Preliminary Exploration of Relationships between NDVI, Surface Temperature, Precipitation, and Population in the Ganges-Brahmaputra-Meghna Delta Region

GIS in Water Resources - Fall 2017
Jayaram Hariharan

December 7, 2017

Contents

1 Introduction

1.1 Site Information

The locations being studied are in Bangladesh. Two study sites were chosen: one is centered on the city of Dhaka, the capital of Bangladesh, and the other is located in the Pirojpur District, a rural area.

Study Sites A and B

Figure 1: Locations of Study Sites A and B relative to county of Bangladesh

Landsat data from the satellite Path 137 and Row 44 was collected and analyzed. Both study site locations are within Row 44. So the satellite data collected will be directly comparable between the two locations. See Figure 2 for the locations of these sites relative to Row 44.

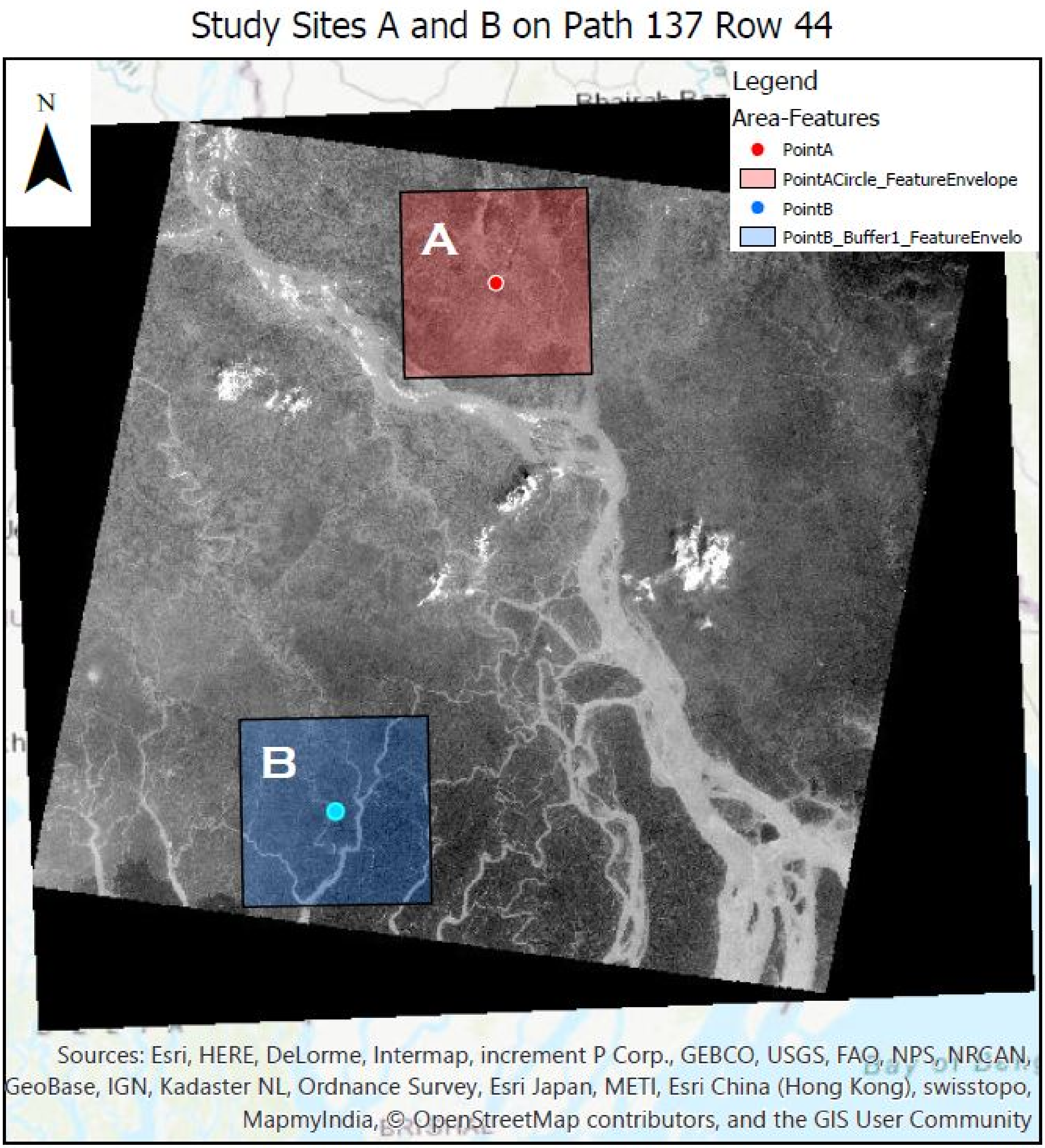

Figure 2: Locations of Study Sites A and B relative to Path 137 Row 44 (Nov. 15, 1993)

1.2 Previous Research

Many researchers have focused on this region of the world in their work (e.g. Kim et al. (2006) Passalacqua et al. (2013) Wilson and Goodbred (2014) & Hoitink et al. (2017)). The Ganges-Brahmaputra-Meghna Delta is one of the largest river deltas in the world, and many people live in the region (Wilson and Goodbred, 2014). As a result it is valuable to know and understand how processes occur and are related to each other in this region.

In their paper, Chaudhuri and Mishra (2016) work to compare the effects of land use changes in two areas of the delta and resulting land surface temperature changes. Their paper outlines the process by which thermal band data from Landsat can be used to estimate the surface temperature. Some of the early work done to generate methods for estimating temperature from Landsat thermal sensors was performed by Malaret et al. (1985), and has later been validated by others such as Anuta et al. (2000).

Researchers have been able to relate NDVI to rainfall with enough confidence to use NDVI to measure precipitation (Birtwistle et al., 2016). However, several of these predictions have been limited to arid regions (Birtwistle et al. (2016) & Yarleque et al. (2009)). This may be due to a stronger and more direct relationship between rain and vegetation growth in drier and arid regions. Tropical climates rarely experience drought conditions and due to their temperature and the abundance of rain, it may be more difficult to find a relationship between NDVI and rainfall. It may also be important in farmed areas to exclude crop areas from NDVI analysis as the harvesting of crops could mask effects of rain on natural vegetation (Fu and Burgher, 2015).

Population growth in Dhaka and general urbanization of Bangladesh has occurred since the country was liberated in 1971 (Hossain, 2008). The "heat island" effect has supposedly already impacted Dhaka and the temperatures there are several degrees warmer than the surrounding areas (Alam and Rabbani, 2007). By analyzing surface temperatures estimated from Landsat data, this temperature difference should be noticeable for study area A. Using both Landsat data and data from the Bangladesh Meteorological Department, Ahmed et al. (2013) were able to show that there has been an increase in temperature in the greater Dhaka area between the years 1989 and 2009. Their Landsat analysis was limited to 3 satellite images, but significant efforts were taken to identify land uses and ensure accurate estimations of land surface temperatures.

1.3 Project Objectives

The overall purpose of the project was to harness ArcGIS Pro to analyze and process satellite data so that comparisons and relationships between the satellite data and other types of data can be made. In this project, the following datasets were collected and compared:

- Normalized difference vegetation index (NDVI)
- Surface temperature estimates
- Precipitation data
- Population data

By using two different study sites, the impacts of land use (urban vs rural) on these datasets can also be compared. The initial expectation was that a relationship would exist between the rainfall data and the satellite derived NDVI and surface temperature estimates. There was also an expectation that the study area around Dhaka, a more urban area, would produced lower NDVI and higher surface temperature values when compared with the study area in the rural Pirojpur District.

2 Methods

2.1 Data Collection

The two study sites shown in Figures 1 are centered upon the coordinates 23.70°' N 90.37° E and 22.58° N 89.97° E for areas A and B respectively. After setting these two points, circular buffers with radii of 15 miles were generated. From these buffers, rectangular envelopes were created and became the study site areas.

2.1.1 Landsat Data

The USGS Earth Explorer tool was used to collect Landsat data (USGS, 2017). In order to generate the NDVI index values and the surface temperature estimates, several types of satellite data were required. Fortunately, the Landsat data is provided by spectral band, where each band represents a range of wavelengths. There is data available from 8 different Landsat satellites at this time. Landsat satellites 1, 2 and 3 lacked a thermal sensor to capture the spectral band ranging from $\sim 10.40 - 12.50\,\mu$m. The Landsat 4 and 5 satellites were launched with a thematic mapper sensor which captured this thermal band.

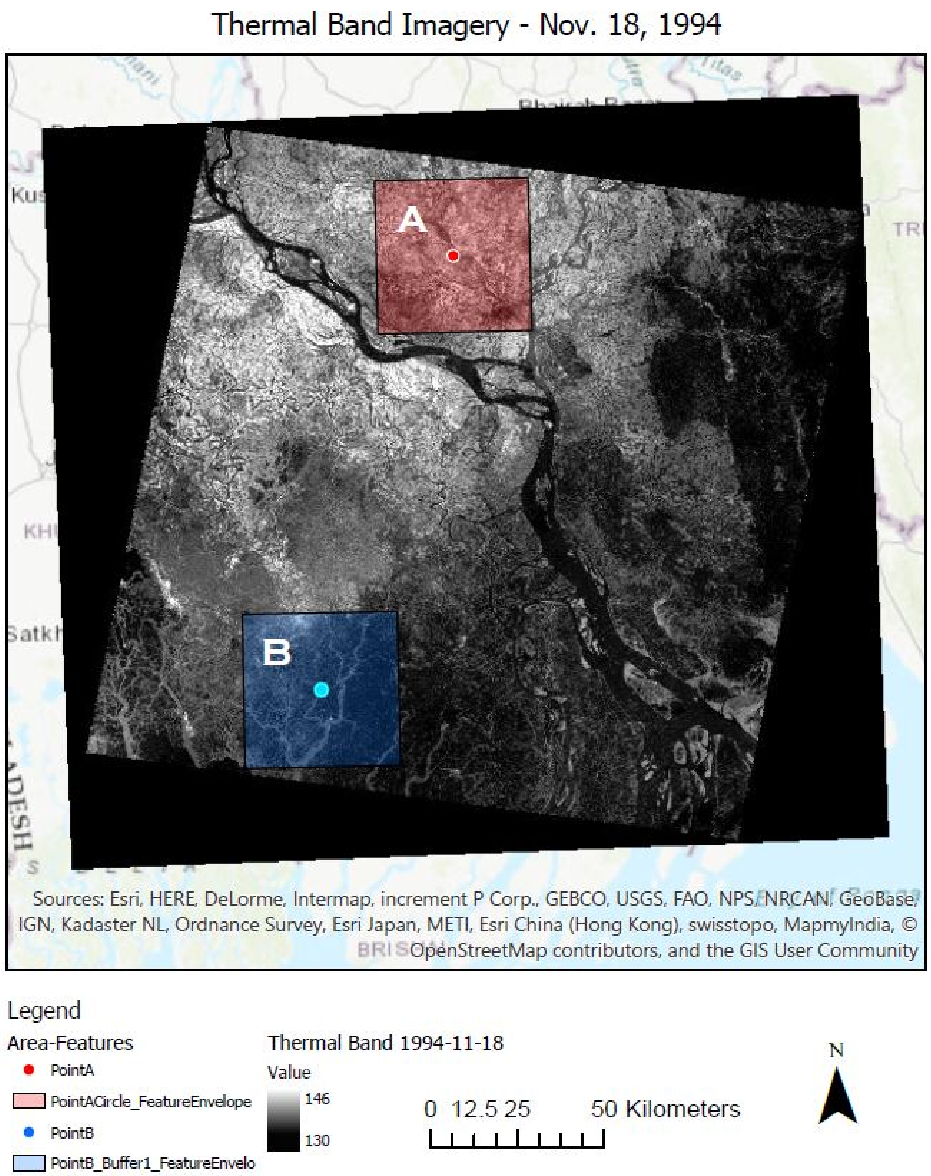

Figure 3: Example of Thermal Band Imagery (Nov. 18, 1994)

Using the information captured by the thematic mapper, the surface temperature can be estimated using procedures outlined in (Chaudhuri and Mishra, 2016). The computation of the NDVI index requires the red visible band and the near infrared band. These two bands were collected by all Landsat missions. For this project, the data collected was from the satellite Path 137 and Row 44.

