii) Auto Regression Integrated Moving Average (ARIMA) for forecasting. Forecast of NDVI values are compared with observed NDVI data for year 2016-2017 using mean absolute percentage error (MAPE). To investigate seasonal impacts on forecast of NDVI, average seasonal MAPE is calculated which is found to be 5-10% in Rabi (Winter season; Mid-October to March) and Zaid (Summer season; April to June) seasons over Gujarat region of India. This seasonal study of NDVI forecasting can be used to supplement the routine monitoring of environmental conditions for wide range of applications. The developed forecasting model is currently operationalized on VEDAS web portal (https://vedas.sac.gov.in).

Keywords: NDVI forecasting, ARIMA, MODIS, VEDAS, Web-GIS

1. Introduction

India is an agricultural dependent economy where farming is main activity. The impact of uncertainty of nature on the farming cannot be eliminated but the impact of nature can be reduced through proactive techniques. To keep track of progress of crop, remote sensing (RS) data can be used with its multi-spectral and multi-temporal observations. In tropical country like India where climate conditions varies a lot spatially, it becomes a challenging task to monitor vegetation status regularly. In such a scenario, it becomes evident to use satellite data through remote sensing to simplifying monitoring on both temporal and spatial scales.

Researches have derived a number of indices for showing vegetation status judiciously. One such index is Normalised Difference Vegetation Index (NDVI), a satellite derived indicator of vegetation vigour and its healthiness (Tucker, 1979). NDVI is derived as:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

where, NIR denotes near infrared region reflectance and red denotes reflectance in red region of electromagnetic spectrum, respectively. Agricultural community has shown its uses in various applications viz. vegetation health monitoring (Zhou et al., 2001, Sharma and Mishra, 2012), crop growth assessment (Oza, 2014), crop assessment and production forecasting (Sridhar et al., 1994, Parihar and Oza, 2006) etc.

For a country like India with multi-season crops, a seasonal understanding of vegetation becomes important apart from overall vegetation understanding. In India, there are mainly three crop seasons, Kharif (Monsoon season; July-Mid October), Rabi (Winter season; Mid-October to March) and Zaid (Summer season; April - June). Seasonal understanding also becomes more important in satellite based agriculture studies due to effects of seasonal climatic variation on quality of optical multispectral imagery.

Another important challenge is to develop a web based data visualization platform as decision support system that can ingest huge satellite data and help decision makers to judge their decision criteria in best available manner. A decision support system that not only gives present status of conditions but also shows picture of future conditions with certain confidence level is always preferred. It can be developed using Geographical Information System (GIS) technologies. Web-GIS is a special form of GIS that often uses web technologies to communicate among different components in its architecture containing at least one server and one client. A web-GIS based solution provides global reach, easy to use functionality and better cross platform capabilities. It enables diverse data at a unified place for convenience of decision makers.

2. Study area and dataset details

Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices are designed to provide consistent spatial and temporal comparison of vegetation conditions. The 16-day Maximum Value Composite (MVC) NDVI product from MODIS-Terra (MOD13A1) is used in the present study (Didan, 2015). The MODIS

NDVI provides systematically processed data series. These are computed from atmospherically corrected bi-directional surface reflectances that have been masked for water, clouds, aerosols and cloud shadows. The spatial resolution of MODIS product used in the present study is 500 m. To determine the NDVI, red reflectance (0.645μm) and NIR (0.858 μm) reflectance are used.

For this study, Gujarat region is chosen which is located at the western side of India as shown in figure 1. It is preferred due to its diversified land use land cover (LULC) and triple crop cycle system (Mishra et al., 2017). Hence, it gives extra benefit to statistical models to learn in term of variance.

