

## Spatiotemporal analysis of LANDSAT Data for Crop Yield Prediction

Sandeep Kumar Singla\*, Rahul Dev Garg and Om Prakash Dubey

Geomatics Engineering Group, Department of Civil Engineering, Indian Institute of Technology (IIT), Roorkee, Uttarakhand, India - 247667.

Received 26 December 2017; Accepted 6 June 2018

### Abstract

Crop growth monitoring and yield estimation is an important process for agricultural economic return prediction and food security. Improvement in the accuracy and timeliness of the information about pre harvest prediction of crops by blending of ancillary data and remotely sensed data in the temporal domain lead to the effective and optimized decision making. Although previous studies have found strong correlation between observed and predicted yield based on Normalized Difference Vegetation Index (NDVI), yet there is a pressing need to provide more accurate and reliable yield in the small areas. The objective of the proposed model is to extract the information related to the crop yield for the sugarcane planted in the small fields of Himalayan foothills region. Relation of crop yield information with the different stages of crop growth has been considered in the proposed model. The process based on correlation analysis of spatiotemporal data to identify of the best periods for the reliable estimation of the sugarcane is presented. The best period for the crop predictability is during 210 to 270 days after plantation. Based on the historical data of 10 years, it has been found that predictability of nonlinear modelling is significant and is of the order of 0.6. By Tukey test and RMSE the best fit regression models are identified for the study area.

*Keywords:* Spatial and Temporal Analysis; Remote Sensing; LANDSAT; NDVI; Satellite information.

### 1. Introduction

Food security is the cradle of social and economic sustainability of a nation [1]. Crop growth monitoring and pre harvest estimation of crop production provide relevant forewarnings to the various agencies regarding the food deficit regions. Timely and reliable estimation of crop growth and yield information is essential for the development of handy tool for the assessment [2]. The use of remotely sensed data in the temporal domain together with the use of sophisticated statistical techniques goad to the effective and optimized decision making [3].

Satellite data in combination with Geographical Information System (GIS) and Global Positioning System (GPS) has been used by the researchers in the past as well as today, to gather the data required for monitoring of crops, management of resources and make decisions on farming practices [4]. There is a pressing need to generate and update the data required for analysis and to automate the procedure of crop inventory using the recent technologies such as geoinformatics [5]. This is an appropriate blending of modules like GPS, GIS, Remote Sensing and Relational Data Base Management System. Combination of geoinformatics tools and digital image processing algorithms has a crucial role in the prediction of yields and the monitoring of crop growth status [6].

Crop yield monitoring and estimation based on the remotely sensed data as well as historical records has been carried out over the last few decades so as to support the

agricultural statistics in an effective manner [7]. Various crop yield models on the basis of meteorological data, historical crop yield records and satellite data or the appropriate blending of these types of data has been considered for the prediction of crop yield, but only few models exhibit a strong match between predicted yield and observed yield [8].

The ultimate aim of this work is to propose a model to extract the information required to predict the accurate and near real time crop yield of sugarcane (*Saccharum officinarum* L.) planted in a small field. To obtain the reliable estimate of the crop yield, the analysis of the correlation has been done between temporal profile of Normalized Difference Vegetation Index (NDVI) data and historical crop yield records. Regression models to predict the crop yield have been developed on the basis of statistical analysis and extracted numerical NDVI values collected for the best predicted period. The obtained information can then be useful for the policy makers and agricultural scientists to support their decisions regarding the regional agricultural risks in the near future.

### 2. Literature Review

Information obtained from the satellite images may be used effectively and economically for crop yield estimation and growth monitoring [9]. The authors of [10] explored that information extraction about crop parameters based on remotely sensed data is most prominent research topic after the launch of civilian remote sensing program since 1960s. The authors also discussed two experiments such as Large Area Crop Inventory Experiment (LACIE) and Crop

\*E-mail address: sandy.dce2014@iitr.ac.in

Identification Technology Assessment for Remote Sensing (CITARS) related to the crop inventory and yield estimation. The authors in [11] proved that the association between Vegetation Indices (VI) derived from remotely sensed data and sugarcane yield is significant. The authors of [11] also suggested that there is low correlation between NDVI and yield due to the mixture of spectral signatures of different crop varieties and mixture of plant and ratoon crops. However the resolution of the spatial data obtained from the remote sensing images may be enhanced significantly using fusion and pan-sharpening methods and alternatively by multi-temporal information [12,13].

The researchers in [14,15] revealed that single date LANDSAT (Land Satellite) data is not sufficient to predict the sugarcane crop yield accurately and they suggested the use of multi-temporal satellite data. The authors in [15] developed a model based on multi-temporal LANDSAT Enhanced Thematic Mapper Plus (ETM+) data and yield of sugarcane at Umfolozi in South Africa. The authors of [16] presented a regression based model with coefficient of determination value ( $r^2 = 0.75$ ) between the sugar content present in the stalk and the numerical values of NDVI collected during the maturity stage of the sugarcane. The model also explored that lengthy and extended season of the sugarcane harvest leads to the variation of crop productivity in spatial as well as temporal domain. The authors of [17] compared the various methods of crop yield forecasting on the basis of remotely sensed data and proposed a model based on empirical relation of NDVI with yield information.

The method proposed by [18] estimated the sugarcane yield with an acceptable error of 1% to the actual yield. In addition to the yield estimation the authors also suggested the use of remotely sensed data to identify the optimal time period in the growth stage to predict the yield. The research work of [19] proposed a decision tree based model with coefficient of determination ( $r^2 = 0.56$ ) between multi-temporal NDVI and yield in the Sao Paulo State of Brazil. The research work of [20] suggested the use of ground truth data with remote sensing information for the quality analysis. A model based on Leaf Area Index (LAI) has been proposed by [21] with strong association between crop production and LAI. A technique based on temporal remote sensing data (2003 to 2012) proposed by [22] predicted the yield with an error of only 0.8% from actual yield. The proposed technique demonstrated the yield monitoring for sugarcane on the basis of combined use of historical yield records and NDVI information at the municipality level. Authors also concluded that the proposed technique may be employed to other type of crops as well as the crops grown in small areas. The research work of [23] classified the sugarcane planted area with an accuracy of 90% and concluded that Temporal values of NDVI from LANDSAT data may be fruitful for monitoring of crop growth stages.

