v1: 11 February 2024

Research Article

Yield Forecasting Model for Maize Using Satellite Multispectral Imagery Driven Vegetation Indices

Peer-approved: 11 February 2024

Qeios, Vol. 6 (2024)
ISSN: 2632-3834

Nur Mohammad[1], Mohammad Amirul Islam[2], Mohammad Mukhlesur Rahman[1], Istiak Ahmed[1], Golam Mahboob[3]

1. Agricultural Statistics & ICT Division, Bangladesh Agricultural Research Institute, Dhaka, Bangladesh; 2. Department of Agricultural and Applied Statistics, Bangladesh Agricultural University, Bangladesh; 3. Bangladesh Agricultural Research Council (BARC), Dhaka, Bangladesh

In recent decades, the cultivation of maize (*Zea mays* L.) has witnessed remarkable growth in Bangladesh, particularly in the northern regions. Maize, as a high-yielding grain crop with diverse applications, plays a pivotal role in the country's agricultural landscape. Traditional methods of yield prediction involve time-consuming and subjective on-site field visits, resulting in significant errors and delayed information dissemination to government authorities and decision-makers. This study explores the potential of remote sensing technology to predict maize yields before harvest, thereby enhancing agricultural decision-making processes. The research utilizes 16-day (~30 m) Landsat 8 and 10-day (~10 m) Sentinel 2A imagery from two years, November 2018 to February 2019 and November 2019 to February 2020, to forecast maize yields in the Kaharole upazila of the Dinajpur district, Bangladesh. Four cloud-free images, representing the maximum normalized difference vegetation index (NDVI) for each maize growing season, are selected from the Landsat 8 and Sentinel 2A data. Regression models are established to relate NDVI values to the maize yields across 20 individual farmers' fields. The results reveal that the prediction models based on mean NDVI values for the combined growing seasons outperform those based on single growing seasons, and the finer spatial resolution of Sentinel 2A contributes to its superior performance in comparison to Landsat 8, offering valuable insights for improved agricultural management and food security.

Corresponding author: Nur Mohammad, nurbari2011@gmail.com

Introduction

Timely and accurate prediction of crop yields is necessary for effective agricultural land management, decision-making, and sustainability of agricultural food production (Masson et al., 2018). Remote sensing technology plays a vital role in the agriculture sector by providing timely and accurate information (Atzberger, 2013). Maize (*Zea mays L.*), also known as corn, is the world's fourth major staple food crop after rice, wheat, and potato. Maize is initially grown for grain and secondly for fodder and raw material for industrial purposes. Maize is one of the important coarse cereal crops grown in distinct agronomical conditions of Bangladesh. Maize cultivation has mainly increased rapidly in the northern part of Bangladesh. It has the highest potential for per day carbohydrate productivity.

During the last decades, maize cultivation in Bangladesh has shown an increasing trend. During 2011-12, the total maize cultivation area and yield were 1,97,000 ha (0.2 million) and 12,98,109 (1.3 million) metric tons, respectively, while in 2019-20, these figures rose to 4,71,900 ha (0.47 million) and 4,015,306 (4.02 million) metric tons, respectively (BBS, 2020). These indicate that the future prospects of maize production in Bangladesh are quite bright. However, maize growth monitoring and its yield estimation have become major issues of consideration.

The process of collecting crop data in the field using traditional methods is characterized by inefficiency, high expenses, significant time requirements, and the potential for inaccuracies (Reynolds et al., 2000). Furthermore, the utilization of conventional approaches for yield calculations has become increasingly disadvantageous for planners due to the excessive time required. In recent years, numerous empirical models have been devised to forecast crop production prior to harvest. However, a significant portion of these models has shown to be impractical, particularly those that rely on the collection of field data. Satellite-based remote sensing is widely recognized as a highly effective technology for acquiring crucial data pertaining to crop distribution and the prevailing growing conditions across extensive regions. Consequently, it may be effectively employed for the purpose of monitoring maize growth and forecasting crop yields. The integration of data obtained from the Landsat 8 and Sentinel 2 remote sensing satellites offers a significant temporal resolution (3-5 days) (Li et al., 2017), which is of utmost importance for various applications that necessitate a dense time series of satellite data. The study conducted by Segarra et al. (2020) provides a conclusion that highlights the diverse range of valuable applications of Sentinel-2 in the agricultural sector. However, the researchers also acknowledge that there is still potential for further enhancements in this regard. Despite the increased frequency of observations offered by the integration of Landsat 8 and Sentinel 2, it is important to note that disparities in the availability of cloud-free data will persist. Numerous investigations have been conducted to examine the association between the normalized difference vegetation index (NDVI) and crop yield (Liu et al., 2002). In recent research, the utilization of Landsat and Sentinel 2 data has been employed to effectively address the task of crop production forecasting at a moderate geographic resolution. As an illustration, Lambert et al. (2018) employed Sentinel 2 data and a peak LAI (Leaf area index) methodology to forecast the agricultural outputs of cotton, maize, millet, and sorghum in Mali. The coefficient of determination (R2) exhibited a range of 0.48 to 0.80 when applied to different crops within the training dataset. In their study, Lai et al. (2018) utilized time-integrated Landsat NDVI data to estimate wheat yield in Australia. The researchers employed an asymmetric bell-shaped growth model in order to accurately describe the relationship between the Normalized Difference Vegetation Index (NDVI) and time. Shakun et al. (2019) applied the combination of Landsat 8 and Sentinel 2 high frequency of observations (3–5 days) at moderate spatial resolution (10–30 m), which is important for crop yield studies, which were executed for the model with near-infrared (NIR) and red spectral bands and derived AUC, constant, quadratic, and linear coefficients of the quadratic model. The best model yielded a root mean square error (RMSE) of 0.201 t/ha (5.4%) and a coefficient of determination R^2 = 0.73 on cross-validation. Rahman et al. (2020) used the Simple Linear Machine Learning (ML) algorithm; the extracted Landsat-derived average green normalized difference vegetation index (GNDVI) values for each of the blocks were converted to Sentinel GNDVI and found strong correlations (R^2 = 0.92 to 0.99) in the Bundaberg growing region of Australia. Lima et al. (2019) observed that both satellites showed the same performance in terms of accuracy for Sentinel-2 and Landsat 8, respectively. However, Landsat 8 mapped 36.9% more area of selective logging compared to Sentinel-2 data for mapping small-scale logging in the Brazilian Amazon. In some research, to predict grain yield 2–3 months before the harvest, more advanced regression models for yield prediction apply time series of NDVI, which allow one to obtain better forecasting (Panek et al., 2021). In Bangladesh, Bala & Islam (2009) expanded potato yield estimation models by using NDVI, LAI (leaf area index), and fraction of photosynthetically active radiation (fPAR) vegetation indices for the Munshiganj District of Bangladesh by applying Moderate Resolution Imaging Spectroradiometer (MODIS) (with the lowest resolution greater than 250m) 8-day composite surface reflectance data and noticed that an average error of estimation is about 15% for the study location. Islam et al. (2011) used the Normalized Differentiate Vegetation Index (NDVI) indicator developed from time series MODIS satellite images; the phenological growth of wheat was monitored during the Rabi season of 2007-2008 for the greater Dinajpur area of Bangladesh. A strong correlation between the wheat production and the satellite-represented wheat area was found (R^2=0.71), which represents the effectiveness of the remote sensing tools for crop monitoring and production estimation. Rahman et al. (2012) applied NOAA-AVHRR data for the prediction of potato yield in Bangladesh. However, a high-resolution (~ 30 m) satellite image from Landsat has been cost-free available since 1984. The availability of Landsat 8 images contributes an ample opportunity for long-term frequent environmental monitoring (Mandanici et al., 2016). Newton et al. (2018) improved a potato yield prediction model by applying 16-day high-resolution (~ 30 m) Landsat surface reflectance data to identify the maximum normalized difference vegetation index (NDVI) value of a potato growing season in the Munshiganj District of Bangladesh. The maximum coefficient of determination (R^2) of the yield forecasting equation was found to be 0.81 between the mean NDVI and potato yield, and the result revealed that the difference between predicted and actual field yield is about 10.4%. However, very few studies have been conducted on the relationship between high-resolution (~ 30 m) Landsat 8 and Sentinel 2A (~ 10 m) satellite data and maize yield in Bangladesh. Even though this study made an attempt to construct a maize yield prediction model based on NDVI at Kaharole Upazila in the Dinajpur District of Bangladesh, respectively, using high-resolution Landsat 8 Operational Land Imager (OLI) and Sentinel 2A Multi-Spectral Instrument (MSI) surface reflectance data. The combined use of high-resolution Landsat 8 and Sentinel 2A images has been applied in this study to improve the yield assessment model for the maize crop at Kaharole Upazila of the Dinajpur district in Bangladesh.