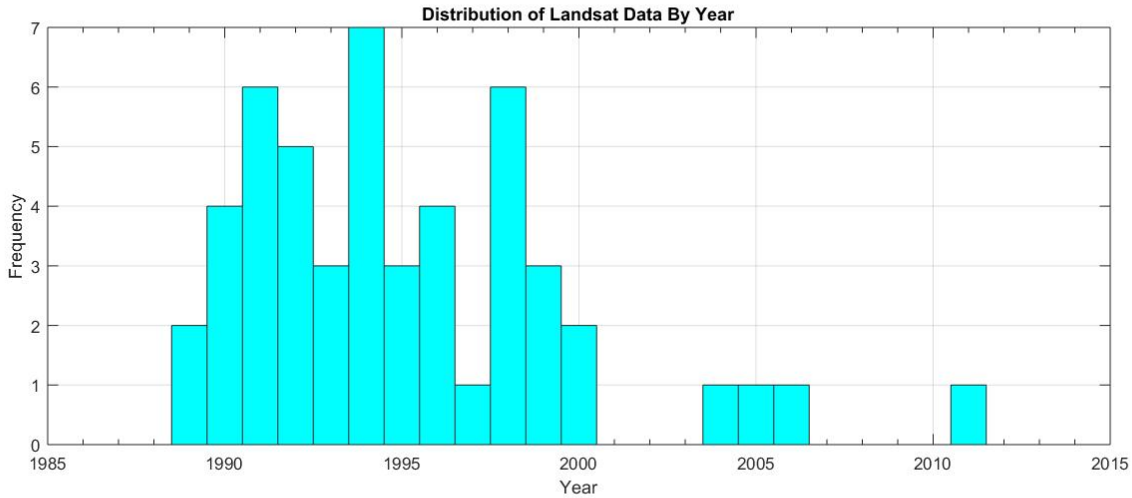

Figure 4: Histogram of Landsat scenes collected by year

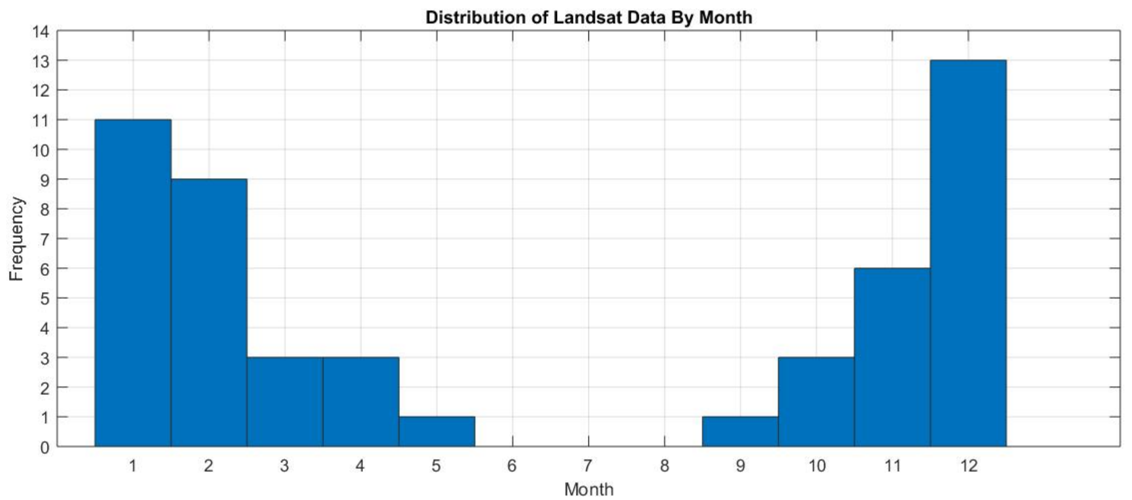

Figure 5: Histogram of Landsat scenes collected by month

Figures 4 and 5 display the distribution of the Landsat scene data collected sorted by year and by month.

Much of the Landsat 7 data is spotty and the quality of the images is lower than what was captured by Landsat 5. In the interest of using higher quality data, and avoiding any transformations or adjustments to data collected from different sensors, only Landsat 5 data was collected and used for this project. 50 satellite images were obtained between the years 1989 and 2011, but the bulk of the data is between the years 1990 and 2000.

2.1.2 Precipitation Data

The World Bank has a collection of monthly precipitation data available on a country by country basis (The World Bank Group, 2016). For Bangladesh, the monthly precipitation data available covers the years 1901 to 2015. The precipitation data is given in millimeters, and for this project, it is assumed that there is no spatial variation in precipitation. Therefore the precipitation values are assumed to be accurate and applicable for both study sites.

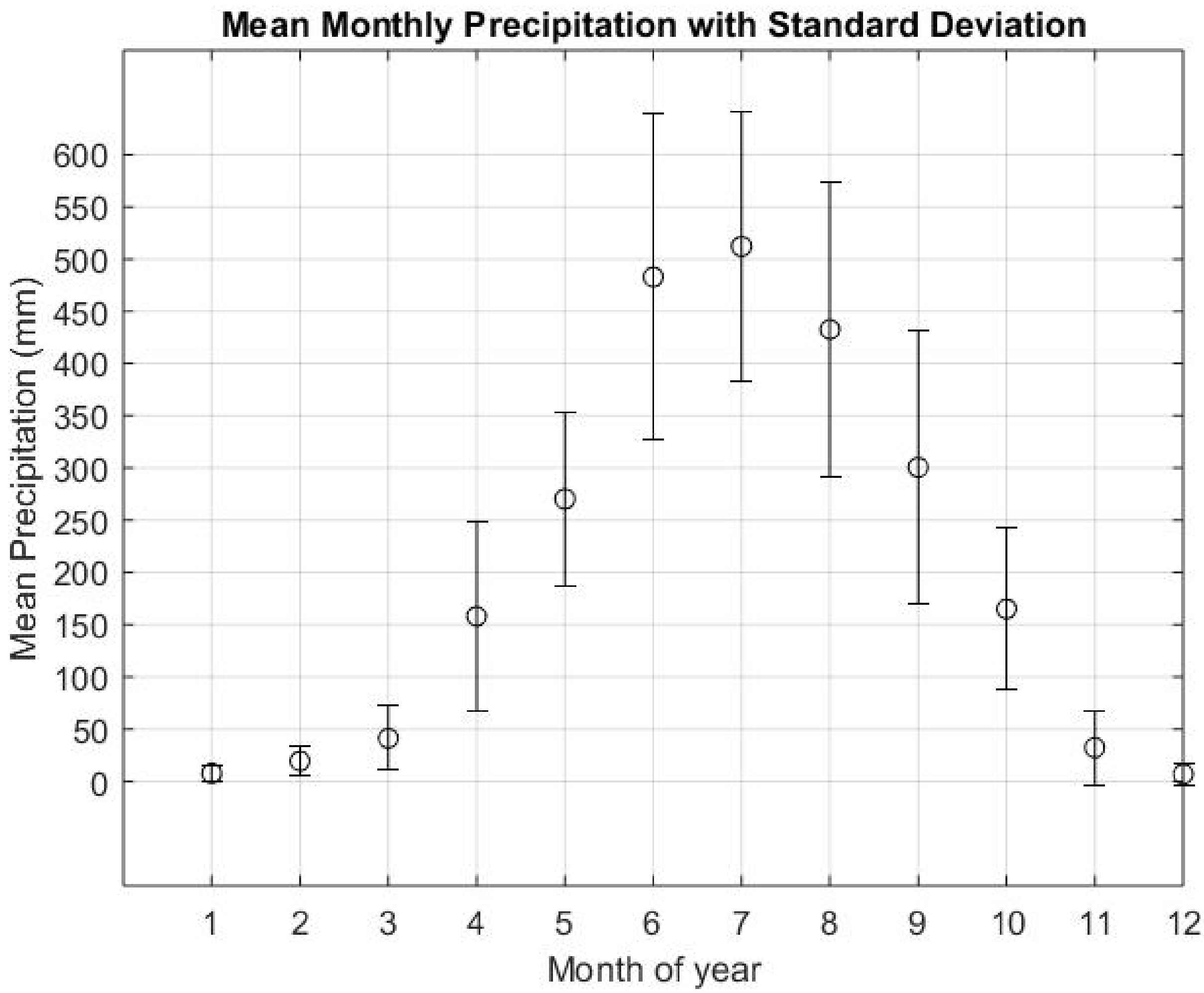

Figure 6: Average Monthly Precipitation, Standard Deviation displayed for each average value

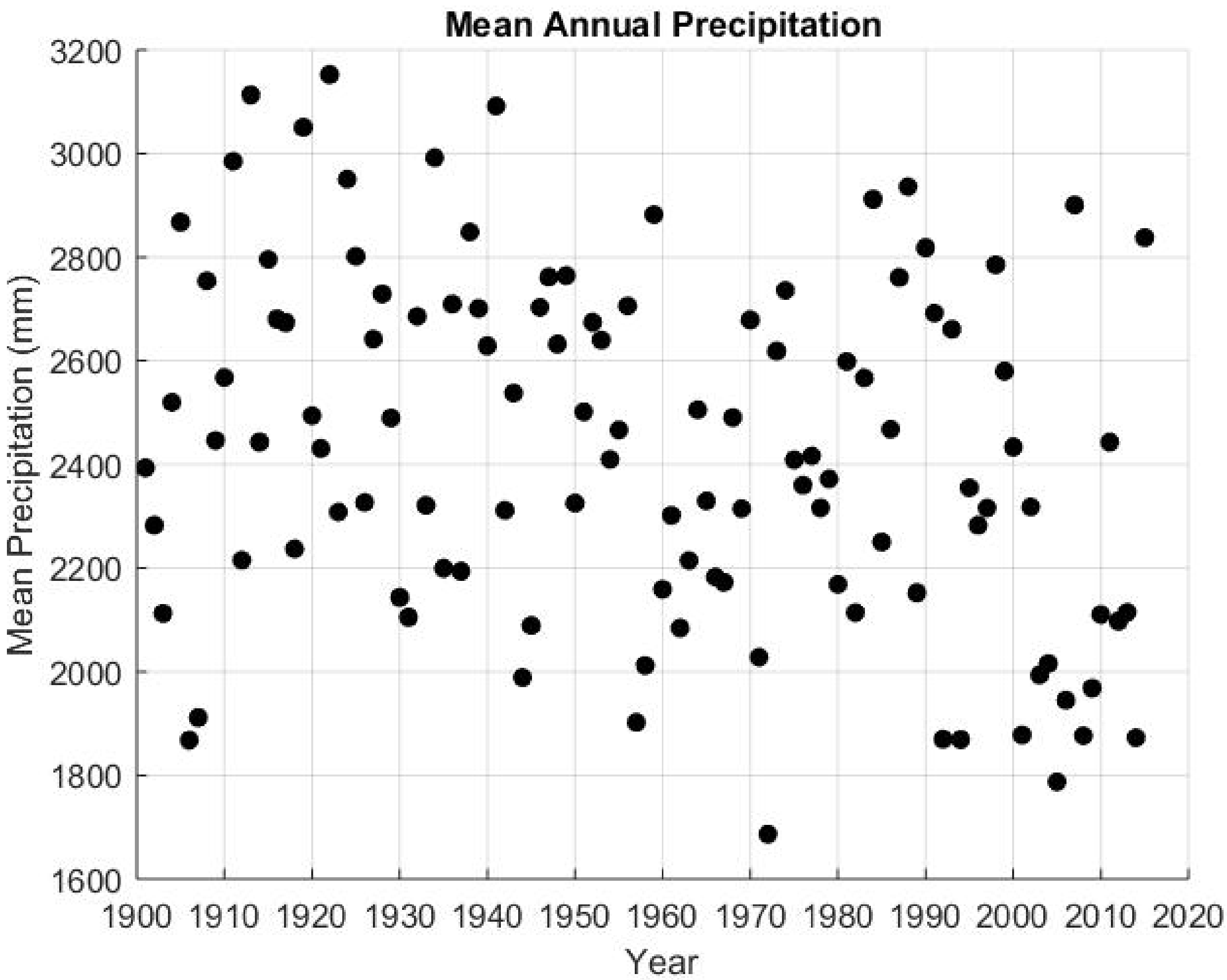

Figure 7: Average Annual Precipitation (1901-2015)

By looking at Figures 6 and 7, it seems reasonable to say that there is a trend to the monthly precipitation values within a year, but from year to year there is no obvious trend in the annual precipitation.

