

To determine the variety of Potato crop using Time series analysis and NDVI Index by implementing CCDC Equation



Rochitansh Singh

Supervisor: **Prof. S. Sundar**
and
Nikhil Sasi Rajan and Vaibhav Chippa
from



Department of Mathematics
Indian Institute of Technology Madras

This dissertation is submitted for the degree of
Master of Technology

Dated: 6 June, 2022

Certificate

This is to certify that the project report titled "**To determine the variety of Potato crop using Time series analysis and NDVI Index by implementing CCDC Equation**" has been carried out by **Mr. Rochitansh Singh (MA20M018)**, at the Department of Mathematics, Indian Institute of Technology Madras, for the partial fulfilment of the degree of **Master of Technology (Industrial Mathematics and Scientific Computing)**, and has been carried out under my guidance and supervision, during the academic session 2021 - 22. This is an original work and has not been presented elsewhere for any degree as far as my concern.

Prof. S. Sundar
Head of Department
Department of Mathematics

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Rochitansh Singh

27 May, 2022

Acknowledgements

I would like to thank my supervisor **Prof. S. Sundar**. I would also like to thank **Mr. Nikhil Shasirajan** and **Vaibhav Chippa** from Cropin technology solutions for their constant help and support, without which the completion of my project for this academic year would be impossible. They have always been there to guide me throughout the course and have patiently solved all the queried associated with the project. Lastly, I would like to thank the Department of Mathematics, Indian Institute of Technology Madras and Cropin for providing me the opportunity to work in this exciting field.

Abstract

Crops growth and yield monitoring in agricultural areas is a critical method for ensuring food, nutritional security and projecting agricultural economic advantages. Potatoes (*Solanum Tubersum L.*) are indeed the world's largest non-cereal food crop. Evaluating end-of-season tuber harvest using in-season data can help farmers make more productive decisions while minimizing impacts on the environment. Earlier researches has already shown that utilizing Landsat images from two distinct years with two different phenological stages to detect changes yields unsatisfactory results and can result into many misclassifications. New algorithm for continuous change detection and classification (CCDC)^[1] of land cover using all available Landsat data is being developed. Many types of land cover changes can be continuously recognized as new images collected and provides a land cover map at any time. So CCDC equation is being implemented in identifying the variety of potato provided the Sowing Date and Harvesting Date is given. For greater generality, we have acknowledged that distinct indicators and entities may also have different starting and ending points for different time spans. Difference Variety of Potato Crops are Sown and harvested in different seasons and their maximum productivity also varies from variety to variety. Since the dataset has data points of different lengths, So instead of re-establishing those data points by any mean value or zero, We are calculating the fixed number of coefficients for each time series data using CCDC equation eliminating the drawbacks of the ability to distinguish between objects due to reduction in the length of the available data.

Table of contents

1.	Introduction	12
1.1.	General	
1.2.	Monitoring land cover changes with remote sensing	13
1.2.1.	What is remote sensing?	13
1.2.2.	Orbits	13
1.2.3.	Electromagnetic spectrum	15
1.2.4.	Sensors	16
1.2.5.	Resolution	17
1.3.	Variety of Potatoes	20
1.3.1.	Sanatana Plain	20
1.3.2.	Bafana potatoes	20
1.3.3.	kennebac potatoes	20
1.3.4.	kufri frysona	21
1.3.5.	Innovator potato	21
1.3.6.	Sarpo mira	22
1.3.7.	Santana potato	23
2.	Implementation	24
2.1.	CCDC equation	24
2.2.	NDVI Value	25
2.2.1.	How NDVI is calculated?	
2.2.2.	Example of NDVI in agriculture	
2.2.3.	Where do we use NDVI?	
2.2.4.	Which satellite imagery have near infra red for NDVI?	
2.3.	Analysis Ready Data	30
3.	Data and processing	31
3.1.	Dataset	31
3.2.	Processing	32
4.	Results and conclusions	36
4.1.	Result	36
4.2.	Conclusions	37

Chapter 1

Introduction

1.1 General

Geographic information can be represented in the form of paper maps all around the world. It's widely known that far more maps were already available in digital format. Computer map updates is a critical undertaking that has been pending for almost three decades. Satellites with optical or microwave sensors, active or passive, and high or low resolution have all been launched into orbit recently. Satellite pictures are a useful source of information for describing urban land use/cover types and their changes across time and place. The process of detecting alterations in the status of an object or phenomenon is known as change detection. Because many distinct physical qualities change as a function of land cover, including as albedo, emissivity, roughness, photosynthetic capacity, and transpiration, land cover has an impact on the balance of energy, carbon budget, and hydrological cycle^[2]. Land cover change can be natural or artificial, however with increased human activity, numerous types of land cover change have drastically altered the Earth's surface in recent years. Modeling the climate and biogeochemistry of the Earth system, as well as many types of management, require knowledge about land cover and land cover change. It is critical that we increase our understanding of the levels, varieties, origins, and repercussions of the changes occurring so that scientists, resource managers, and other decision-makers may respond effectively.

1.2 Monitoring land cover changes with Remote Sensing

1.2.1 What is Remote Sensing?

The practise of identifying and monitoring an area's unique structure and physical attributes by monitoring its reflected and emitted radiation from a distance is known as remote sensing (typically from satellite or aircraft). Here are several examples:

- Satellites and aeroplanes use cameras to photograph broad sections of the Planet's surface, enabling us to see far more than we can from the ground.
- Shipboard sonar systems can be used to create photographs of the sea bed without having to go to the bottom.
- Images of temperature variations in the waters can be captured using satellite cameras.

The following are some examples of applications for remotely sensed images of the Earth:

- Massive forest fires can indeed be monitored from space, giving rangers a much wider view than they would have on the ground.
- Watching erupting volcanoes and tracking cloud cover to help make accurate predictions.
- Over several years or decades, following the rise of a city and changes in agriculture or forests.
- Discovery and charting of the ocean floor's severe terrain (e.g., vast mountain ranges, deep gorges, and "magnetic striping").

Remote sensing is ideal for defining land cover change information to assess land changes throughout all scales, from local to global. Various regional and worldwide land use / cover mapping solutions have been generated as technology advanced and the cost of satellite data has decreased. The United States Geological Survey (USGS) has an ancient legacy of classifying landscape change using moderate spatial resolution remote sensing data (such as Landsat Satellite data, the Moderate Resolution Imaging Spectroradiometer, and the Advanced Resolution Radiometer) to assist regional and national evaluations for both systems and management concerning the extent, position, magnitude, influence, and trajectory of land cover change.

1.2.2 Orbits

Satellites could be deployed into various orbits around the Earth. Low-Earth orbit (about 160 to 2,000 km above Earth), medium-Earth orbit (approximately 2,000 to 35,500 km above Earth), and high-Earth orbit (above 35,500 km above Earth) are the three most prevalent orbit classifications .