Materials and Methods

Study Area

The research was carried out in Kaharole upazila, located in the Dinajpur district of Bangladesh. This region is particularly known for its high maize cultivation, especially during the Rabi season. According to the Bangladesh Bureau of Statistics (BBS, 2020), the northern region of Bangladesh accounted for 26.36% of the total maize-cultivated area, while the remaining 14.15% was distributed throughout the rest of the country. The Kaharole upazila is situated within the geographical coordinates of 25° 44′ to 25° 53′ N latitude and 88° 30′ to 89° 43′ E longitude, as depicted in Figure 1. The Kaharole upazila encompasses an area of 205.54 square kilometers, with cultivable land accounting for around 59% of this total area. The climatic conditions in this region are characterized by high temperatures and humidity throughout the period from April to October, commonly referred to as summer, and by cooler temperatures and lower humidity from November to March, known as winter. The Kaharole upazila experiences an average yearly rainfall of 1965 mm and 2417 mm, respectively. It is worth noting that approximately 90% of this rainfall is concentrated between the months of May and October. In the Kaharole upazila, the typical temperature range throughout the winter season is from 23.6 to 16.8°C for the maximum value and from 24 to 16.8°C for the minimum value. During the summer season, the average maximum and minimum temperatures exhibit a range of 33.2 to 26.0 and 29.8 to 25.6°C, respectively. The Kaharole upazila experiences a range of humidity levels throughout the year, with an average monthly minimum of 68% and an average monthly maximum of 86%. The highest humidity levels are typically recorded during the summer season, while the lowest humidity levels are observed during the dry season. The agricultural pattern observed in this region is characterized by the presence of two distinct growing seasons, namely Rabi and Kharif. The primary growing season in this upazila is known as Rabi, characterized by the cultivation of maize and wheat. It typically commences in late October or early November and concludes in April. In contrast, the Kharif season is characterized by the predominant cultivation of rice and jute, commencing in May and concluding in September. Potato, pepper, onion, legumes, sugarcane, and oilseed are among the additional food crops planted in both regions. The prevailing soil conditions in this area consist of non-calcareous brown floodplain soils and grey floodplain soils.

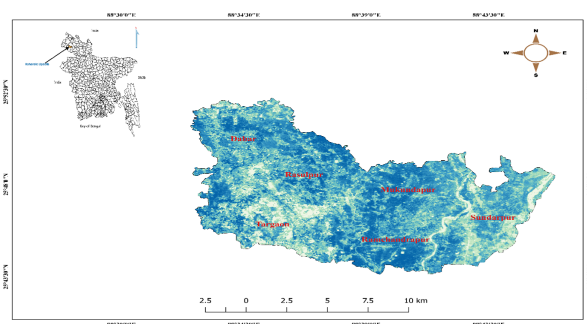

Figure 1. Map of the study location: Kaharole upazila map, Dinajpur.

Yield data collection from farmers' fields

Twenty farmers' maize fields were selected for the three maize growing seasons 2018-2019, 2019-2020, and

2020-2021, respectively, with the agreement of the farmers at Kaharole upazila of Dinajpur district (Figure 2). A total of 20 different farmers' maize field data were collected from Kaharole upazila for each season. Crop information data such as field GPS locations, planting and harvesting times, and yield were collected from those selected upazila farmers' fields.