2.1.3 Population Data

Population data related to the Dhaka urban area was obtained from the World Population Review which aggregated data from the Bangladesh Bureau of Statistics and the United Nations population estimates to compile this dataset (Worldpopulationreview, 2017).

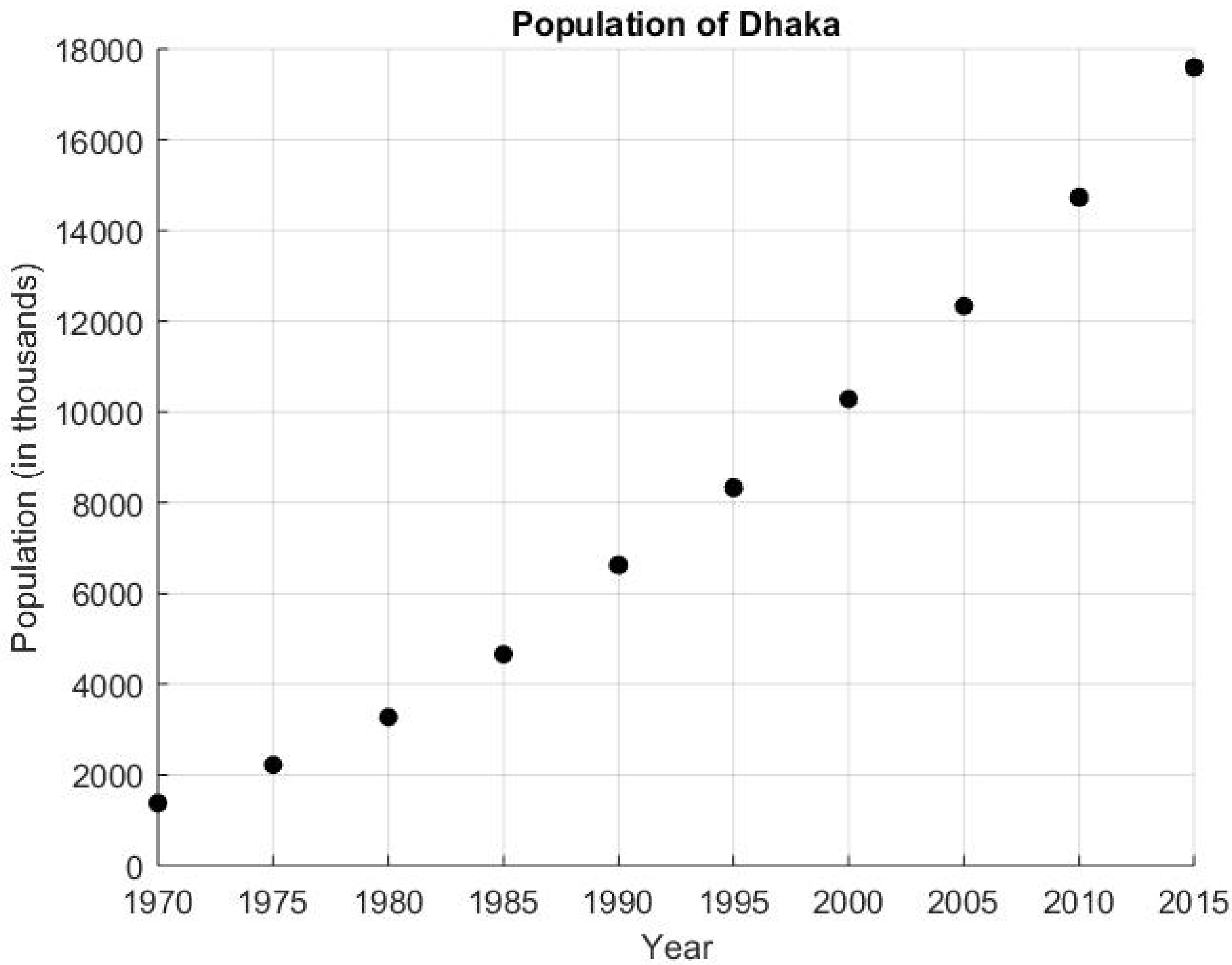

Figure 8: Dhaka population from 1970 to 2015

The population in Dhaka reflects the global trend of urbanization. A rapidly growing urban area is expected to have less vegetation and also higher temperatures due to the "urban heat island effect."

2.2 Data Processing

2.2.1 NDVI Computation

Computing the normalized difference vegetation index (NDVI) for each image was performed in ArcGIS Pro using the raster calculator. The following computation was performed on each scene:

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

Where NIR represents the near infrared band, and RED represents the red band from the visible spectrum. This raster calculation produces a new NDVI raster with values that can range from -1 to 1.

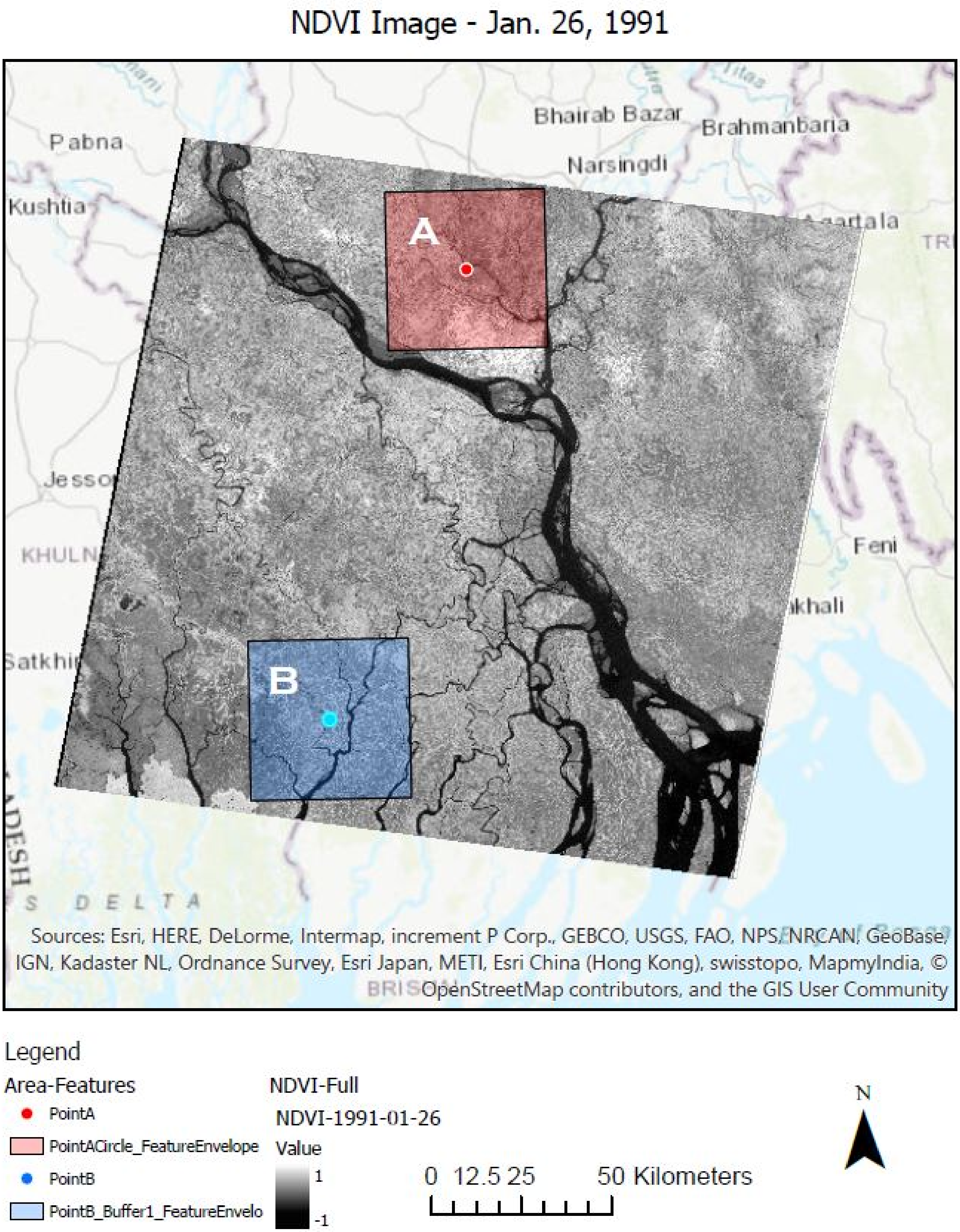

Figure 9: NDVI Image (Jan. 26, 1991)

NDVI values closer to 1 are supposed to be indicative of higher vegetation water content and a higher vegetative fraction. Statistics were computed for each raster and a mean value was obtained to represent the averaged NDVI for each scene. The rasters were also clipped to each study area and an average value for each study area was obtained as well.

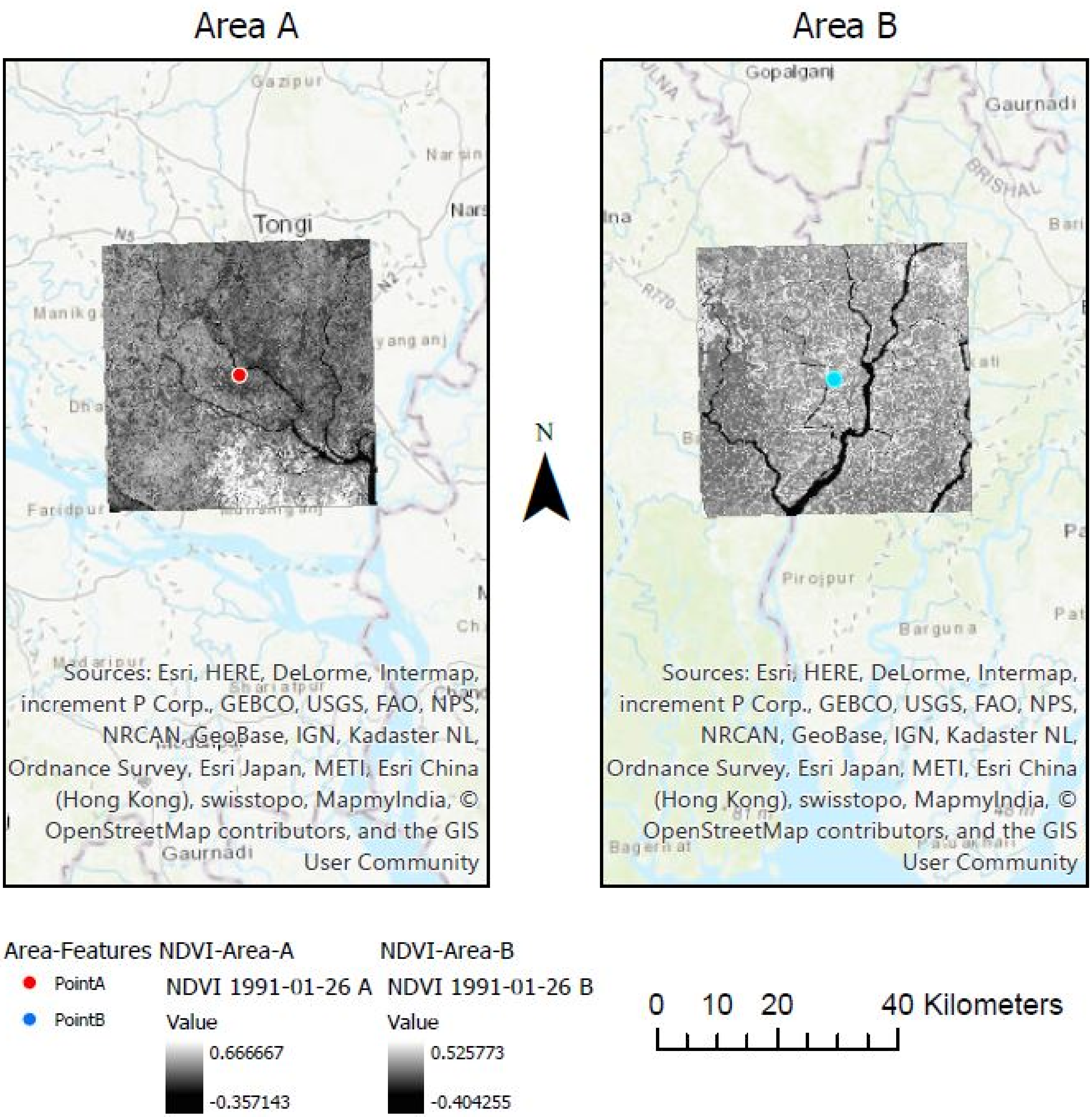

Figure 10: NDVI rasters for each study site (Jan. 26, 1991)

