

MACHINE LEARNING BASED CROP CYCLE MAPPING USING MULTI TEMPORAL SPACE DATASETS

Prasun Kumar Gupta ^{1,*}, Prabhakar Alok Verma ¹, Prakash Chauhan ¹

¹ Indian Institute of Remote Sensing, Indian Space Research Organisation, 4 Kalidas Road, Dehradun, India - (prasun, prabhakar, director)@iirs.gov.in

Commission III, WG III/10

KEY WORDS: Machine Learning, NDVI Time Series, Phenology, Cropping Cycles

ABSTRACT:

The Indo-Gangetic region is an important source of food grain produced in India. With advent of advanced crop management practices, there has been a recent increase in production. This could be due to improved yield and/or multiple crops per year. Understanding cropping patterns is, therefore, a fundamental requirement to strategize food security in the country. The long running space based earth observation programmes, provides a powerful tool to study the change in vegetation, through use of indices, synoptically and temporally. This study proposes a machine learning based technique to identify distributed (pixel-wise) cropping cycles, based on the temporal trend in normalized difference vegetation index (NDVI). The technique is applied to several time series NDVI data sets (OCM, MODIS and GIMMS) and results are discussed. There is tremendous scope to further extend this study to predict drought conditions, change in cropping pattern, and estimate production statistics, by using advanced artificial intelligence methods (such as deep learning).

1. INTRODUCTION

1.1 Background

World leaders at historic United Nations summit in September 2015 adopted 17 Sustainable Development Goals of the 2030. Out of 17 goals, Goal 2 is 'Zero Hunger' which focuses on ending hunger & malnutrition, double agricultural productivity and ensure sustainable food production system etc. This sustainable development goal can be achieved by providing proper food security to the people. As defined by United Nation's World Food Conference in 1975, food security is the adequate food supply of basic food to sustain a steady expansion of food consumption and to offset fluctuations in production and prices. As per Census of India 2011, total population of India is 1.21 billion and it is increasing with decadal growth of 17.64%. Hence, to cater this situation and achieve 'Zero Hunger' goal of United Nations, agricultural production of India need to be increased, also the agricultural pattern may need attention.

Therefore, it is important to know the distribution of land cover, for various applications like natural resources management, agricultural/vegetation monitoring, environmental studies etc. Temporal availability of Remote Sensing data from space borne sensors such as Moderate Resolution Imaging Spectroradiometer (MODIS), Ocean Color Monitor (OCM), Advanced Very High Resolution Radiometer (AVHRR) etc., helps us to perform analysis over time and monitoring. Remote sensing data for more than 20 years at regular time interval helps to understand such pattern using time series analysis. Time series is a sequence of observations listed in the order of time. Mostly observations are made at regular time interval.

Time series analysis of remote sensing images is important to identify and track the changes on the ground throughout a time period.

In this study, an attempt has been made to analyse time series data of MODIS, OCM and NOAA-AVHRR based Global Inventory Modeling and Mapping Studies (GIMMS) NDVI datasets, to find out the various types of cropping cycles, changing scenario of cropping pattern in India and forecast of pattern using machine learning techniques. Machine learning is opted in this study because of high dimensionality and non-linear nature of time series data. These factors obstruct traditional techniques like regression, causing loss of efficiency and performance on large datasets. The primary objective is to perform phenology based mapping of number of cropping seasons in India.

1.2 Previous Studies

A number of studies have attempted to identify cropping pattern using time series analysis with different approaches (Li et al. 2014, Estel et al. 2016, Roy, D.P. and Yan, L., 2018).

It is difficult to discriminate between the crops based on the phenology due to several reasons like 1) multiple crops may be following similar growth pattern 2) crop growth pattern is strongly affected by local weather conditions 3) technical factors on the limitation of resolution of satellite imagery (Peña-Barragán et al. 2011). Therefore, accurate mapping of cropping pattern in India is still a challenge due to diversity in crops and agroclimatic zones.

* Corresponding author

Time series analysis of different vegetation indices derived from space borne data, has been reported to provide good results for detection of type of crop (Arvor et al., 2011) and even to map single, double and mixed cropping (Maus et al. 2016).