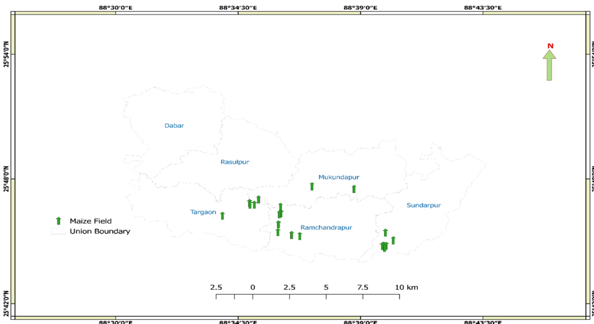

Figure 2. Location of selected 20 maize fields (red triangles) over Kaharole, Dinajpur.

Landsat-8/OLI and Sentinel-2A /MSI Datasets

Landsat 8 images (OLI) were obtained from the United States Geological Survey (USGS) Earth Explorer website (http://earthexplorer.usgs.gov/). Landsat 8 (OLI) is a sun-synchronous satellite staying at an altitude of 705 km above the Earth with a 16-day repeat cycle. Landsat 8 has two types of sensors, especially the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). The OLI sensor is equipped with nine spectral bands, including a pan band, and TIRS produces two spectral bands. Sentinel-2A images (MSI) were obtained from the European Space Agency (ESA) Copernicus portal (https://scihub.copernicus.eu). Sentinel-2A carries a multispectral instrument (MSI). Images furnished by Sentinel-2A are publicly available for free and have 13 spectral bands with a spatial resolution ranging from 10 m to 60 m (depending on the band) and a current temporal resolution of about 10 days (depending on the latitude). We downloaded a total of 6 images which were maximum cloud-free, viz., 3 images were collected from Landsat 8 OLI and the rest of 3 images were collected from Sentinel 2A MSI satellite data for Kaharole upazila from Dinajpur district in the consecutive years 2018-19, 2019-20, and 2020-21, respectively. A total of 6 images (three from Landsat 8 and three from Sentinel-2A) were suitable and used in the subsequent analyses. The single date of image acquisition based on maximum greenness was used for each growing period, i.e., 2018-2021, for maize cultivation. The maize sowing date was considered to be the last week of November and the first week of December for Kaharole upazila for each growing season 2018-19, 2019-20, and 2020-21, respectively, for the entire study site based on the information taken from the location visits. Every single image was calculated from the starting day of the plantation. The dates of image acquisition of Landsat 8/OLI and Sentinel 2A/MSI for the maize growing seasons 2018-19, 2019-20, and 2020-21, respectively, for Kaharole, Dinajpur, used for this study are presented in Table 1.

Satellite Images	Landsat 8			Sentinel 2A		
Growing Season	2018-19	2019-20	2020-21	2018-19	2019-20	2020-21
IAD	10/03/2019	12/03/2020	15/03/2021	16/03/2019	20/03/2020	05/03/2021
DAP	95	97	103	101	105	93

Table 1. Model development using Landsat 8 and Sentinel 2A satellite images at Kaharole upazila, Dinajpur

IAD=Image acquisition date; DAP= Days after plantation

Satellite Image Pre-Processing

For Landsat data, raw digital numbers (DN) were adjusted to top-of-atmosphere (TOA) reflectance values following reference (Simonetti et al., 2015). Two techniques were used to preprocess the satellite images: (1) radiometric calibration and (2) atmospheric correction. Remote sensing data adopted from satellite sensors are influenced by several factors, such as atmospheric scattering and absorption, sensor-target-illumination geometry, sensor calibration, and also by the data processing procedures (Teillet, 1986). For that, radiometric calibration is needed. Radiometric calibration means a set of correction techniques that are associated with correction for the sensitivity of the satellite sensor, topography and sun angle, atmospheric scattering, and absorption (Kim et al., 1990). The radiometric calibration was done by transforming the digital numbers (DNs) to surface reflectance by radiance conversion. The open source-based Quantum Geographic Information System (QGIS) software, version 2.18.13, allows a plugin, and the plugin provides a tool for atmospheric correction, which is known as dark object subtraction (DOS-1) level 1. In this study, this tool was used on the radiometrically calibrated images to minimize the atmospheric scattering effect. DOS-1 searches each pixel of a band to find the darkest value. The scattering is eliminated by subtracting this value from every pixel in the band.

Two approaches were followed to download and process Sentinel-2A imagery. The first was a simplified process for farmers and advisors to monitor the crop status during the season. For this purpose, we used the free and open source QGIS 2.18.13 version software together with the Semi-automatic Classification Plugin (SCP) (Congedo, 2016). The advantage of using the SCP is that the user can preview and download per date and per tile single bands and correct the Sentinel-2 images in the same interface. Afterward, the vegetation indices can be computed, stored, and compared with other dates within the same QGIS environment. The only limitation is that in the conversion of top-of-atmosphere (TOA) reflectance values into bottom-of-atmosphere (BOA), the image-based Dark Object Subtraction (DOS-1) technique is applied. This process is less accurate than the physically-based correction that could be applied to Sentinel-2 images using the attached metadata (Congedo, 2016). The second was a more accurate, though complex, criterion. It should be recommended to extract vigour indices to be correlated with yield data and to be compared with subsequent seasons. The images were pre-processed with the open-source ESA Sentinel Application Platform (SNAP), which covers the Sentinel-2 Toolbox. In addition, the third-party plugin Sen2Cor was applied (http://step.esa.int/main/third-party-plugins-2/sen2cor). Sen2Cor is a processor for Sentinel-2 Level-2A product generation and formatting; it operates the physical atmospheric, terrain, and cirrus correction of TOA Level-1C Sentinel-2 products and creates, among other products, BOA reflectance corrected bands. Its output product format is equivalent to the Level 1C User Product: JPEG 2000 images with bands with three different resolutions, 60, 20, and 10 m.

Normalized Difference Vegetation Index (NDVI)

Prediction and estimation of yield are closely associated with the capability of identifying crop species and certain agronomic parameters, such as maturity, density, vigour, and disease, which can be used as yield indicators. Remote sensing can arrange these types of information to a great extent. There are distinct types of vegetation indices (VIs) generated from different spectral reflectance that are specially used to get these types of information. The NDVI is generally applied extensively around the world to monitor vegetation quality, growth, and distribution over a large area.

