Interpolation of NDVI in areas covered by clouds using Landsat 8 image through Space-time IDW

Roberto Luna[1], Manuel Mendoza[2] William Martinez[3]
IMS Unideversidade Nova Lisboa
Campolide, 1070-312 Lisboa, Portugal
[1] M20170523@segi.unl.pt
[2] M20170515@segi.unl.pt
[3] M20170525@segi.unl.pt

Abstract

This paper propose a method for filling pixel values of NDVI covered by clouds in optical remote sensing images, accounting for the variability of previous images. In this sense, the fast performance of IDW and the configuration of the problem lead us to adapt this versatile algorithm to the spatio-temporal domain to predict the lost pixels values. The prediction is performed over three synthetic clouds over the region of Nicaragua using three different scenes of the sensor Landsat 8 through one year. Various IDW-based spatio-temporal interpolation trials with different parameter configurations are evaluated by cross-validation. In the end, the sensitive analysis of the errors lead us to apply IDW only considering the pixel values through the time. In this way, we show the results of the prediction and the error assessment of the estimations.

Keywords Clouds,NDVI,IDW

1 Introduction

Clouds can be considered as one of the most common atmospheric interferes in optical remote sensing imagery. The existence of clouds not only cause some difficulties for image processing, but also cause visual recognition and classification problems. Hence, in order to overcome that obstacle, there are several approaches whose effectiveness vary depending on the sensor, the size and how thick the cloud is. For example, in hyperspectral instruments is applied mixture analysis using different channels to retrieve lost pixels values. We can find several works, such as Gao et al. (2009) or Schroeder and Fischer (2003), fitting parametric and not parametric models with several bands of the scenes obtained by Aviris and Meris sensors. In this sense, this methodology takes advantage of the high spectral resolution for solving our main concern. However, the spatial resolution of those images is very low, making the use of these images impractical for forestry and agricultural purposes.

Nowadays, in the same framework of the algorithms that use stored bands of lower frequency able to go through the atmosphere without scattering, we can find works using multispectral sensor applying the same methodology in, for example, Sentinel 2 images (Cerra et al. (2016)). Although Sentinel 2 does not count with the same range than those hyperspectral sensors, it has bands that are not affected for water vapor. Although novelty, and useful for the applications that this sensor

has in the area of agriculture and forestry, its main drawback for our main concern is that is used for retrieving pixels only affected by thin clouds or cirrus clouds.

Other example, that it is not very common but it is worth mentioning, it is the use of SAR. SAR operates in the radar part of the spectrum and particles in the atmosphere are much smaller than the wavelength of radar waves, as a result, it is not scattered and thus does not get affected by the atmosphere. Therefore, there are studies using the combination of SAR and Landsat Images to correct areas covered by clouds as long as the objects do not change so much in the area (Hoan and Tateishi (2009)). However, this last condition makes it useless in areas with high variety of objects.

Finally, one of the most common and practical methodologies for retrieving pixels values affected by clouds is considering the temporal resolution of the satellite missions. Accounting for the historical scenes of certain area, there are works, such as Dozier et al. (2008), where the pixels covered by clouds in areas with snow is obtained using cubic spline interpolation on previous pixel values from MODIS images.

In this sense, in the same framework of the last methodology, we propose to work with images available associated to a previous periods to predict the pixel values covered by clouds, but using IDW and considering both the spatial and temporal domain simultaneously. Moreover, the changing trends of the ecosystem activity in the study area will be quantified using Normalized Difference Vegetation Index (NDVI).

2 Methodology

The methodology is based mainly in two steps. Firstly, the calculation of the NDVI, and secondly, the estimation of NDVI values covered by clouds using space-time IDW.

2.1 Normalized Difference Vegetation Index

The normalized difference vegetation index (NDVI) is a simple graphical indicator that can be used to analyze remote sensing measurements. The NDVI is calculated from these individual measurements as follows:

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

where red and NIR stand for the spectral reflectance measurements acquired in the red (visible) and near-infrared regions, respectively (Tucker (1979)).