Satellites orbiting at 35,786 km are at an altitude at which their orbital speed matches the planet's rotation, and are in what is called *geosynchronous orbit* (GSO). In addition, a satellite in GSO directly over the equator will have a *geostationary orbit*. A geostationary

orbit enables a satellite to maintain its position directly over the same place on Earth's surface. Because satellites can follow multiple orbital courses around the planet, low-Earth orbit is a common choice. Polar-orbiting satellites, for example, go from pole to pole as Earth rotates at roughly 90 degrees to the equatorial plane. This allows the satellite's sensors to quickly collect data for the whole world, including the polar areas.

A medium-Earth orbit satellite estimate takes 12 hours to complete an orbit. The satellite crosses over the same two spots on the equator every day in 24 hours. This orbit is highly predictable because it is consistent. Hence, this orbit is used by many telecommunications and GPS satellites.

While both geosynchronous and geostationary satellites revolve around the orbit at 35,786 km above Earth, geosynchronous satellites have orbits that can be tilted above or below the equator. On the other hand, Geostationary satellites orbit around the Earth on the same plane as the equator. Hence, These satellites capture identical views of Earth with each observation and provide almost continuous coverage of one area.

1.2.3 Electromagnetic Spectrum

EMR is a dynamic form of energy that propagates as wave motion at a velocity of $c = 3 \times 10^{10}$ cm/sec. The parameters that characterize a wave motion are wavelength (λ), frequency (v) and velocity (c) (Fig. 2). The relationship between the above is $c = v\lambda$.

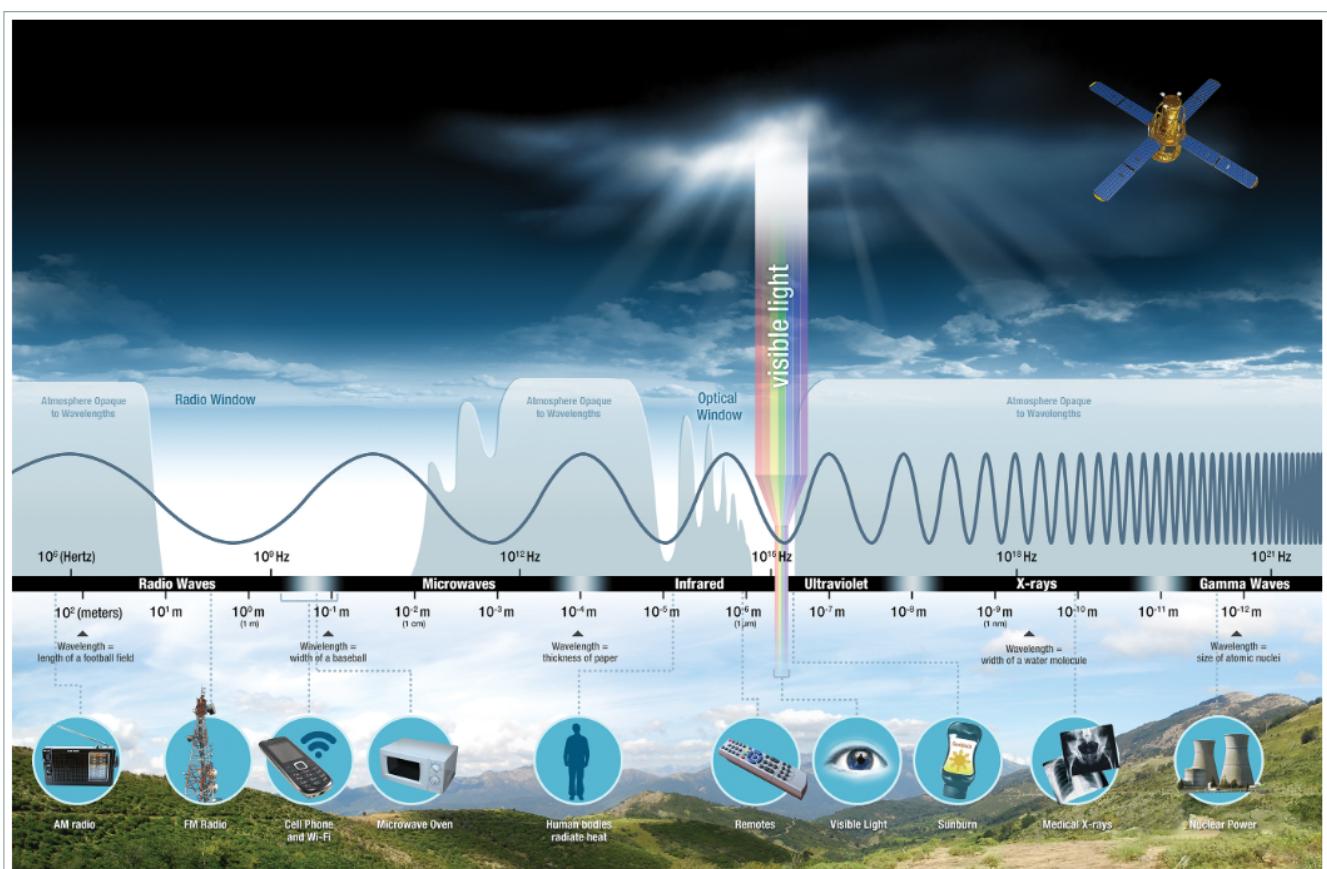


Fig 1: Distribution of electromagnetic spectrum with its wavelengths and applications

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

Some waves are absorbed or reflected by atmospheric components, like water vapor and carbon dioxide, while some wavelengths allow for unimpeded movement through the atmosphere; visible light has wavelengths that can be transmitted through the atmosphere. Microwave energy has wavelengths that can pass through clouds, an attribute utilized by many weather and communication satellites.

Some waves are either absorbed or reflected by atmospheric constituents such as water vapour ,carbon dioxide and other micro elements, while others can pass through the atmosphere unobstructed. Visible light has wavelengths that can travel through the atmosphere. Many meteorological and telecommunication satellites use microwave energy because its wavelengths can flow through clouds. The Sun is the principal source of energy detected by satellites. The amount of solar energy reflected is determined by the surface roughness and its albedo (the ability of a surface to reflect light rather than absorb it). Snow, for instance, has a high albedo and can reflect up to 90% of incident solar radiation. The ocean, on the other hand, merely reflects about 6% of solar radiation and absorb the remaining. When energy is absorbed, it is frequently re-emitted at longer wavelengths. The energy received by the ocean, for example, is re-emitted as infrared radiation.