Differences in the phenological growth stages of different plants are reflected in the temporal NDVI profiles, since NDVI can measure growth conditions (greenness of vegetation) (Belgiu & Csillik, 2018; Croitoru et al., 2012). It is a dimensionless index, which is derived from the ratio between the surface reflectance of the NIR and RED bands of the spectrum as follows (Equation 1) (Rouse et al., 1974).

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

Where RED (Visible red) and NIR (Near infrared) are reflectance measurements for RED and NIR bands, respectively. Here, for Landsat 8/OLI, band 4 and band 5 represent RED and NIR bands, and for Sentinel 2A/MSI, band 4 and band 8 represent Red and NIR bands (Tucker, 1979). Factors like strong reflectance in NIR and strong absorption in Visible Red of specific vegetation distinguish the vegetation from bare soil. NDVI for a given pixel can always output a number that ranges from -1 to +1; however, for natural surfaces, NDVI values are within the 0 to +1 range. Negative values of NDVI, i.e., values approaching -1, correspond to water. An NDVI close to 0 corresponds to no vegetation, while values lying between -0.1 to 0.1 generally correspond to barren areas of rock, sand, or snow.

Maize yield estimation by satellite-based remote sensing technique

The red band and NIR of the calibrated images were selected from each dataset and exported into QGIS 2.18.13. A simple raster calculation was done by QGIS 2.18.13 using Equation 1 to find the NDVI images. Finally, the NDVI images were masked using the shapefile from different locations like Kaharole upazila of Dinajpur district. The field points of the location were imported, and the mean NDVI values for each point were extracted from the satellite image considering a 3 × 3 matrix surrounded by each point on the image.

The relationship between NDVI and the maize growing period was established by plotting the respective values in terms of single days from the start date of maize plantation to the harvesting period. The day of the maximum NDVI was selected from their relationship with crop yield. To establish this relationship, NDVI data from the growing seasons 2018-2020 were used. Then, a total of four satellite images, viz., each image collected from Landsat 8 as well as Sentinel 2A from Kaharole upazila, Dinajpur, depending on the date of the maximum NDVI, were selected from two growing seasons, namely 2018-2019 and 2019-2020, to build a relationship between the NDVI values and field-level maize yield. This relationship, based on each farmer's field point NDVI values, was validated using the satellite image of the 2020-2021 growing season. NDVI values less than 0.25 and more than 0.95 were removed from the listed fields to reduce the influence of the reflectance of other objects like bare soil, settlements, water bodies, non-agricultural crops, and infrastructure.

Yield prediction model

The final step is to determine the relationship between NDVI and maize yield from farmers' fields and BBS-selected fields with the equation below:

$$y = f(x) \tag{2}$$

Where y and x are maize yield data collected from farmers' fields and NDVI, respectively. The relationship between NDVI and crops like maize yield has been observed through the linear regression model, where the response variable is denoted by maize yields and the explanatory variables by NDVIs. Several studies applied a linear regression model to describe the relationship between NDVI and crop (wheat) yield in distinct locations (Ren et al., 2008). To develop the maize yield estimation model for both fields, the data of maize yield and Landsat 8 (OLI) and Sentinel 2A (MSI) images were used for 2018-2021.

Results and Discussion

Maize yield from farmers' fields and corresponding NDVI values for different locations

Maize yield data and NDVI values from different satellite images like Landsat 8 and Sentinel 2A for corresponding farmers' fields have been collected from Kaharole upazila during the maize growing seasons 2018-19 and 2019-20, respectively. Twenty farmers′ field yield data collected from Kaharole upazila and their corresponding NDVI values for two satellite images have been presented in Tables 2 and 3 for consecutive maize growing seasons 2018-19 and 2019-20, respectively.

Farmer’s Field	Longitude	Latitude	NDVI values		Yields (t/h)
			Landsat 8 OLI	Sentinel 2A MSI	
1	88.665283	25.757000	0.55	0.77	11.35
2	88.664166	25.746660	0.70	0.82	12.72
3	88.663333	25.746100	0.67	0.80	11.42
4	88.665833	25.746383	0.69	0.84	12.35
5	88.664783	25.744716	0.71	0.85	12.12
6	88.670109	25.750864	0.63	0.78	10.95
7	88.601321	25.772369	0.66	0.82	11.25
8	88.600303	25.771734	0.57	0.78	10.85
9	88.599752	25.763595	0.58	0.77	10.65
10	88.607792	25.755088	0.68	0.76	10.94
11	88.612763	25.754269	0.58	0.80	11.27
12	88.599452	25.757273	0.60	0.83	12.27
13	88.600953	25.777712	0.59	0.72	9.45
14	88.582059	25.780769	0.63	0.75	10.68
15	88.587503	25.783616	0.55	0.75	10.52
16	88.582345	25.779255	0.65	0.81	11.69
17	88.584909	25.779358	0.61	0.72	10.8
18	88.565401	25.770503	0.28	0.33	6.80
19	88.620312	25.794201	0.35	0.60	8.61
20	88.646041	25.792031	0.56	0.69	9.25

Table 2. NDVI values of satellite images and yields of corresponding farmers’ fields at Kaharole, Dinajpur during the season of 2018-19

Table 2 shows that NDVI values are 0.71 and 0.85, which were the highest for Landsat 8 and Sentinel 2A, and the yield was 12.12 t/ha for farmer’s field 5, but the yield is a maximum of 12.72 t/ha for field 2; and NDVI values are 0.28 and 0.33, which were the lowest, and the yield was 6.8 t/ha for farmer’s field 18 at Kaharole upazila during 2018-19.