2. MATERIAL AND METHODS

2.1 Study Area and Dataset

Mainland India (Figure 1) is selected as study region for this study. Total land area of the country is ~3.29 M sq km out of which ~1.59 M sq km is under various types of agricultural activities. It is very challenging to map all the croplands in India using time series analysis due small size of land parcels & crop diversity and various agro-climatic zones. Also India is classified in 15 agro-climatic regions namely Wester Himalaya, Eastern Himalaya, Lower Gangetic planes, Upper Gangetic Planes, Trans Gangetic Planes, Eastern & Hills, Central Plateau & Hills, Western Plateau & Hills, Southern Plateau & Hills, East Coast Plains & Hills, west Coast Plains & Hills, Gujarat Plains & Hills, western Dry region and Islands as shown in Figure 1.

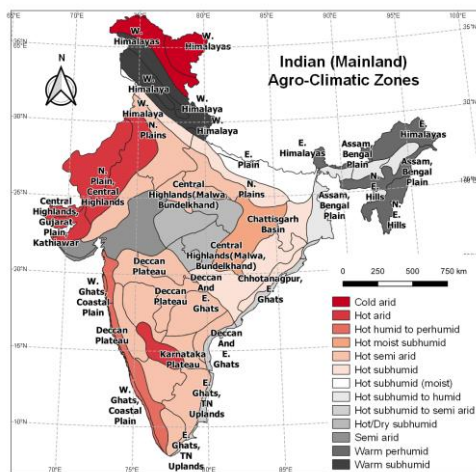


Figure 1: Study Area (Source: NBSSLUP)

Sensor (Product)	Spatial Resolution (m)	Temporal Resolution (days)	Availability
MODIS ¹ (MOD13Q1)	250	16	2000-2006
OCM ²	1000	15	2012-2017
AVHRR ³ (GIMMS)	8000	15	1981-2006

Source:
¹ <https://earthengine.google.com/datasets/> (Sep 2018)
² <http://bhuvan.nrsc.gov.in/data/download/index.php> (Sep 2018)
³ <http://glcf.umd.edu/data/> (Jun 2015)

Table 1: Datasets Used

Three sensor products were used in this study as provided in Table 1. Normalised Difference Vegetation Index (NDVI) derived from the MODIS instrument data, on-board the Terra satellite, was used to perform the analysis from February 2000 till December 2006 (16 day temporal resolution). A purpose-specific Javascript code written on the Google Earth Engine platform was used to download the required data (in space and time).

The Global Inventory Modeling and Mapping Studies (GIMMS) NDVI data product are available for a 25 year period (1981-2006). This data set is derived from imagery obtained from the AVHRR instrument onboard the (National Oceanic and Atmospheric Administration) NOAA satellite series 7, 9, 11, 14, 16 and 17. The GIMMS NDVI dataset is reported to be corrected for calibration, view geometry, volcanic aerosols, and other effects not related to vegetation change.

The NDVI Filtered Product, derived from OCM sensor on-board the OceanSAT-2 satellite, is being provided every 15 days using a special compositing technique based on Fourier adjustment (NRSC, 2015). The OCM NDVI have been found to be in good correlation ($R^2 \sim 0.9$, RMSE ~ 0.08) with the MODIS NDVI. The missing product (2nd fortnight of Dec 2013) was replaced by averaging the temporally nearest data.

To focus only on agricultural areas, the following masks were also applied to OCM and GIMMS datasets:

1. Indian Landmass
2. Permanent Snow (from Nikam et al., 2017)
3. Forest (from Roy et al., 2012).

The forest data originally had 153 classes, out of which 95 classes were marked as different types of forest. Pixels belonging to the 95 forest classes and to water body, wet lands, settlement, snow and reject classes were masked out. Pixels belonging to the remaining 53 classes (comprising of “other” types of vegetation) was considered for this work. No masks were applied on the MODIS dataset.

2.2 Methodology

The datasets were temporally smoothed using several types of Gaussian and Savitzky Golay (SavGol) filtering (Figure 2). All the experiments were performed on a pixel-wise (i.e., spatially distributed) manner. SavGol filter was found to be most appropriate as it induced the least amount of error in the data (Rodrigues, A et.al, 2012; Li, L et al., 2014).