2.2.2 Surface Temperature Estimation

Estimating surface temperature from satellite data is performed by transforming the data recorded by the thematic mapper sensor which captured radiation between wavelengths of 10.4 and 12.5 μm.

To estimate the surface temperature, a procedure a paper titled *Spatio-temporal dynamics of land cover and land surface temperature in the Ganges-Brahmaputra delta: A comparative analysis between India and Bangladesh* (Chaudhuri and Mishra, 2016) was utilized. First, an equation to convert the Band 6 data values into radiant temperatures is used. This equation was derived in a paper which examined the data quality of the thematic mapper data (Malaret et al., 1985)

$$\mathrm{T}\,(k) = 209.831 + 0.834\,(\mathrm{DN}) - 0.00133\,(\mathrm{DN})^2$$

Where DN is the digital number from the satellite data, and the $\mathrm{T}(k)$ value is the radiant temperature in Kelvin.

The radiant temperature value is then transformed into a surface temperature by using a correction which involves the use of an assumed emissivity. This equation is from Chaudhuri and Mishra (2016):

$$\mathrm{T}\,(s) = \frac{\mathrm{T}\,(k)}{1 + (\lambda \times \mathrm{T}\,(k)\,/\rho)\ln(\epsilon)}$$

Where λ is the wavelength of emitted radiance (11.5 nm), ρ is $1.4*10^{-2}$ mK, and ϵ is the emissivity (assumed to be 0.96). The result is an estimation of $\mathrm{T}(s)$ or the surface temperature in Kelvin, which can be easily converted into degrees Celsius. The emissivity assumption decided by looking at the Chaudhuri and Mishra (2016) paper and choosing a value that appeared to best represent the scene captured in Path 137, Row 44.

To perform these conversions using ArcGIS Pro, the band 6 data from the Landsat missions was first loaded into ArcGIS. A model routine was generated to perform the raster calculations which change the band 6 data into radiant temperature and ultimately into the surface temperature estimate. This routine was applied to all 50 scenes. Then raster clipping was performed to generate average statistics for the two study areas. The surface temperature estimate results provide a mean value for the whole area of Path 137, Row 44 as well as average estimates for study areas A and B.

2.3 Data Analysis

2.3.1 NDVI Analysis

The average NDVI data for the entirety of the Path 137, Row 44 scene and the average NDVI values for each of the study sites can be compared. On average, it should be expected that study area A would have a lower average NDVI value than study area B because urban areas are expected to have less vegetation than rural areas.

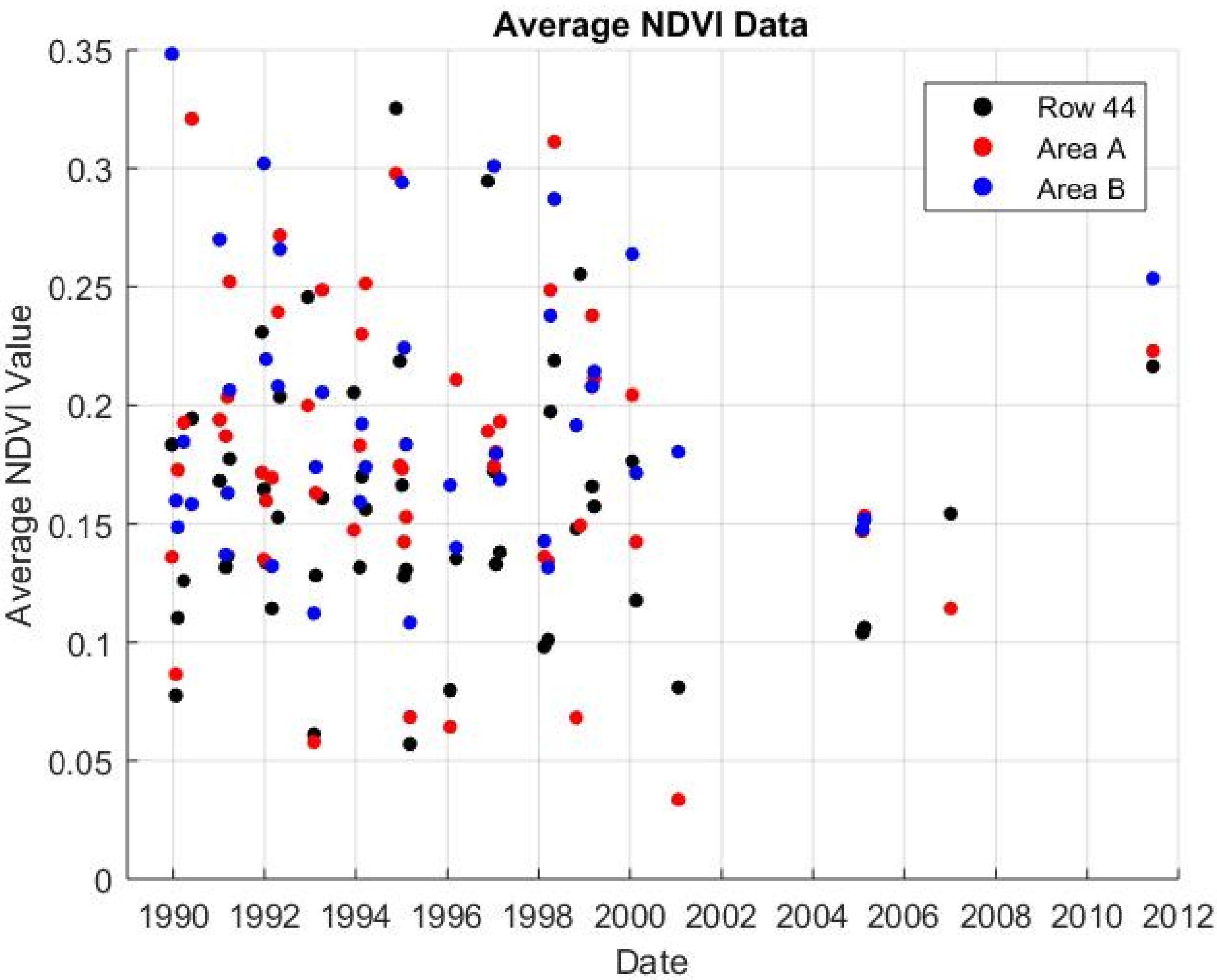

Figure 11: Scatter plot of average NDVI values for the entirety of Path 137-Row 44, study area A and study area B for all 50 data points

The data plotted in Figure 11 can be averaged in time to provide single average values for each area. The entirety of Path 137, Row 44 has an average NDVI value of 0.1567 for these data points. Similarly, the averaged NDVI value for study area A is 0.1781, and for study area B the average NDVI value is 0.2251.

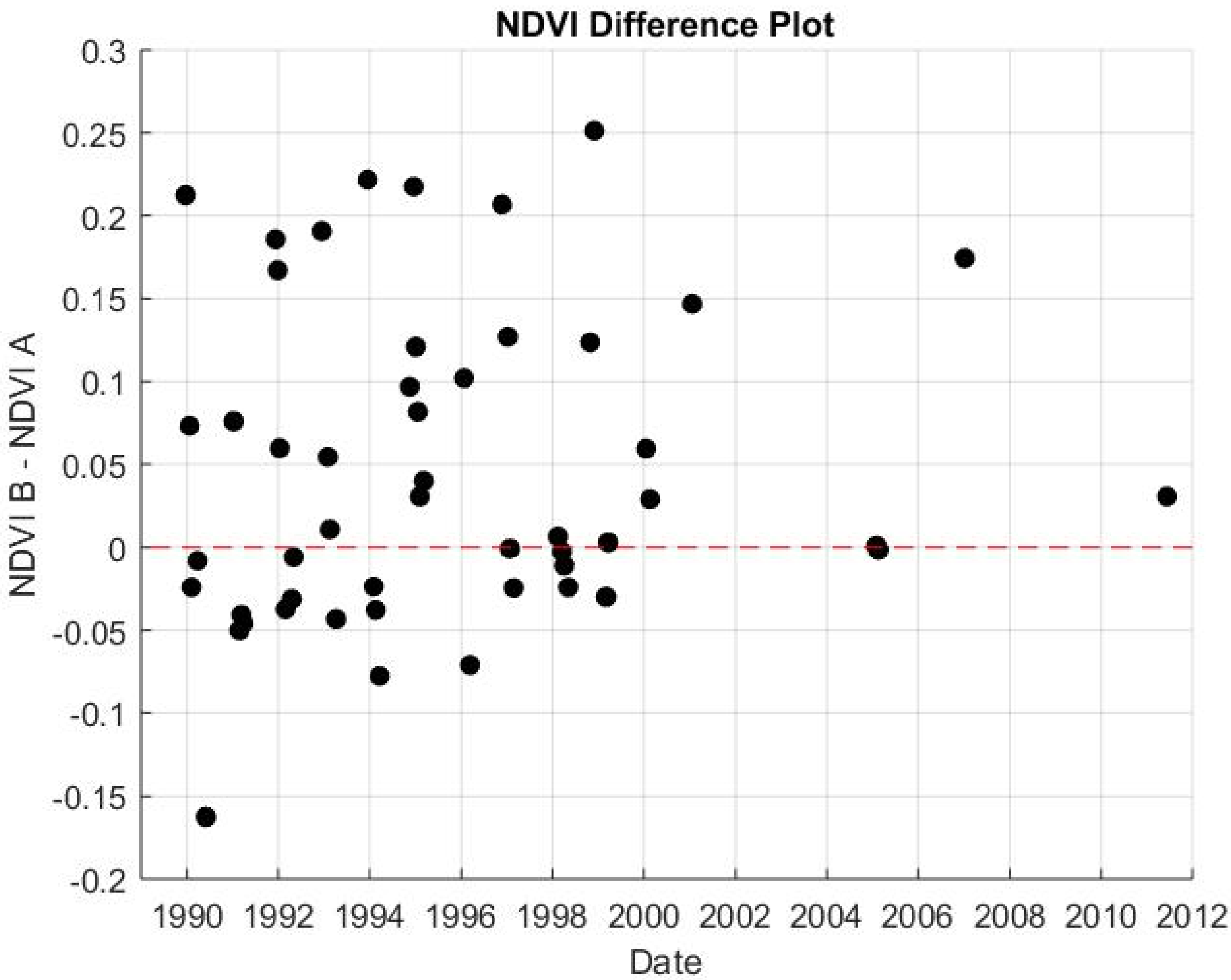

Figure 12: Plot of difference values when the average NDVI values in area A are subtracted from the average NDVI values in area B, a dashed red line has been shown at $y = 0$.

