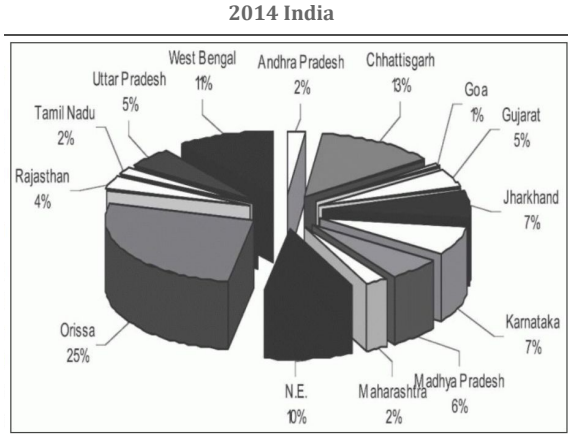
Disease Surveillance

Aditi Singla (2014CS50277)
Prachi Singh (2014CS50289)

Advisor: Prof. Aaditeshwar Seth

Background

- Spread of various diseases like Malaria, Dengue, Typhoid etc. across India
- Dependance of spread of such diseases on various factors
  - Population, education, employment, urbanisation, migration patterns, media, government interventions, heat strokes, climate change



Tracking longitudinal relationships between diseases and urbanisation patterns

Objectives

- Extracting and classifying built-up & non-built-up areas for disease prone areas
  - Using Satellite Imagery Data to see urbanisation trends over the years
- Tracking relationships between diseases and extent of urbanisation
  - Understanding their correlation strength using statistical tests

Literature Survey

- Tatem, Andrew J., *et al.* "**Defining approaches to settlement mapping for public health management in Kenya using medium spatial resolution satellite imagery.**" Remote Sensing of Environment 93.1-2 (2004): 42-52.
- Bischke, Benjamin, *et al.* **"Multi-task learning for segmentation of building footprints with deep neural networks."** arXiv preprint arXiv:1709.05932 (2017).

Classification of built-up & non-built-up areas

DATA:

- 20K datapoints by Goldblatt et al., marked as built-up & non -built-up
- LANDSAT Satellite Imagery: “*Landsat 8 Collection 1 Tier 1 TOA Reflectance*”
- Available for the years (2013-18) - 16 days Revisit Interval
- 11 Bands with resolution of 30m, based on wavelengths (0.43µm - 12.51µm)

CLASSIFIER:

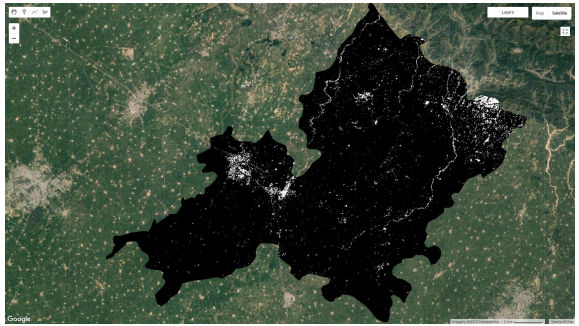
- Using Google Earth Engine
- CART Classifier (Classification And Regression Trees)
- Features Used: 11 Available Bands & customised bands (NDVI, NDBI, GSW, etc)

Issues with the available LandSat Images

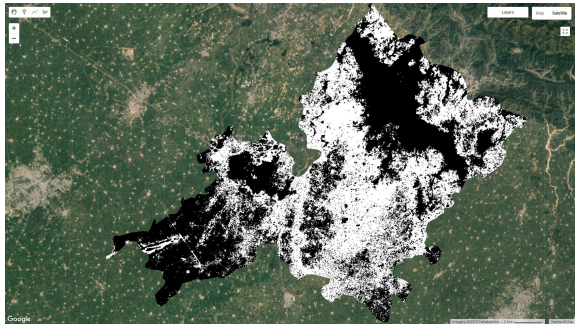
- Initial Experimentation:
  - Filter out comparable number of built-up & non-built-up points, to remove bias
  - Limit number of images to minimum cloud cover
- Urbanisation for images from 2015 unexpectedly high
- Further Experimentation:
  - For 2015, cloud cover significant, with high values of cloud scores
  - Use of Cloud masks, by Rodrigo E. Principle

AMBALA

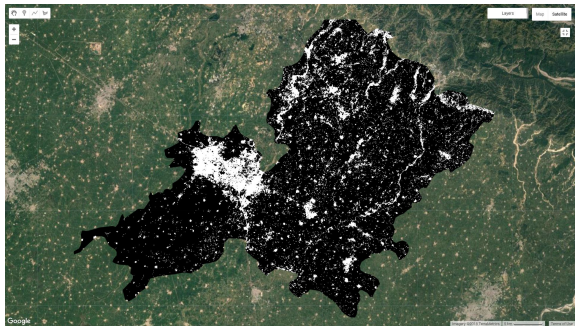
Simple classification



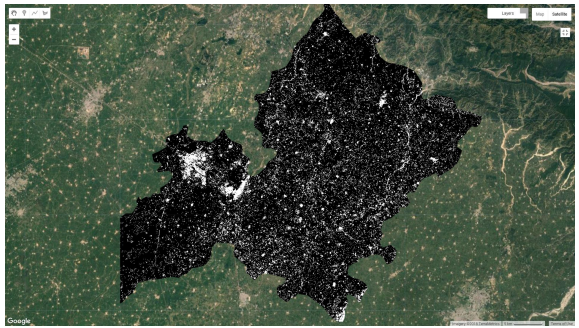
2014



2015

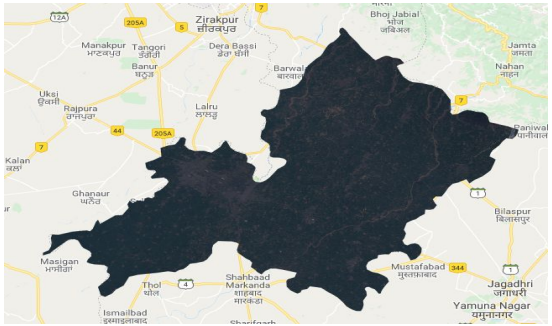


2016



2017

Original LandSat images



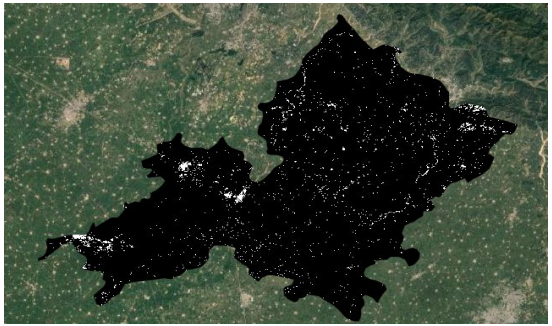
2014

2015

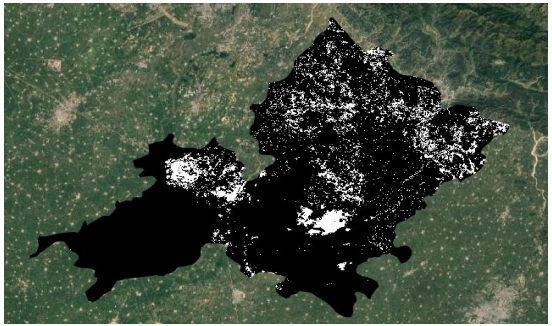
2016

2017

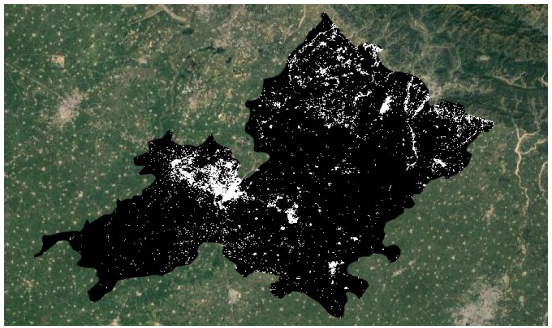
Classification after accounting for cloud cover using mosaic