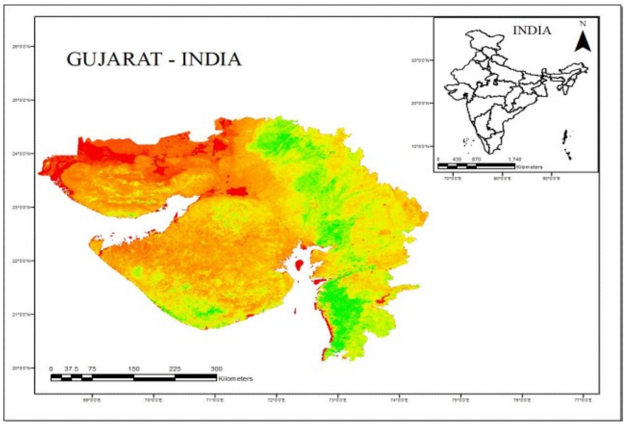

Figure 1: Classified image of Gujarat region of India based on NDVI values showing diversified vegetation in region

3. Methodology

Time series analysis in remote sensing helps in identifying the nature of the phenomenon represented by the sequence of the observation of RS data and exploit it to forecast future values of time series variables. MOD13A1 data has a temporal resolution of 16 days. To forecast NDVI values two methods are used: (1) Conditional mean and variance based statistical method (Oza, 2014), and (2) Auto Regression Integrated Moving Average (ARIMA) (Box and Jenkins, 1976).

3.1 Conditional mean and variance based statistical method

Let Y be a vector of length (n X 1) representing temporal profile of data, Y_1 be vector of length (n_1 X 1; such that $n_1 < n$) corresponds to observed temporal series and Y_2 be a vector of length (n_2 X 1; such that $n_2 = n - n_1$) corresponds to missing temporal profile. Then, mean temporal profile (M) using k samples can be calculated as given in eqn. (1),

$$M = k^{-1} \sum_{i=1}^{k} Y_i \qquad ...(1)$$

and dispersion matrix (S) can be computed by using eqn. (2) as,

$$S = (k-1)^{-1} \sum_{i=1}^{k} [(Y_i - M)(Y_i - M)^T] \qquad ...(2)$$

where, T denotes matrix transpose operation. Dispersion matrix, S is a square, symmetric and positive definitive matrix with size of n X n. It can be decomposed as,

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

where S_{11}: submatrix of size n_1 X n_1, corresponds to observed data

S_{12}: submatrix of size (n_1 X n_2)

S_{21}: submatrix of size (n_2 X n_1)

S_{22}: submatrix of size (n_2 X n_2)

From mean temporal profile and dispersion matrix, unobserved data values (Y_2) can be conditionally computed from (Y_1) as given in eqn. (3),

$$E(Y_2/Y_1) = M_2 + S_{21}S_{11}^{-1}(Y_1 - M_1) \qquad ...(3)$$

Here, E stands for expected value vector. The mean vector M_2 from observed time series corresponding to unobserved data period contains “apriori” information. The term $S_{21}S_{11}^{-1}$, derived from “hidden” pattern from data, controls the correction that is applied to “apriori” information. The more rough its history is more changes are required in “apriori” information.

3.2 Auto Regression Integrated Moving Average (ARIMA)

The ARIMA models are one of the classes of stochastic models for describing time series (Box and Jenkins, 1976). ARIMA models include three basic types: autoregressive (AR) models, moving average (MA) model and combined MA and AR model which is ARMA model. In ARIMA model approach, the past observations are analyzed to formulate a model describing the inner correlation among them. The acronym in ARIMA is descriptive and capturing key aspect of model itself. Briefly they are,

AR (p): Auto Regression. A sub-structure that uses dependent relationship between the current observation and number of lagged previous observations.

$$X_t = \sum_{j=1}^{p} \emptyset_j X_{t-j} + \omega_t$$

I (d): Integrated. The letter “I” in ARIMA indicates that the modeling time series has been transformed into a stationary time series. The ‘d’ value is used to remove the trend from given time series so that overall series becomes stationary.