Farmer's Field	Longitude	Latitude	NDVI values		Yields (t/h)
			Landsat 8 OLI	Sentinel 2A MSI	
1	88.665283	25.757000	0.69	0.84	11.50
2	88.664166	25.746660	0.79	0.90	12.74
3	88.663333	25.746100	0.74	0.83	11.45
4	88.665833	25.746383	0.80	0.86	12.48
5	88.664783	25.744716	0.79	0.86	12.50
6	88.670109	25.750864	0.68	0.81	11.20
7	88.601321	25.772369	0.69	0.84	11.50
8	88.600303	25.771734	0.38	0.76	10.65
9	88.599752	25.763595	0.71	0.78	10.87
10	88.607792	25.755088	0.79	0.89	11.62
11	88.612763	25.754269	0.63	0.81	11.42
12	88.599452	25.757273	0.74	0.86	12.63
13	88.600953	25.777712	0.61	0.83	11.45
14	88.582059	25.780769	0.70	0.84	11.30
15	88.587503	25.783616	0.52	0.77	10.24
16	88.582345	25.779255	0.71	0.82	11.45
17	88.584909	25.779358	0.72	0.82	10.15
18	88.565401	25.770503	0.30	0.48	7.60
19	88.620312	25.794201	0.32	0.65	8.71
20	88.646041	25.792031	0.68	0.76	10.62

Table 3. NDVI values of satellite images and yields of corresponding Farmer's fields at Kaharole, Dinajpur during the season of 2019-20

Maximum NDVI for Kaharole were 0.80 and 0.90 for farmer's fields 4 and 2, and minimum NDVI were 0.30 and 0.48 for Landsat 8 and Sentinel 2A for farmer's field 18, but the highest and lowest yields were 12.74 t/ha and 7.6 t/ha for farmer's fields 2 and 18, respectively, during 2019-20 in Table 3.

Regression analysis of the NDVI values over the field locations

A total of four satellite images (2 images each for Landsat 8 and Sentinel 2A) from two growing seasons during 2018-2019 and 2019-2020 were selected. Based on available images, those showing the maximum NDVI in each growing season were found on the 95th and 97th days after plantation for Kaharole upazila from Landsat 8 images, as well as on the 101st and 105th days after plantation for Kaharole upazila from Sentinel 2A images for the 2018-2019 and 2019-2020 growing seasons, respectively. The spatial distribution of the NDVI varies from year to year. Spatial distribution of the NDVI over the selected location for selected distinct satellite images against different growing seasons is presented in Figure 3.

For Landsat 8 data, NDVI distribution was maximum during 2019-2020 and minimum during the season 2019-20 at Kaharole upazila in Figure 3. On the other hand, for Sentinel 2A data, NDVI distribution was maximum during the season 2019-2020 and minimum during the season 2019-20 at Kaharole upazila in Figure 3. From Figure 3, we can see that the NDVI distribution from different locations of Sentinel 2A data is better than Landsat 8 data during the maize growing seasons 2018-2019 and 2019-2020, respectively.

Figure 3. Spatial distribution of the NDVI for different satellite images during the growing seasons 2018-19 and 2019-20. a. 95th day after plantation during 2018-19; b. 97th day after plantation during 2019-20 for Landsat 8; c. 101st day after plantation during 2018-19; d. 105th day after plantation during 2019-20 for Sentinel 2A at Kaharole upazila.

For two satellites, viz., Landsat 8 and Sentinel 2A, the NDVI values and their corresponding yields for twenty farmers' maize fields from different locations, i.e., Kaharole upazila during the individual maize growing seasons 2018-2019 and 2019-2020, respectively, are shown in Table 2 and Table 3. From Tables 2 and 3, we have calculated the mean NDVI and mean yield for two satellite data, like Landsat 8 and Sentinel 2A, for Kaharole upazila during the combined seasons 2018-2019 and 2019-2020, respectively, which are represented below in Table 4. Mean NDVI and mean yield are calculated for two satellite images for the combined two seasons because maize yield in each season, i.e., 2018-19 and 2019-20, is mostly the same for Kaharole upazila. The mean NDVI and mean yield of the combined maize season performed better than the NDVI and its corresponding yield of individual maize seasons for each satellite image at Kaharole upazila because, according to best model criteria, viz. Multiple determination of coefficient (R2), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), it fitted best for the mean NDVI and mean yield of the combined maize season rather than a single maize growing season like 2018-19 and 2019-20 (Figure 3). The mean NDVI is the largest, which is 0.75 for farmers' fields 2, 4, and 5, as well as the smallest, which is 0.29 for farmers' field 18 for Landsat 8, and for Sentinel 2A, the largest and lowest are 0.86 and 0.41 for farmers' fields 2, 5, and 18 for Kaharole upazila for the combined year, respectively. The highest and lowest mean yields of the two satellite data are 12.73 (t/ha) and 7.2 (t/ha) for Kaharole upazila for the combined year, respectively, in Table 4.

doi.org/10.32388/COEBSC.2

Farmer’s Field	Mean NDVI		Mean Yield (t/ha)
	Landsat 8 OLI	Sentinel 2A MSI	
1	0.62	0.81	11.43
2	**0.75**	**0.86**	**12.73**
3	0.71	0.82	11.44
4	**0.75**	0.85	12.42
5	**0.75**	**0.86**	12.31
6	0.66	0.80	11.08
7	0.68	0.83	11.38
8	0.48	0.77	10.75
9	0.65	0.78	10.76
10	0.74	0.83	11.28
11	0.61	0.81	11.35
12	0.67	0.85	12.45
13	0.6	0.78	10.45
14	0.67	0.80	10.99
15	0.54	0.76	10.38
16	0.68	0.82	11.57
17	0.67	0.77	10.48
18	**0.29**	**0.41**	**7.20**
19	0.34	0.63	8.66
20	0.62	0.73	9.94

Table 4. Mean NDVI values for two satellite images and corresponding mean yields of Farmer’s Maize Fields at Kaharole, Dinajpur during the combined seasons of 2018-19 and 2019-2020

Maize yield and NDVI relationship using regression model

Regression analysis of maize yield against the single season basis NDVI and combined season basis mean NDVI for Kaharole was performed for two satellite images, such as Landsat 8 and Sentinel 2A, and is graphically presented in Figure 4.