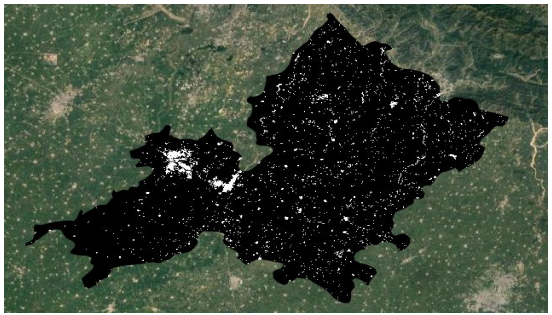
2014



2015

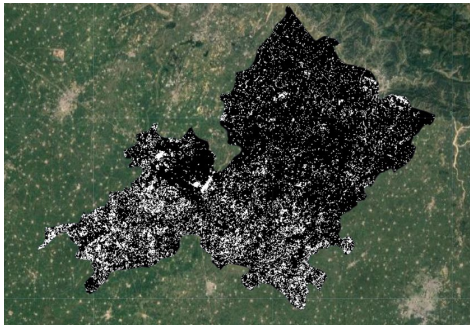


2016

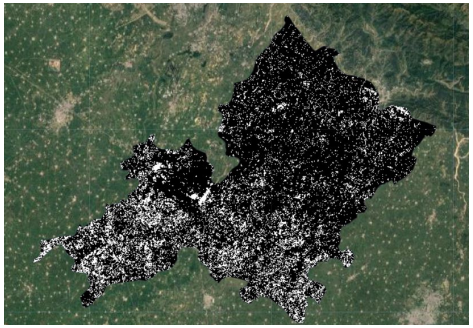


2017

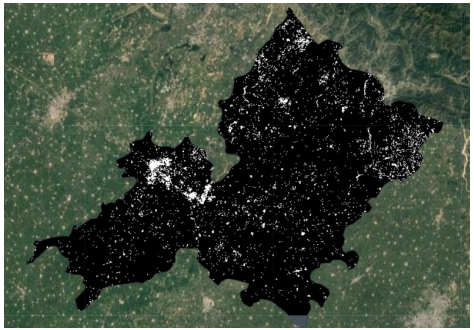
Classification after accounting for cloud cover using median



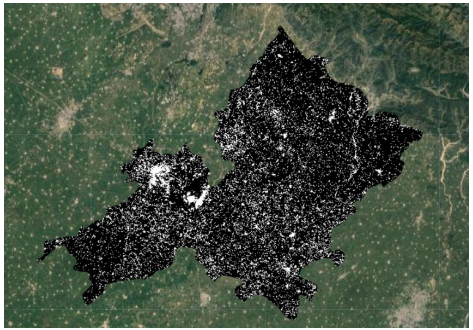
2014



2015



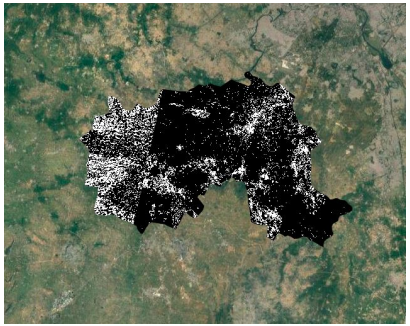
2016



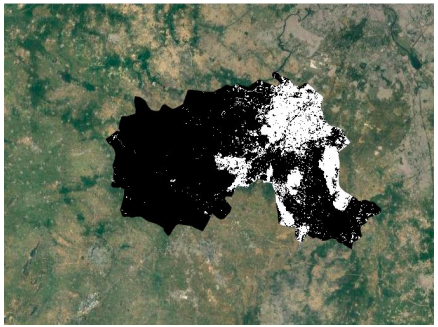
2017

GURGAON

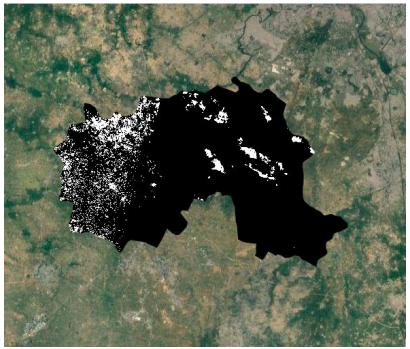
Simple classification



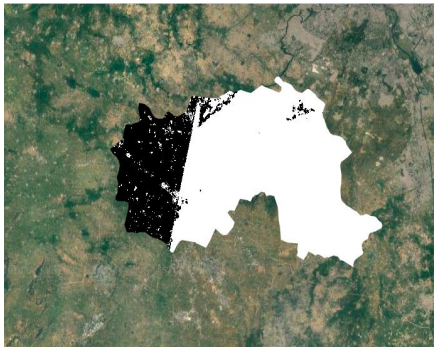
2014



2015

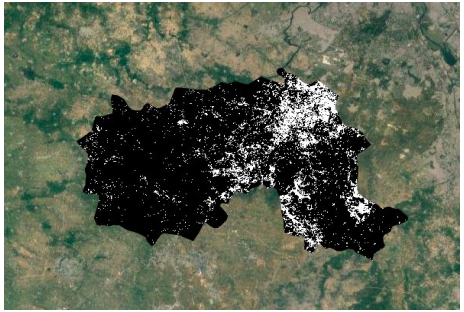


2016

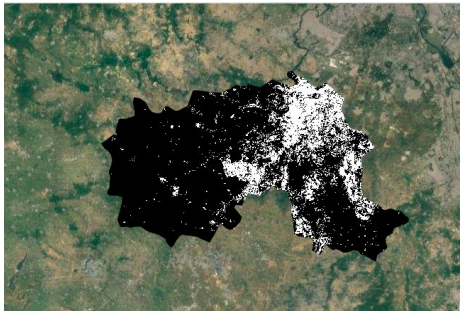


2017

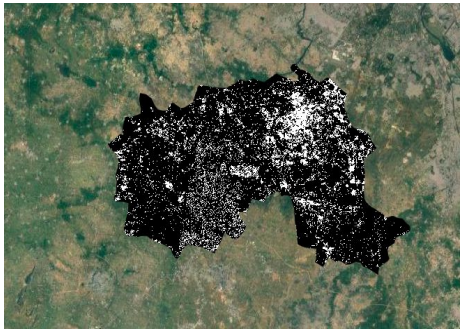
Classification after accounting for cloud cover using mosaic