MA (q): Moving Average. A sub-structure that uses the dependency between a current observation and a residual error ($\in_j$) from a lag observation.

$$X_t = \sum_{j=1}^{q} \emptyset_j \in_{t-j} + \in_t$$

Each of these components are explicitly specified in the model as a parameter. “ARIMA (p,d,q)” is the generalized notation where parameters are substituted in the integer value indicating which ARIMA model is being used.

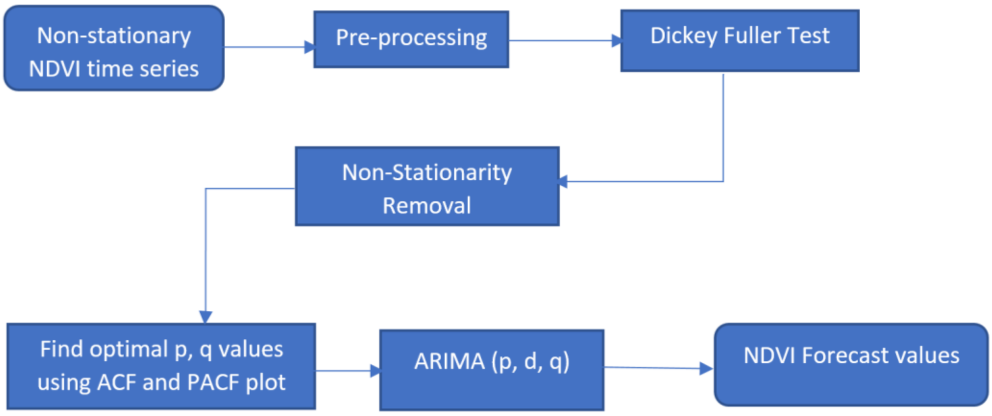

Figure 2: ARIMA methodology flow chart for NDVI forecast

3.3 Dickey Fuller Test of stationarity

A stationary time series can be defined when the data has a constant mean and no trend overtime. However, sometime time series may not be stationary and have trend line. In such cases, stationary is introduced by first order derivative, logarithmic operations etc. Stationarity of time series is tested by Dickey Fuller test (Said and Dickey, 1984).

Suppose there is an auto regressive time series:

$$y_t = \emptyset y_{t-1} + \epsilon_t$$

It can be transformed by first order derivative into:

$$y_t - y_{t-1} = (\emptyset - 1)y_{t-1} + \epsilon_t$$

If delta operator i.e. $\Delta y_t = y_t - y_{t-1}$ and β = $(\emptyset - 1)$ are set, above equation can be rewritten as,

$$\Delta y_t = \beta y_{t-1} + \epsilon_t$$

This makes null and alternative hypothesis of this test as follows:

H_0 : Unit root is present in autoregressive model i.e. $\beta = 0$

H_1: The series has no unit root i.e. $\beta \neq 0$

The null hypothesis is rejected when p value of test is less than critical value i.e. series is stationary and vice versa. Here, critical value of 5% is selected for rejecting null hypothesis.

For making NDVI time series stationary, first derivative of series is used i.e. d=1. After removing seasonality, Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) are plotted in order to get appropriate values of p and q in ARIMA (p, d, q) respectively. ACF plot is a bar chart of coefficients of correlation between a time series and lag of itself whereas PACF plot is a bar chart of partial correlation coefficients between series and its lag. From ACF and PACF plots, highest values of p and q values are figured out as 3 and 2 respectively.

A flow chart depicting ARIMA methodology is shown in figure 2.

For selecting best p and q values, randomly 1000 locations are selected in Gujarat state as shown in figure 3 and Akaike Information Criterion (AIC) value is calculated for each p and q pair. These AIC values are used as estimator of relative quality of statistical models. From AIC values of each pair, highest AIC value pair is selected as best model. Figure 4 shows p and q pair values with its frequency of occurrences as best model.