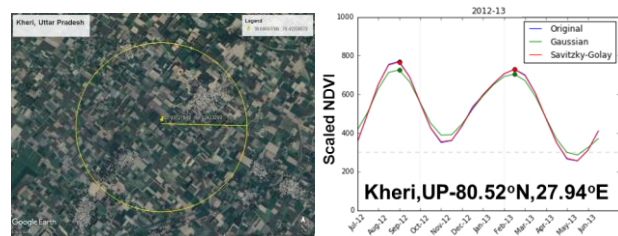


Figure 2: (Left) Zoomed in Google Image for 1-km pixel in Uttar Pradesh, found to have 2 cropping cycles in 2012-13. (Right) Effect of smoothening on time-series OCM NDVI data

There were 2 main category of techniques applied to find the number of cropping cycles. The first category (hereafter referred to as the “moving window” technique) applied a moving window to identify the local min/max value of NDVI. Based on the local min/max and difference between two peaks, number of crops in year are identified at a location/pixel. The second category (hereafter referred to as “peak detection” technique) was based on finding the cycle with the minimum length of season (75 days) and the minimum value of NDVI achieved (0.30). These other techniques based on Amplitude, Threshold, Distance, Prominence and Width were experimented with. Both

the techniques were calibrated, by training the model iteratively based on heuristics.

3. RESULTS & DISCUSSION

The techniques discussed were applied incrementally on larger areas to test its efficacy. This section will first report and discuss the results of “moving window” technique applied in parts of Gujarat and Punjab states, using MODIS data. Next the results using “peak detection” technique will be reported and discussed - which was done for the entire Indian landmass, using OCM and GIMMS data products.

3.1 MODIS data Analysis

3.1.1 Potato Farming in Deesa, Gujarat: Figure 3(a) shows the NDVI time series plot for Deesa in Northern Gujarat. Potato research station, agricultural university etc. are established here to study the potato plantations in this area. The plot clearly shows 3 peaks in each year (2000–2006), implying 3 cropping cycles. This technique is then applied to all the pixels in that area, and a distributed map of cropping cycles per year map is generated (Figure 3(b-c)). The information on growth period, peak NDVI value and month of peak NDVI, for potato plantations, was also derived (Table 2).

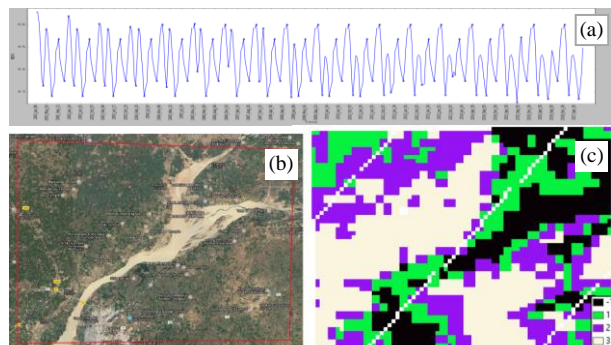


Figure 3: (a) Time Series NDVI for Deesa; (b) Zoomed-in Google Image of the area; (c) Re-classified map showing pixels having 1, 2 and 3 cropping cycles (-1 shows no vegetation)

Growth period	Peak NDVI	Time of Peak
~78 days	0.57	May/June
~96 days	0.53	October
~81 days	0.63	January

Table 2: Characteristics of 3 potato crops in Deesa

3.1.2 Intensive cropping in Punjab: Figures 4(a-d) show temporal NDVI data for 4 pixels in the Punjab state which have (a) no vegetation or cropping (deduced by consistently low NDVI values); (b-c) three & two cropping cycles (deduced by number of local maxima/minima); and (d) forest areas (deduced by consistently high NDVI values, with seasonal dips). After iteratively fitting the technique, it is applied to each pixel in the study area. Figure 4 (e) shows the spatial distribution of cropping cycles. The area covered by each class of vegetation is shown in Table 3. Majority of the area is dominated by 2 crop agriculture. Some parts of Northern Punjab (East of Amritsar and around Jalandhar) show intensive 3 cropping phenomena.