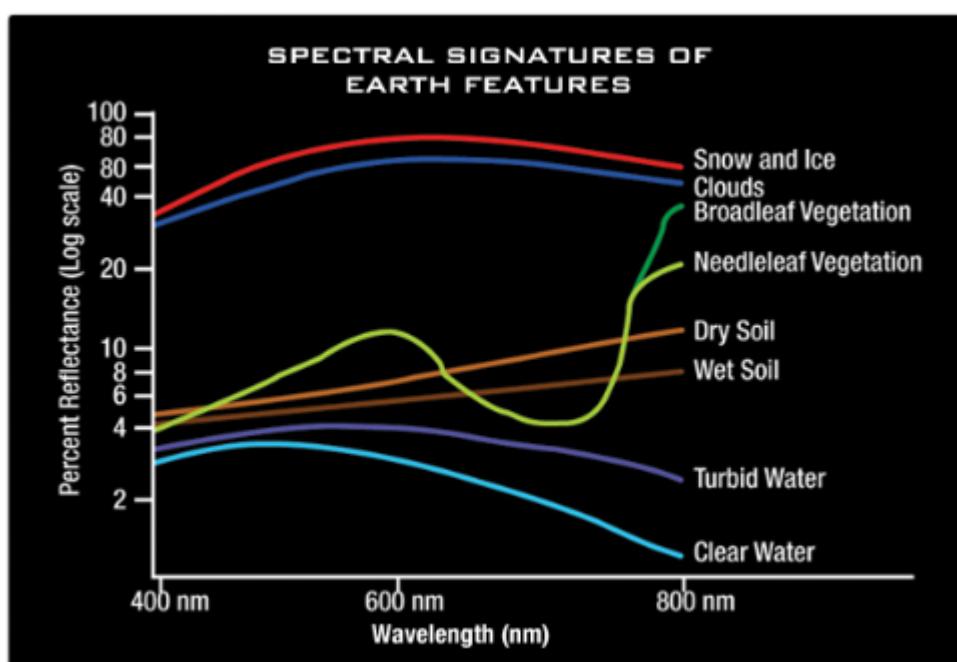


Figure 2: Spectral signatures of different Earth features within the visible light spectrum.

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

1.2.4 Sensors

Sensors, the ones placed on satellites or planes utilizes the illuminating energy of the Sun and delivering their own illumination source by measuring the energy that is reflected back.

There are two types of sensors:

- 1) **Passive sensors:** Sensors that use natural energy from the sun. Example: Radiometers (instruments quantitatively measuring the intensity of electromagnetic radiation in selected bands) and Spectrometers (devices that are designed to detect, measure, and analyze the spectral content of reflected electromagnetic radiation). These sensors work on the visible, infrared, thermal infrared, and microwave portions of the electromagnetic spectrum and cannot penetrate through the dense clouds.
- 2) **Active sensors:** Sensors that provide their own energy. Examples include different types of radio detection and ranging (radar) sensors, altimeters, and scatterometers. They work in the microwave region of the spectrum and hence can penetrate through the atmosphere under most conditions.



Figure 3: pictorial representation of active and passive sensors

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

1.2.5 Resolution

Resolution, depending on orbit and sensor design, determines how data from a sensor can be utilized. There are four types of resolution to consider for any dataset—radiometric, spatial, spectral, and temporal.

- 1) **Radiometric Resolution:** The amount of information in each pixel, or the number of bits indicating the energy captured, is referred to as radiometric resolution. Each bit represents an exponent of power 2. An 8 bit resolution, for example, is 2⁸, indicating that the sensor can register 256 distinct digital values (0-255). As a result, the higher the radiometric resolution, more and more values are large enough to hold information, allowing for improved detection of even minor energy changes. For example, to differentiate between tiny variances in ocean hue, radiometric resolution is required for analysing water quality.

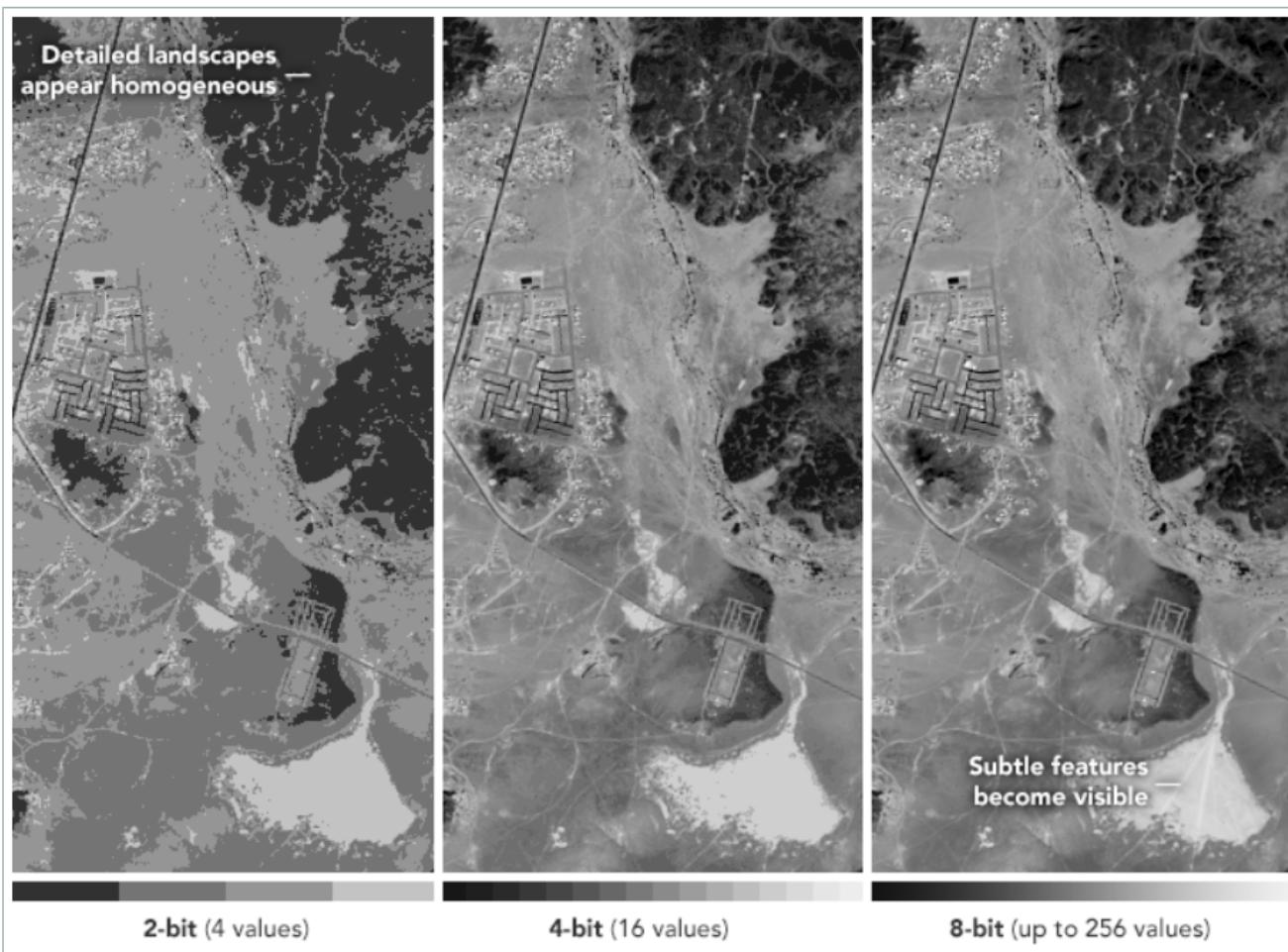


Figure 4: Radiometric resolution

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

- 2) **Spatial Resolution:** The size of each pixel in a digital image and the area of Earth's surface represented by that pixel constitute spatial resolution. The majority of bands detected by the Moderate Resolution Imaging Spectroradiometer (MODIS), for example, have a spatial resolution of 1 km; each pixel represents a 1 km x 1 km area on the earth.