Researchers in [5] discussed about recent advancements in information technology and use of RS, GIS, GPS and spatial databases that can assist the policy makers to extract the information about the estimation of crop yields more accurately. Researchers also suggested that combination of geoinformatics tools and digital image processing algorithms exhibits a huge potential for the reliable estimation of sugarcane yield [ton of stalk per hectare ( $t\ ha^{-1}$ )]. On time, accurate and cost effective information obtained through different vegetation indices is a prerequisite for efficient crop growth monitoring [24]. Crop growth status such as Nitrogen content of the plant may be calculated with the use of Near Infrared (NIR) reflectance values in combination

with Artificial Neural Network (ANN) and statistical data [25]. Spatiotemporal trend analysis of land surface temperature (LST) may be very fruitful to explore the effect of climate change [26] on crop production. Remote sensing classification according to spectral and temporal information is affected by the presence of mixed pixels. Classification and extraction accuracy may be significantly enhanced by the use of some boosting techniques of data mining [27,28].

Review of the literature as well as the detailed review presented by the authors of [29] and [30] suggested the use of techniques based on remote sensing to liberate reliable, robust, on time and accurate information to assist in the quality management and production of the sugarcane crop. The review also concluded that selection and blending of appropriate spatial as well as spectral information alongwith suitable processing methods for the extraction of information related to sugarcane is most important particularly for the small areas. Based on the review of literature and findings, this work aims to extract the optimal period for the crop predictability and the estimation of the crop yield of sugarcane planted in small fields with proposed statistical models on the basis of the temporal NDVI information collected from satellite data.

### 3. Material and Methods

This research has been focused on providing the timely and accurate estimates about the sugarcane crop and by products so as to enable the policy makers to take decisions about the food grain production. Data and information from multiple sources is fused to further guide the analysis process.

#### 3.1. Study Area

The most common data for projects related to earth resources are satellite-derived data. They become very popular in recent years because of their better spatial and spectral resolutions and their capacity to generate multi temporal products more cheaply than aerial photographs. Ground data related to the proposed model is important information.

The research work has been carried out in a small experimental area in Himalayan foothill (Fig. 1). The area is dominated by agricultural practices in around Bhadarabad region, Haridwar district, Uttarakhand State, India. The area is enveloped by  $78^{\circ} 0' 0''$  E,  $29^{\circ} 52' 30''$  N to  $78^{\circ} 5' 0''$  E,  $29^{\circ} 57' 30''$  E. This area has a fair network of irrigation canals with more than a century old Upper Ganga canal as a main canal with highly fertile lands. Under this canal two distributaries, Bhadarabad and Ahmadpur carry irrigation water. The experimental area is part of Northern Division Ganga Canal (NDGC) command area. It spreads over the Khelri minor command. The minor passes through a small village Khelri and provides irrigation to about 200 ha.

The major soil textures have been found to be Sandy Loam which is suitable for all types of crops. Major crops grown are Wheat and Sugarcane in Rabi season and Paddy and Sugarcane in Kharif season. Other minor crops are oil seeds, barseem, fodder, and some vegetables. Fig. 1 shows the location map for the experimental area.

#### 3.2. Historical Data

Historical sugarcane yield data have been collected from State Agriculture Department, sugarcane industries and by interaction with farmers. The sowing and harvesting periods of the different crops grown in the Khelri area is shown as Fig. 2.