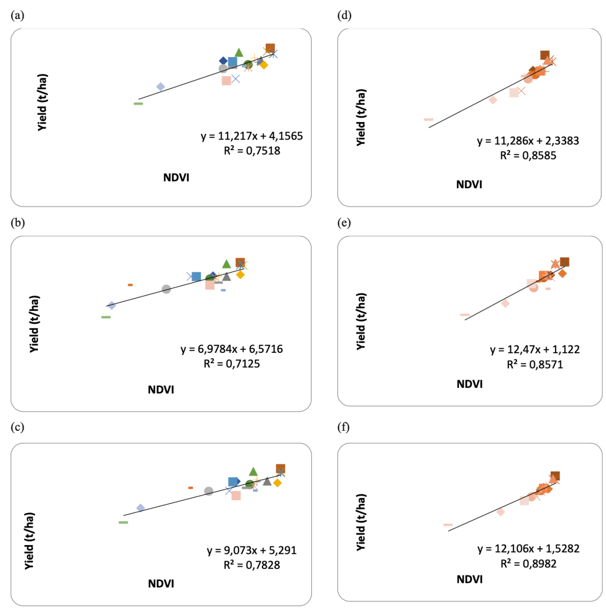

Figure 4. Yield prediction model established from regression analysis between yield data collected from 20 farmers' maize fields for different images; [a. yield vs NDVI, 2018-19 b. yield vs NDVI, 2019-20 c. yield vs NDVI (mean), combined 2018-19 and 2019-20 for Landsat 8.]; [d. yield vs NDVI, 2018-19 e. yield vs NDVI, 2019-20 f. yield vs NDVI (mean), combined 2018-19 and 2019-20 for Sentinel 2A] at Kaharole upazila.

The yield vs. NDVI relationship for Landsat 8 and Sentinel 2A satellite images, as shown in Figure 4, revealed that the multiple determination coefficients, which are highest, and other accuracy assessments like the minimum values of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) of mean NDVI for the combined maize growing season, i.e., 2018-2019 and 2019-2020, are better fitted than the single maize growing season, i.e., 2018-2019 and 2019-2020, respectively, for Kaharole upazila, Dinajpur. The parameters of the regression analysis estimated from the yield vs. NDVI relationship for the combined season, together with the values of R2, MAPE, and RMSE, are presented in Table 5. The relationship between mean NDVI for the combined two maize growing seasons and yield is provided almost as well compared to the single season basis NDVI vs. yield relationship for the two satellite images.

doi.org/10.32388/COEBSC.2

Location	Satellite Data	Regression parameter for mean value				Best Model Criteria		
		β_0	β_1	$SE(\beta_1)$	P-Value	MAPE	RMSE	R^2
Kaharole	**Landsat 8**	5.291	9.073	1.126	0.0000	0.046	0.589	0.782
	Sentinel 2A	1.528	12.10	0.961	0.0000	0.031	0.403	0.898

Table 5. Regression parameters and R^2 for combined season at farmers' maize fields

Here, the regression coefficients of all the fitted models of the two satellite images show a highly significant effect for Kaharole upazila in Table 5. From Table 5, we have also seen that the multiple determination coefficient (R2), along with other parameters, is better for the Sentinel 2A satellite image than for the Landsat 8 satellite image for Kaharole upazila, Dinajpur.

Development and validation of the yield prediction model

The yield prediction model based on the regression analysis was developed based on the yield data collected from the 2018–2019 and 2019–2020 maize growing seasons. To evaluate the performance of the model, validation is essential. Based on the deviation from the estimated and model prediction, model performance can be determined. Hence, the model has been further validated using yield data from the 2020-2021 maize growing season. After the 103rd and 93rd days after plantation, an NDVI image was selected for the two satellite images, viz., Landsat 8 and Sentinel 2A, from the 2020 to 2021 growing season at Kaharole upazila, Dinajpur, as shown in Figure 5.

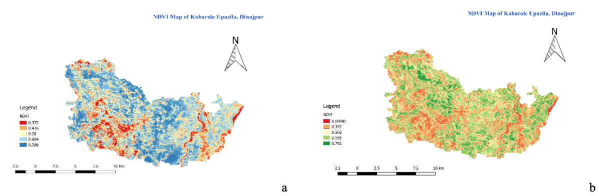

Figure 5. Spatial distribution of the NDVI over the different locations for Landsat 8 and Sentinel 2A satellite images during the growing season of 2020-2021. (a+b): 103rd and 93rd days after plantation for Kaharole upazila, Dinajpur, for different satellite images.

The NDVI value was extracted from each of the 20 farmers' fields, each for two satellite images from Kaharole upazila of Dinajpur during the maize growing season 2020-2021, which is presented in Table 6. As the coefficient of determination (R2) was found to be high from the relationship between the mean value of NDVI and yield of the combined season, the validation was done using the general mean value equation (NDVI and yield relationship of the combined season) for two satellite images at the maize location, such as Kaharole upazila, Dinajpur. The general mean value equation, which was elaborately defined in Table 5, has been used separately to develop two regression models for validation along with two satellite images for Kaharole upazila. The general regression equation from the mean NDVI (combined maize growing season) is presented in equation 3.

doi.org/10.32388/COEBSC.2

$$Yield = \beta_0 + \beta_1 \times NDVI_{mean(Combined\ maize\ growing\ season)} \tag{3}$$

The actual yield (t/ha) of maize and the predicted yield (t/ha) using equation 3 for two different satellite images for Kaharole upazila during the 2020-2021 maize growing season were also presented in Table 6.