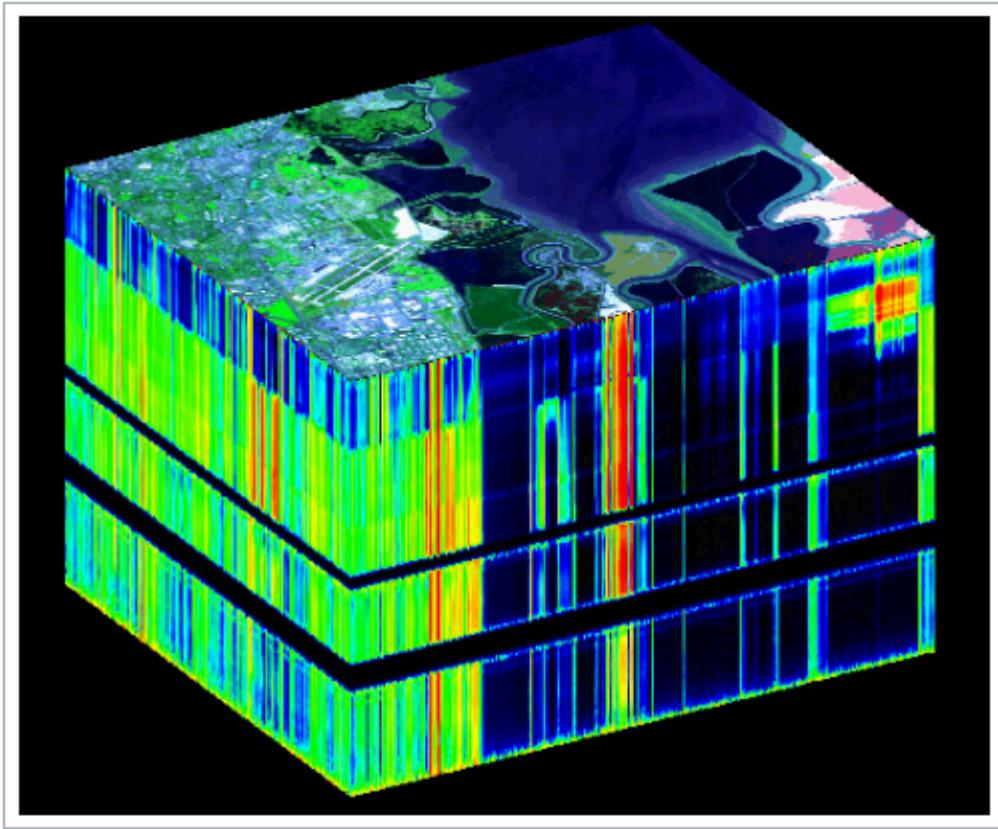
Landsat 8 image of Reykjavik, Iceland, acquired July 7, 2019, illustrating the difference in pixel resolution. Credit: NASA Earth Observatory.

Figure 5: Spatial resolution

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

- 3) **Spectral Resolution:** This refers to a sensor's capacity to distinguish finer wavelengths, or having more and narrower bands. Many sensors are multispectral, indicating they have three to ten bands.



The top of the cube is a false-color image made to accentuate the structure in the water and evaporation ponds on the right. The sides of the cube are slices showing the edges of the top in all 224 of the AVIRIS spectral channels. The tops of the sides are in the visible part of the spectrum (wavelength of 400 nanometers), and the bottoms are in the infrared (2,500 nanometers). Credit: NASA JPL.

Figure 6: Representation of Spectral resolution

Source:

<https://earthdata.nasa.gov/learn/backgrounders/remote-sensing#observing-with-the-electromagnetic-spectrum>

- 4) **Temporal Resolution:** The time a satellite takes to complete an orbit and return to the same monitoring region is known as temporal resolution. The orbit, sensor parameters, and swath width significantly determine the resolution. The temporal resolution is substantially finer since geostationary satellites rotate at the same rate as the Earth. The temporal resolution of polar orbiting satellites can range from 1 to 16 days.

1.3 Variety of Potatoes

1.3.1 Santana Plain

- Table potatoes Santana do not boil and remain dense under heat treatment, thus they are used to make chips and fries.
- Santana's production varies depending on planting and care circumstances, but it can reach 419 kg per hectare. The average yield is between 164 and 384 c/h.
- Santana potatoes are considered mid-season - their tubers can be harvested 80-95 days after planting

1.3.2 Bafana Potatoes

- They become fully mature in 115-120 days after there germination.
- They are seeded in may under the depth of 8-10 cm.
- Keep the soil in a loose state and systematically destroying the weeds

1.3.3 Kennebac Potatoes

- Kennebec potatoes are medium to large in size and are long and ovate in shape with rounded ends.
- Kennebec potatoes are available year-round, with peak season in the late summer through early winter.
- Kennebec potatoes contain potassium, niacin, riboflavin, thiamin, vitamin C, and fiber.



Figure 7: Kennebac Potatoes

1.3.4 Kufri frysona

- India lacks a suitable potato variety for French fry production. The variety Kufri Frysona has been developed to fill this gap to provide suitable raw material to French fry industry.
- Kufri Frysona gives tuber yields higher than the presently used Indian processing varieties Kufri Chipsona-1 and Kufri Surya.
- It also excels them in giving high proportion of oblong to long French fry grade tubers and better French fries in taste, texture and colour.
- Kufri Frysona possesses high field resistance to late blight disease and keeps well under country storage conditions for over 8 weeks



Figure 8: Kufri frysona

1.3.5 Innovator Potato

- Innovator is potato variety that is oblong in shape with a smooth skin. It is a popular potato variety in Europe and is gaining popularity in North America as a frying and baking potato.
- The skin of the potato variety is russeted, similar to that of a Russet Burbank Potato.

- Innovator also has shallow eyes with a cream coloured flesh



Figure 9: Innovator Potato

1.3.6 Sarpo Mira

- In built dormancy. This means that as long as they are kept at relatively low temperatures and in darkness, they can be harvested and stored without refrigeration or anti sprouting treatments for considerable periods
- Slow to chit. As a result of the inbuilt dormancy that allows easy over winter storage, Sarpo varieties do not chit as fast as others. It is still advantageous to do this but only expect short spouts to appear before it is time to plant
- Full foliage early in the season. Sarpo potato varieties tend to be very vigorous. This results in the crop “closing cover” very quickly, stopping almost all weed competition. This vigorous growth results in sometimes very large yields.



Figure 10: Sarpo Mira

1.3.7 Santana Potato

- **Maturity:** (90-110 days)
- **Yield potential:** 30-40 t/ha
- **Shape:** Elongated Oblong
- **Size:** 45mm to 90mm
- **Skin:** Smooth Yellow
- **Flesh:** Pale Yellow
- **Availability (Harvest Period):** January Ending to March
- **Availability (CIPC -Stored in Cold Storage):** February to September with sugar level controlled.
- **Processing Quality:** Boiling, Baking and Chipping

Chapter 2

Implementation

2.1 CCDC Equation

For a given research area, CCDC employs all cloud-masked Landsat surface reflectance data. Zhu and Woodcock discuss the original implementation in detail (2014). CCDC is a generalised technique that can be used to track various sorts of land change. As a result, it does not filter changes based on specific spectral directional changes, nor does it rely on a single spectral band or index.