2.2 Inverse distance weighting in spatio-temporal data

we propose to apply space-time IDW for predicting lost NDVI values. The algorithm is based on ideas of Li et al. (2014), where there is a consideration of the evolution of time and space together of the particulate matter $PM_2.5$ in the

United States, showing computational performance and enhancing prediction of this phenomenon.

Therefore, the IDW will take into account the distances of the closest neighbors both in space and time. This interpolation method considers that the value of a pixel can be obtained from the weighted sum of digital numbers of the closest neighbors.

$$\hat{Z}(\mathbf{s}_0, t_0) = \sum_{i=1}^{n} \lambda_i \cdot z(\mathbf{s}_i, t_i) \tag{2}$$

Where:

- $\mathbf{s}$ is the spatial locations of a pixel using the columns (a) and rows (b).
- $\hat{Z}(\mathbf{s}_0, t_0)$ is the vale to predict in the new location $(\mathbf{s}_0, t_0)$, where $\mathbf{s}_0$ is the spatial location of the pixel in $\Re^2$ and t_0 is the time of interest in $\Re^1$
- n is the number of neighbor pixels with NDVI values around the pixel $(\mathbf{s}_0, t_0)$.
- λ_i are the weights assigned to each measured pixel value that are going to be used for the prediction.
- $z(\mathbf{s}_i, t_i)$ are the NDVI values in the location $(\mathbf{s}_i, t_i$.

The expression for calculating the weights is the following:

$$\lambda_i = \frac{{d_{i0}}^p}{\sum_{i=1}^{n} {d_{i0}}^p} \tag{3}$$

The weight is controlled by a factor p that control the weights over the distance of the neighbors; d_{i0} is the distance between the prediction position and each of the measured positions. The expression $di0$ can be obtained by the equation 4:

$$d_{i0} = \sqrt{(a_i - a_0)^2 + (b_i - b_0)^2 + \theta \cdot (t_i - t_0)^2} \tag{4}$$

The equation 4 has a new parameter θ that makes comparable the time and distance differences for predicting the NDVI values.

2.2.1 leave-one-out cross-validation

The sample values are deleted from the dataset, one at the time, and then the interpolation method is applied to predict the missing value using the remaining observed values. This process is repeated for all pixels.

3 Predicting NDVI in areas covered by synthetic clouds

3.1 Study region and data

We followed the evolution of crop parcels in the center of Nicaragua, a region characterized for high humidity and high levels of precipitation. Located in the

district of Leon to 17 km from the maritime coastline (see Figure 1). The region of the image is characterized mainly for two land cover, forest and agricultural land.

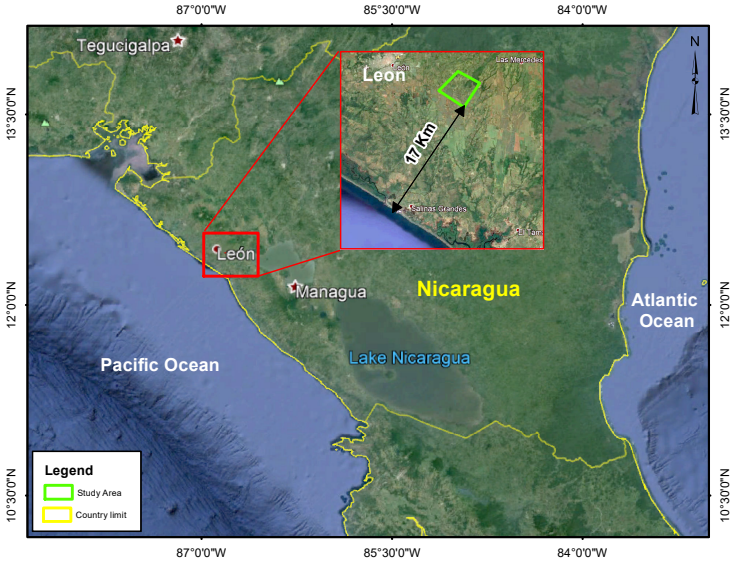


Figure 1: location of the study area.

For developing the methodology, we gathered the red and infrared bands (see Table 1) of Landsat-8 images taken by the months of March, July and December. The images are masks of $4km^2$ of the real scenes, each image is a matrix of 133 pixels.

