Change detection and trend analysis for Oceansat-2 Ocean Color Monitor (OCM-2) time series data

Ravi Kamal Choudhary*, Tushar Shukla, Sampa Roy and Debajyoti Dhar
Space Applications Centre, ISRO, Ahmedabad, India
*Email: choudharyravi@sac.isro.gov.in

(Received: Jan 05, 2019; in final form: Jun 06, 2019)

Abstract: Satellite ocean-colour observations are now widely recognized as an important component of international remote sensing programs. Ocean Color Monitor (OCM-2), one of the payloads aboard Oceansat-2 (2009), is designed to obtain quantitative information of ocean-colour variables e.g. chlorophyll-a concentration, etc. The two important tools for analyzing multi-temporal data obtained from earth observation satellite are Change Detection and Time series Trend Analysis. The remote sensing data has become a heart of change detection and trend analysis techniques because of its high temporal frequency and wider selection of spatial and spectral resolution. The general objectives of change detection in remote sensing include recognizing the geographical location and type of changes, quantifying the changes, and assessing the accuracy of change detection results. Change detection is useful in many applications such as land use changes, habitat fragmentation, rate of deforestation, coastal change, urban sprawl, and other cumulative changes. In time series analysis, the goal is to estimate the future value using the behaviors in the past data. Trend Analysis predicts the values over cloudy and missing data, thus helping in generating a uniform time series. This paper discusses remote sensing techniques namely Differencing and Principal Component Analysis (PCA) for Change Detection and ARIMA based method for Trend Analysis on multi-temporal OCM-2 data.

Keywords: Change Detection, Trend Analysis, OCM-2, Oceansat-2, Vegetation fraction, Chlorophyll-a, ARIMA, Image Differencing, PCA, Time series.

1. Introduction

OCEANSAT-2 spacecraft of Indian Space Research Organization (ISRO) is the second satellite in ocean series, which carried three main instruments namely i) Ku band pencil beam Scatterometer, ii) Ocean Colour Monitor (OCM) and iii) Radio Occultation Sounder of Atmosphere (ROSA) instrument of Italian Space Agency (ASI). The main objectives of OceanSat-2 are to study surface winds and ocean surface strata, observation of chlorophyll concentrations, monitoring of phytoplankton bloom, study of atmospheric aerosols and suspended sediments in the water. The OCEANSAT-2 OCM is mainly designed to provide continuity to the OCEANSAT-1 OCM instrument and OCM collects data in 8 spectral bands operating in the Visible-Near IR spectral range and the imaging principle of OCM is based on push-broom technique. The OCM application lies in ocean application like identifying potential fishrie zone (PFZ). The configuration of the OCM-2 payload is identical to the OceanSat-1 OCM except that the spectral band is modified for band 6 and band 7. For band 6, the central wavelength is shifted from 670 nm to 620 nm to improve the reflectance from suspended sediments; for band 7, the cenetral wavelength is shifted from 760 nm to 740 nm to avoid oxygen absorption. Ocean Color Monitor (OCM-2) instrument is designed to obtain quantitative information of ocean-colour variables e.g. chlorophyll-*a* concentration, vertical diffuse attenuation of the light, (*Kd*) and total suspended matter (TSM) concentration in coastal waters, apart from ocean-colour information OCM data is also useful for studying the aerosol transport and terrestrial bio-sphere.

Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times. It involves the ability to quantify temporal effects using multi temporal data sets (Singh, 1989). It involves the analysis of two registered multispectral remote sensing images acquired in the same geophysical area at different times to measure how the attributes of a particular area have changed between two or more time periods. Two remote sensing techniques, including Image Differencing and Principal Component Analysis (PCA) were used to detect the changes in multi-temporal OCM-2 data, such as vegetation fraction (VF), land surface water (LSW), chlorophyll-*a* concentration (CLO) and aerosol optical depth (AOD).

A Time Series (TS) is a collection of observations made sequentially through time. The TS data used to provide visual information to the unpredictable nature of the data. It is an ordered sequence of observations of a variable or captured object at equally distributed time interval. TS is anything which is observed sequentially over the time at regular interval like hourly, daily, weekly, monthly, quarterly etc. TS data is important when you are predicting something which is changing over the time using past data. To carry out the change detection techniques, trend analysis and ARIMA forecast, Oceansat-2 (OCM-2) satellite data were used.