CCDC is made up of two parts: a change detection component and a classification component. All available Landsat imagery and a user-defined collection of spectral bands or indices are used to detect changes. For change detection, at least the Green, Red, NIR, SWIR1, and SWIR2 bands are typically employed.^[3]

$$\hat{p}(i, t) = c_{0i} + c_{1i}t + \sum_{n=1}^3 \left(a_{ni} \cos \frac{2\pi nt}{T} + b_{ni} \sin \frac{2\pi nt}{T} \right)$$

where t is the ordinal date where January 1 of the year 1 has ordinal 1 (proleptic Gregorian calendar)

i is the ith Landsat band. Since we are working for only one band So it can be neglected

T denotes the average number of days per year, 365.2425

a_{ni} , b_{ni} represents the estimated nth order seasonal harmonic coefficients for the ith Landsat band

c_{0i} , c_{1i} represents the estimated intercept and slope coefficients for the ith Landsat band

and $p(i, t)$ denotes the predicted value for the ith Landsat band at ordinal date t

2.2 NDVI Value

NDVI stands for Normalized Difference Vegetation Index. It measures the difference between near-infrared (which plant strongly reflects) and red light (which vegetation absorbs) to quantify vegetation. Range of NDVI value is between -1 and +1. However, each form of land cover does not have its own distinct boundary.

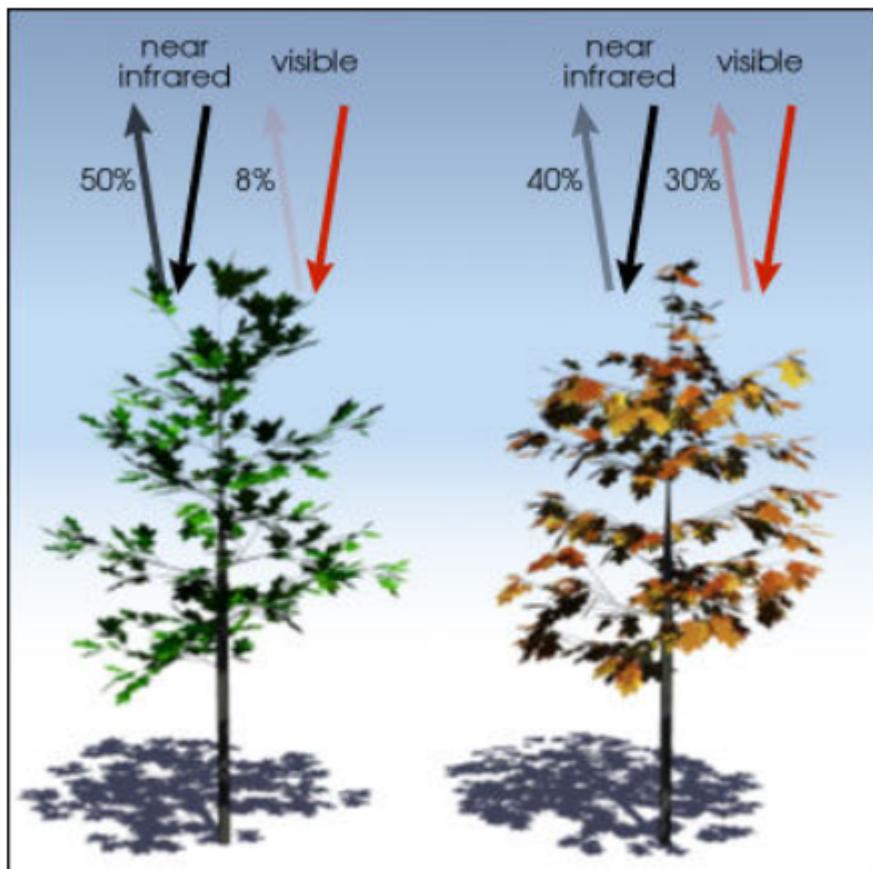
When you have negative values, for example, it's almost certainly water. If your NDVI number is near to +1, though, it's most likely dense green foliage. When the NDVI is close to zero, however, there are no green leaves and the area may be urbanised.

2.2.1 How NDVI is calculated?

NDVI formula works on the electromagnetic spectrum of Near Infra Red and Red bands. The formula goes by

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

When compared to other wavelengths, healthy vegetation (chlorophyll) reflects more NIR and green light. However, more red and blue light is absorbed. This is why vegetation seems green to human sight. It would be strong for vegetation as well if you could see near-infrared. Satellite sensors such as Landsat and Sentinel-2 have the appropriate NIR and red bands.



$$\frac{(0.50 - 0.08)}{(0.50 + 0.08)} = 0.72$$

$$\frac{(0.4 - 0.30)}{(0.4 + 0.30)} = 0.14$$

Image courtesy of NASA.

Figure 12: Explanation of ndvi

Source:

<https://gisgeography.com/ndvi-normalized-difference-vegetation-index/#:~:text=We%20see%20several%20sectors%20using,a%20good%20indicator%20of%20drought>.

The result of this formula generates a value between -1 and +1. If you have low reflectance (or low values) in the red channel and high reflectance in the NIR channel, this will yield a high NDVI value. And vice versa.

Overall, NDVI is a standardized way to measure healthy vegetation. When you have high NDVI values, you have healthier vegetation. When you have low NDVI, you have less or no vegetation. Generally, if you want to see vegetation change over time, then you will have to perform atmospheric conditions.

2.2.2 Example of NDVI in agriculture

Let's look at the NDVI for a farm with centre pivot irrigation. Pivot irrigation creates a circular crop pattern by rotating on a point.

Here's how the red, green, and blue bands appear in full colour. True colour refers to the way our eyes see colour. Hence the below represents the true colors image.



Figure 13: Satellite image seen from the eye (real image)

Source:

<https://gisgeography.com/ndvi-normalized-difference-vegetation-index/#:~:text=We%20see%20several%20sectors%20using,a%20good%20indicator%20of%20drought>.

You can see how NDVI uses near-infrared(NIR) light in the formula . As a result, when we set the NIR band to red, we obtain colour infrared. Near-infrared is in the red channel, hence we call it colour infrared. The pivot irrigation vegetation, as seen below, should already be screaming at you in bright red!