Table 1: Specifications Landsat 8

Spectral Band	Wavelength	Resolution
Band 4 - Red	0.630 0.680 m	30 m
Band 5 - Near Infrared	0.845 0.885 m	30 m

3.2 NDVI Calculation

The scenes were assembled to create three scenes of NDVI for the months of March, July and December using the equation 4. In Figure 2 we can see how heterogeneous is the space in the region. Although this zone is mainly characterized for forest and agricultural land (sugar cane), the crops depicts different values of NDVI, which translate in different phenology stages.

This paper is part of one experiment for predicting pixel values, hence, our main concern is not the identification of clouds (consult Champion (2016) for further research). Therefore, we propose to estimate the pixels values of NDVI covered by three synthetic clouds in the month of December. The sizes of the clouds are shown

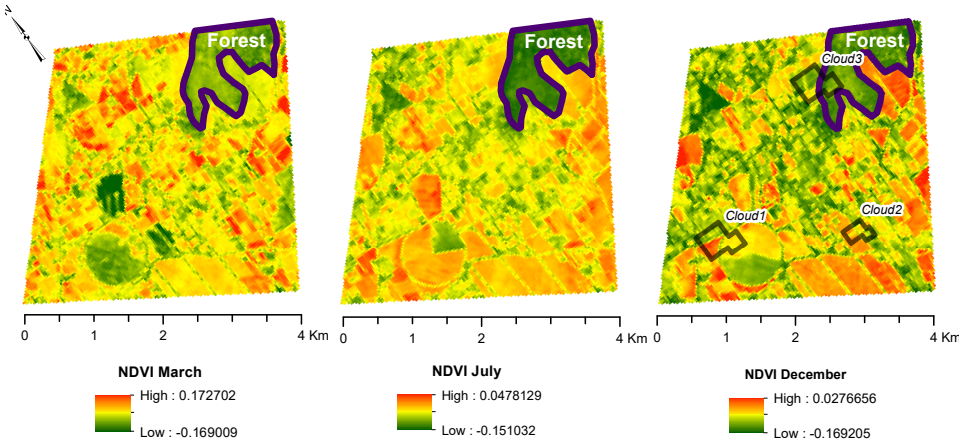


Figure 2: NDVI for March, July and December months

in table 2. Their position and sizes were subjectively chosen in order to estimate pixels in heterogeneous areas (different land cover and phenological estates of the crops).

Table 2: Dimensions syntetic clouds

Cloud	Area (Ha)	Number of pixels
Cloud 1	19	213
Cloud 2	8	82
Cloud 3	19	213

3.3 Exploratory Spatial Data analysis

The Figure 3 shows the distribution of the NDVI values by each month, the concentration of the values is symmetric for March and December. However, the month of July has low frequency of high values that make the empirical distribution to drag slightly to the left (positively skewed). Additionally, the mean of the histograms are different among the months, but the range stays more or less constant (see table 3).

Table 3: Descriptive statistics

Month	Min	1st Qu	Median	Mean	3rd Qu	Max	SD
March	-0.169	-0.048	-0.016	-0.019	0.006	0.172	0.043
July	-0.151	-0.083	-0.057	-0.061	-0.035	0.047	0.034
December	-0.169	-0.116	-0.089	-0.085	-0.056	0.027	0.039

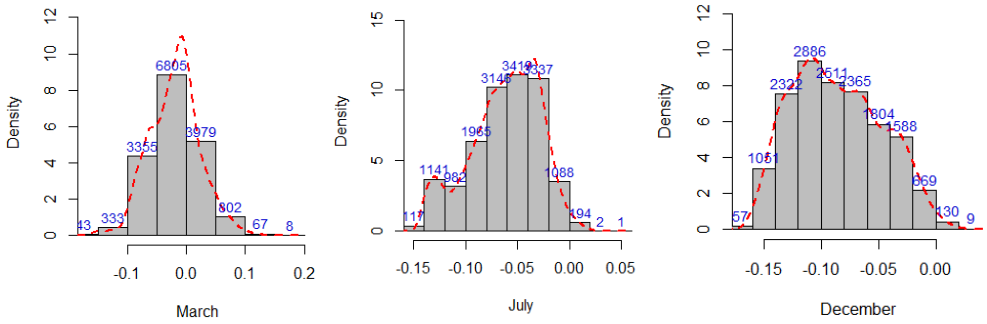


Figure 3: Histograms

The scatter plot of the Figure 4 is showing in the bottom left, the dispersion among the values of the 3 months, and in the top right, their respective correlation coefficients. Therefore, the month of December shows strong correlation with the period that is closer in time, July. However, as the interval increases this linear correlation becomes more diffuse, for example December versus March.