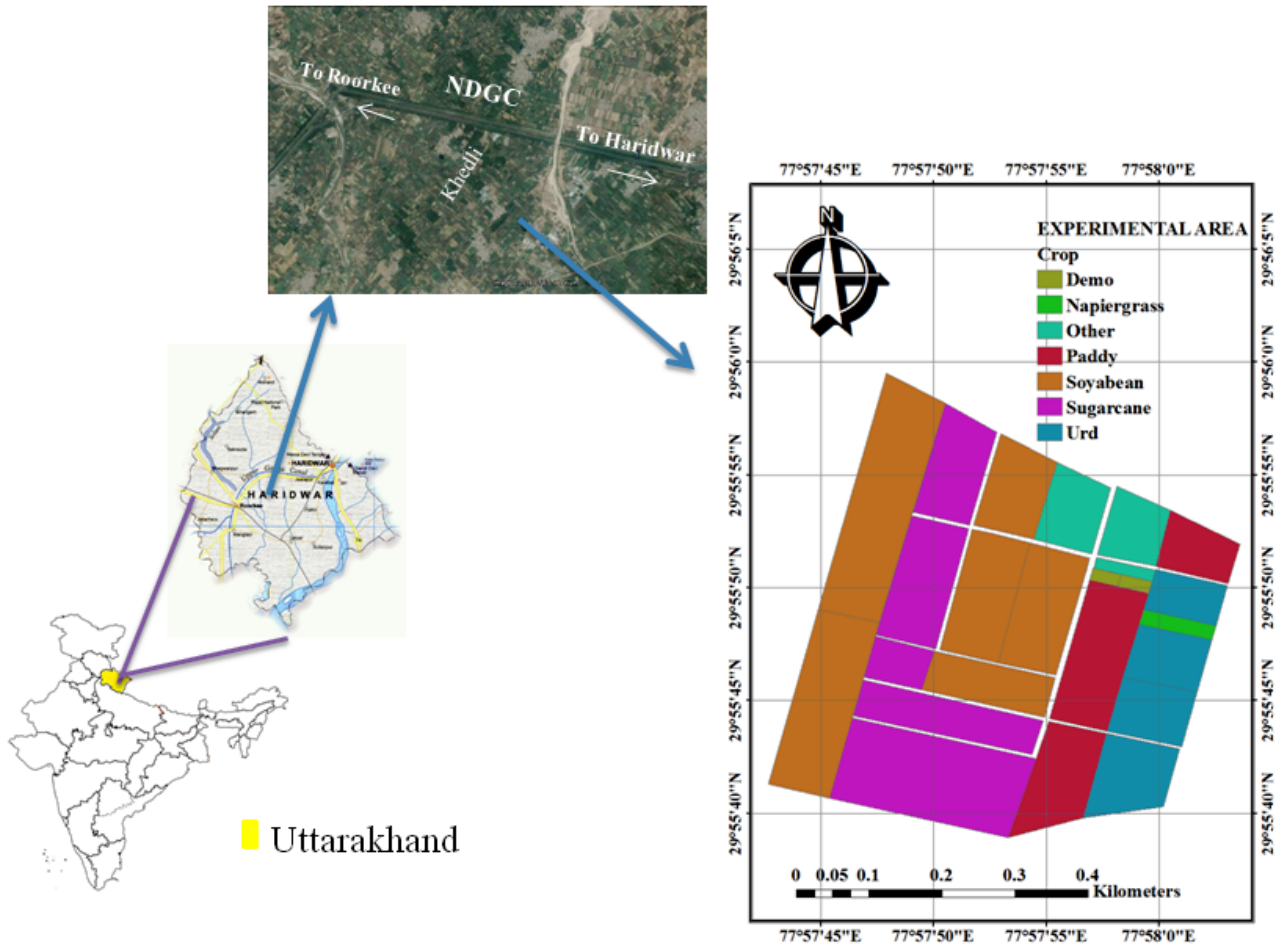


Fig. 1. Study Area

Crop Sowing and Harvesting Periods in Study Area

| Jan                    | Feb                     | Mar | Apr                | May                   | June | July | Aug | Sep                       | Oct | Nov                 | Dec |
|------------------------|-------------------------|-----|--------------------|-----------------------|------|------|-----|---------------------------|-----|---------------------|-----|
|                        |                         |     |                    | Sowing - Kharif Crops |      |      |     | Harvesting - Kharif Crops |     |                     |     |
|                        | Harvesting - Rabi Crops |     |                    |                       |      |      |     |                           |     | Sowing - Rabi Crops |     |
| Harvesting - Sugarcane |                         |     | Sowing - Sugarcane |                       |      |      |     |                           |     |                     |     |

Fig. 2. Crop Types in Khelri Region

### 3.3. Meteorological Data

Meteorological data pertaining to the experimental area has been collected from year 2000 onward. Preliminary analysis of the data yields the following inferences. The temperature has been found to vary between 2.40°C to 42.60°C, Relative Humidity has been found to vary between 11% to 100%. Average Vapour pressure has been found to vary between 5.6 mm to 27.7 mm, monthly Rainfall has been found to vary between 0.0 mm to 134.8 mm, Evaporation has been found to vary between 0.2 mm/day to 10 mm/day, whereas, the wind velocity has been found to vary between 0.4 km/hr to 5.9 km/hr.

### 3.4. Cadastral Survey Map

The Cadastral survey map has been obtained from the Revenue Department. Georeferencing has been done so that the map gets integrated with GIS interface. Integration of Geographical and agro meteorological data available in various formats is not possible. Integration of these data in

GIS are only possible when these data are geographically referenced which is known as Georeferencing. Georeferencing requires transformation for which 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> order polynomial can be computed. 1<sup>st</sup> order transformation requires minimum 3 control points, minimum six control points are required for 2<sup>nd</sup> order transformation and minimum ten points are required in 3<sup>rd</sup> order transformation. 1<sup>st</sup> order transformation causes rotation, translation and scaling while 2<sup>nd</sup> and 3<sup>rd</sup> order transformation will bend and curve straight lines. Georeferencing is essential step in process of digitization. Digitization allows us to assign additional spatial and non spatial attributes to features. The digitization of georeferenced Cadastral survey map helps in computation of the area of existing plots.

### 3.5. Satellite Data

LANDSAT 8 (Operational Land Imager-OLI) and LANDSAT 7 (Enhanced Thematic Mapper Plus -ETM+)

satellite data from the year 2004 to 2013 have been used for the proposed model. The collected data was put to preliminary analysis followed by detailed analysis. The collected remotely sensed data was put to detailed analysis for creation of Normalized Difference Vegetation Index (NDVI) images in the temporal domain.

### 3.6. Methods

In this work an attempt was made first to transform the pixel (DN) values in to surface reflectance. This is based on radiance rescaling parameters obtained from the metadata (MTL) file [31]. Mathematically the transformation is expressed as:

$$\rho\lambda' = M_{\rho}Q_{cat} + A_{\rho} \quad (1)$$

where:

|                |   |   |
|----------------|---|---|
| $\rho\lambda'$ | = | Top of Atmosphere (TOA) planetary reflectance |
| $M_{\rho}$     | = | multiplicative rescaling parameter            |
| $A_{\rho}$     | = | additive rescaling parameter                  |
| $Q_{cat}$      | = | Calibrated and Quantized pixel values         |

It is important to mention here that  $\rho\lambda'$  does not contain a required correction for the sun angle. Multiplicative rescaling parameter ( $M_{\rho}$ ) can be obtained from the metadata file using the variables REFLECTANCE\_MULT\_BAND\_x and REFLECTANCE\_ADD\_BAND\_x respectively, where x is the band number.

Reflectance value obtained in Eq. 1 is then corrected for the required sun angle using:

$$\rho\lambda = \frac{\rho\lambda'}{\cos(\theta_{SZ})} = \frac{\rho\lambda'}{\sin(\theta_{SE})} \quad (2)$$

where:

|                     |   |                           |
|---------------------|---|---------------------------|
| $\rho\lambda$       | = | TOA planetary reflectance |
| $\cos(\theta_{SZ})$ | = | Local sun elevation angle |
| $\sin(\theta_{SE})$ | = | Local solar zenith angle  |

The Normalized Difference Vegetation Index (NDVI) was computed with the formula:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3)$$

where NIR represents the reflectance in the near-infrared band while R is the red band reflectance [32]. The values of this index vary between -1 and +1. Generally, the negative values and values near to zero are specific to soil with no vegetation or sparse vegetation whereas the surfaces covered by dense and healthy vegetation have the values 0.7 to 1.0.