Vegetation Class	Area (M ha.)	%
No Vegetation	0.18	3.6
Forest	0.35	6.9
2 crop	4.47	88.8
3 crop	0.03	0.6

Total	5.03	
-------	------	--

Table 3: Area covered by each vegetation class in Punjab

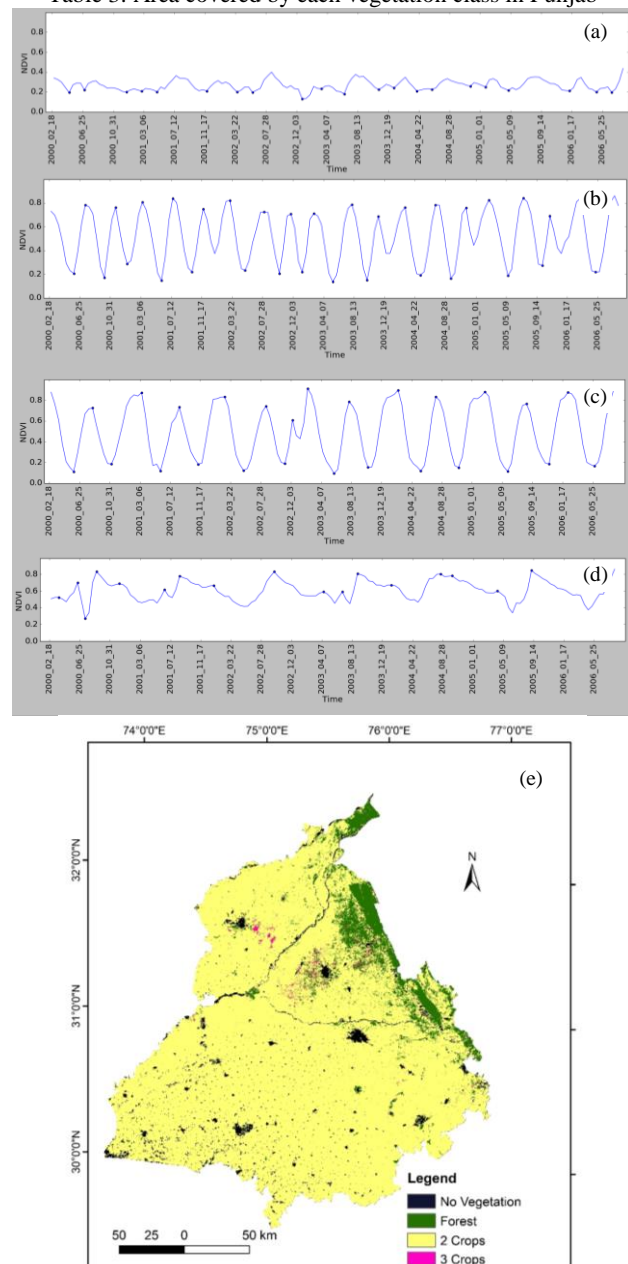


Figure 4: NDVI Time Series Plots in Punjab (a) No Vegetation (31.33°N, 75.56°E); (b) Three crops (31.59°N, 74.99°E); (c) Two crops (31.02°N, 75.06°E); (d) Forest (31.81°N, 75.90°E); (e) Cropping cycles map for Punjab

3.2 OCM data Analysis

Figure 5 shows the availability of high quality OCM filtered NDVI 15-day composites in agricultural areas. With nearly 77% of the agricultural area having 100% data available during 2012-17, makes OCM a reliable source of data for agriculture monitoring and management.

The “peak detection” technique was applied to the OCM data, masked-off by the permanent snow and forest pixels. To best detect agricultural cycles, the crop year was fixed from first fortnight of July to second fortnight of June in the succeeding year. The technique was repeated for all available OCM filtered NDVI products i.e., 2012-13 till 2016-17.