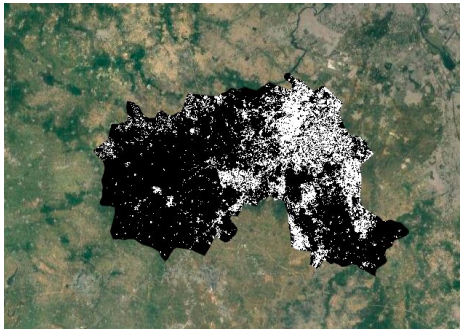
2014



2015

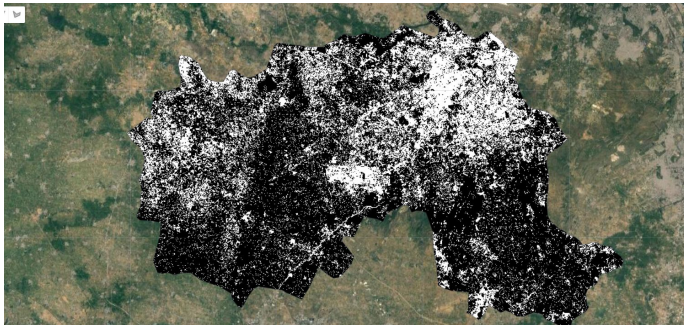


2016

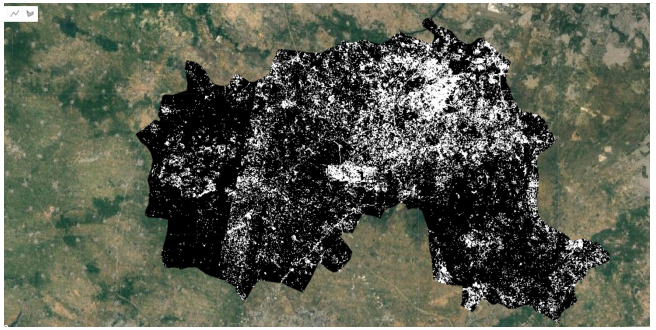


2017

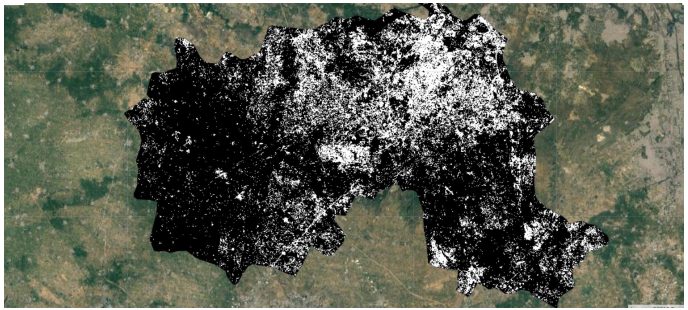
Classification after accounting for cloud cover using median