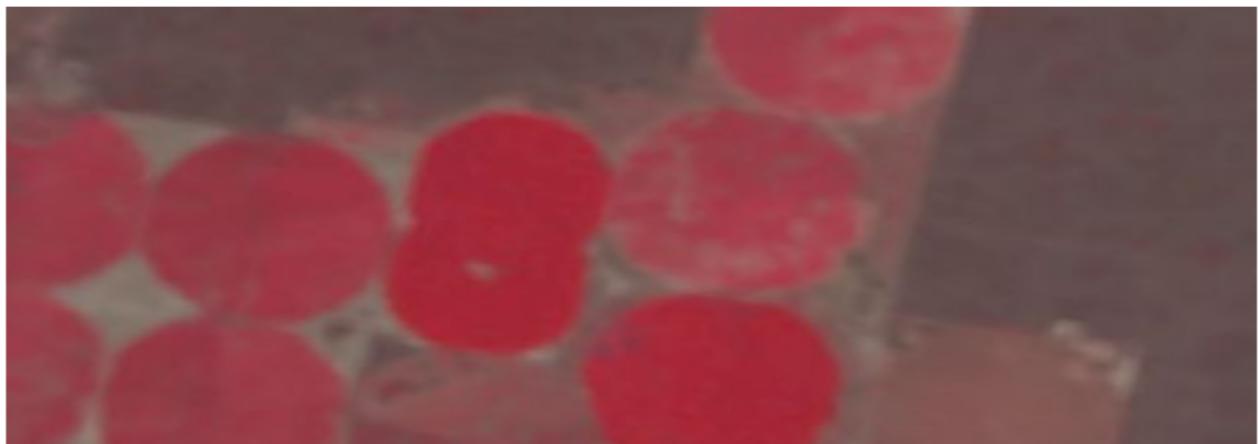


Figure 14: Representation in Infra-red

Source:

<https://gisgeography.com/ndvi-normalized-difference-vegetation-index/#:~:text=We%20see%20several%20sectors%20using,a%20good%20indicator%20of%20drought.>

Bright green indicates a high NDVI when using the formula. The NDVI of red is low. So it's measuring the difference between near-infrared (which foliage heavily reflects) and red light to quantify vegetation (which vegetation absorbs).

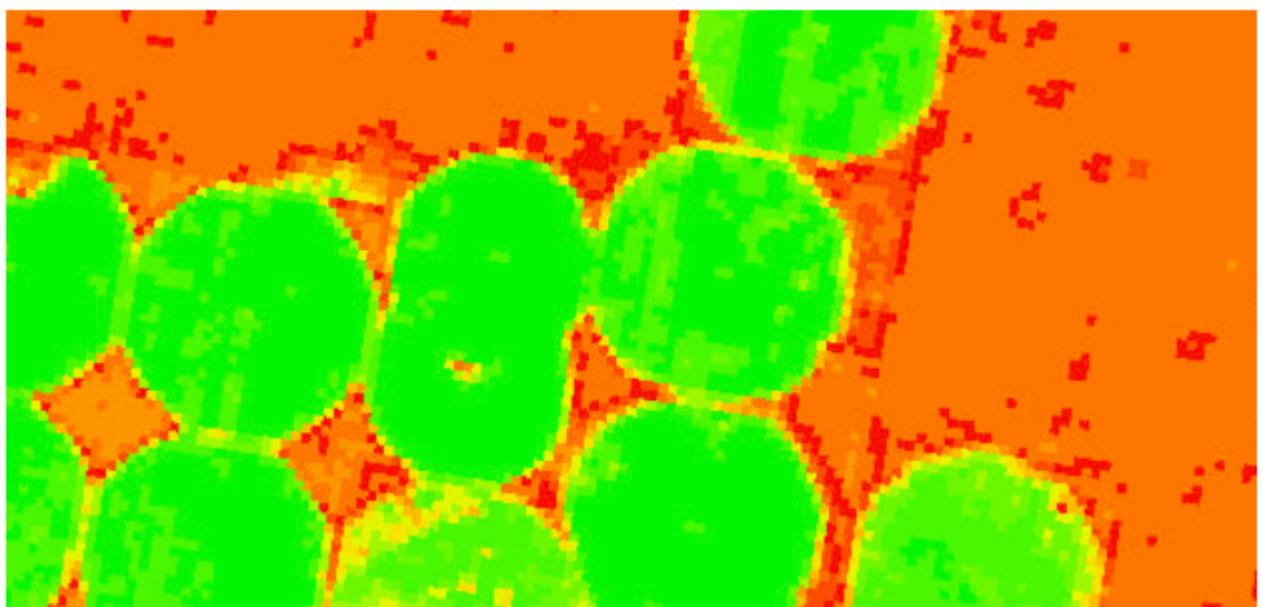


Figure 15: Image after application of the difference formula

Source:

<https://gisgeography.com/ndvi-normalized-difference-vegetation-index/#:~:text=We%20see%20several%20sectors%20using,a%20good%20indicator%20of%20drought.>

2.2.3 Where do we use NDVI?

NDVI is used in numerous industries. Farmers, for example, employ NDVI for precision farming and biomass measurement in agriculture.

Foresters utilise NDVI to quantify forest supply and leaf area index in forestry. NDVI is also a good indicator of drought, according to NASA. When water restricts vegetation growth, the relative NDVI and density of vegetation are reduced.

In actuality, NDVI and other remote sensing technologies are used in a wide range of applications in the real world.

2.2.4 Which satellite imagery have Near Infra-red for NDVI?

- 1) **Sentinel 1:** With the objectives of Land and Ocean monitoring, Sentinel-1 will be composed of two polar-orbiting satellites operating day and night, and will perform Radar imaging, enabling them to acquire imagery regardless of the weather. The first Sentinel-1 satellite was launched in April 2014.
- 2) **Sentinel 2:** The objective of Sentinel-2 is land monitoring, and the mission will be composed of two polar-orbiting satellites providing high-resolution optical imagery. Vegetation, soil and coastal areas are among the monitoring objectives. The first Sentinel-2 satellite was launched in June 2015.

Spectral bands of sentinel 2A are:

Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Ultra blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Figure 16: Spectral bands of sentinel 2A

It's temporal resolution is 10 days and has 13 different bands.

- 3) **Landsat:** Nine Landsat satellites have been launched. One of them failed to reach orbit. The other eight satellites, however, did not. The most current mission, Landsat 9, was launched in September of 2021. The Landsat Data Continuity Mission is another name for Landsat (LDCM). This is due to a commitment to ensure Landsat's long-term viability.