Correlation analysis and regression modelling has been employed independently for each of the crop in the experimental fields by collecting the numerical values of NDVI with 16 days repeat frequency and the historical crop yield data from the year 2004 to 2013. Database for each

crop type in the experimental area is prepared on the basis of the Pearson's correlation coefficient ( $r$ ). The coefficient  $r$  is the relation between the historical crop yield ( $y$ ) and mean NDVI( $x$ ) values for each crop in the temporal domain of year 2004 to 2013 in the experimental area. The dates having maximum value of NDVI for each pixel are stored in another database. The corresponding time period with the highest NDVI has been recognized as optimal period for the prediction of crop yield.

After successfully identification of the optimal period for prediction of crop yield, the regression analysis has been done between mean NDVI values of that period and historical crop data. Results from different regression models such as linear model, logarithmic model, polynomial model, exponential models and power series are analyzed and evaluated so as to identify the most appropriate model for the experimental area.

The methodology adopted to develop the proposed model is graphically depicted by the Fig. 3 while all the steps used in the methodology are represented by following algorithm.

#### **Algorithm 1: Proposed algorithm for yield estimation**

##### **Step 1: Initialization:**

-Collection and Compilation of relevant data: meteorological, LANDSAT images, historical crop yield, cadastral map and ground truth

##### **Step 2: Pre-processing:**

-Generation and Analysis the Metadata Files: Checking of data sufficiency, discarding the images with cloud cover more than 30% and planning field data collection program wherever necessary

-Creation of the layer stack for all the temporal images  
-Conversion of DN values to reflectance values (using Eq. (1) and Eq. (2))

-Generation of NDVI images based on NIR and Red Band (using Eq. (3))

##### **Step 3: Preliminary Statistical Analysis:**

-Generation of Correlation Matrices based on the Pearson Correlation Coefficient between yield data and NDVI values obtained in Step 2

-On yearly basis the date having maximum value of NDVI for each pixel is stored in another database

##### **Step 4: Advanced Statistical Analysis:**

-Mean NDVI values is calculated for all the pixels obtained in Step 3

-Correlation Analysis of mean NDVI values and Crop Yield data is recalculated for the planting areas

-Dates containing the maximum values of NDVI are grouped together to identify the best period for the estimation

##### **Step 5: Crop Yield Modelling:**

-Development of crop yield prediction models based on the mean NDVI profile and historical crop yield data obtained in Step 4 above