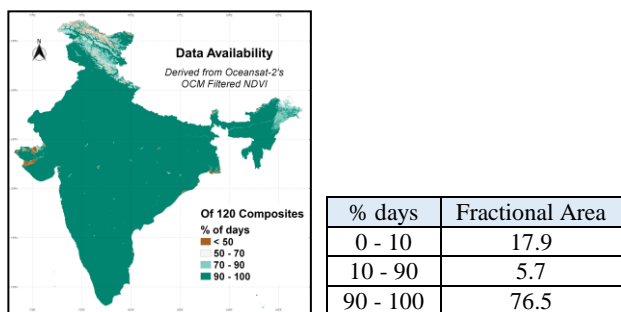


Figure 5: Data Availability in Agricultural Areas

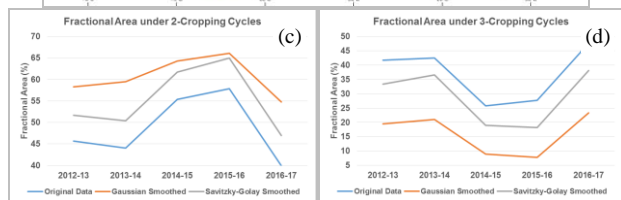
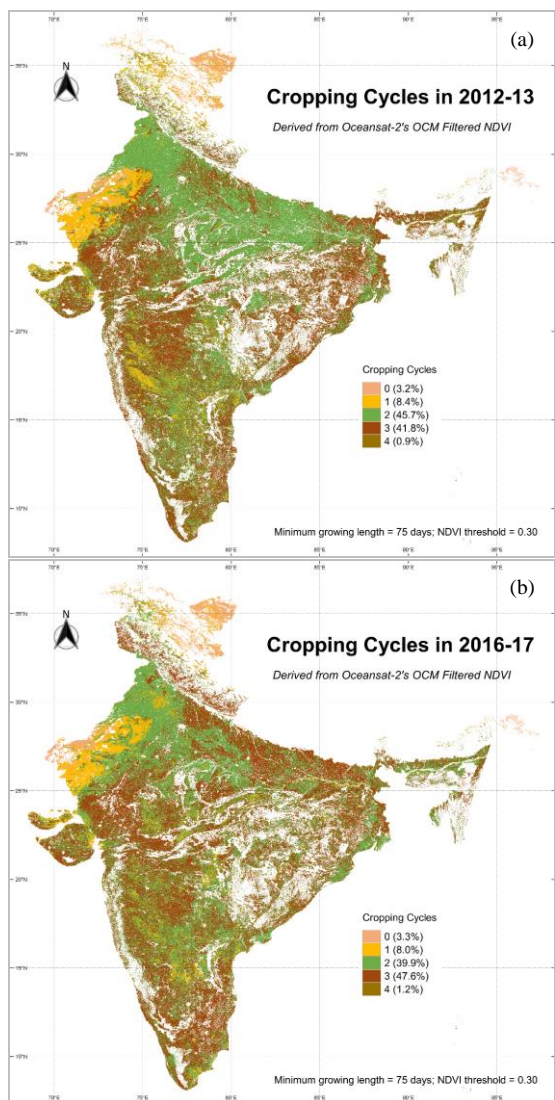


Figure 6: Cropping cycles derived from "peak detection" algorithm for (a) 2012-13 and (b) 2016-17; (c-d) shows the change in fractional area under 2-3 cropping cycles across 5 cropping years (2012-17) using original OCM NDVI time series & smoothed NDVI time series datasets

Figure 6(a-b) shows the spatially distributed cropping cycle maps obtained. The technique could capture the no vegetation (harsh conditions) in hot and cold deserts in Rajasthan and Jammu & Kashmir respectively. Large tracts of North East India and Central / Eastern India were masked out due to forests. The most important outcome is that large parts of the Indo-Gangetic plains seem to be converting from 2 cropping cycles to 3 cropping cycles.

Figure 6(c-d) show the change in percentage or fractional area in India, performing zero or more cropping cycles. Analysis done using the SavGol smoothed data gives the best representation, as the original data and Gaussian smoothed data either overestimate or underestimate the area statistics. From the 2 graphs it is also found that there was increase in area under 2-cropping cycles and decrease in area under 3-cropping cycles from 2013-14 till 2015-16. Also coastal areas (only 0.9-1.2% of fractional area) was found to be performing 4-cropping cycles, which is the case due to high water availability in these regions.

3.3 GIMMS data Analysis

The 8km NDVI product derived from NOAA-AVHRR data provides a longer time span data to analyse the regional changes. The "peak detection" algorithm was applied on the GIMMS dataset for 1981 to 2006 agricultural years (July-June). The increased use of crop management practices is evident from the results (Figure 7), which show that large tracts of the Indian mainland moved from 1 crop to 2 crop cycles, over the 25 year period. The trend can also be validated with Area-Production-Yield statistics, released annually by the Govt. of India.