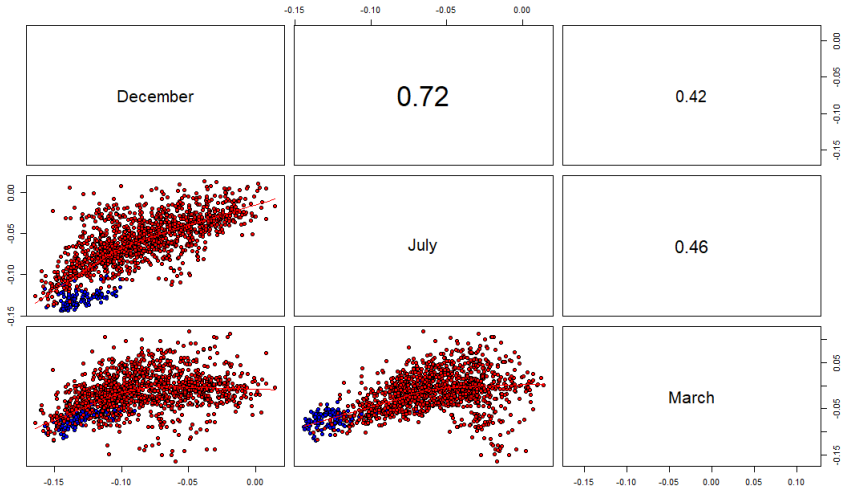


Figure 4: Scatter plot NDVI by month, blue points represent forest and red ones agricultural land

Generally, the vegetation has different phenological characterization of NDVI through the year. Therefore, according with the Figure 5a we can see how the interquartile range of NDVI values stays more or less constant in each month and the global mean values decreases from March to December. In this sense, in order to perform the interpolation, we will remove this seasonal behavior of the data by simply fitting a linear function. The figure 5b shows the result of that process.

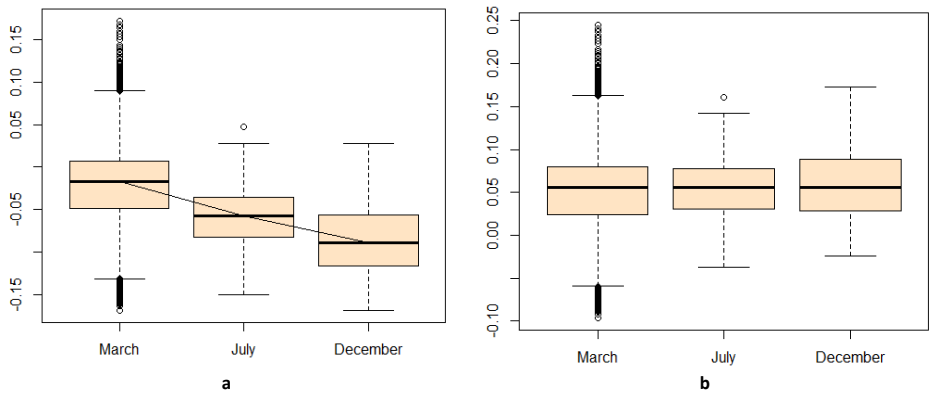


Figure 5: Removing seasonal behavior of the NDVI values

3.4 Inverse distance weighting for predicting pixels values of NDVI

Despite the different unities of the space and time, we did not perform any normalization over the spatial coordinates, we did not consider necessary since the parameter θ will control that difference. On the other hand, the time index was considered as a sequence of integers after removing he seasonal behavior, thus the indexes are December=1, July=2 and March=3.