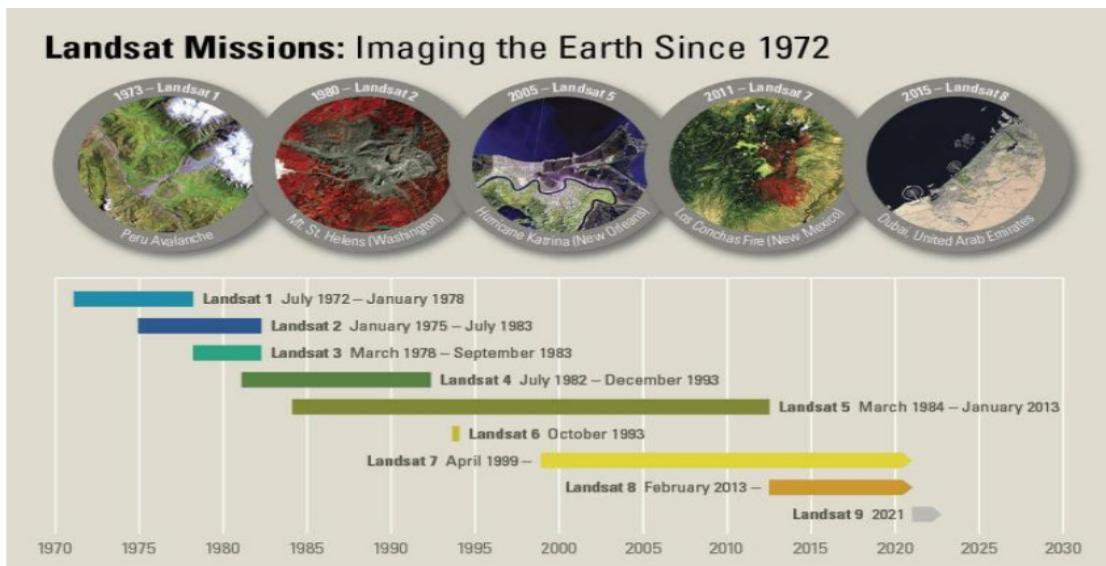


Figure 17: All Landsat 1-9 satellites launch dates.

Below represents the Information of two recent Landsat satellites

A. Landsat 8:

Launch Date	September, 2021
Sensor 1	Operational Land Imager 2 (OLI-2)
Sensor 2	Thermal Infrared Sensor 2 (TIRS-2)
Resolution	15m, 30m and 100m
Bands	Coastal, Blue, Green, Red, NIR1, SWIR-1, SWIR-2, Cirrus, Panchromatic, TIR-1, TIR-2
Launch Base	Vandenberg Air Force Base

Figure 18: Landsat 8 description

B. Landsat 9:

Launch Date	September, 2021
Sensor 1	Operational Land Imager 2 (OLI-2)
Sensor 2	Thermal Infrared Sensor 2 (TIRS-2)
Resolution	15m, 30m and 100m
Bands	Coastal, Blue, Green, Red, NIR1, SWIR-1, SWIR-2, Cirrus, Panchromatic, TIR-1, TIR-2
Launch Base	Vandenberg Air Force Base

2.3 Analysis Ready Data

The USGS creates Landsat ARD for the CONUS using imagery from the Landsat 4 and 5 TM, Landsat 7 ETM+, and Landsat 8 OLI. These data have been treated to the maximum level of geometric and radiometric quality possible, resulting in data products appropriate for time series analysis.^[3]

Analysis Ready Data (ARD) from the United States are pre-packaged and pre-processed bundles of Landsat data products that make the Landsat archive more accessible and easier to examine, while also reducing the time users spend on data processing for time-series analysis.

Chapter 3

Data and Processing

3.1 Dataset

Data is provided by the company Cropin Technologies. The dataset consist of different species of Potato crop, the date on which they are sown, the day on which they are harvested and the NDVI values of the dates in between. The resolution of NDVI values available between two data points varies from 3 to 8 days.

The variety of potatoes is being kept confidential.

The description of the characteristic of each species given above is already discussed earlier.
The dimensions of the dataset is 223719 x 33.

The number of entries are as follows for each variety of potato

Variety A	:	44313
Variety B	:	41632
Variety C	:	34924
Variety D	:	23912
Variety E	:	19189
Variety F	:	18570
Variety G	:	17798
Variety H	:	17675
Variety I	:	5541
Variety J	:	165

3.2 Processing

What we are trying to infer here is that how well coefficients of the CCDC equation determine the time series.

since the lengths of data points can vary, representation of time series by using fixed set coefficients can be a good way of implementation.

The below graph represents the variation of NDVI over the period of time for variety Variety C. The peak is obtained during the period when the plant is at its maturity.