-Performance Evaluation of the developed models using Tukey Test and RMSE

-Selection of best performer model

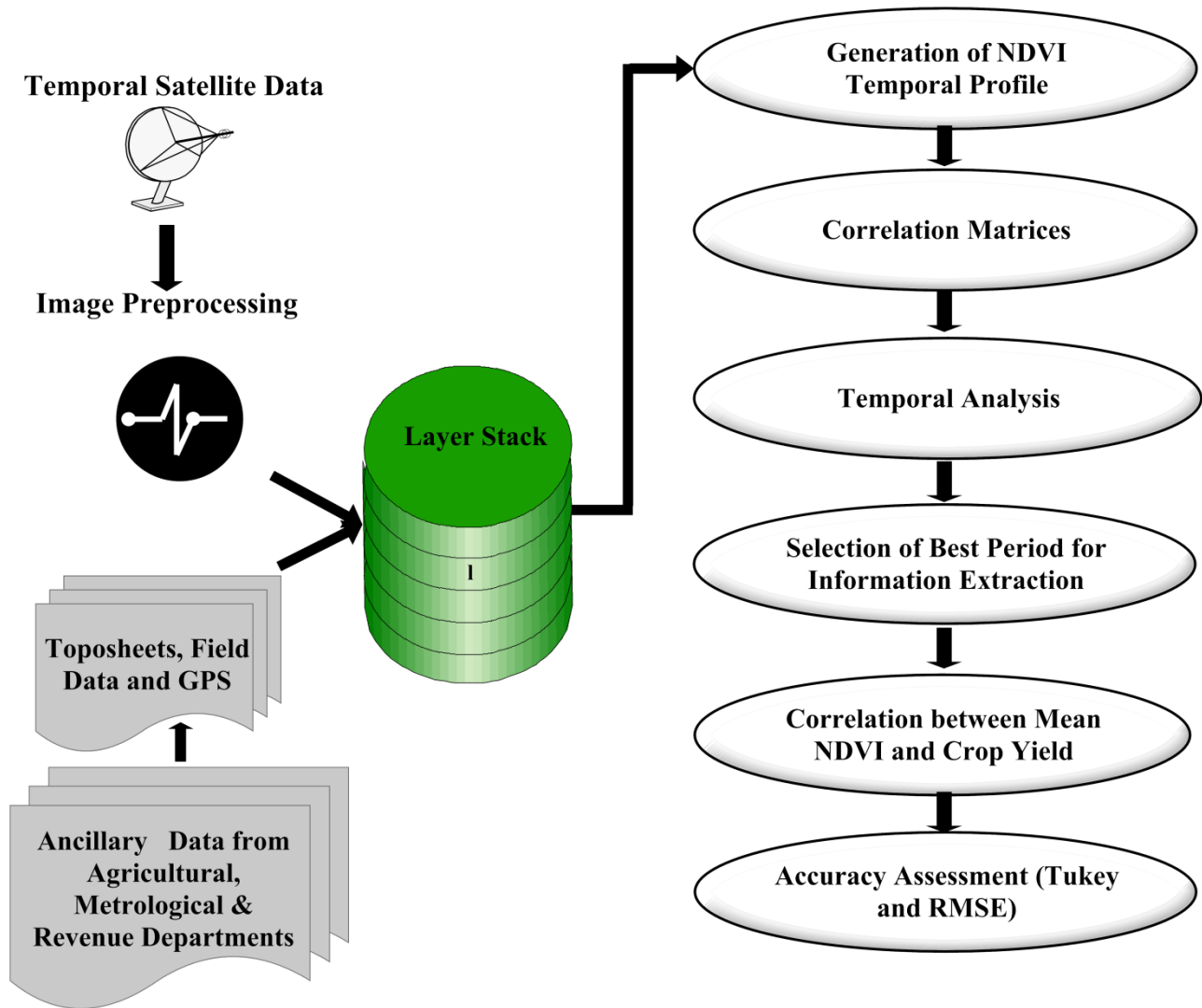


Fig. 3. Flow diagram of the methodology

#### 4. Results

Tab. 1 presents the correlation coefficient of crop yield records with averaged numerical NDVI values for the different crops in the experimental area. The cavernous study of the Tab. 1 brings out the fact that the correlation of sugarcane yield and averaged numerical NDVI values is strong, having the mean value of 0.75 whereas, the correlation of wheat and rice is below 0.6 in the area. The correlation coefficient achieved the maximum value of 0.77 around the maturity stage of sugarcane. The sugarcane yield values are strongly correlated with the averaged numerical NDVI values taken from the grand growth stage (210-270 days after plantation (DAP)) of crop. This result has been attributed to the fact that during this period the NDVI values of the sugarcane fields also found to be on the higher side. The mean NDVI values during 2004 to 2013 of the extracted planting areas in Khelri Minor demonstrate the behaviour of the sugarcane crop in each year (Fig. 4). It has been observed from the graph that the NDVI values increase in the initial stage i.e. 50 to 55 days after the sowing and shoots up again towards the grand growth stage and finally dips again around the ripening stage. These observations are in agreement with the findings of [18] to observe the relations between NDVI and yield data during eight to ten months of crop growth.

Table 1 Result of Correlation Analysis

| Sr. No. | Crop Type | Correlation Coefficient (r) | Best Period          |
|---------|-----------|-----------------------------|----------------------|
| 1       | Sugarcane | 0.75                        | 210-270 DAP (Fig. 4) |
| 2       | Wheat     | 0.59                        | -                    |
| 3       | Rice      | 0.55                        | -                    |

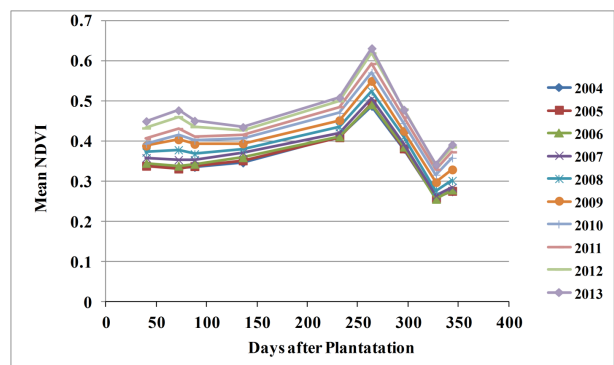


Fig. 4. Mean NDVI values sugarcane planting areas

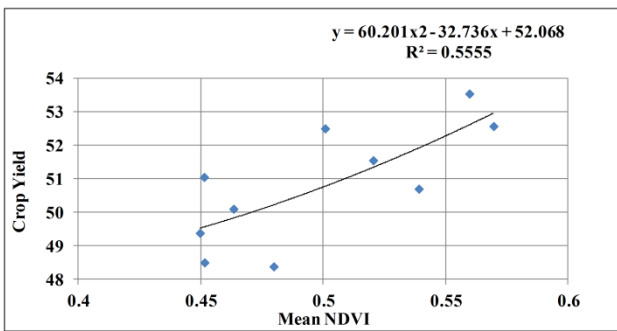
Results of the regression analysis based on mean NDVI values of sugarcane planting areas in the optimal period and

the historical crop yield data of the Khelri is presented in Tab. 2. Regression equations, the coefficient of determination ( $r^2$ ) and the corresponding models are also shown in Fig. 5. The relation of sugarcane yield with mean NDVI values is represented effectively by the polynomial model in Khelri area. However the coefficient of determination for the regression models related to the proposed algorithm has been found very similar.

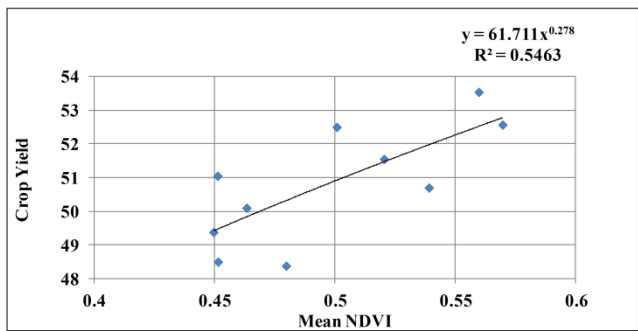
**Table 2.** Regression Equations for the yield estimation

| S. No | Model | Equation | Coefficient of Determinati |
|-------|-------|----------|----------------------------|
|-------|-------|----------|----------------------------|

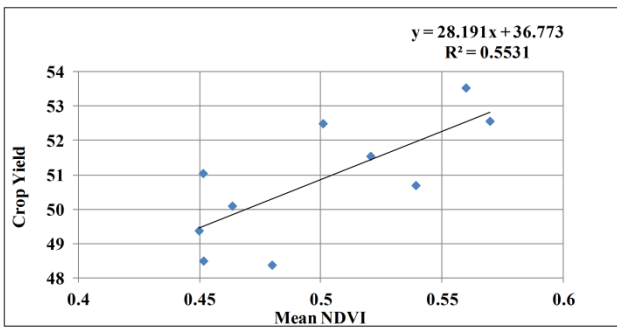
|   |              |                                  | on    |
|---|--------------|----------------------------------|-------|
| 1 | Polynomial   | $y = 60.201x^2 - 32.736x + 52.0$ | 0.555 |
| 2 | Linear       | $y = 28.191x - 36.773$           | 0.553 |
| 3 | Exponential  | $y = 38.559e^{0.553x}$           | 0.549 |
| 4 | Power Series | $y = 61.711x^{0.278x}$           | 0.546 |
| 5 | Logarithmic  | $y = 14.169\ln(x) + 60.745$      | 0.550 |