The average difference between the NDVI values in study areas A and B is 0.0470, which when compared to the standard deviation (0.0947), and the range (0.4139) of the differences in NDVI, does not appear to be indicative of any clear and consistent relationship. Figure 12 does not reveal much beyond the fact that the majority of these differences are greater than 0 which indicates that the NDVI value for area B is usually larger than the NDVI value for area A.

2.3.2 Surface Temperature Analysis

The estimated surface temperature values can be averaged for each area just like the NDVI data was. This allows for similar plots to be generated for the surface temperature data. The expectation, is that study area A should have higher surface temperature values than study area B.

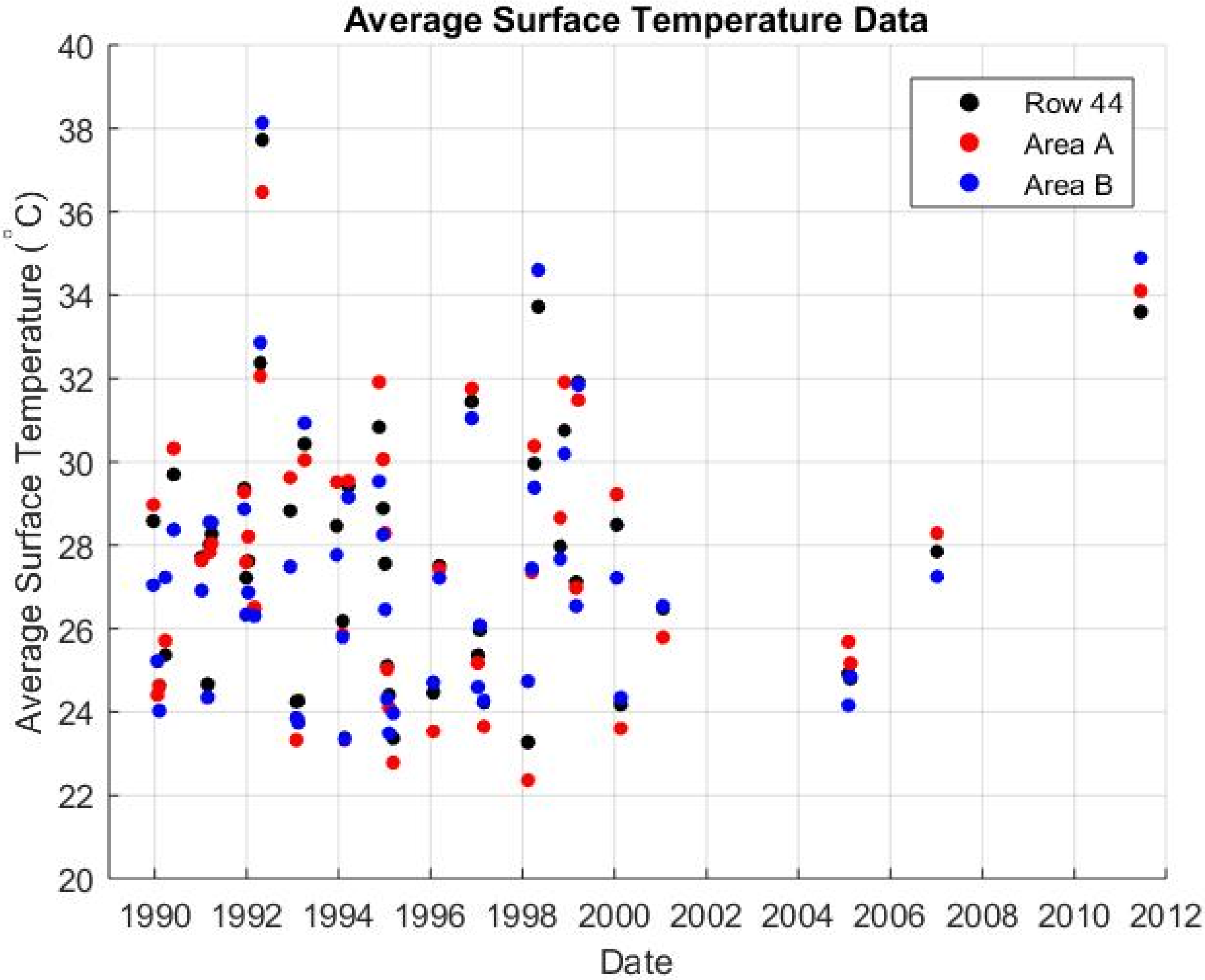

Figure 13: Scatter plot of average surface temperature values for the entirety of Path 137-Row 44, study area A and study area B for all 50 data points

Taking the average of all 50 values, the whole scene has an average surface temperature value of 27.5909°C, study area A has an average surface temperature value of 27.6460°C and study area B has an average surface temperature value of 27.3440°C.

These findings might suggest that study area A has a marginally higher temperature on average than study area B. The plot of the differences is shown in Figure 14.

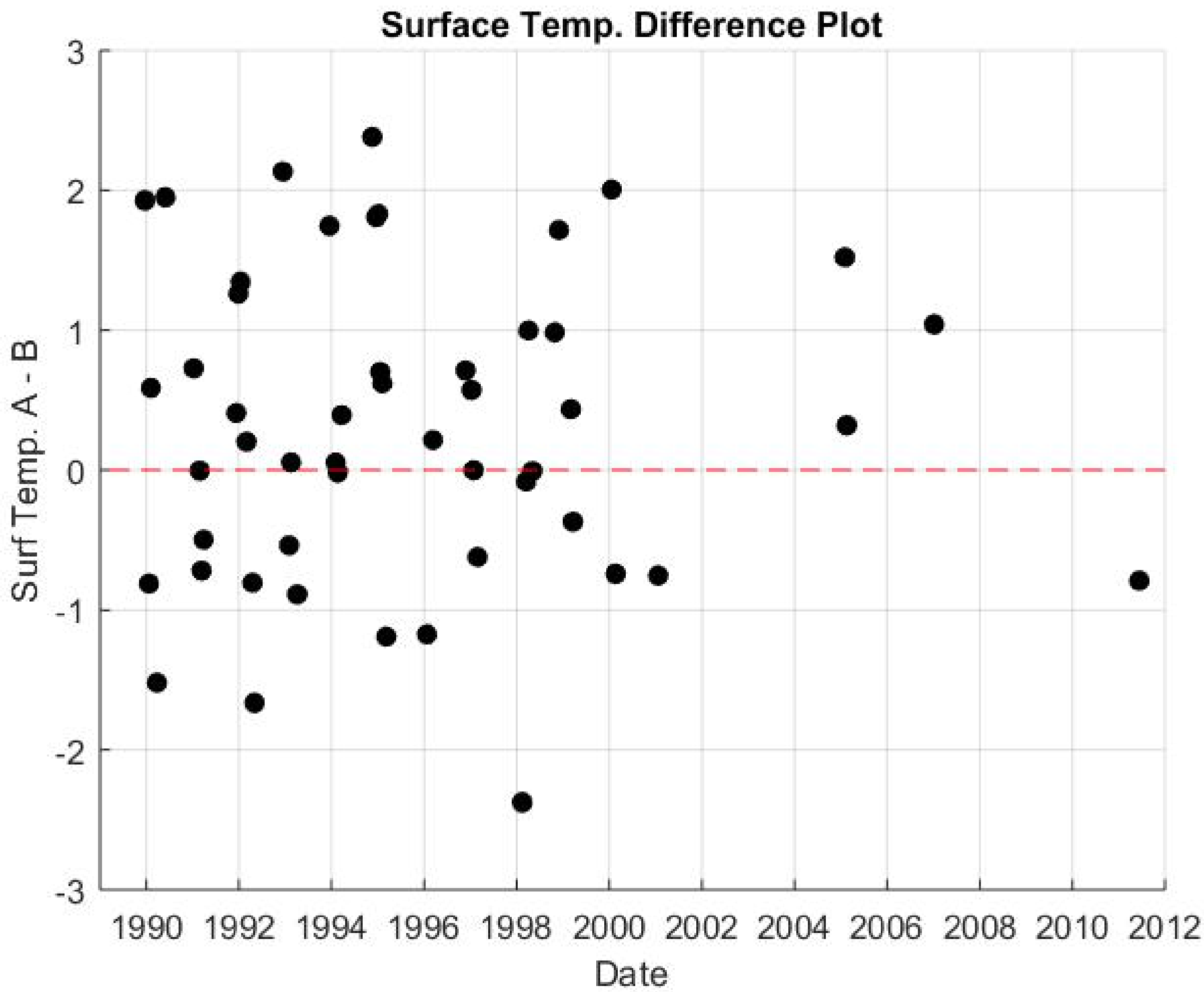

Figure 14: Plot of difference values when the average surface temperature values from area B are subtracted from those in area A, a dashed red line has been shown at y = 0.

Note the location of the 0 difference value on the y-axis. The temperature data estimates do not appear to show with any real significance a difference between the temperatures in study area A and study area B. This would be surprising given the land use in each area, if a singular emissivity value had not been assumed which means that the land use was assumed to be the same for both scenes.

2.3.3 Precipitation Analysis

The precipitation data collected from the World Bank provided a value in millimeters of precipitation for the entire country of Bangladesh for every month from the year 1901 to the year 2015. For this project, it has been assumed that spatial variation in precipitation can be ignored and the rain is equally distributed in time and space across Bangladesh. Figures 6 and 7 show average precipitation values for each month and annual averages.

2.3.4 Dhaka Population Analysis

The population data collected for Dhaka only provided population values at 5 year intervals. To interpolate and create a population estimate for the time of each Landsat image, a curve was fit to the population data. The curve fit in relation to the population data is shown in Figure 15.

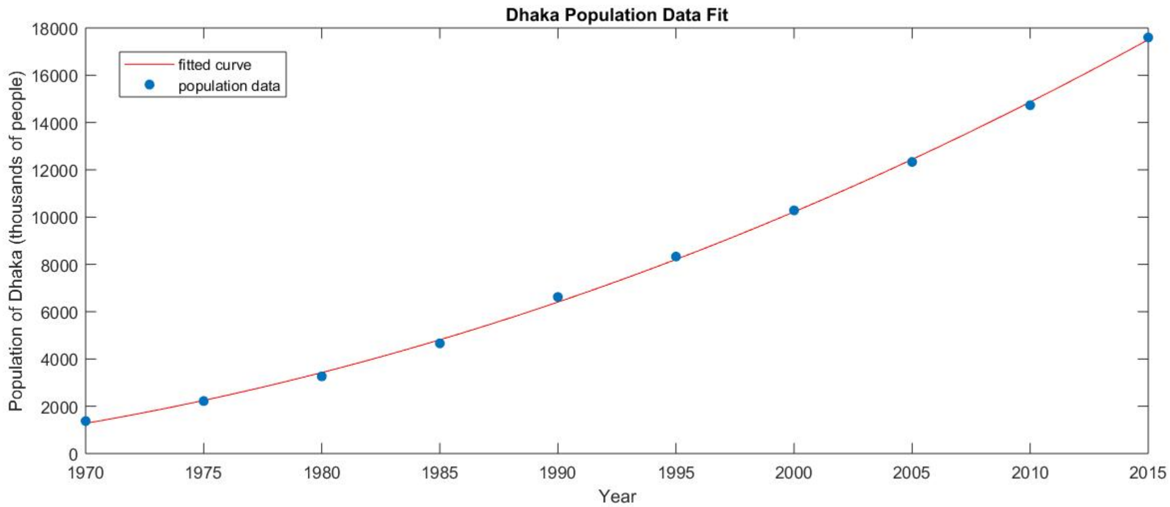

Figure 15: Curve fit to the population data

The equation of the curve fit to this data is as follows:

$$\text{pop}\,(\text{year}) = \left[954.6 \times (\text{year})^2\right] + [5459 \times (\text{year})] + 7283$$

Population data alone is not terribly exciting, but by comparing and analyzing it in concert with other data collected, hopefully relationships between the data will emerge.