Farmer's Field	Actual Yield(t/ha)	NDVI Values		Predicted Yield (t/ha)		Error of Yield (%)	
		Landsat 8	Sentinel 2A	Landsat 8	Sentinel 2A	Landsat 8	Sentinel 2A
1	11.25	0.53	0.73	10.1	10.36	10.22	7.91
2	11.65	0.57	0.74	10.46	10.48	10.21	10.04
3	11.95	0.61	**0.77**	10.82	10.85	9.46	9.20
4	**12.55**	0.67	**0.78**	11.37	10.97	9.40	12.59
5	11.53	0.53	0.71	10.09	10.12	12.49	12.23
6	10.67	**0.44**	0.66	9.28	9.51	13.02	10.87
7	12.37	0.61	**0.77**	10.83	10.85	12.44	12.28
8	12.2	0.6	**0.77**	10.73	10.85	12.04	11.06
9	12.15	**0.61**	**0.77**	10.83	10.85	10.86	10.69
10	10.89	**0.52**	**0.75**	10.01	10.6	8.08	2.66
11	12.35	0.58	**0.76**	10.55	10.72	14.57	**13.19**
12	10.35	**0.48**	0.69	9.64	9.88	6.85	4.54
13	10.87	0.5	**0.71**	9.83	10.12	9.56	6.89
14	11.42	**0.58**	**0.76**	10.55	10.72	7.61	6.12
15	11.1	0.44	**0.67**	9.28	9.64	**16.39**	13.15
16	11.25	**0.54**	**0.72**	10.19	10.24	9.42	8.97
17	10.34	**0.41**	0.65	9.01	9.39	12.86	9.18
18	**9.1**	0.39	**0.62**	8.82	9.03	**3.07**	**0.77**
19	9.5	**0.43**	**0.64**	9.19	9.27	3.26	2.42
20	11.04	**0.54**	**0.72**	10.19	10.24	7.69	7.25

Table 6. Estimated yield and predicted yield from selected Farmer's Maize Fields at Kaharole, Dinajpur during the season of 2020-2021

The highest and lowest observed yields of maize were 12.55 (t/ha) and 9.1 (t/ha), respectively, as well as the maximum and minimum NDVI values, which were 0.67 and 0.39 for Landsat 8; 0.78 and 0.62 for Sentinel 2A for farmers' fields 4 and 18, respectively, at Kaharole upazila, Dinajpur, during the maize growing season 2020-2021 (Table 6). The largest and smallest predicted yields of maize for Landsat 8 were 11.37 (t/ha) and 8.82 (t/ha), as well as for Sentinel 2A, these were 10.97 (t/ha) and 9.03 (t/ha) for farmers' fields 4 and 18, respectively, at Kaharole upazila, Dinajpur, during the maize growing season 2020-2021 (Table 6). The maximum and minimum yield gaps (%) for Landsat 8 were 16.39 and 3.07 for farmers' fields 15 and 18, respectively, and also for Sentinel 2A were 13.19 and 0.77 for farmers' fields 11 and 18, respectively, at Kaharole upazila, Dinajpur, during the maize growing season 2020-2021 (Table 6). From Table 6, we showed that the predicted yields of maize for the two satellite data were less than the actual yields of maize (underestimated) in each farmer's field for Kaharole upazila, Dinajpur district. Depending on the values in Table 6, validation of maize yield for two satellite images, viz., Landsat 8 and Sentinel 2A, at Kaharole upazila was presented in Table 7 during the maize season 2020-2021.

Location	Actual Yield (Mean)	Predicted Yield (Mean)		Mean Error of Yield (%)	
		Landsat 8	Sentinel 2A	Landsat 8	Sentinel 2A
Kaharole, Dinajpur	11.23	10.09	10.23	10.15	8.82

Table 7. Validation of maize yield at Kaharole Upazila during the season 2020-2021

The estimated farmers' field yield (mean) was 11.23 (t/ha), and the predicted yields (mean) for Landsat 8 and Sentinel 2A were 10.09 (t/ha) and 10.23 (t/ha), respectively, at Kaharole upazila, Dinajpur, during the maize growing season 2020-2021. The percentage of mean yield error was 10.15 for Landsat 8 and 8.82 for Sentinel 2A. The error of mean yield (%) of Sentinel 2A performed better than that of Landsat 8 at Kaharole upazila, Dinajpur, during the maize growing season 2020-2021 (Table 7).

Conclusions

The study has investigated the prediction capacity of remote sensing NDVI data for maize yield in the selected location, viz., Kaharole upazila, Dinajpur, of Bangladesh, which is known as a maize-dominating district. The study has also investigated the relationship between NDVI and yield for the study region. Here, two satellite images, Landsat 8 (OLI) and Sentinel 2A (MSI), which are high spatial resolution data, were used in this study. We downloaded a total of 6 images, which were the maximum cloud-free data collected from Landsat 8 OLI and Sentinel 2A MSI satellite data for Kaharole upazila from Dinajpur district in the consecutive years 2018-19, 2019-20, and 2020-21, respectively. The single date of cloud-free image acquisition based on maximum NDVI was used for each cropping period, i.e., 2018-2019 and 2019-2020, respectively, for maize cultivation to develop the yield prediction model. These equations were validated by using data from the 2020-2021 maize growing seasons.

Mean NDVI and mean yield of the combined maize season performed better than NDVI and its corresponding yield of a single maize season for each satellite image and for Kaharole upazila. The yield prediction equations were found based on mean values of NDVI for the combined growing season against the yield of maize. The yield against NDVI relationship for both satellite images showed that the multiple determination coefficient (R^2) along with Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) of mean NDVI for the combined maize growing season almost performed better than the single maize growing season for the location Kaharole, Dinajpur. We have also shown that the regression model is fitted very well based on a higher value of R^2 for Sentinel satellite 2A data than for Landsat 8 satellite data at Kaharole upazila, Dinajpur district. The estimated farmers' maize field yield (mean) was 11.23 (t/ha), and the predicted yield (mean) for Landsat 8 and Sentinel 2A were 10.09 (t/ha) and 10.23 (t/ha), respectively. The absolute mean error of prediction was found to be about 10.15% for Landsat 8 and 8.82% for Sentinel 2A compared to the actual yield at Kaharole upazila, Dinajpur district during the maize growing season 2020-2021. We observed that the predicted yield (mean) of Sentinel 2A is 1.33% closer to the actual yield than Landsat 8 at Kaharole upazila in Dinajpur district. We revealed that in this research, the yield prediction model for Sentinel 2A images performed better than Landsat 8 because of the high spatial resolution (~10m). It was found that NDVI data extracted from Sentinel 2A high-resolution satellite images can be successfully used to predict the maize yield over Kaharole upazila with appreciable accuracy. So, the high-resolution Sentinel 2A images can be an effective means for early prediction of maize yield.

Statements and Declarations

Conflict of Interest

The authors whose names are listed immediately below the title certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

Funding Statement

This research received funding from the Bangabandhu Science and Technology Fellowship Trust, Government of the People's Republic of Bangladesh.