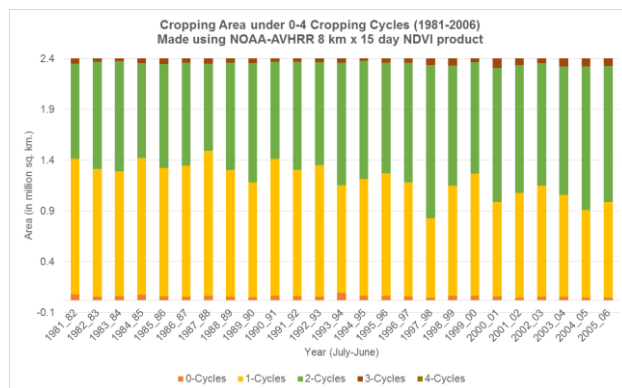


Figure 7: Area under different cropping cycles across 25 cropping years (1981-2006) using smoothed GIMMS NDVI time series dataset

4. CONCLUSIONS

The importance of food security is highlighted, regularly, at national (both civilian and strategic) and international (United Nations, etc.) levels. Therefore, it is imperative to understand and quantify the agricultural practices and predicted yield of the farmlands. The current study presented robust machine learning algorithms for detection of cropping patterns in India. The 2 algorithms showcased effectively captured local (3 potato crop) to regional (India moving from 1 crop to 2 crop) phenomena. The outputs of the study can also be used as input for advanced "data intensive" machine learning techniques (such as deep learning) for preparation of labelled training/testing datasets to perform non-linear data fitting using neural networks. The study can also be extended to generate type of crop maps, based on

the min/max NDVI, duration of crop, and by measuring the Full Width at Half Maximum (FWHM).

ACKNOWLEDGEMENTS

PKG and PAV wish to thank Dr. S. K. Srivastav and Dr. Sameer Saran for their encouragement and constructive comments. The support provided by scientists at the Agriculture and Soils Department at IIRS is duly acknowledged.

REFERENCES

- Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V. and Durieux, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, 32(22), pp.7847-7871.
- Estel, S., Kuemmerle, T., Levers, C., Baumann, M. and Hostert, P., 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. *Environmental Research Letters*, 11(2), p.024015.
- Li, L., Friedl, M.A., Xin, Q., Gray, J., Pan, Y. and Froking, S., 2014. Mapping crop cycles in China using MODIS-EVI time series. *Remote Sensing*, 6(3), pp.2473-2493.
- Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F.M. and de Queiroz, G.R., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), pp.3729-3739.
- Nikam, B.R., Garg, V., Gupta, P.K., Thakur, P.K., Kumar, A.S., Chouksey, A., Aggarwal, S.P., Dhote, P. and Purohit, S., 2017. Satellite-based mapping and monitoring of heavy snowfall in North Western Himalaya and its hydrologic consequences. *Current Science*, 113(12).
- NRSC, 2015. OCM-2 NDVI Filtered Product: Version-01 for Bhuvan NOEDA. URL: https://bhuvan-noeda.nrsc.gov.in/data/download/tools/document/OCM2%20NDVI%20Filtered%20Product_for_bhuvan.pdf (Jan 03, 2019)
- Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E. and Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6), pp.1301-1316.
- Rodrigues, A., Marcal, A. R., & Cunha, M. (2012). Phenology Parameter Extraction from Time-Series of Satellite Vegetation Index Data using Phenosat. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* (pp. 4926-4929). IEEE.
- Roy, D.P. and Yan, L., 2018. Robust Landsat-based crop time series modelling. *Remote Sensing of Environment*. In Press.
- Roy, P.S., Kushwaha, S.P.S., Murthy, M.S.R., Roy, A., Kushwaha, D., Reddy, C.S., Behera, M.D., Mathur, V.B., Padalia, H., Saran, S. and Singh, S., 2012. Biodiversity Characterisation at Landscape Level: National Assessment. Indian Institute of Remote Sensing, Dehradun, India. ISBN 81-901418-8-0.
- United Nations (1975). Report of the World Food Conference, Rome, 5-16 November 1974. *New York: UN*