In this paper, Mean Absolute Percentage Error (MAPE) is used as a measure of accuracy of statistical forecast methods. MAPE is calculated as:

$$MAPE = \frac{\sum_{t=1}^{n} \left| {(A_t - F_t)}/{A_t} \right|}{n}$$

where, A_t is original value, F_t is forecast value for n number of sample points.

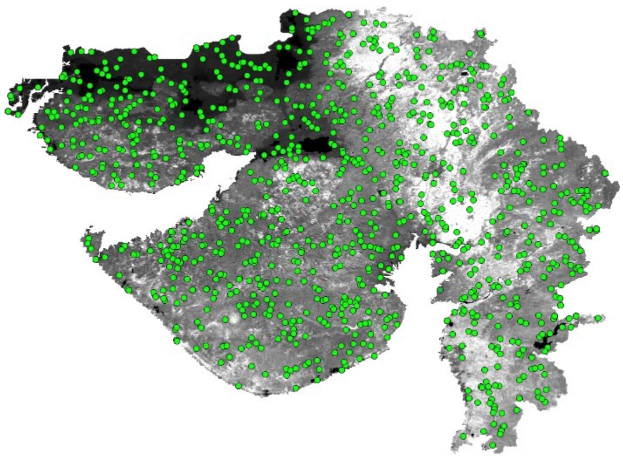

Figure 3: Randomly selected 1000 sample points over Gujarat region

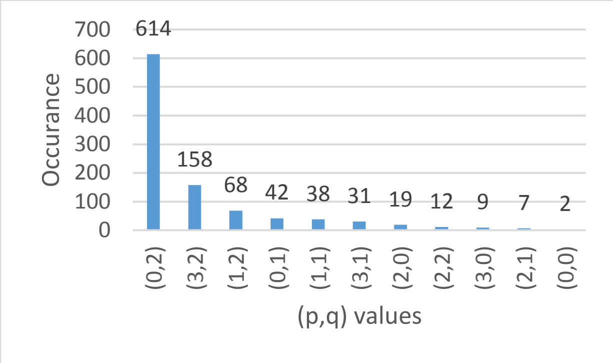

Figure 4: Frequency distribution for best model selection for each (p, q) pair

4. Results and discussion

Based on above discussed methodology, conditional mean and variance method as well as ARIMA are applied to MODIS NDVI time series data from 2002 to 2016, which was later verified with the NDVI data of 2016-2017. NDVI Forecast are done for up to next 3 forecast values corresponding to 16 days, 32 days and 48 days forecasts. Table 1 consolidates average MAPE results of both methods based on annual time series data for a sample date over Gujarat region. From table 1, it can be inferred that both the methods perform equally well in annual forecast.

Table 1: Mean Absolute Percentage Error (MAPE) for ARIMA and conditional mean and variance based method using annual time series for 08 JAN 2017

Forecast Days	Avg. ARIMA (0,1,2) MAPE	Avg. Conditional Mean & Variance MAPE
16	**3.02**	3.57
32	7.8	**7.2**
48	**9.47**	9.74

Table 2: Mean Absolute Percentage Error (MAPE) for ARIMA and conditional mean and variance based method using seasonal time series

Season	Forecast Days	Avg. ARIMA (0,1,2) MAPE	Avg. Conditional Mean & Variance MAPE
Rabi	16	**6.758**	8.74
	32	**8.936**	9.47
	48	12.457	**11.25**
Kharif	16	**19.871**	23.24
	32	**21.249**	26.95
	48	**26.478**	37.95
Zaid	16	**5.412**	8.82
	32	**6.451**	9.33
	48	**8.969**	9.89

For checking robustness of forecast methods under seasonal variations, average MAPE is calculated for Gujarat region. Table 2 summarizes seasonal effects on forecasting models.

It can be seen and analysed from table 2 that both forecasting models underperform in Kharif season. The reasons may be limitations of optical multispectral data in cloudy and foggy conditions. In other seasons, both methods give a satisfactory forecast error (MAPE) of 5-10% averaged over Gujarat region. In general, ARIMA performs well as compared to conditional mean and variance based forecast model under seasonal variations.