For applying space-time IDW, we used the library Melo and Melo (2015). This library is part of the packages in **R project** for spatio-temporal analysis. It is mainly characterized for having functions to perform *Inverse Distance Weighting and Radial Basis Functions with Distance-Based Regression.*

3.4.1 Leave one out cross validation

With the purpose of explaining the final parameters used for interpolation, we prepared 6 trials. Each trial present the mean square errors using different combinations of K-neighbors (K:50,30,20,10,5,2), powers and θ parameters. Therefore, Figure 6 allowed us to conclude the following:

- Since the area is quite heterogeneous for the presence of different kinds and sizes of crops, we expected high values of the *power* in presence of high number of neighbors. In this way, the mean square error tend to be minimum when the predictor only take into account the measurements of pixels around the NDVI to predict
- Particularly, the lowest mean square error was obtained using the parameters k=2, p=1.6 and $\theta = 7 \times 10^{-9}$. This configuration, lead us to think that the predictor is only taking into account the contribution of the pixels throughout the time (two images to explain).

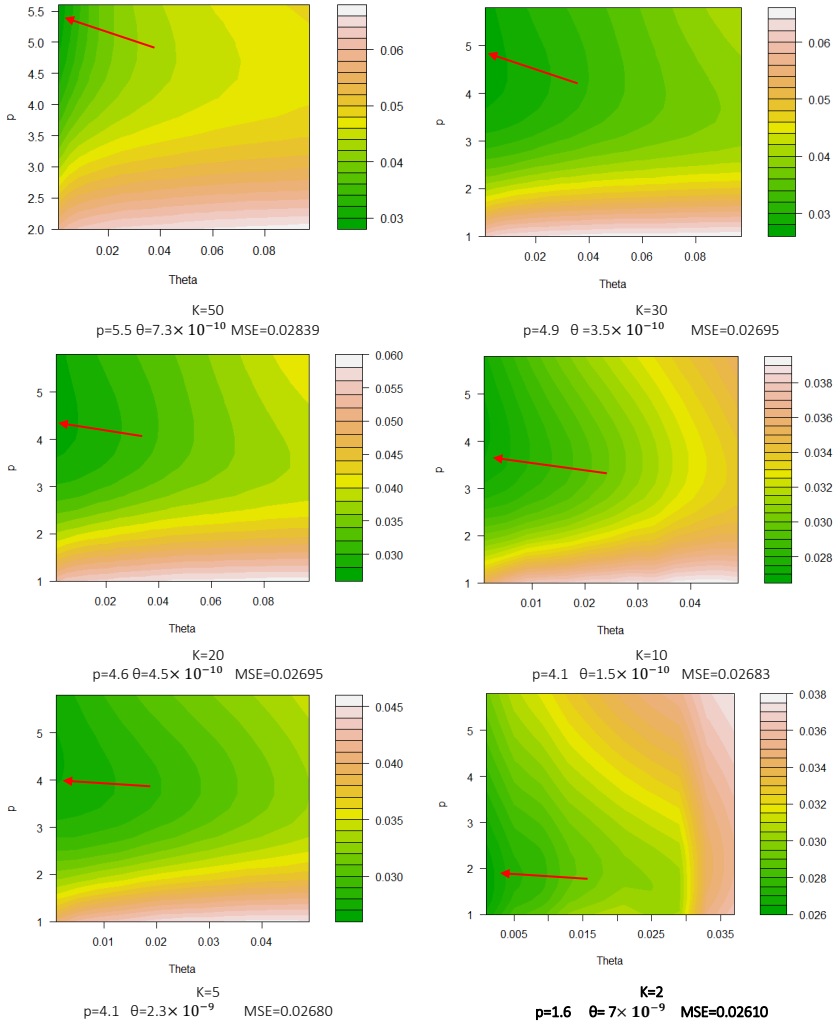


Figure 6: Cross validation

- In the scenario of lowest MSE the power converge to a value of 1.6. Which mean that the predictor is considering also the influence of the farther month, that is to say March. This is possibly due to December in cyclical terms is closer to March than July.

3.5 Prediction of NDVI values

Once performed the optimization of the parameters, we applied IDW to predict the pixel values covered for synthetic clouds. Thereafter, we transformed again the values to the original units of NDVI and finally we obtained the predictions of the Figure 7.