Author's Contribution

A.B and E conceived of the presented idea, developed the theory, and performed the computations. C and D verified the analytical methods. B and E encouraged and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Availability of data and material

All data and materials supporting the findings of this study are available from the corresponding author on request.

Code availability

All the code used in the analysis of this study is available from the corresponding author on request.

Ethics approval

On behalf of all authors, the corresponding author consciously assures that for this study the following is fulfilled:

- This material is the authors' own original work, which has not been previously published elsewhere.
- The paper is not currently being considered for publication elsewhere.
- The paper reflects the authors' own research and analysis in a truthful and complete manner.
- The paper properly credits the meaningful contributions of co-authors and co-researchers.
- The results are appropriately placed in the context of prior and existing research.
- All sources used are properly disclosed (correct citation). Literally copying text must be indicated as such by using quotation marks and giving proper reference.
- All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

Consent to participate

The study uses data from secondary data sources not involving any human subjects.

References

- Atzberger, C. (2013). Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sens* 5(2):949–981
- Bala, S. K., Islam, A. S. (2009). Correlation between potato yield and MODIS-derived vegetation indices. *Int J Remote Sens* 30(10): 2491–2507
- BBS (2020). Yearbook of Agricultural Statistics-2020, 32nd series Statistics and Information Division, Ministry of Planning, Dhaka, Bangladesh
- Belgiu, M., and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment*, 204, 509–523.
- Congedo, L. (2016). Semi-Automatic Classification Plugin Documentation. https://fromgistors.blogspot.com/p/semi-automatic-classification-plugin.html (retrieved 5/12/2016).
- Croitoru, A.-E., Holobaca, I.-H., Lazar, C., Moldovan, F., & Imbroane, A. (2012). Air temperature trend and the impact on winter wheat phenology in Romania. *Climatic Change*, 111, 393–410.
- Islam, A. K. M. S. and Bala, S. K. (2011). Estimation of yield of wheat in greater dinajpur region using MODIS data. 3rd International Conference on Water & Flood Management (ICWFM-2011).
- Kim, H. H. and Elman, G. C. (1990). Normalization of satellite imagery. *Int J Remote Sens* 11(8):1331–1347.
- Lai, Y. R.; Pringle, M. J.; Kopittke, P. M.; Menzies, N. W.; Orton, T. G. and Dang, Y. P. (2018). An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *Int. J. Appl. Earth Obs. Geoinf*. **2018**, 72, 99–108.
- Lambert, M. J.; Traoré, P. C. S.; Blaes, X.; Baret, P. and Defourny, P. (2018). Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens*. Environ. **2018**, 216, 647–657.
- Li, J.; Roy, D.P. (2017). A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. *Remote Sens*. **2017**, 9, 902.
- Lima, T. A.; Beuchle, R.; Langner, A.; Grecchi, R. C.; Griess, V. C. and Achard, F. (2019). Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian

Amazon. *Remote Sens*. 2019, 11, 961p. 1-21; doi:10.3390/rs11080961

- Liu, W. T. and Kogan, F. (2002). Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. *Int J Remote Sens* 23(6):1161–1179
- Masson-Delmotte, V., Zhai, P., Po¨rtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Pe´an, C., Pidcock, R. (2018). Global warming of 1.5 OC: an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization Geneva, Switzerland.
- Newton, I. H, Islam, A. F. M T., Islam, A. K. M. S., Tarekul, G. M., Anika, T. and Razzaque, S. (2018). Yield Prediction Model for Potato Using Landsat Time Series Images Driven Vegetation Indices. *Remote Sensing in Earth Systems Sciences* (2018) 1:29–38.
- Panek, E., and Gozdowski, D. (2021). Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries. *Agronomy* 2021, 11, 340. https://doi.org/10.3390/agronomy11020340
- Rahman, M. M. and Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 Time Series Data for Yield Prediction of Sugarcane Crops at the Block Level. *Remote Sens.* 2020, 12, 1313; doi:10.3390/rs12081313.
- Rahman. A., Khan, K., Krakauer, N. Y., Roytman, L. and Kogan, F. (2012). Using AVHRR-based vegetation health indices for estimation of potato yield in Bangladesh. *J Civil Environ Eng* 2:111
- Ren, J., Chen, Z., Zhou, Q., and Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI datain Shandong, China. *International Journal of Applied Earth Observation and gioinformation*, 10(4), 403-413.resolution earth observations: The Global Agriculture Monitoring (GLAM) project.
- Reynolds, C. A., Yitayew, M. and Slack, D. C. (2000). Estimating crop yields and production by integrating the FAO crop specific water balance model with real-time satellite data and ground-based ancillary data. *Int J Remote Sens* 21(18):3487–3508 *Sens* 2: 1589-1609.
- Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In: Proceedings of the ERTS-1 3rd symposium, vol 1. NASA SP-351. NASA, Washington, pp 309–317.
- Segarra, J.; Buchaillot, M. L.; Jose Luis Araus, J. L. and Kefauver, S. C. (2020). Remote sensing for precision agriculture: sentinel-2 improved features and applications. *Agronomy* 2020, 10, 641; doi:10.3390/agronomy10050641.
- Simonetti, D.; Marelli, A.; Eva, H. (2015). IMPACTool Box: Portable GIS Toolbox for Image Processing and Land Cover Mapping; Publications Oce of the European Union: Luxembourg, 2015; ISBN 978-92-79-50115-9.
- Skakun, S., Vermote, E., Franch, B., Roger, J. C., Kussul, N., Ju, J. and Masek, J. (2019). Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models. *Remote Sens*. 2019, 11, 1768:1-19; doi:10.3390/rs11151768. www.mdpi.com/journal/remotesensing.
- Teillet, P. M. (1986). Image correction for radiometric effects in remote sensing. *Int J Remote Sens* 7(12):1637–1651.
- Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ*. 1979, 8, 127–150.
- USGS. Earth Explorer. (2016). Available online: http://earthexplorer.usgs.gov/ (accessed on 20 February 2019).

Declarations

Funding: This research received fundings from Bangabandhu Science and Technology Fellowship Trust, Government of the People's Republic of Bangladesh.

Potential competing interests: No potential competing interests to declare.