2.3.5 Surface Temperature and Population

So for study area A, the surface temperature and the population data can be compared. One might hypothesize that the surface temperature would increase as the population increased and the region became increasingly urban. To make this comparison without being confounded by seasonal temperature fluctuations, only temperature data from the month of December will be used. December is being chosen because it is the month with the greatest number of Landsat data points (see Figure 5).

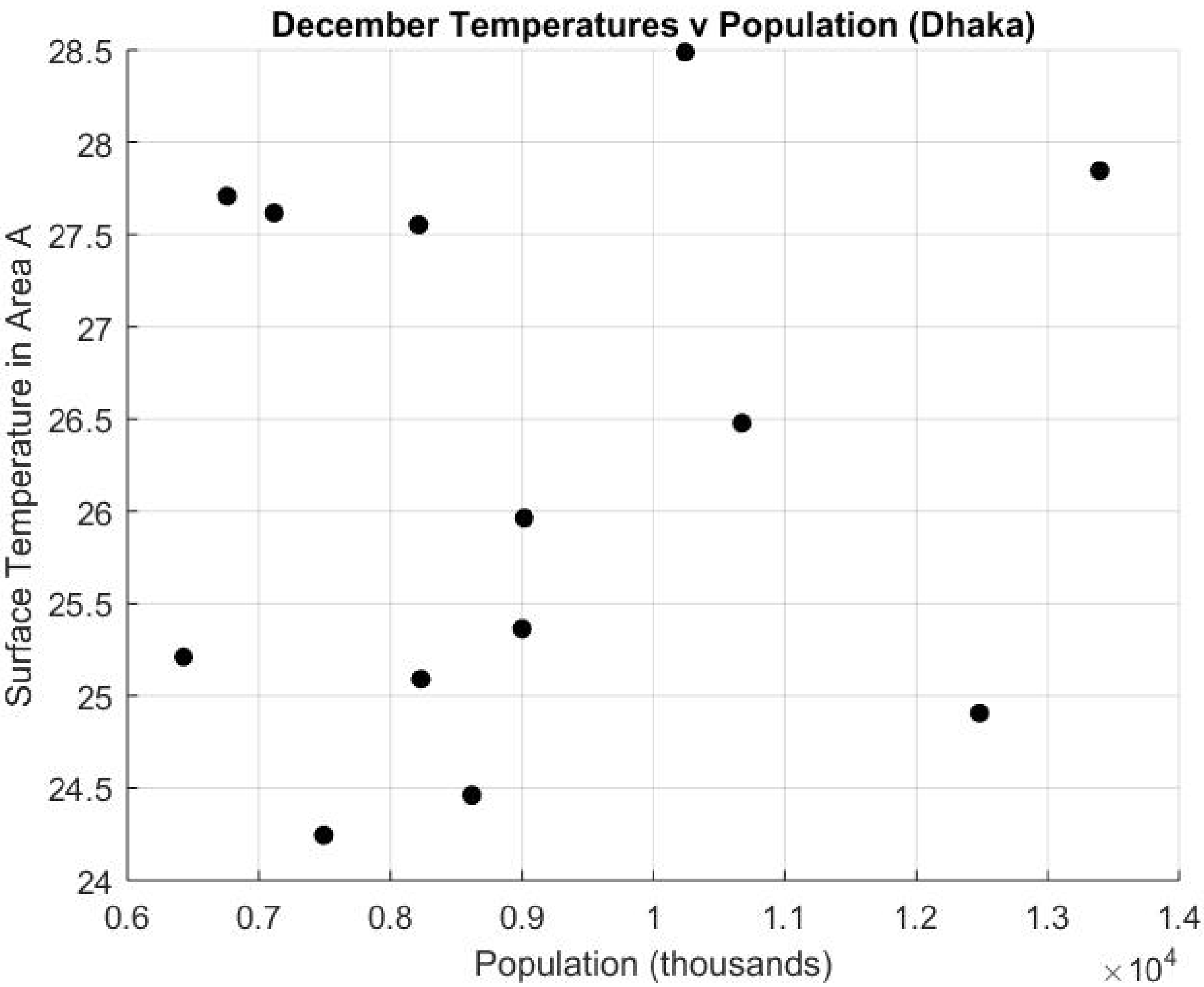

Figure 16: Plot of temperature v population

Looking at the December temperature data plotted against the corresponding population estimates in Figure 16, there is no clear trend or correlation between the two. Therefore it does not seem wise or prudent to fit any sort of trend line or curve to this data.

2.3.6 Precipitation and NDVI

If NDVI is a measure of vegetative health, it stands to reason that some relationship between the precipitation and NDVI may be observed. To check for this relationship, the NDVI data collected is plotted against the precipitation data in Figure 17.

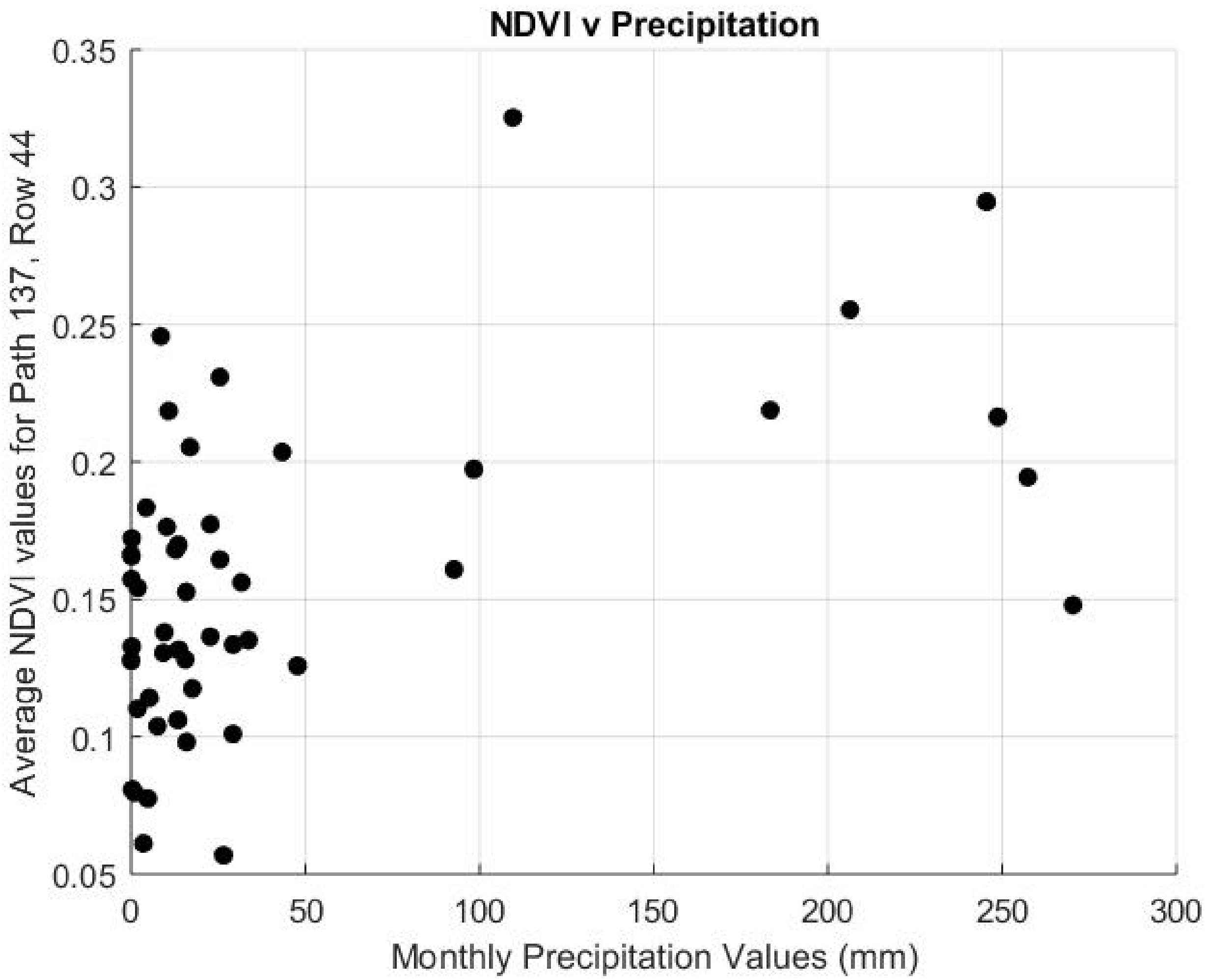

Figure 17: NDVI v Precipitation

Again there does not appear to be any clear correlation between these two sets of data. This lack of a clear relationship makes it unwise to force a fit or trend line.

2.3.7 Surface Temperature and NDVI

A plot of the averaged surface temperature values against the averaged NDVI values is provided in Figure 18.

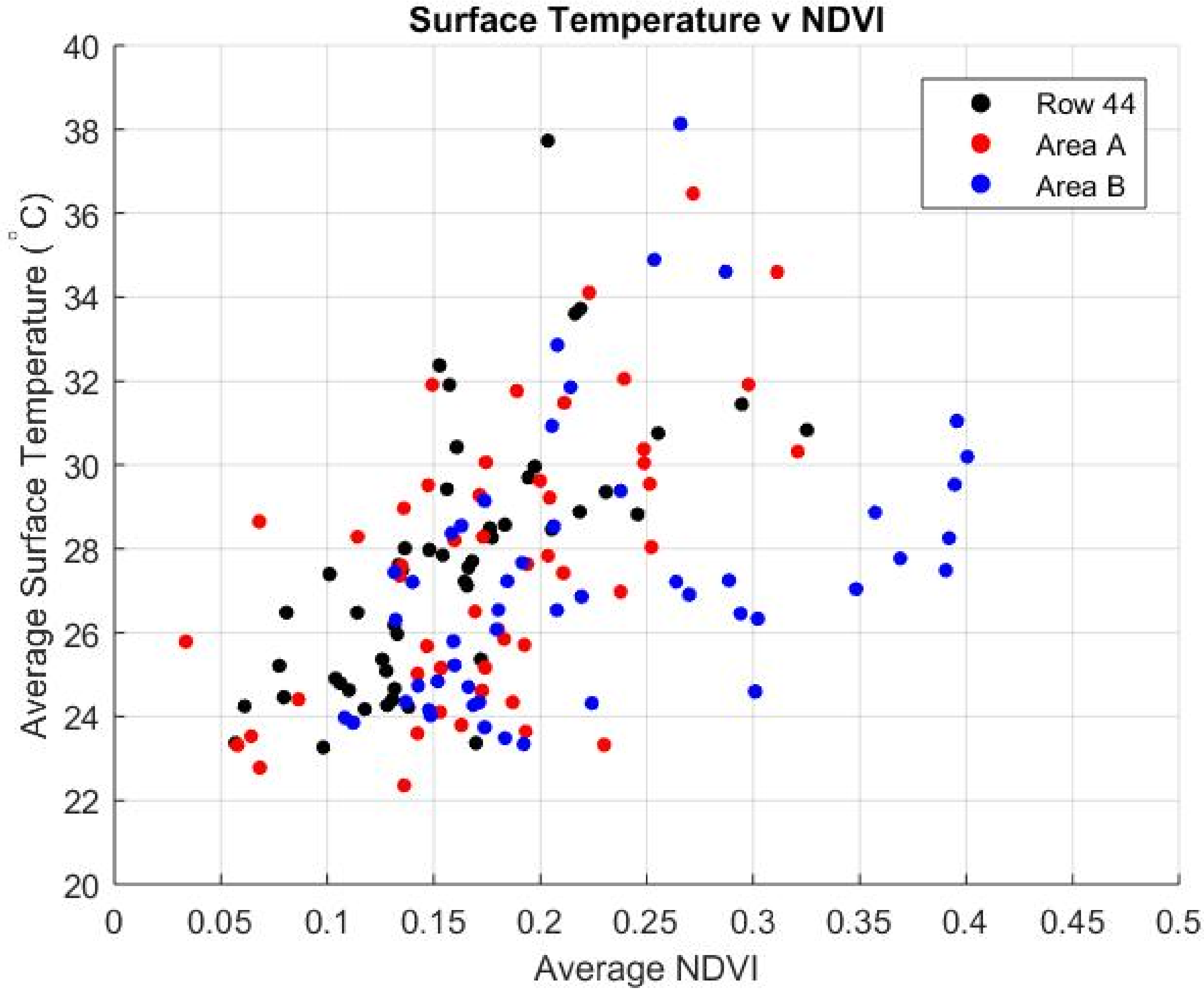

Figure 18: Surface Temperature v NDVI

At a glance, there appears to be a possible correlation between the NDVI values and the surface temperature estimates in study area A. This data for study area A is plotted along in Figure 19. A line of best fit is also plotted using a linear model and its equation is provided within the figure.