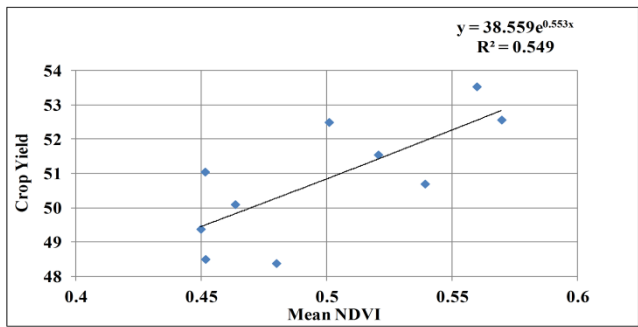
(a)



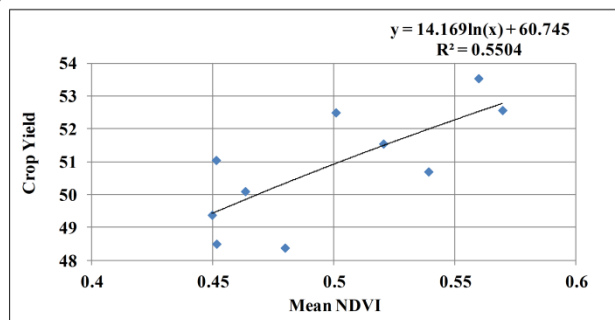
(b)



(c)



(d)



(e)

**Fig. 5.** Regression models for Sugarcane Yield (t/ha) in Khelri Area (a) Polynomial (b) Power (c) Linear (d) Exponential (e) Logarithmic Models

A detailed statistical analysis had been performed to analyze the efficiency of the results obtained using various models of our proposed algorithm. The most conservative multi comparison ‘Tukey Test’ [33] has been implemented to carry out this analysis. This test is based upon the two-tailed t-test with equal means, i.e.  $\mu_i = \mu_j$ , where  $i$  and  $j$  represent all possible combinations of the available sample means. The test is performed at 5% level of significance. This test assesses whether the means of two considered groups differ significantly from each other. After performing

this test, the results demonstrate that the difference between the chosen models was not highly significant.

Statistical indicator root mean square error (RMSE) has been employed to evaluate the performance and behaviour of the proposed models to identify the best period for prediction of the crop yield as well as crop yield estimation in the experimental area. This indicator enumerated the weighted variations between the estimated and actual yields. Mathematically, it is given as:



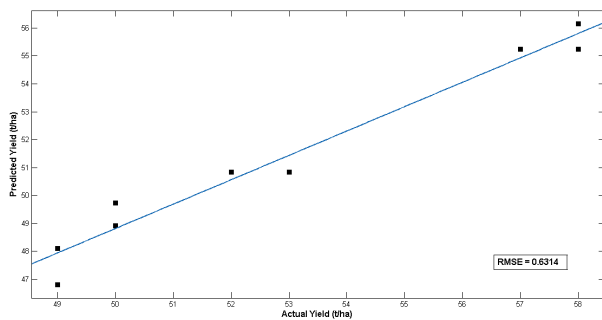
$$RMSE = \sqrt{\frac{1}{n} \sum (py_i - oy_i)^2} \quad (4)$$

where  $n$  is the total number of observations,  $py$  denotes estimated or predicted yield  $oy$  is the observed or actual yield.

To test the obtained models, the sample data from 10 different fields has been used from the experimental area. The sample data was tested for the analysis of differences between the predicted and observed yield. The values obtained from the analysing using RMSE are given in Tab. 3. Fig. 6 depicts the plots between actual and predicted yield and with minimum RMSE for the selected model for the crop yield.

**Table 3.** Model validation and performance indicator (\* Minimum, † Maximum values of RMSE in the experimental area)

| S. No. | Model        | RMSE (t/ha)   |
|--------|--------------|---------------|
| 1      | Polynomial   | 0.673         |
| 2      | Linear       | 0.658         |
| 3      | Exponential  | <b>0.631*</b> |
| 4      | Power Series | 0.787         |
| 5      | Logarithmic  | <b>0.852†</b> |



**Fig. 6.** Actual yield vs predicted yield in the experimental area

## 5. Discussion

An attempt has been made for the prediction of crop yield by establishing correlation between crop yield and numerical values of NDVI in the temporal domain. The preliminary analysis goals that the correlation coefficient varies in the temporal domain. The variation of correlation coefficient is also found to be multimodal in the temporal domain. Coefficient of correlation with an average value of NDVI for the single date data is also not significant and useful. This may be attributed to the fact that every individual crop has some crucial important crop growth period.

Field traversing revealed that the prevalent farm practice in the area is intercropping or mixed cropping in many of the fields. The best estimate of crop yield is only possible by using crop specific NDVI values. This becomes the crucial for prognostic crop yield modelling and estimation. This pushes the need for precise identification of the crop type in spatial domain. For a particular model, generally one single crop is considered. For this the spatial average values of NDVI may be used as a basic input to the crop yield model. The authors in [34] also proposed a model to estimate winter wheat crop yield on the basis of the numerical values of NDVI. It is worth mentioning that spatial average NDVI values are not only robust but dependable to reflect crop

yield estimation in the experimental area. However, many of the times these types of models may not predict accurately, simply due to the fact that spatial average NDVI may not be representative of the crop under consideration. In contrast, high correlation coefficient has been obtained from the single crop in experimental fields.