The following basic things have been addressed in this paper

1. Change Detection methodology

2. Change Detection techniques- Image Differencing and Principal Component Analysis (PCA)

3. Trend Analysis and Auto-Regressive Integrated Moving Average (ARIMA) forecast model

2. Change detection

The remote sensing is very important part of the change detection techniques. The time and accuracy of change detection on the earth's surface can provide a better understanding of the relationship and interaction between human and natural phenomena. The change detection problem discussed in this paper is as follows: we are given two images of the same scene taken at several different times. The goal is to identify the set of pixels that are "significantly different" between these two images, these pixels comprise the change mask. The change mask may result from a combination of underlying factors, including appearance or disappearance of objects, motion of objects relative to the background, or shape changes of objects.
There are many different methods for doing change detection (Jianya et al., 2008); this paper analyses the Image Differencing and PCA change detection techniques using OCM-2 satellite images. Image differencing is pixel-based method while principal component analysis is transformed based method of change detection. Classification based methods such as post classification methods and two-date image clustering specify the occurrence of changes in image pixels and label them. Various methods have been proposed to detect changes captured by satellite images. The algebraic methods like image differencing is relatively simple and easy to understand. The disadvantage of these methods are in choosing an appropriate threshold. However, this method does not extract details of the changes completely. Transformation methods focus on reducing the data between bands, and value different information in derived components.
A basic change detection algorithm takes the image sequence as input and generates a binary image 'B' containing [0, 1] called a change mask that identifies changed regions in the last image according to the following generic rule (Radke et al., 2005):

$$B(x) = \begin{cases} 1, if\ there\ is\ a\ significant\ change\ at \\ \quad pixel\ x\ between\ consecutive\ images \\ 0, otherwise \end{cases}$$

Before implementing change detection analysis, the following conditions must be satisfied (Lu et al., 2004):
1. Precise registration of multi-temporal images;
2. Precise radiometric and atmospheric correction or normalization between multi-temporal images;
3. Region/ Area of Interest: same geographic location, free of clouds in the area of analysis;
4. Remote sensing system consideration- spatial, spectral, radiometric and temporal: whenever possible, select images acquired from the same type of sensors, with the same spectral and spatial resolutions, and the same seasonal timeframe in order to minimize unwanted variances.

2.1 Change detection methodology

The methodology used for analysis of different change detection methods using satellite images is divided into five major parts as shown in figure 1. The method involves processing of Oceanasat-2 (OCM-2) images that are detecting the change in study area. The image processing techniques using first preprocessed the images then image processing is used for different change detection algorithms. The block diagram of proposed system as shown in figure 1.

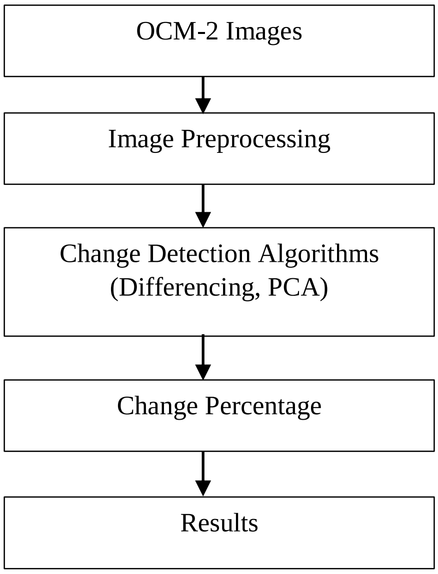

Figure 1: Block diagram of Change Detection System

2.2 Change detection techniques

After the pre-processing stage was completed, two different change detection algorithms were applied to both the images. These were Image Differencing and Principal Component Analysis. The first is pixel based method while the second is transformation based method of change detection.

2.2.1 Image differencing

Image differencing is very intelligible method of the change detection techniques. It is applied to a wide variety of the images and geographical data. It is generally conducted on the basis of gray level images. Image differencing is used widely because of its simplicity to implement and interpret. It involves absolute subtraction of the second-date image from a first-date image, pixel by pixel. The changed and unchanged area is determined by selecting the appropriate threshold value of gray level subtraction image using Otsu's method (Otsu, 1979).