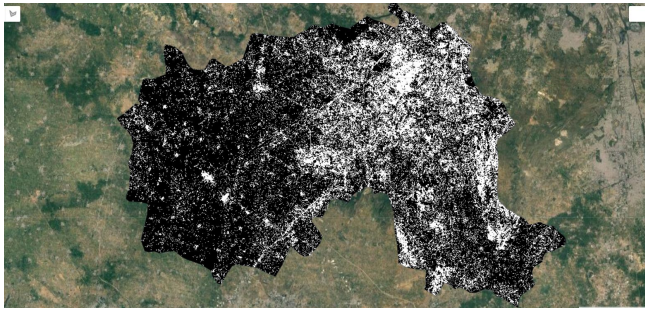
2014



2015



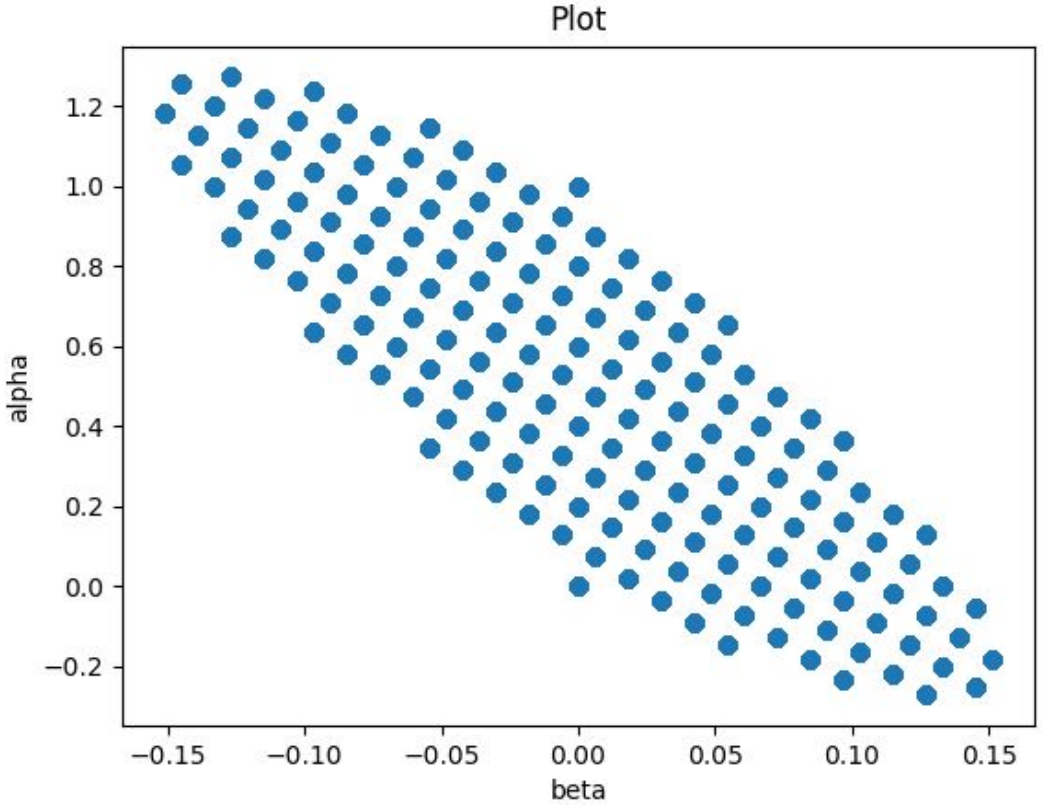
2016



2017

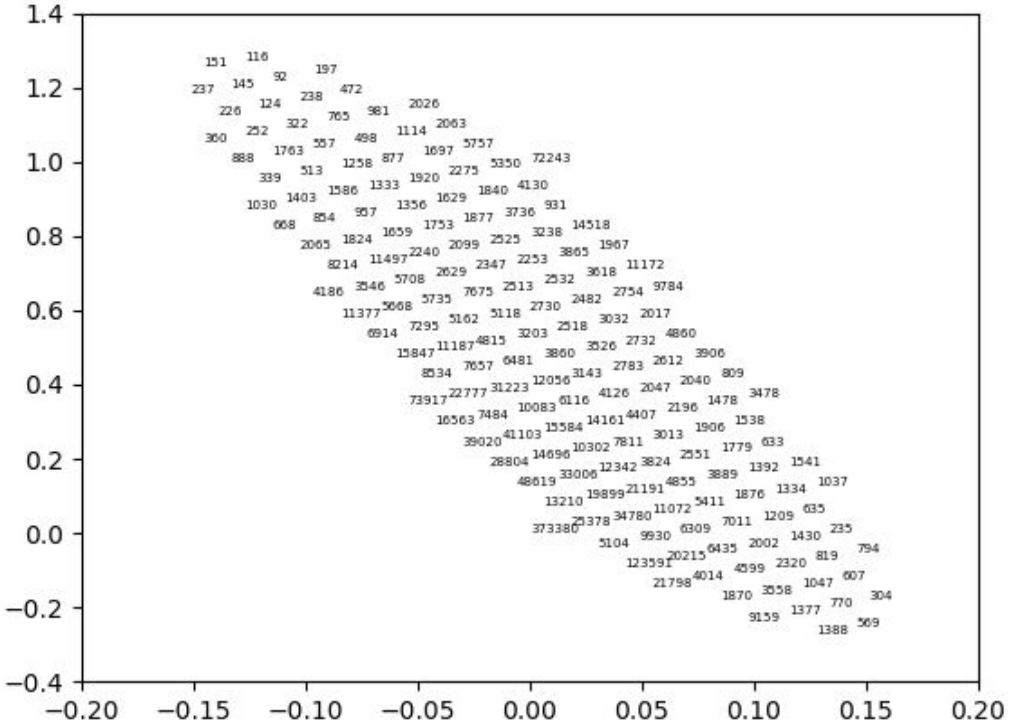
ANALYSING THE CLASSIFICATION RESULTS

Using Regression Analysis



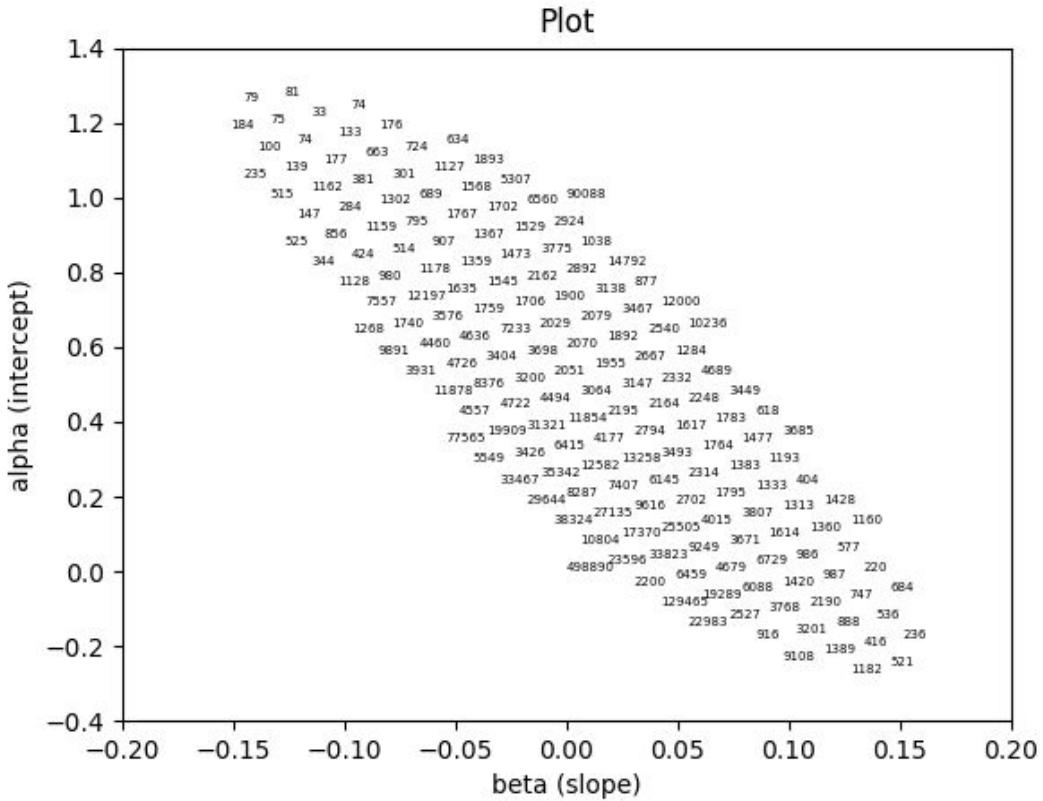
- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting alpha (intercept) vs beta (slope) for each regression line obtained, on a graph
- **Many points have same values of alpha and beta due to same combination of values**

Using Regression Analysis



- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting the alpha (intercept) vs beta (slope) for each regression line obtained, on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **The counts are quite evenly distributed across most points**

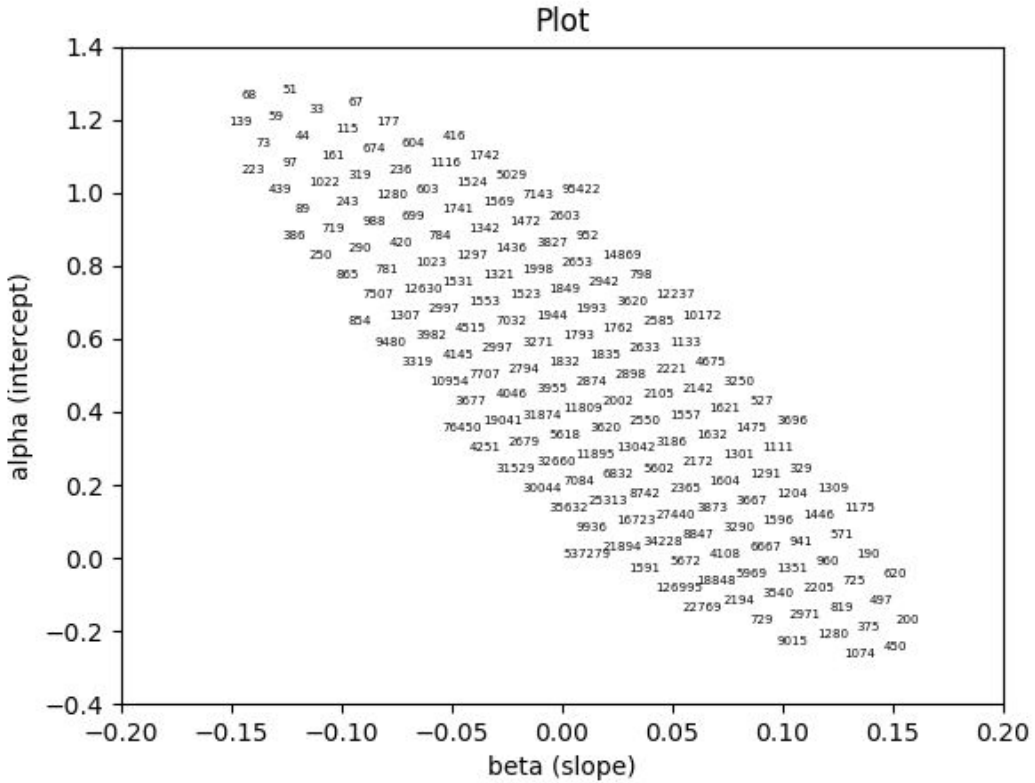
Spatial smoothing using Neighbourhood count



MAJORITY COUNT THRESHOLD 2:

- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting alpha (intercept) vs beta (slope) for each regression line obtained, on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **With thresholding, less points occur with negative slope**

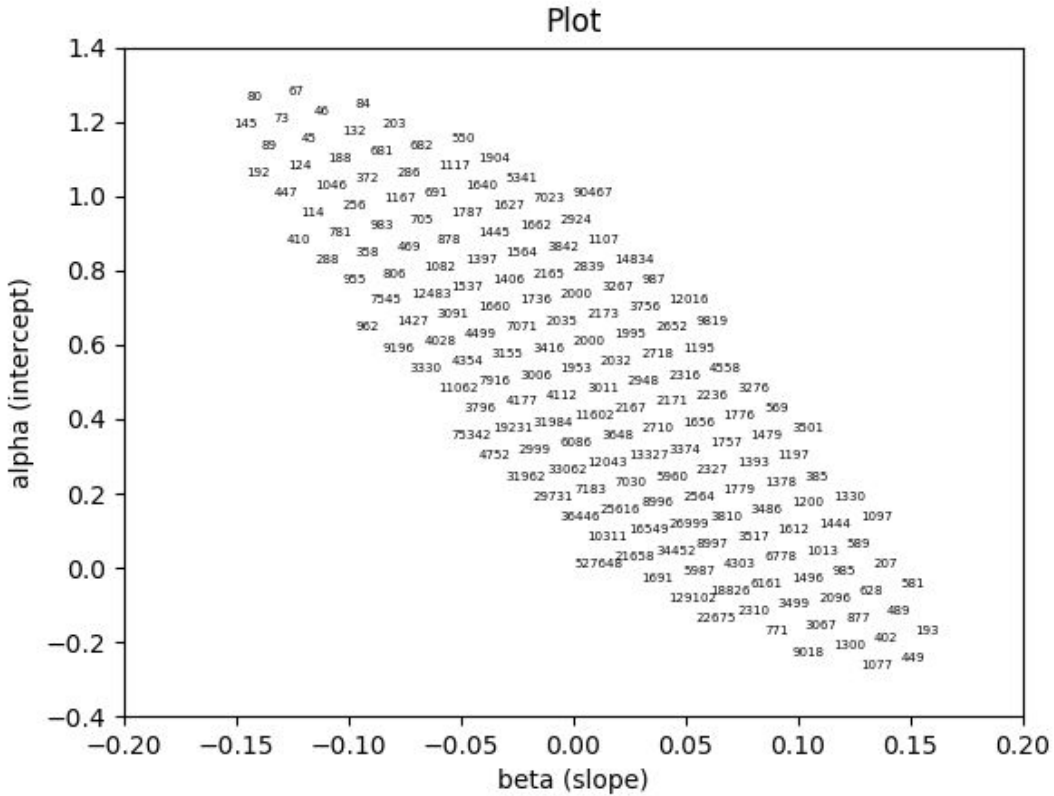
Spatial smoothing using Neighbourhood count