Although visually the interpolations looks very similar to the original, the Figure 8 depicts clearly some errors in the estimation by each month. Generally, through the dispersion plots and the statistics for assessing the error, we can say that there is a underestimation of low values and overestimation of high values of NDVI. .

4 Conclusion

Although images in optical remote sensing can be considered as a continue function, the radiance that is stored in each pixel for the sensor belongs to different targets. As we have seen, these targets in the surface vary in size and reflect in different and unique way the radiance, so the application of the spatial approximation for predicting pixel values must be taken with caution, since the variable is not necessarily continue in space in this kind of rasters.

Finally, trough this paper we have seen how IDW is a versatile tool for interpolation, fast and easy to adapt to new perspectives over the data. Although in the end the spatial component was not representative for our prediction, we think that is an important tool that should be explored deeper, especially in scenarios where the phenomena is continuous and the time domain representative for the analysis.

References

Cerra, D., Bieniarz, J., Müller, R., Reinartz, P., 2016. Cloud removal from sentinel-2 image time series through sparse reconstruction from random samples. In: XXII ISPRS Congress, Technical Commission III. No. XLI-B3. ISPRS Archive, pp. 469–473.

Champion, N., 2016. Automatic detection of clouds and shadows using high resolution satellite image time series. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences 41.

Dozier, J., Painter, T. H., Rittger, K., Frew, J. E., 2008. Time–space continuity of daily maps of fractional snow cover and albedo from modis. Advances in Water Resources 31 (11), 1515–1526.

Gao, B.-C., Montes, M. J., Davis, C. O., Goetz, A. F., 2009. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. Remote Sensing of Environment 113, S17–S24.

Hoan, N. T., Tateishi, R., 2009. Cloud removal of optical image using sar data for alos applications. experimenting on simulated alos data. Journal of The Remote Sensing Society of Japan 29 (2), 410–417.

Li, L., Losser, T., Yorke, C., Piltner, R., 2014. Fast inverse distance weighting-based spatiotemporal interpolation: a web-based application of interpolating daily fine particulate matter pm2. 5 in the contiguous us using parallel programming and kd tree. International journal of environmental research and public health 11 (9), 9101–9141.

Melo, C., Melo, O., 2015. geosptdb: Spatio-Temporal; Inverse Distance Weighting and Radial Basis Functions with Distance-Based Regression. R package version 0.5-0.
URL geosptdb.r-forge.r-project.org/

Schroeder, T., Fischer, J., 2003. Atmospheric correction of meris imagery above case-2 waters. In: Proceedings of the 2003 MERIS User Workshop.

Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment 8 (2), 127–150.