Threshold value (T) is determined from difference image, which decides the quality of change detection. Choosing this suitable threshold value can be maximum separated for the areas of real change. A critical aspect of image differencing method is deciding where to place the threshold boundaries between change and no-change pixels displayed in the histogram. Otsu's automatic thresholding algorithm is used to find out the threshold. Change mask is then prepared by assigning '1' to pixels which are significantly changed in second-date image as compare to first-date image and '0' otherwise representing change and no-change regions respectively. Mathematically,

Difference image, $D(x) = Image1 - Image2$

Change mask, $B(x) = \begin{cases} 1\ \ if\ \ |D(x)| > T \\ 0\ \ otherwise \end{cases}$

2.2.2 Principal Component Analysis (PCA)

The principal components transformation is a linear transformation that defines a new, orthogonal co-ordinate system such that that data can be represented without correlation. PCA is a technique to emphasize variation and bring out strong patterns in a data set. It converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (Gong, 1993; Fung and LeDrew, 1987). It is a widely used method for dimensionality reduction. The three main steps towards implementing PCA algorithm are:

1. Difference image generation and Eigen vector space (EVS)
2. Building the feature vector space (FVS)
3. Clustering of the feature vector space and change map

2.2.3 Difference image generation and Eigen vector space (EVS)

The difference image has the absolute valued differences of the intensity values of the corresponding pixels of the 2-grayscale images. The computed difference image would hence be such that the values of the pixels associated with land changes will have values significantly different from those of the pixels associated with unchanged pixels.

$$difference\ image\ (i,j) = |\ Image1\ (i,j) - Image2\ (i,j)|$$

To construct the EVS, we take non-overlapping blocks of size N X N from the difference image and flatten them into row vectors to make vector set from the difference image. PCA takes this vector set and determines its co-variance matrix after performing mean normalization on it. The Eigen vectors and Eigen values of the co-variance matrix are computed (giving us the EVS) and then the Eigen vectors are sorted in the descending order of Eigen values.

2.2.4 Building the Feature Vector Space (FVS)

Taking N x N blocks from the difference image, flattening them, and lastly projecting them onto the EVS, forms feature vector space. This time the blocks are overlapping. A vector space (VS) is first made by constructing one vector for each pixel of the difference image such a way that one N x N block is actually a pixel's N x N neighborhood. FVS is constructed by projecting the vector set on to the EVS, simply means to perform the following matrix multiplication:

$$Feature\ Vector\ Set\ (FVS) = Vector\ Set(VS).Eigen\ Vector\ Space(EVS)$$

2.2.5 Clustering of the feature vector space and change map:

The feature vectors for the pixels carry information whether the pixels have characteristics of a changed pixel or an unchanged one. Having constructed the feature vector space, we now need to cluster it so that the pixels can be grouped into two disjoint classes, changed and unchanged class. To do that we used K-means algorithm. Thus, each pixel will get assigned to a cluster in such a way that the distance between the cluster's mean vector and the pixel's feature vector is the least. Each pixel gets a label from 1 to K, which denotes the cluster number that they belong to. It can be postulated that the cluster that has the highest mean is the cluster that belongs to the changed class. The reason behind the cluster which has the highest value of mean belongs to changed class is that the values of the difference image pixels in a region where some changes have occurred are higher than the values of pixels in the regions where there is no change.

Then we build a change map – a binary image to show the output of change detection. We are showing the background blackcurrant, i.e., intensity value of those pixels is zero and the changes in yellow, i.e., intensity value of those pixels is 1. Thus

$$Change\ map\ (i,j) = \begin{cases} 1, & if\ (i,j) \in\ highest\ mean\ cluster \\ 0, & otherwise \end{cases}$$

3. Trend analysis

A time series is said to be continuous when observations are made continuously through time and is said to be discrete when observations are taken only at specified times, usually equally spaced. The special feature of TS analysis is the fact that successive observations are usually not independent and that the analysis must take into account the time order of the observations. The first step in the analysis is usually to plot the observations against time to give what is called a time plot, and then to obtain simple descriptive measures of the main properties of the series. A graph will not only show up trend and seasonal variation, but will also reveal any wild observations or outliers that do not appear to be consistent with the rest of the data. Other features to look for in a time plot include sudden or gradual changes in the properties of the series. If there is some sort of discontinuity in the series, then different models may need to be fitted to different parts of the series. The prediction of the future values of the series is another important task of time series analysis.