MAJORITY COUNT THRESHOLD 3:

- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting alpha (intercept) vs beta (slope) for each regression line obtained, on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **Count of points with negative slope decreases which means less discrepancies**

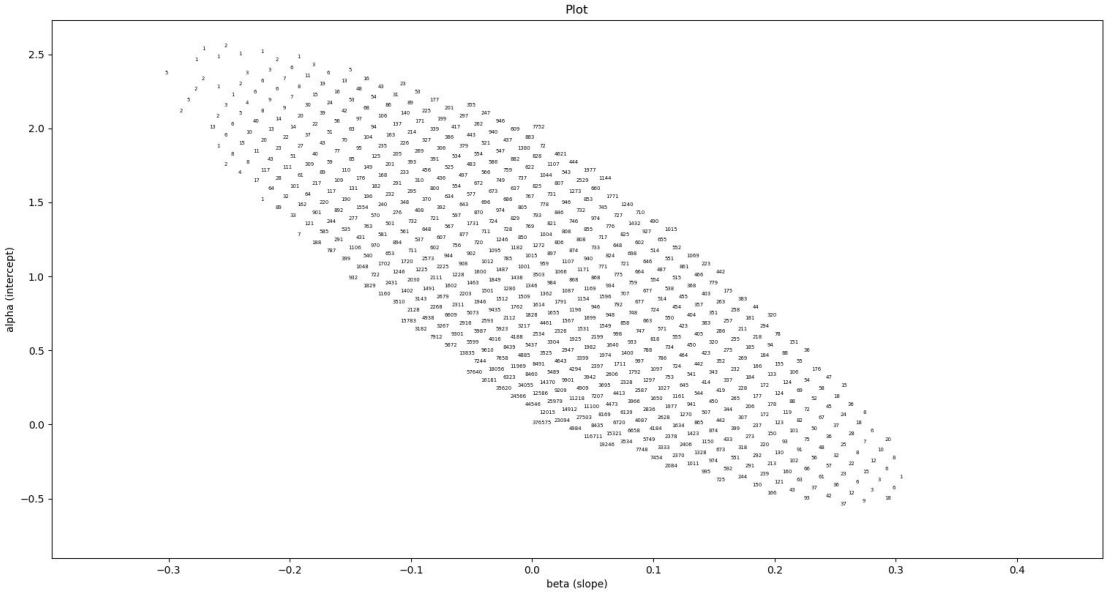
Spatial smoothing using Neighbourhood count



MAJORITY COUNT THRESHOLD 4:

- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting the alpha (intercept) and beta (slope) for each regression line obtained on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **Count of points with negative slope should decrease even further, but aggressive threshold leads to more noise**

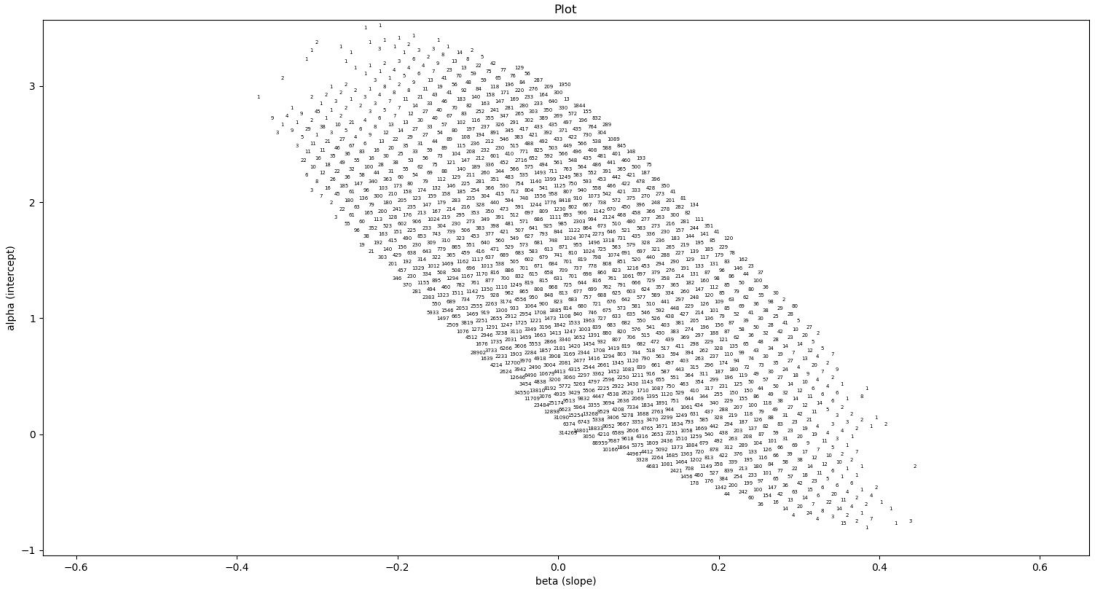
Spatial smoothing using Built-up score



BUILT-UP SCORE USING 8 NEIGHBOURS:

- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting alpha (intercept) vs beta (slope) for each regression line obtained on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **Number of combinations of alpha and beta increases as regression score becomes continuous than distinct**

Spatial smoothing using Built-up score



BUILT-UP SCORE USING 16 NEIGHBOURS:

- Plotting score (0 for NB & 1 for B) vs year (1 - 10 for images from 2013 II, 2014 I, 2014 II, ...2017 II, 2018 I)
- Generating a regression line for each pixel
- Plotting the alpha (intercept) and beta (slope) for each regression line obtained on a graph
- Plotting the counts of each combination of alpha-beta values obtained
- **Number of combinations of alpha and beta increases even further**