Currently, Annual ARIMA based NDVI forecast is implemented using MODIS NDVI time series pixel drilling. It is available at VEDAS web portal (https://vedas.sac.gov.in) through web GIS as Open Geospatial Consortium (OGC) compliant web map service. Figure 5 demonstrates a screenshot of operationalized model showing forecast result along with previous years NDVI time series data.

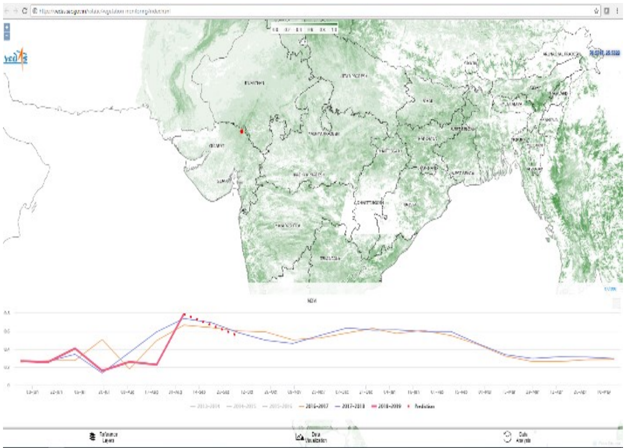

Figure 5: Forecast results of operationalized statistical model at VEDAS web portal through web-GIS

5. Conclusion and future work

In this paper, two standard statistical methods are used and found them suitable for NDVI forecast purpose with 5-10% seasonal MAPE in Rabi and Zaid seasons when averaged over Gujarat region for up to three forecasts. It was found that MAPE in Kharif season is more compared to other seasons which may be due to limitation of optical multispectral data in cloudy and foggy conditions. While forecasting NDVI values using annual timeseries data, both methods, (i) conditional mean and variance based statistical method as well as (ii) ARIMA method, are found to be performed at par to each other. NDVI forecast using statistical methods can be applied using remote sensing data at both temporal and spatial scales. These methods can be applied for forecast related to homogeneous cropping area for better understanding of phenological process of crop in future.

References

Box, G.E. and G. M. Jenkins (1976). *Time Series analysis: forecasting and control, revised ed*. Holden-Day.

Didan, K. (2015). *MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD13A1.006

Mishra, S., M. P. Oza and S. A. Sharma (2017). Crop phenology identification using NDVI time-series and its dissemination using WebGIS, Proceedings of 38th Asian Conference on Remote Sensing (ACRS-2017), 23-27 October, New Delhi, India.

Oza, M. (2014). Use of Statistical Method to Remote Sensing Data for In-season Crop Growth Assessment. Journal of the Indian Society of Remote Sensing, 42(1), 243-248

Parihar, J. S., and M. P. Oza (2006). FASAL: an integrated approach for crop assessment and production forecasting. In *Agriculture and Hydrology Applications of Remote Sensing* (Vol. 6411, p. 641101). International Society for Optics and Photonics.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 71(3), 599-607.

Sharma, S. A. and S. Mishra (2012). Web-GIS based monitoring of vegetation using NDVI profiles. Journal of Geomatics, 6(2), 109-112.

Sridhar, V. N., V. K. Dadhwal, K. N. Chaudhari, R. Sharma, G. D. Bairagi and A. K. Sharma (1994). Wheat production forecasting for a predominantly unirrigated region in Madhya Pradesh, India. Remote Sensing, 15(6), 1307-1316.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment, 8(2), 127-150.

Zhou, L., C. J. Tucker, R. K. Kaufmann, D. Slayback, N. V. Shabanov and R. B. Myneni (2001). Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. Journal of Geophysical Research: Atmospheres, 106(D17), 20069-20083.