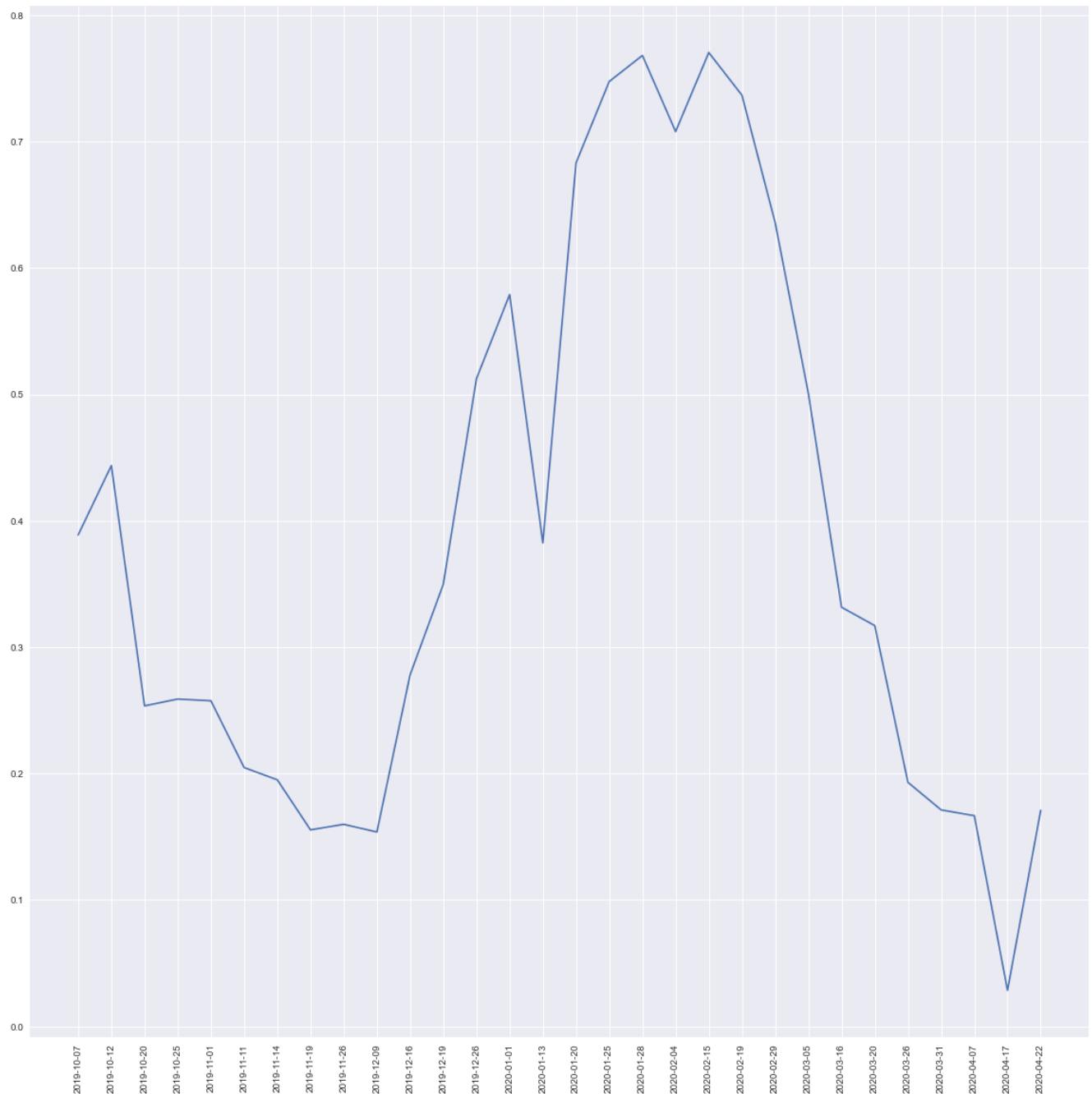


Figure 23: Graph 1



Figure 24: Graph 2

The above graph shows the difference in predicted NDVI value using the coefficients that are found out by implementing CCDC equation and the actual NDVI for the Variety C.

```

pivot1 = 0
pivot2 = 2
from tqdm import tqdm

for pivot1, row in tqdm(df2.iterrows(), total = len(df2)):
    for pivot2 in range(2,len(df2.columns)-1):
        if df2['SowingDate'][pivot1].days <= int(df2.columns.values[pivot2]) and df2['Har_date'][pivot1].days >= int(df2.columns.values[pivot2]):
            popt, _pcov = curve_fit(CCDC, listdateArr, row[2:-1])
            variety = df2["Variety_x"][pivot1]

            c0,c1,a1, b1, a2, b2, a3, b3 = popt
            coeff_dict_new['variety'].append(variety)
            coeff_dict_new['c0'].append(c0)
            coeff_dict_new['c1'].append(c1)
            coeff_dict_new['a1'].append(a1)
            coeff_dict_new['b1'].append(b1)
            coeff_dict_new['a2'].append(a2)
            coeff_dict_new['b2'].append(b2)
            coeff_dict_new['a3'].append(a3)
            coeff_dict_new['b3'].append(b3)

```

Figure 25: Code snippet 4

The value of coefficients are appended in a dictionary. I have converted the timestamps into number of days from a common reference date. One method is to take sowing date as reference date and other is to not take sowing date as reference.

I applied the CCDC algorithm here by not taking Sowing date as reference. Since there are dates before the sowing date and after the harvesting date. So we have to consider only those dates that lie between Sowing date and Harvesting date.

Once the CCDC coefficients are calculated for each row, We are appending that into a dictionary.

We will now predict the value of NDVI for each row by putting back these coefficients into the CCDC equation

to get the generalised curve defining the variation in the value of NDVI over the period of time.

```
pivot1 = 0
pivot2 = 2
from tqdm import tqdm

for pivot1, row in tqdm(df2.iterrows(), total = len(df2)):
    ndvi_list = []
    ndvi_list = list(df2.iloc[pivot1,2:-1])

    for pivot2 in range(2,len(df2.columns)-1):
        if df2['SowingDate'][pivot1].days <= int(df2.columns.values[pivot2]) and df2['Har_date'][pivot1].days >= int(df2.columns.values[pivot2]):
            popt, _pcov = curve_fit(CCDC, listdateArr, row[2:-1])

            variety = df2["Variety_x"][pivot1]

            c0,c1,a1, b1, a2, b2, a3, b3 = popt
            coeff_dict_new['variety'].append(variety)
            coeff_dict_new['c0'].append(c0)
            coeff_dict_new['c1'].append(c1)
            coeff_dict_new['a1'].append(a1)
            coeff_dict_new['b1'].append(b1)
            coeff_dict_new['a2'].append(a2)
            coeff_dict_new['b2'].append(b2)
            coeff_dict_new['a3'].append(a3)
            coeff_dict_new['b3'].append(b3)
            p_ndvi = CCDC(listdateArr, c0,c1,a1, b1, a2, b2, a3, b3)
            plt.figure(figsize = (20,10))
            plt.plot(newDays, ndvi_list, label = 'observed ndvi')
            plt.plot(listdateArr, p_ndvi, label = 'Predicted ndvi')
            plt.legend()
            plt.show()
```

Figure 26: Code snippet 5

This code snippet will generate the 223719 different graphs that shows how CCDC is a good fit for the curve prediction by extracting coefficients of the CCDC equation(as seen in figure 24)

Chapter 4

Result and Conclusions

4.1 Result

We applied the Random forest classifier to upto depth of 2. The results that we find out were astonishing. The accuracy of our model drills out to be 0.31296, that is, 31.3%. Which is way better than the random prediction of the available data between any of the 10 different varieties of potato whose value is $1/(\text{number of varieties})$, that is, $\frac{1}{10} = 0.1$ which is 10%.

```
In [39]: from sklearn.model_selection import train_test_split  
X_train, X_test, Y_train, Y_test = train_test_split(x_train, y_train, test_size=0.33, random_state=42)
```

```
In [40]: from sklearn.ensemble import RandomForestClassifier  
clf = RandomForestClassifier(max_depth=2, random_state=0)  
clf.fit(X_train, Y_train)
```

```
Out[40]: RandomForestClassifier(max_depth=2, random_state=0)
```

```
In [41]: Y_preds=clf.predict(X_test)
```

```
In [42]: from sklearn.metrics import accuracy_score  
accuracy_score(Y_test, Y_preds)
```

```
Out[42]: 0.3129678854290984
```

Figure 27: Code snippet 6

```
In [43]: from sklearn.metrics import mean_squared_error  
from math import sqrt
```

```
#calculate RMSE  
sqrt(mean_squared_error(ndvi_list,p_ndvi))
```

```
Out[43]: 0.06389612081289801
```

Figure 28: Code snippet 7

The **RMSE**(Root MEan Square Error) will be 6.38%.

4.2 Conclusions

This concludes that our implementation of CCDC algorithm to the dataset has enable our model to predict the variety of the unseen data with higher accuracy(increment of nearly 22 percent). Hence we conclude that representing time series as a fixed set of coefficients oF CCDC equation is a better way of representing and processing data than working with the time series of various lengths. This must be noted that the above conclusion holds by taking some individual date in the past as reference(Not taking Sowing date as a reference date).

Chapter 5

References

- [1] Zhe Zhu, Curtis E. Woodcock, Continuous change detection and classification of land cover using all available Landsat data, *Remote Sensing of Environment*, Volume 144, 2014, Pages 152-171, ISSN 0034-4257,
<https://doi.org/10.1016/j.rse.2014.01.011>.
- [2] Coppin, Pol R., and Marvin E. Bauer. "Digital change detection in forest ecosystems with remote sensing imagery." *Remote sensing reviews* 13.3-4 (1996): 207-234.
- [3] Brown, Jesslyn & Tollerud, Heather & Barber, Christopher & Zhou, Qiang & Dwyer, John & Vogelmann, Jim & Loveland, Thomas & Woodcock, Curtis & Stehman, Stephen & Zhu, Zhe & Pengra, Bruce & Smith, Kelcy & Horton, Josephine & Xian, George & Auch, Roger & Sohl, Terry & Sayler, Kristi & Alisa, Luo & Zelenak, Daniel & Rover, Jennifer. (2019). Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Project (LCMAP) approach. *Remote Sensing of Environment*. 238. 10.1016/j.rse.2019.111356.