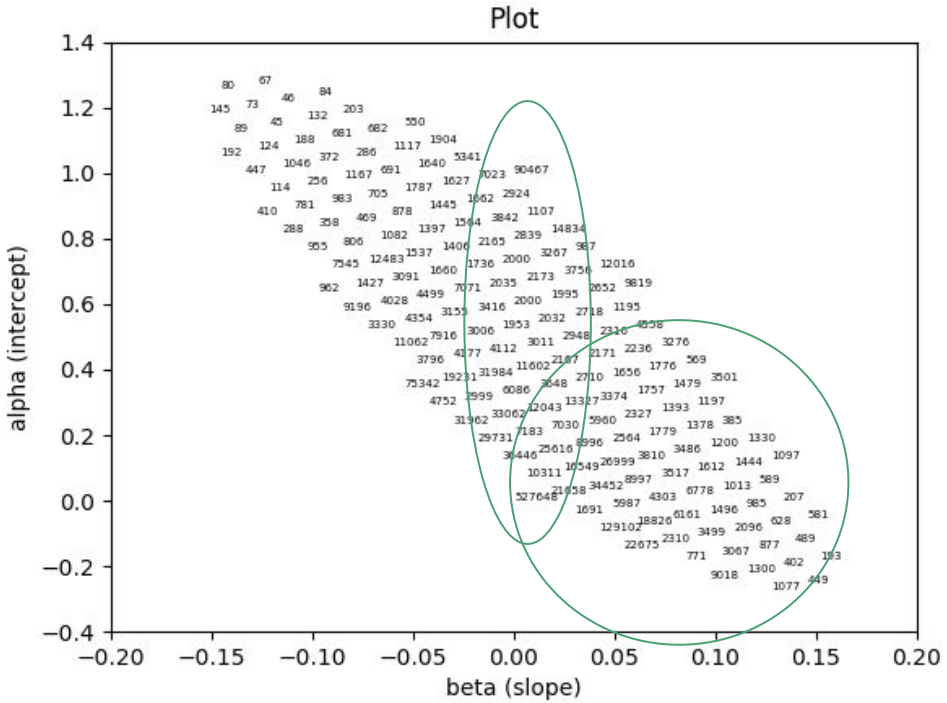
Analysis of alpha-beta points

Favourable points:

- **Positive beta & small values of alpha**: Transition of non-built-up to built-up across the years.
- **Zero beta:** No change in pixel throughout, remains as it is.

Unfavourable points:

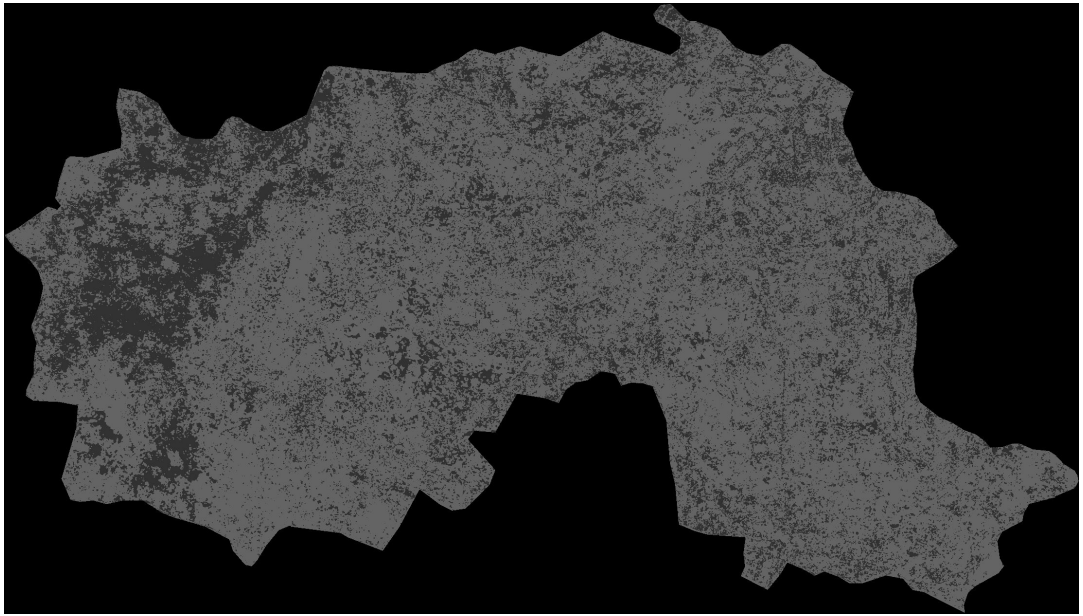
- Negative beta, especially when alpha is too high



Identifying the points wrongly classified

Using Thresholding on neighbour count

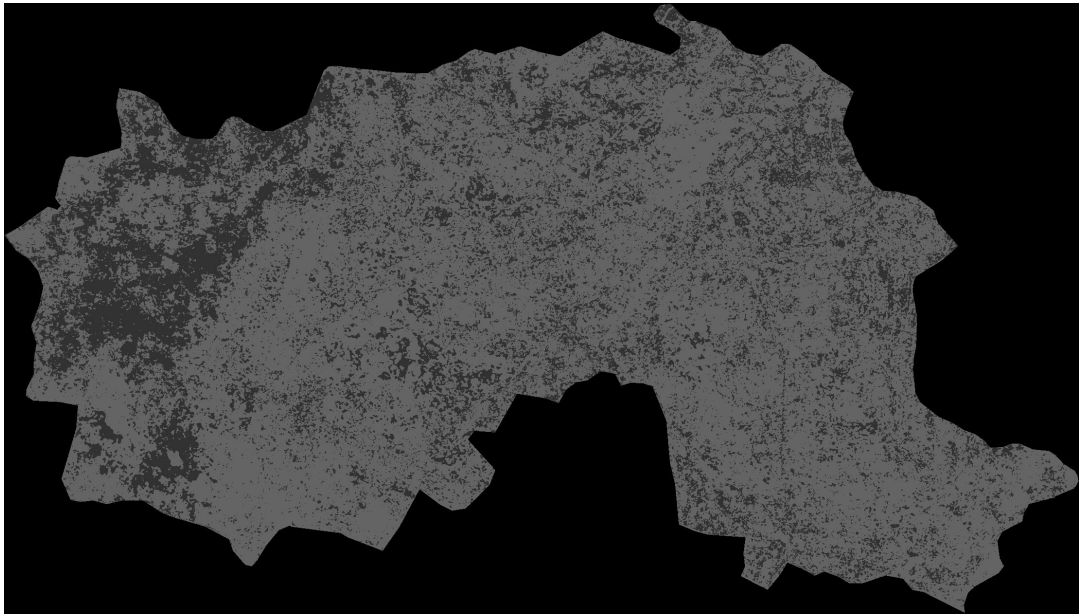
- Threshold = 2
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density is way too high
- High noise zones
- Counts:
    - Positives: 1173105
    - Negatives: 450816



Identifying the points wrongly classified

Using Thresholding on neighbour count

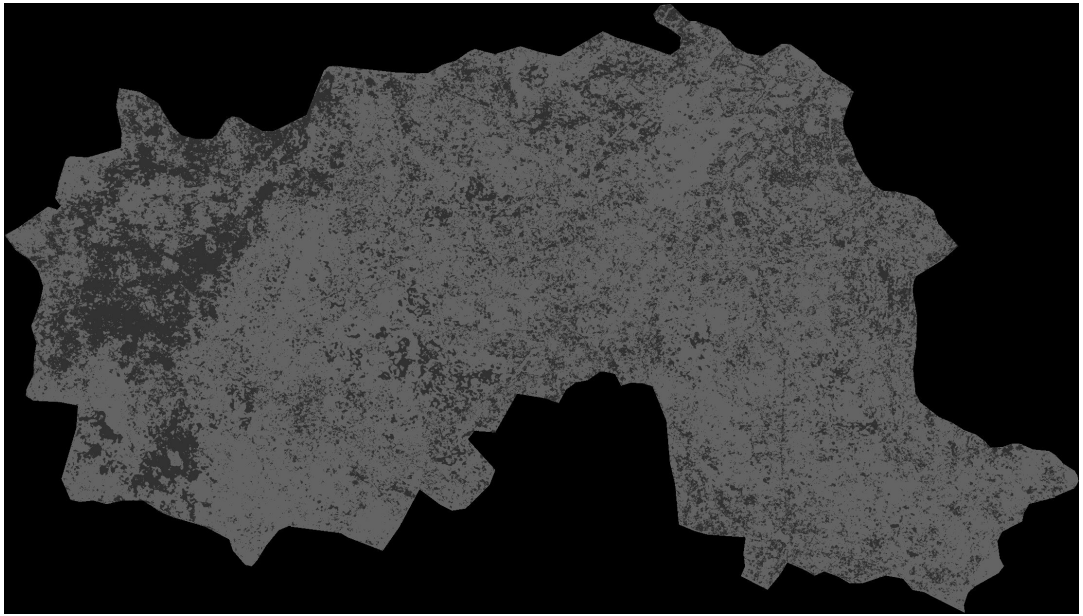
- Threshold = 3
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density is way too high
- High noise zones
- Counts:
    - Positives: 1197145
    - Negatives: 426776