This paper includes plotting the data and looking for trends, seasonal fluctuations and forecasting of time series data based on model called autoregressive integrated moving average (ARIMA). Traditional methods of time series analysis are mainly concerned with decomposing the variation in a series into components representing trend, seasonal variation and other cyclic changes. Any remaining variation is attributed to 'irregular fluctuations'. This approach is not always the best but is particularly valuable when the variations dominated by trend seasonality. Vegetation fraction over a specific area for year 2012 is choosen for trend analysis. Median is choosen to remove effect of extreme values in time series and to capture spatial variation coming from experimental area. The missing value in time series is estimated by applying cubic spline fitting to the known data. Trend and Seasonal variations are removed from cubic spline fitted time series to make it stationary before applying ARIMA forecasting model.

The components, by which time series is composed of, are called component of time series data. The four basic component of time series data are: Trend, Seasonal variations, Cyclical variations and Irregular (random) variations.

3.1 Time series forecast models

Most of the Time series (Das, 1994) models work on the assumption that the time series is stationary. TS is said to be stationary if its statistical properties such as mean, variance remain constant over time. Intuitively, we can say that if TS has a particular behavior over time, there is a very high probability that it will follow the same in the future. Almost none of the TS are stationary. There are two major reasons behind the non-stationarity of TS:

1. Trend: varying mean over time.
2. Seasonality: variations at specific time frames.

Therefore, to make TS stationary, we need to remove trend and seasonal variations from TS. There are different ways to estimate and eliminate trend from a TS, and some of commonly used are:

1. Aggregation: taking average for a time period like monthly/weekly average.
2. Smoothing: taking rolling averages.
3. Polynomial fitting: fit a regression model.

Seasonal variations are eliminated from a time series by doing seasonal adjustment to get seasonally adjusted TS.

3.1.1 Auto-Regressive Integrated Moving Avergae (ARIMA)

ARIMA (Box and Jenkins, 1970; Mondal et al., 2014) is a statistical technique that uses time series data to predict future. The ARIMA model is a combination of autoregressive (AR), integration (I) - referring to the reverse process of differencing to produce the forecast, and moving average (MA) operations (Farhath et al., 2016) . In ARIMA (p, d, q) model; p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive components, the number of differencing operators, and the highest order of the moving average term of the model respectively.
ARIMA modelling will take care of trend, seasonal and cyclic variations of a data set while making forecasts. ARIMA requires stationary TS to forecast. In ARIMA, ‘d’ parameter controls the forecast of TS based on stationarity. In ARIMA model a non-stationary time series is made stationary by applying finite differencing of the data points. The ARIMA forecasting for a stationary time series is nothing but a linear (like a linear regression) equation. The predictors depend on the parameters (p, d, q) of the ARIMA model:

1. Number of AR (Auto-Regressive) terms (p): AR terms are just lags of dependent variable.
2. Number of MA (Moving Average) terms (q): MA terms are lagged forecast errors in prediction equation.
3. Number of Differences (d): These are the number of non-seasonal differences.

To determine the values of ‘p’ and ‘q’, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used.

1. Autocorrelation Function (ACF): It is a measure of the correlation between the TS with a lagged version of itself.
2. Partial Autocorrelation Function (PACF): This measures the correlation between the TS with error term.

Autoregressive models

ARIMA methodology attempts to describe the movements in a stationary time series as a function of "autoregressive and moving average" parameters. An AR model with k parameters may be written as

$$X(t) = A(1) * X(t-1) + A(2) * X(t-2) + \cdots + A(k) * X(t-k) + E(t)$$

Where,
X(t) = time series under investigation
A(k) = the autoregressive parameter of order k
X(t-k) = the time series lagged k period
E(t) = the error term of the model

The p, d,and q parameters are retrieved from experimental area. The best suited ARIMA model p, d and q values are carried out for experimental data based on minimum MAPE value. These values for various parameters are shown in table 1.