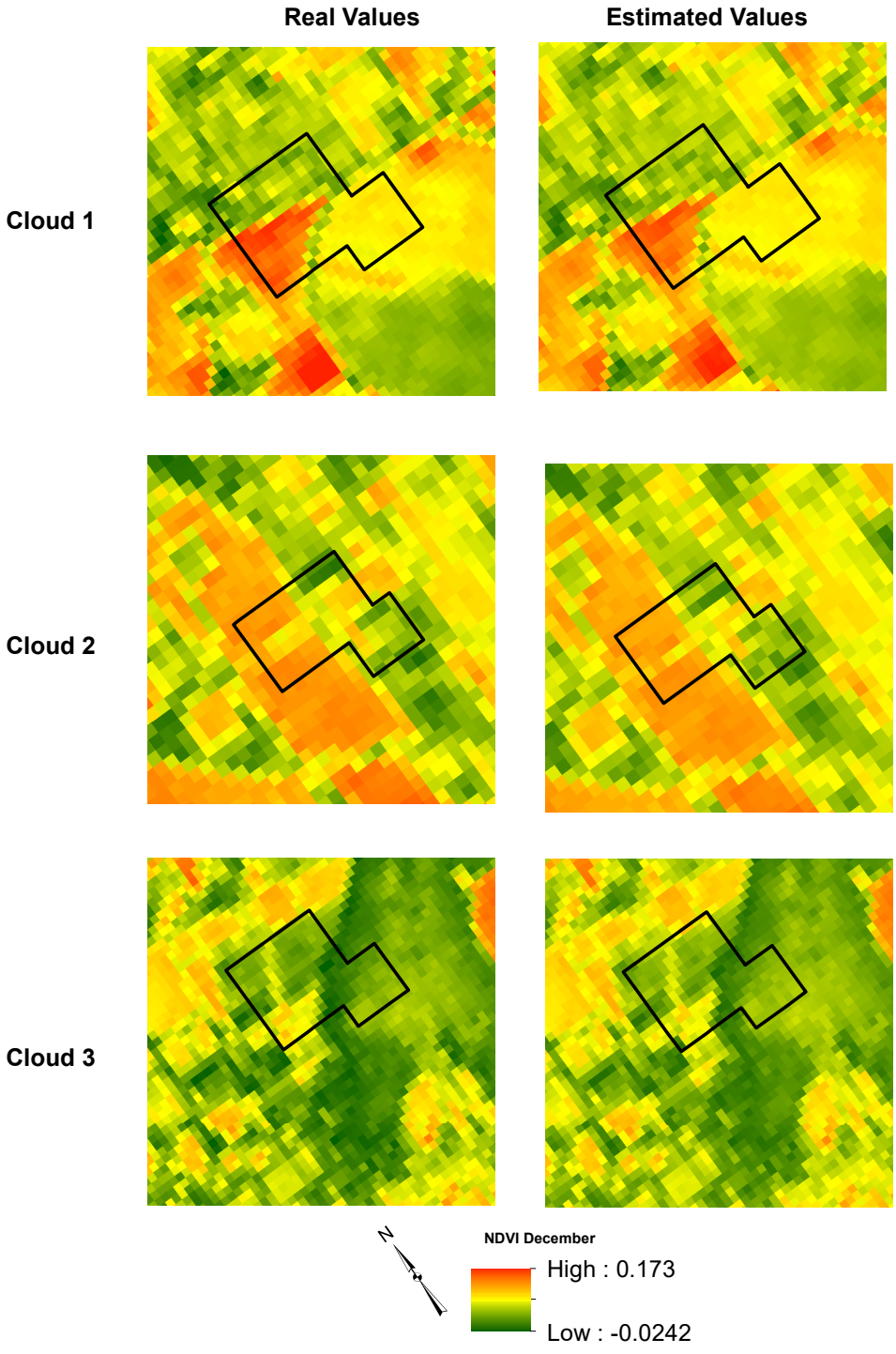
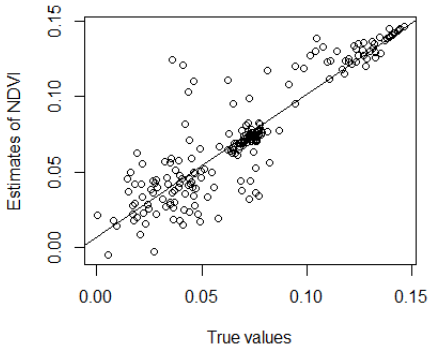
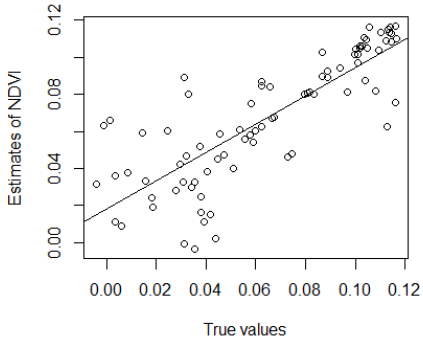


Figure 7: Predictions



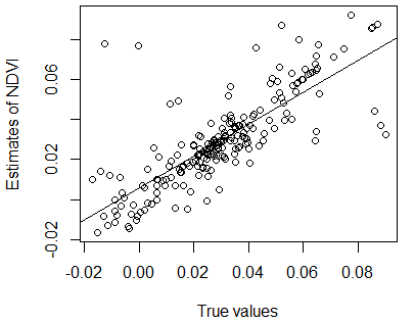
MBE= -7.2e-19 RSME=0.017 Adjusted R^2 0.79

Cloud 1



MBE -3.5e-19 RSME= 0.01815153 Adjusted R^2 0.66

Cloud 2



MBE -8.18e-19 RSME= -8.179e-19 Adjusted R^2 0.63

Cloud 3

Figure 8: Error estimation