Identifying the points wrongly classified

Using Thresholding on neighbour count

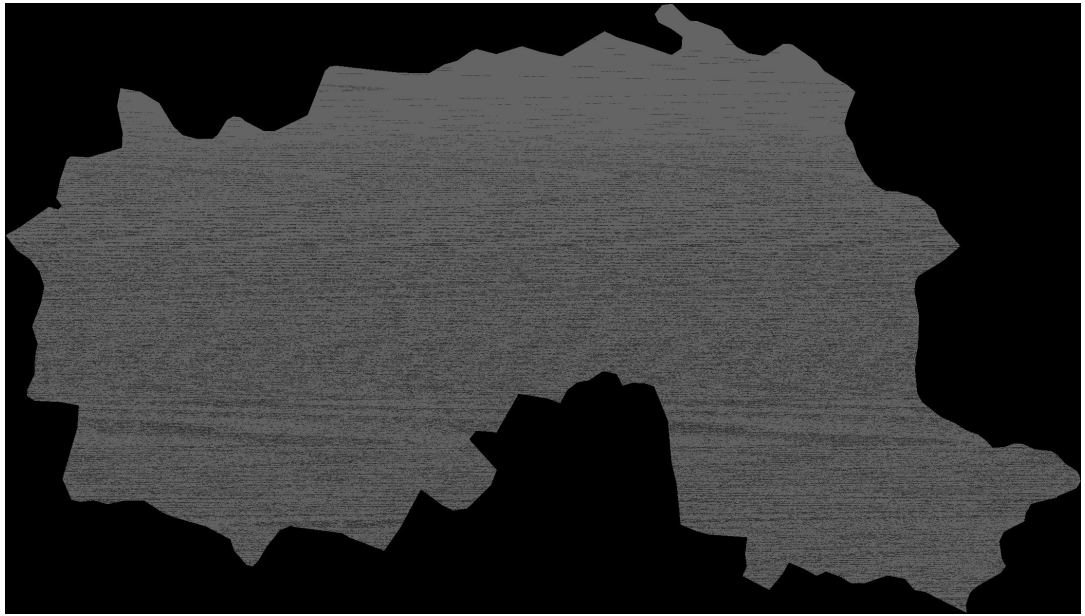
- Threshold = 4
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density is way too high
- High noise zones
- Counts:
    - Positives: 1190460
    - Negatives: 433461



Identifying the points wrongly classified

Using Built-up Score

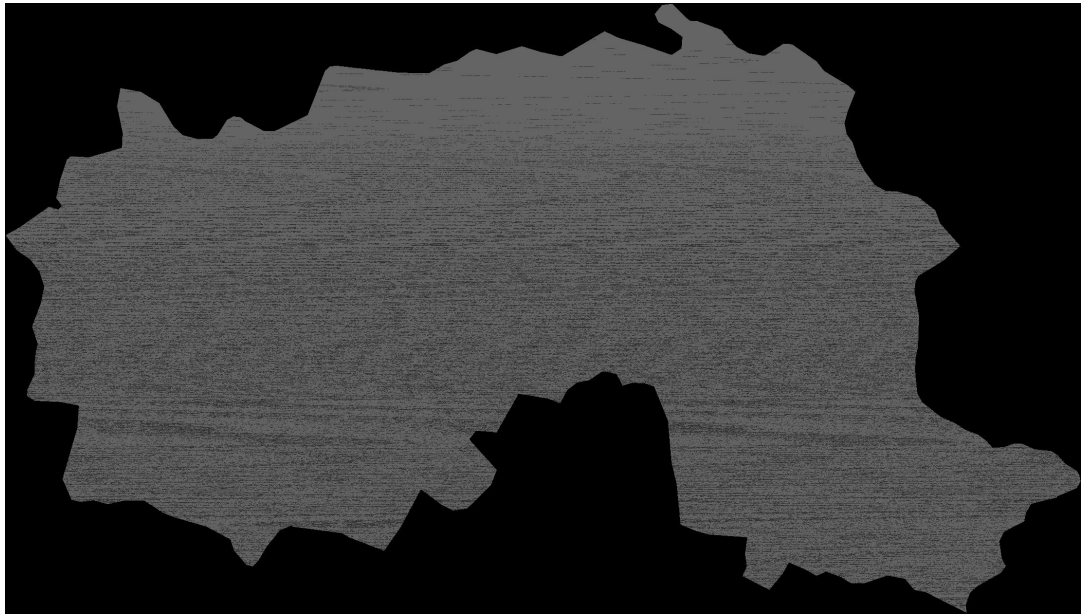
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise is spread uniformly, can be ignored for our further investigation on its correlation with disease spread
- Counts:
    - Positives: 1232998
    - Negatives: 390923



Identifying the points wrongly classified

Using Built-up Score

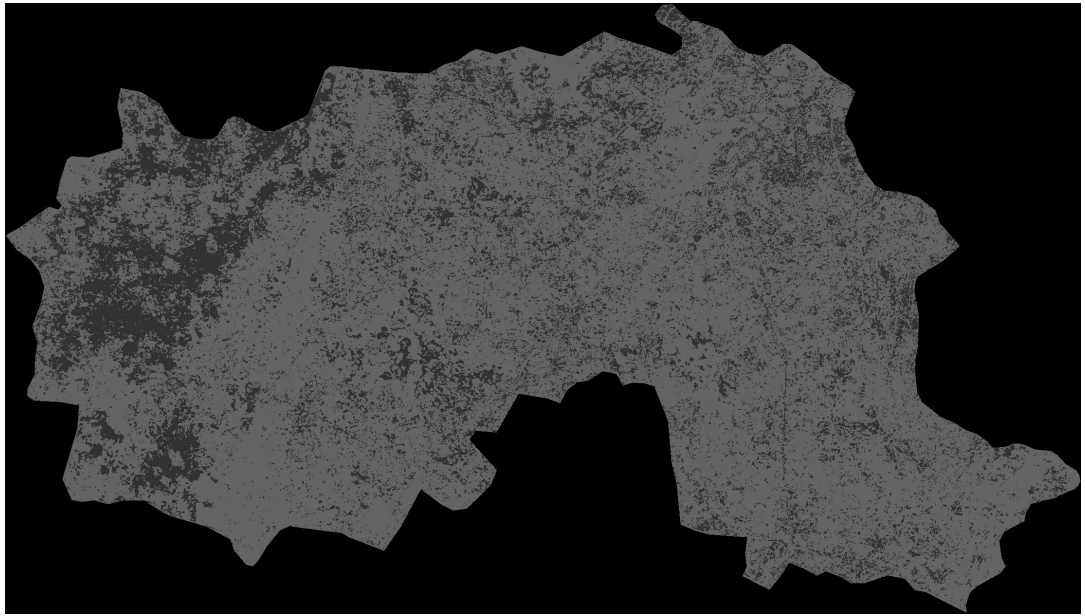
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise is spread uniformly, can be ignored for our further investigation on its correlation with disease spread
- Any possible reasoning for noises?
    - Cloud Cover?