Table 1: Best suited ARIMA (p, d, q) parameters retrieved from experimental data

S. No.	Parameter	ARIMA Model (p,d,q)
1.	Vegetation fraction	(5,1,0)
2.	Land Surface Water	(3,1,0)
3.	Chlorophyll-a concentration	(2,1,0)
4.	Aerosol Optical Depth	(2,1,0)

4. Results

The Change Detection (CD) techniques (Image Differencing, Principal Component Analysis (PCA)) are tested on multiple OCM-2 datasets. In time series forecasting experiments, ARIMA model is fitted with the best-suited model parameters (p, d and q) for experimental data. We tested the original time series with the ARIMA forecasted values and observed that the forecasted values are nearly matching with the original time series. The Mean Absolute Percentage Error (MAPE) was used to measure the prediction accuracy of the forecasting model. Change Detection results for vegetation fraction are shown in figure 2, where change map is binary map containing one for changed pixels and zero for no changed pixels and change map magnitude shows how much value of pixels are changed between these two dates.

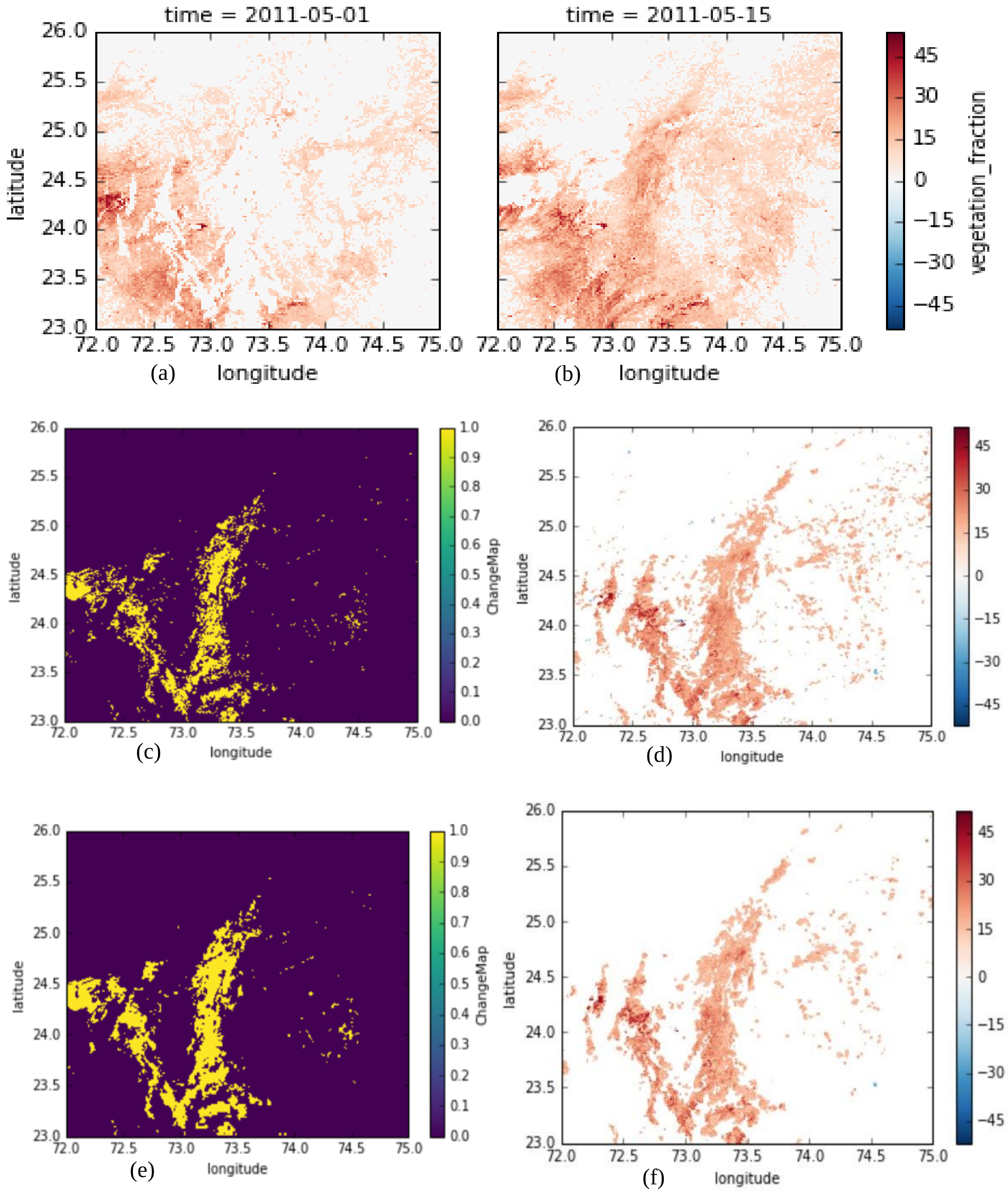

Figure 2: Vegetation fraction of date (a) 01-05-2011, (b) 15-05-2011, (c) Image differencing change Map, (d) Image differencing change map magnitude, (e) PCA change map, (f) PCA change map magnitude

Trend Analysis results are shown in figure (3a - 3d), where plot (a) shows the median time series of a parameter for the choosen dates, (b) shows the cubic spline fitting using non-zero values of time series, (c) shows Train data, Test data; which are 66% and 34% of seasonally adjusted TS respectively and ARIMA forecast for Test data, (d) shows Original TS (Cubic spline fitted TS), Smoothed TS (after removing trend from spline fitted TS), Seasonally Adjusted TS (after removing seasonal variation from smoothed TS) and ARIMA forecast TS for Test data. As it can be seen from figure 3a that OCM-2 data contains lot of missing values, therefore to estimate the missing values the cubic spline fitting is done to non zero data.