It can be safely inferred that crop based NDVI values in place of spatial average NDVI makes crop yield modelling realistic, statistically sound. This leads to reliable crop yield prediction. The researchers in the past also proposed the crop yield estimation based on the identification of a particular crop from mixed crop [35]. In this method the optimum growth period of the crop has been identified by the above mentioned statistical approach. In Indian agricultural practices this may be the most feasible solution for the yield estimation.

A detailed statistical investigation has been carried out in spatial and temporal domain to work out a minimum threshold of NDVI to ascertain whether a particular pixel is having a crop. The threshold numerical value of the coefficient of correlation was found to be 0.6 on the basis of averaged numerical NDVI values and historical crop yield values. In the proposed model the crop discrimination process has also been done using clustering. For this K-means analysis has been carried out for eight possible crop combinations in a mixed cropping pattern. It has been observed that the numerical value of NDVI alone is not sufficient to discriminate some of the crops having more or less similar phenology in a particular growth period. Further, the decision tree concept has also been investigated for the crop discrimination. The proposed model indicated that even decision tree approach is also not sufficient to discriminate the mixed crops effectively. This may be attributed to the variation in the agricultural practices such as variation in sowing date, crop variety, soil characteristics, integrated pest management, spatial and temporal variation of crop growth.

From food management point of view the preferable period for crop yield prediction should be as early as possible. Field experimentation indicated that reliable prediction can only be done if physical phenomena of crop growth cycle and crop yield is studied and modelled. In the Himalayan foothill area, it is clearly observed that for the sugarcane crop the numerical value of NDVI pops to rise around the third week of May, it attains maximum towards the end of September, and thereafter keep on decreasing after the first week of October. This is associated with exposing of the whole planting area passes through productive stage before it attains the maturity. In local conditions, it is harvested during November and December.

From above it has been inferred that the proposed approach is more effective and reliable for the crop yield estimation. The effort has also been made to ascertain the best model for crop yield estimation. The polynomial regression model is found to be better as compared to linear and exponential regression models from the correlation point of view. This finding is in line with the results indicated by [36]. In the study given by [37], the correlation between the numerical value of NDVI and corn yield for the grain was exponential in nature. In contrast to this the authors of [38] demonstrated that the correlation between yield of soybean and numerical values of NDVI can be best represented by the power function.

It has been found that the different models in this work are exhibiting more or less similar performance. Cross validation has not been done for a simple reason due to shortage of data memory. Under these circumstances the

Tukey Test [33] has been used to evaluate the performance of models and come out with the best. After performing this test, the results demonstrate that the difference between the chosen models was not highly significant. The analysis concludes that none of these methods outperforms the other one statistically. The usability and reliability of the proposed method may be enhanced if data with longer memory period can be incorporated in the process.

## 6. Conclusions

Crop yield estimation in small areas based on temporal profile of LANDSAT-NDVI data using correlation and regression analysis has been explored in this paper. The identification technique of the planting areas using the correlation analysis is also presented based on the assumption that strong association between crop yield and NDVI indicates high probability of the vegetation land. Based on previous 10 years data it has been found that predictability of linear and nonlinear modelling is significant and is of the order of 0.7. Coefficient of determination with  $r^2 = (0.55)$  value are found in the experimental area. The best

periods for the reliable and accurate estimation of crop yield are also identified. In comparison with other methods and models related to crop yield estimation, this work conducts correlation analysis to optimize the correlation of historical crop yield values with numerical values of NDVI. The polynomial model exhibits a significant enhanced performance as compared with other models such as linear, logarithmic, power and exponential models. Although, the results of the proposed crop yield prediction method shows potential, yet a more comprehensive model can be developed if input from farming system and climatic characteristics of the area is incorporated in the modelling. It is worth mentioning that if information related to the farming practices and climate is easily accessible then crop growth process can be effectively modelled. This may lead to timely and reliable prediction of crop yield devoid of historical data of the crop yield.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence



## References

- Hayes, M.J.; Decker, W.L. Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *International Journal of Remote Sensing* **1996**, *17*, 3189–3200.
- Labus, M.P.; Nielsen, G.A.; Lawrence, R.L.; Engel, R.; Long, D.S. Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing* **2002**, *23*, 4169–4180.
- Singh, R.; Semwal, D.P.; Rai, A.; Chhikara, R.S. Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing* **2002**, *23*, 49–56.
- Sadrykia, M.; Delavar, M.R.; Zare, M. A GIS-based decision making model using fuzzy sets and theory of evidence for seismic vulnerability assessment under uncertainty (case study: Tabriz). *Journal of Intelligent & Fuzzy Systems* **2017**, *33*, 1969–1981.
- Promburom, P.; Jintrawet, A.; Ekasingh, M. Estimating sugarcane yields with Oy-Thai interface. International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress; Hogarth, D.M., Ed.; , 2001; Vol. 2, pp. 17–21.
- Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sensing* **2010**, *2*, 673–696.
- Wall, L.; Larocque, D.; Leger, P.M. The early explanatory power of NDVI in crop yield modelling. *International Journal of Remote Sensing* **2008**, *29*, 2211–2225.
- Victoria, D.d.c.; Paz, A.R.d.; Coutinho, A.C.; Kastens, J.; Brown, J.C. Cropland area estimates using MODIS NDVI time series in the state of Mato Grosso, Brazil. *Pesquisa Agropecuária Brasileira* **2012**, *47*, 1270–1278.
- Prasad, A.K.; Chai, L.; Singh, R.P.; Kafatos, M. Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation* **2006**, *8*, 26–33.
- Dadhwal, V.; Singh, R.; Dutta, S.; Parihar, J. Remote sensing based crop inventory: A review of Indian experience. *Tropical Ecology* **2002**, *43*, 107–122.
- Chen, S.C.; Spiguel, M. Evaluation of the possibility of using Landsat MSS data for sugarcane yield estimation. *Sao Jose dos Campos* **1989**.
- Luo, X.; Wu, X.; Zhang, Z. Regional and Entropy component analysis based remote sensing images fusion. *Journal of Intelligent & Fuzzy Systems* **2014**, *26*, 1279–1287.
- Thampi, S.M.; El-Alfy, E.S.M. Application of M-band wavelet in pan-sharpening. *Journal of Intelligent & Fuzzy Systems* **2017**, *32*, 3151–3158.
- Rao, P.; Rao, V.; Venkataratnam, L. Remote sensing: A technology for assessment of sugarcane crop acreage and yield. *Sugar Tech* **2002**, *4*, 97–101.
- Gers, C. Relating remotely sensed multi-temporal Landsat 7 ETM+ imagery to sugarcane characteristics. *Proc S Afr Sug Technol Ass*, **2003**, p. 7.
- Begue, A.; Lebourgeois, V.; Bappel, E.; Todoroff, P.; Pellegrino, A.; Baillarin, F.; Siegmund, B. Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing* **2010**, *31*, 5391–5407.
- Morel, J.; Todoroff, P.; Begu, A.; Bury, A.; Martine, J.F.; Petit, M. Toward a Satellite-Based System of Sugarcane Yield Estimation and Forecasting in Smallholder Farming Conditions: A Case Study on Reunion Island. *Remote Sensing* **2014**, *6*, 6620–6635.
- Almeida, T.I.R.; Filho, C.R.D.S.; Rossetto, R. ASTER and Landsat ETM+ images applied to sugarcane yield forecast. *International Journal of Remote Sensing* **2006**, *27*, 4057–4069.
- Fernandes.; Lobato, J.; Rocha.; Vieira, J.; Lamparelli.; Camargo, R.A. Sugarcane yield estimates using time series analysis of spot vegetation images. *Scientia Agricola* **2011**, *68*, 139–146.
- Rembold, F.; Atzberger, C.; Savin, I.; Rojas, O. Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sensing* **2013**, *5*, 1704.
- Chatwachirawong, P.; Kitaura, A.; Srinives, P.; Nawata, E. Construction of a Simple Yield Estimation Model for Productivity Prediction in Sugarcane. *Tropical Agriculture and Development* **2012**, *56*, 113–116.
- Pupin Mello, M.; Atzberger, C.; Formaggio, A. Near real time yield estimation for sugarcane in Brazil combining remote sensing and official statistical data. Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014, pp. 5064–5067.
- Mulianga, B.; Bégué, A.; Clouvel, P.; Todoroff, P. Mapping Cropping Practices of a Sugarcane-Based Cropping System in Kenya Using Remote Sensing. *Remote Sensing* **2015**, *7*, 14428–14444.
- Waldner, F.; Canto, G.S.; Defourny, P. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *110*, 1–13.
- Afandi, S.D.; Herdiyeni, Y.; Prasetyo, L.B.; Hasbi, W.; Arai, K.; Okumura, H. Nitrogen Content Estimation of Rice Crop Based on Near Infrared (NIR) Reflectance Using Artificial Neural Network (ANN). *Procedia Environmental Sciences* **2016**, *33*, 63–69. The 2<sup>nd</sup> International Symposium on LAPAN-IPB Satellite (LISAT) for Food Security and Environmental Monitoring.
- Wang, N.; Xia, J.; Yin, J.; Liu, X. Trend analysis of land surface temperatures using time series segmentation algorithm. *Journal of Intelligent & Fuzzy Systems* **2016**, *31*, 1121–1131.



27. Li, D.W.; Yang, F.B.; Wang, X.X. Crop region extraction of remote sensing images based on fuzzy ARTMAP and adaptive boost. *Journal of Intelligent & Fuzzy Systems* **2015**, 29, 2787–2794.
28. Zhou, Z.; Huang, J.; Wang, J.; Zhang, K.; Kuang, Z.; Zhong, S.; Song, X. Object-Oriented Classification of Sugarcane Using Time-Series Middle-Resolution Remote Sensing Data Based on AdaBoost. *PLoS One* **2015**, 10, 1–16.
29. Abdel-Rahman, E.M.; Ahmed, F.B. The Application of Remote Sensing Techniques to Sugarcane (*Saccharum Spp.* Hybrid) Production: A Review of the Literature. *Int. J. Remote Sens.* **2008**, 29, 3753–3767.
30. Ahamed, T.; Tian, L.; Zhang, Y.; Ting, K. A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy* **2011**, 35, 2455 – 2469.
31. Gunnula, W.; Kositttrakun, M.; Righetti, T.L.; Weerathaworn, P.; Prabpan, M. Evaluating Sugarcane Growth and Maturity Using Ground-Based Measurements and Remote Sensing Data. *Thai Journal of Agricultural Science* **2012**, 45, 17–28.
32. Rouse, J.W.; Haas, H., R.; Schell, A., J.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Goddard Space Flight Center 3<sup>rd</sup> ERTS-1 Symposium*. **1974**, 1, 309–317.
33. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley, 1997.
34. Ren, J.; Chen, Z.; Zhou, Q.; Tang, H. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation* **2008**, 10, 403 – 413.
35. Maselli, F.; Rembold, F. Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogrammetric Engineering & Remote Sensing* **2001**, 67, 593–602.
36. Jiang, D.; Wang, N.b.; Yang, X.h.; Wang, J.h. Study on the interaction between NDVI profile and the growing status of crops. *Chinese Geographical Science* **2003**, 13, 62–65.
37. Teal, R.K.; Tubana, B.; Girma, K.; Freeman, K.W.; Arnall, D.B.; Walsh, O.; Raun, W.R. In-Season Prediction of Corn Grain Yield Potential Using Normalized Difference Vegetation Index. *Agronomy Journal* **2006**, 98, 1488–1494.
38. Ma, B.L.; Dwyer, L.M.; Costa, C.; Cober, E.R.; Morrison, M.J. Early Prediction of Soybean Yield from Canopy Reflectance Measurements. *Agronomy Journal* **2001**, 93, 1227–1234.