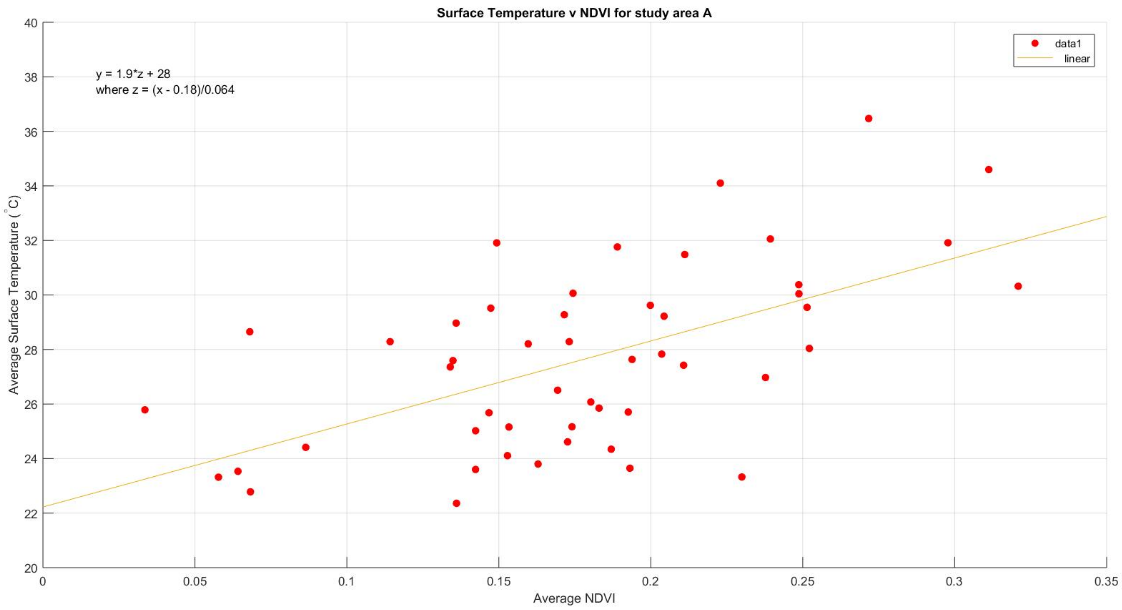

Figure 19: Surface Temperature v NDVI for study area A

It is hard to say from looking at this figure, that a linear fit is appropriate. While a relationship between NDVI and surface temperature appears to exist for study area A, given this data and without any knowledge of the underlying physical relationship between the two metrics, an equation or a fit for the relationship between this data should not be created or used.

2.3.8 Comparing Study Areas

In the event of cloud cover or poor satellite data over a portion of the image, it may be useful to estimate data about one region by using data collected from elsewhere. Since there are two study areas the potential for correlations between the two can be explored.

First, the NDVI data for the two regions will be compared. In Figure 20, the NDVI values from area B are plotted against those from area A.

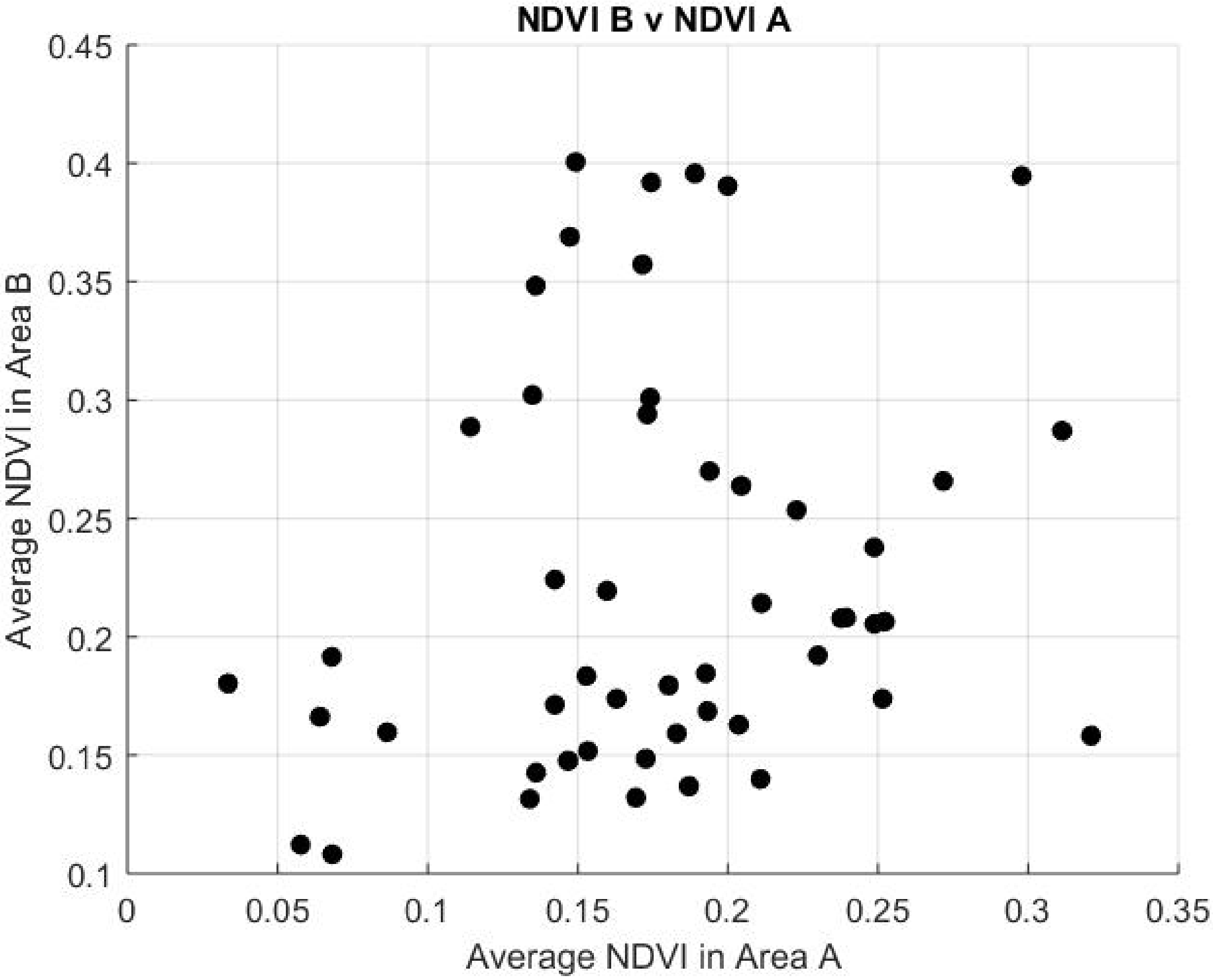

Figure 20: NDVI B v NDVI A

Judging by the lack of values in the upper left hand portion of the plot, it could be argued that there is some relationship between the NDVI values in both regions. This is expected as the two study areas are close spatially, and even exist within the same Landsat scene. Given their proximity in space, the correlation between the two is not strong.

Next the surface temperature values for each study area will be plotted against each other to determine whether or not there is a correlation to that data.

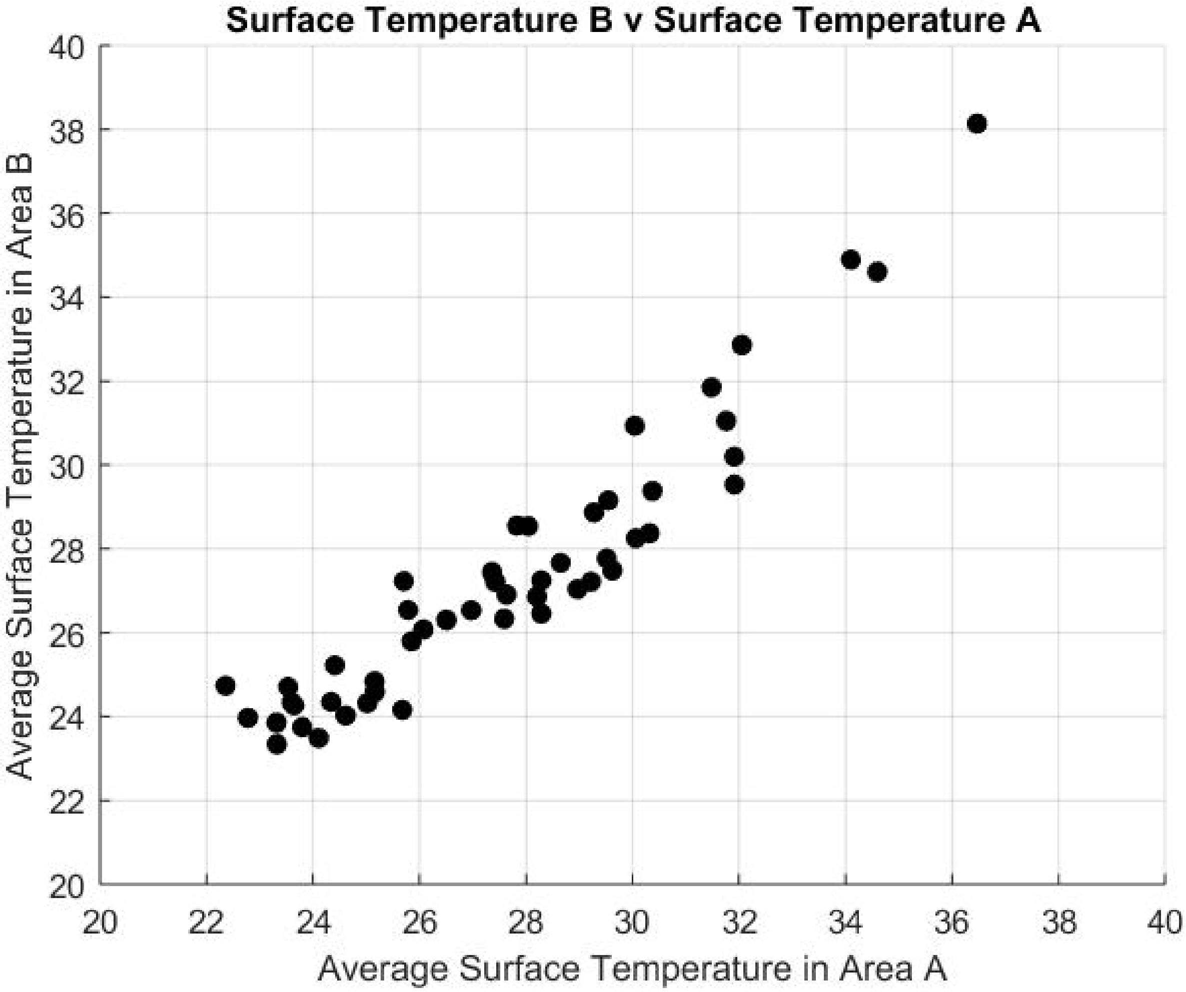

Figure 21: Surface Temperature Area B v Surface Temperature Area A

From Figure 21 it is clear that there is a strong correlation between the surface temperature in area A and the surface temperature in area B. This is the type of strong relationship that can be used to make predictions.