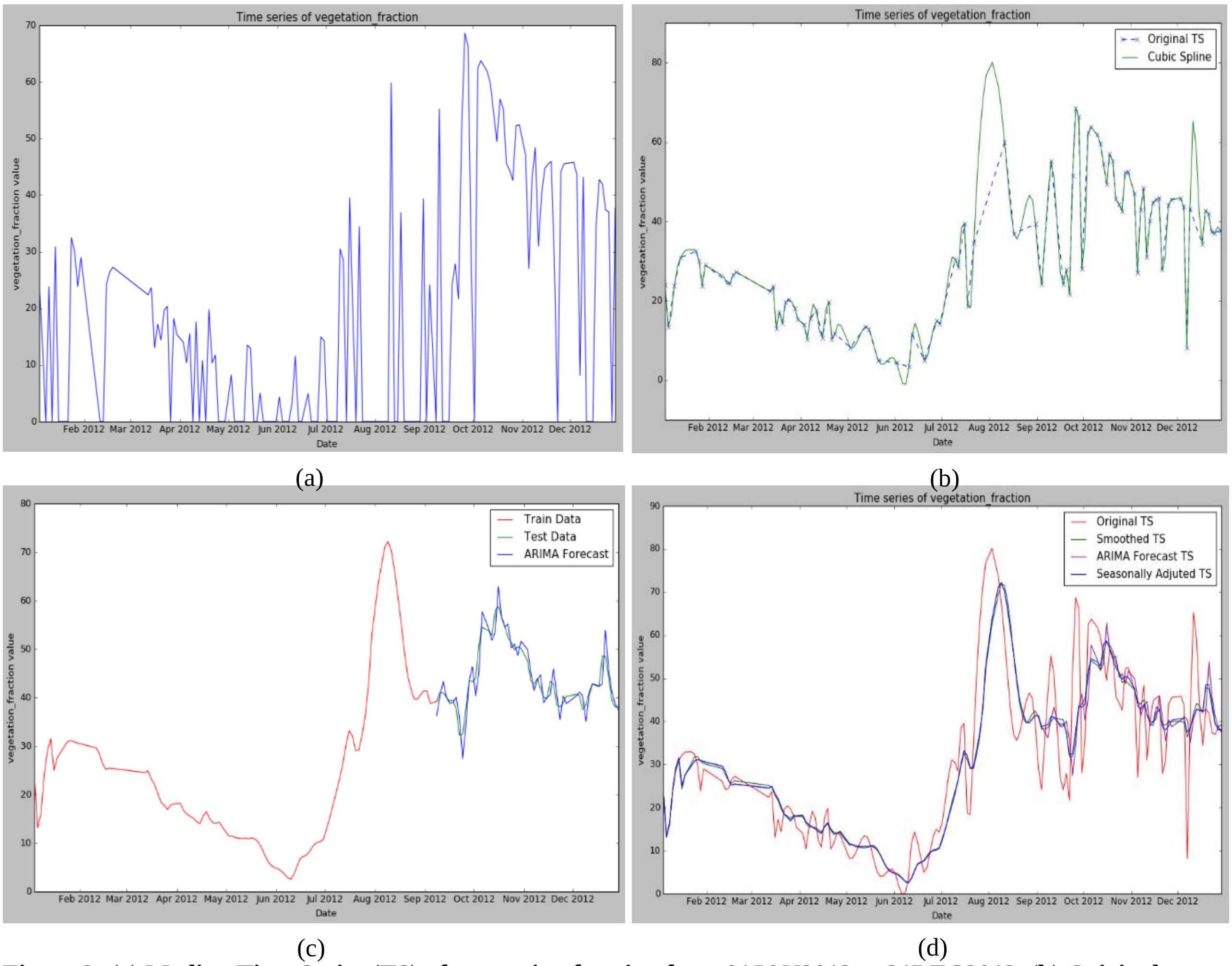

Figure 3: (a) Median Time Series (TS) of vegetation fraction from 01JAN2012 to 31DEC2012, (b) Original non-zero and Cubic Spline fitted TS, (c) Train, Test data and ARIMA forecast for Test, (d) Original TS, Smoothed TS, Seasonally Adjusted TS and ARIMA forecast TS for Test data

MAPE for various parameters are shown in table 2.

Table 2: MAPE for various parameters

S.No.	Parameter	MAPE
1.	Vegetation fraction	4.52
2.	Land Surface Water	0.77
3.	Chlorophyll-a concentration	3.99
4.	Aerosol Optical Depth	2.49

5. Conclusion

Two remote sensing techniques namely Differencing and Principal Component Analysis (PCA) for Change Detection and ARIMA based method for Trend Analysis on multi-temporal OCM-2 time series data is discussed in detail. . The change detection outcomes are change map, change map magnitude and percentage of area changed between two images. The percentages of changed area for Differencing and PCA are 7.4 and 8.8 respectively for experimental data.

This paper also forecast various parameters using statistical method , ARIMA, where best suited (p,d,q) values are estimated from minimum MAPE values. The MAPE values of 4.52, 0.77, 3.99 and 2.49 are observed for Vegetation Fraction, Land Surface Water, Chlorophyll-a concentration and Aerosol Optical Depth respectively for experimental data using best suited (p,d,q) values.

References

Box, G.E.P., and G.M. Jenkins (1970). Time Series Analysis: Forecasting and Control, CA: Holden-Day, San Francisco.