When a curve is fit to this data, an exponential fit appears to be quite strong. For this data, the exponential model that was fit had the following equation for the surface temperature estimate in area B if the surface temperature estimate in area A is known.

$$\text{TempB} = 10.92 \times e^{(0.03299 \times \text{TempA})}$$

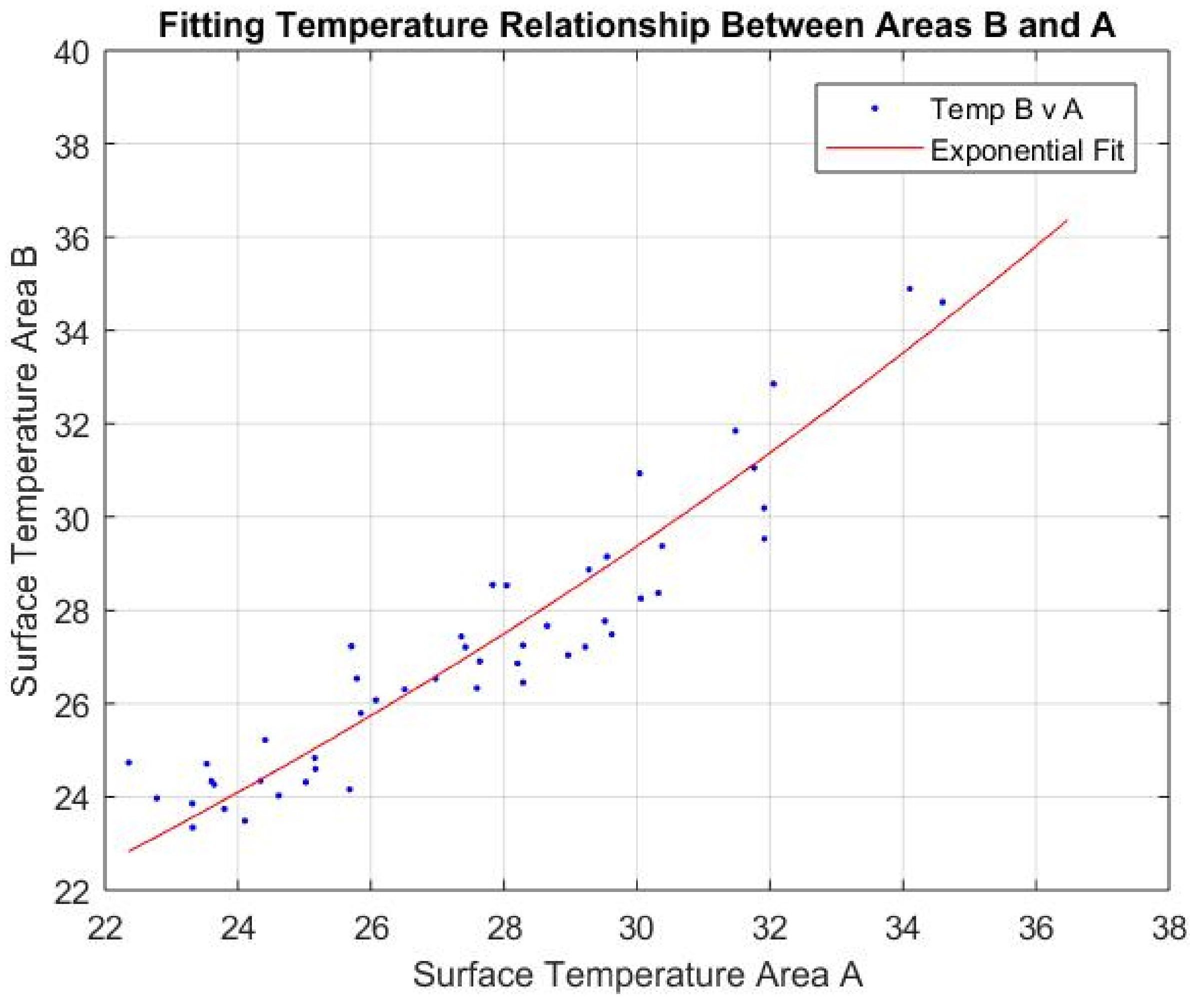

Figure 22: Surface Temperature Area B v Surface Temperature Area A with an exponential curve fit to the data

The fit to the data shown in Figure 22, has an R^2 correlation coefficient value of 0.9087.

3 Results

3.1 Results

Many of the relationships between the datasets that were explored were ultimately inconclusive. Much of this data did not correlate strongly with the other data. The only real exception was the surface temperature in the two study areas. By using ArcGIS to process the data, the analysis was greatly accelerated. ArcGIS also allowed for easy splitting of rasters into the various study sites, and by computing statistics, averaged measurements could be taken for surface temperature and NDVI. The fit to the surface temperature data suggested that the temperature in the area of study site B could be predicted by applying an exponential equation to the surface temperature estimate in study site A. This relationship would also be applicable in the reverse, and so the surface temperature in study area A could be predicted given knowledge of the temperature in study area B.

3.2 Discussion

This project was ultimately not successful in relating these measurements. There may be many reasons for the lack of conclusive results or correlations between the data collected.

In the literature, most land surface temperature using satellite data makes use of land cover information to improve the estimations of surface temperature (Ahmed et al. (2013) & Chaudhuri and Mishra (2016)). By assuming a single emissivity value for this project and simplifying the surface temperature estimation process, any changes in temperature due to land use changes are missed. For rapidly growing urban regions such as Dhaka, this simplification may result in missing temporal increases in surface temperature as land uses change.

Image quality control was limited to the use of the EarthExplorer functionality to filter images with cloud cover. Quality images were not used to verify the images used. So lower quality images may also exist in this 50 image dataset and outliers may be impacting the results of any analysis.

The precipitation data collected was on a monthly basis for the entire country. For this project it was assumed that rainfall occurs uniformly over the entire county and so any spatial variation was ignored. This simplification may have made it more difficult to find any links between NDVI and precipitation. The use of more accurate local measurements such as rain gage data would have helped improve the spatial accuracy of that precipitation data.

The NDVI data was derived using the standard procedure for computing NDVI. Beyond choosing more scenes to analyze, it would be difficult to further enhance the accuracy of the NDVI data.

The population data obtained was only present at a 5 year scale. However, the population data fit well to a polynomial model, and so the interpolated estimates of population for given points in time is relatively accurate. This data should be less critical in analysis where population is primarily being used as a measure of city size and growth. The population data was expected to serve as an indicator of urbanization.

3.3 Future Work

Initially, the expectation was for these datasets to have greater cross-correlations. Despite this not being the case, it still may be possible to view these temporal data points as signals, and potentially apply signal processing procedures and tease out relationships between the data. Yarleque et al. (2009) was able to determine time lag statistics to estimate rainfall data from NDVI information. Their procedure involved treating the rainfall and NDVI information as signals and processing them using Fourier and Wavelet transforms. A similar methodology may apply in this case if the NDVI and rainfall data can be viewed as period and proportional waves.

If viewing these data points as signals appears to be a reasonable strategy, then procedures such as Fourier Transforms and Wavelet Transforms may be applicable for predicting unknown data values. These types of methods will only be applicable if equations relating the data can be determined, and at least one type of data can be collected.

Performing additional statistical tests and methods to modify and manipulate the data may allow some weak relationships to be determined using this data. This report was very much limited to plotting the data as it is and visually attempting to determine whether or not correlations and relationships were present.

3.4 Conclusions

More work needs to be done to investigate relationships and links between precipitation, surface temperature, NDVI, and population. Logically it is reasonable to expect some sort of correlation between these datasets. The strength of the correlation may be weak, but given enough good data, there should be evidence of links between all of these data sets.

In the analysis done for this report, which was limited to using 50 scenes exclusively from Landsat 5, the only strong correlation between data was a relationship between values for surface temperature in the two study areas. Finding a relationship between temperature between these two areas does not come as a surprise since they are both located within this same scene and should be expected to share the same climate and weather conditions. The land uses between the two study areas do vary, and so comparing these two regions over a different time scale may make it possible to understand the effect of urbanization or other changes to land use if this relationship is not preserved. Having a way to relate the surface temperatures of two different places can also be useful in the event of localized cloud cover. Relationships such as these may allow more continuous datasets to be developed if missing data due to cloud obscurities can be reliably estimated by using regional data that was not obscured.

4 Bibliography

References

Ahmed, B., Kamruzzaman, M., Zhu, X., Rahman, M., and Choi, K. (2013). Simulating Land Cover Changes and Their Impacts on Land Surface Temperature in Dhaka, Bangladesh. *Remote Sensing*, 5(11):5969–5998.

Alam, M. and Rabbani, M. G. (2007). Vulnerabilities and responses to climate change for Dhaka. *Environment and Urbanization*, 19(1):81–97.

Anuta, P. E., Member, S., Bartolucci, L. A., Dean, M. E., Lozano, D. F., Malaret, E., Mcgillem, C. D., Valdes, J. A., and Valenzuela, C. R. (2000). LANDSAT-4 MSS and Thematic Mapper Data Quality and Information Content Analysis. (3).

Birtwistle, A. N., Laituri, M., Bledsoe, B., and Friedman, J. M. (2016). Using NDVI to measure precipitation in semi-arid landscapes. *Journal of Arid Environments*, 131:15–24.

Chaudhuri, G. and Mishra, N. B. (2016). Spatio-temporal dynamics of land cover and land surface temperature in Ganges-Brahmaputra delta : A comparative analysis between India and Bangladesh. *Applied Geography*, 68:68–83.

Fu, B. and Burgher, I. (2015). Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater. *Journal of Arid Environments*, 113:59–68.

Hoitink, A. J. F., Wang, Z. B., Vermeulen, B., Huismans, Y., and Kästner, K. (2017). Tidal controls on river delta morphology. *Nature Geoscience*, 10(9):637–645.

Hossain, S. (2008). Rapid Urban Growth and Poverty in Dhaka City. *Urban Studies*, 5(1):1–24.

Kim, W., Paola, C., Swenson, J. B., and Voller, V. R. (2006). Shoreline response to autogenic processes of sediment storage and release in the fluvial system. *Journal of Geophysical Research*, 111(F4):F04013.

Malaret, E., Bartolucci, L. A., Lozano, D. F., Antlta, P. E., and Mcgillem, C. D. (1985). Landsat-4 and Landsat-5 Thematic Mapper Data Quality Analysis. 51(9):1407–1416.

Passalacqua, P., Lanzoni, S., Paola, C., and Rinaldo, A. (2013). Geomorphic signatures of deltaic processes and vegetation: The Ganges-Brahmaputra-Jamuna case study. *Journal of Geophysical Research: Earth Surface*, 118(3):1838–1849.

The World Bank Group (2016). Climate Data.

USGS (2017). EarthExplorer.

Wilson, C. A. and Goodbred, S. L. (2014). Construction and Maintenance of the Ganges-Brahmaputra- Meghna Delta: Linking Process, Morphology, and Stratigraphy.

Worldpopulationreview (2017). Karachi Population 2017 (Demographics, Maps, Graphs).

Yarleque, C., Roberto, Q. A., Yarlequé, C., Posadas, A., Mares, V., and Quiroz, R. (2009). Estimating daily rainfall from NDVI using the wavelet transform. (March 2015).