Das, S. (1994). Time Series Analysis. Princeton University Press, Princeton, NJ.

Farhath, Z.A., B. Arputhamary, L. Arockiam (2016). A survey on ARIMA forecasting using time series model, International Journal of Computer Science and Mobile Computing, 5(8), 104-109.

Fung, T. and E. LeDrew (1987). Application of principal components analysis to change detection, Photogrammetric Engineering and Remote Sensing 53(12), 1649-1658.

Gong, P. (1993). Change detection using principal component analysis and fuzzy set theory, Canadian Journal of Remote Sensing, 19(1), 22-29.

Jianya, G., S. Haigang, M. Guorui and Z. Qiming (2008). A review of multi-temporal remote sensing data change detection algorithms, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37(B7), 757-762.

Lu, D., P. Mausel, E. Brondizio and E. Moran (2004). Change detection techniques, International journal of remote sensing, 25(12), 2365–2401.

Mondal, P., L. Shit and S. Goswami (2014). Study of effectiveness of time series modelling (ARIMA) in forecasting stock prices, International Journal of Computer Science, Engineering and Applications, 4(2), 13.

Otsu, N. (1979). A threshold selection method from gray-level histograms, IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

Radke, R.J., S. Andra, O. Al-Kofahi and B. Roysam (2005). Image change detection algorithms: A systematic survey, IEEE Transactions on Image Processing, 14(3), 294-307.

Singh, A. (1989). Digital change detection techniques using remotely-sensed data, International Journal of Remote Sensing, 10(6), 989–1003.