Identifying the points wrongly classified

Using Thresholding on neighbour count

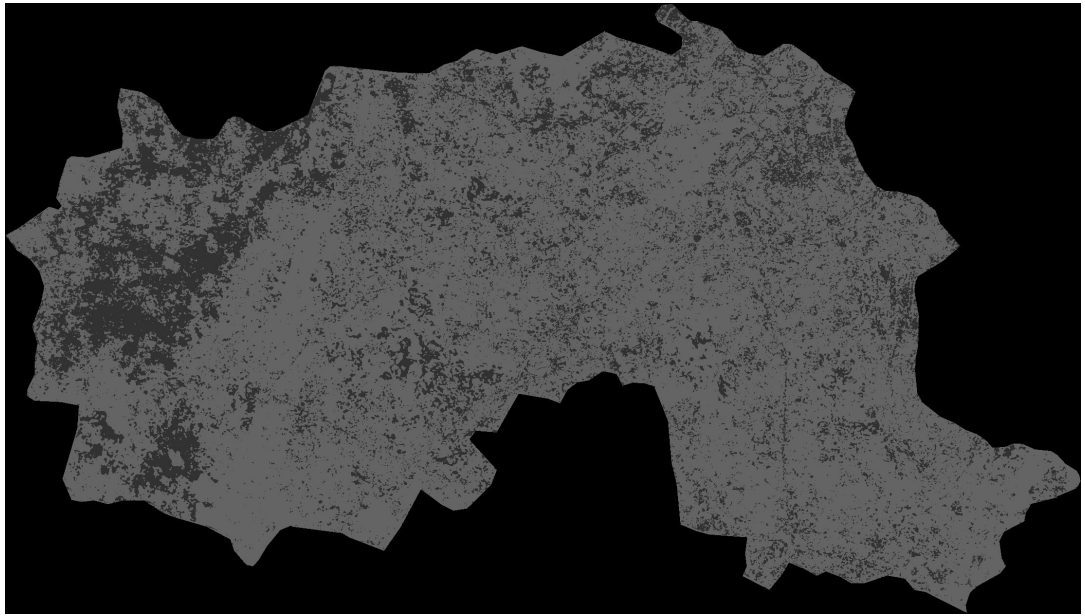
- **Eliminate the images for year 2015**
- Threshold = 2
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density high with noise zones
- Better than previous plots though
- Counts:
    - Positives: 1239672
    - Negatives: 384249



Identifying the points wrongly classified

Using Thresholding on neighbour count

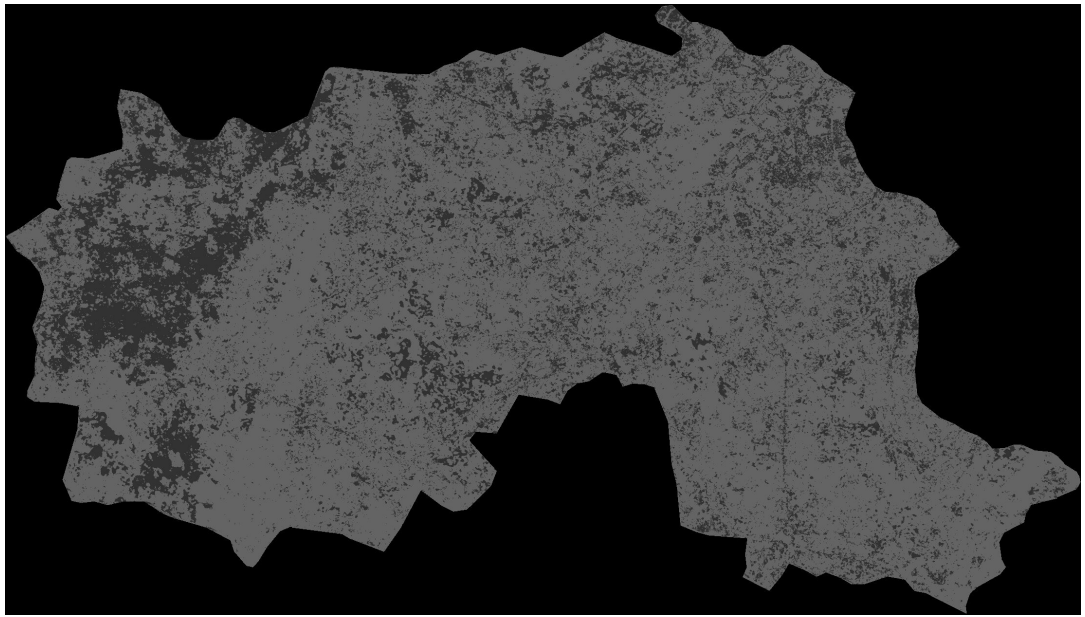
- **Eliminate the images for year 2015**
- Threshold = 3
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density high with noise zones
- Better than previous plots though
- Counts:
    - Positives: 1261377
    - Negatives: 362544



Identifying the points wrongly classified

Using Thresholding on neighbour count

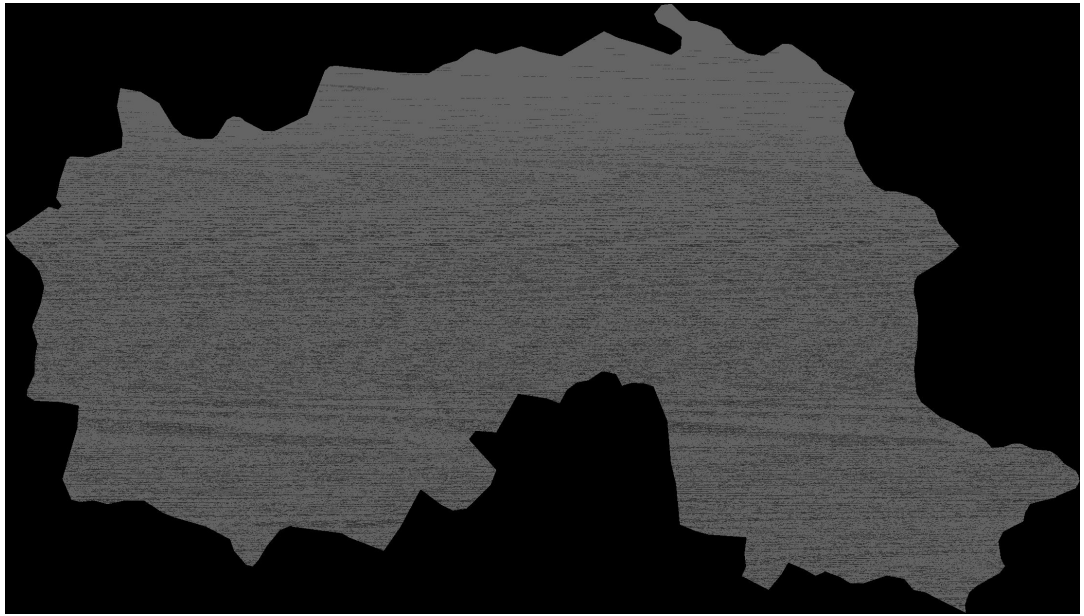
- **Eliminate the images for year 2015**
- Threshold = 4
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise density high with noise zones
- Better than previous plots though
- Counts:
    - Positives: 1255899
    - Negatives: 368022



Identifying the points wrongly classified

Using Built-up Score

- **Eliminate the images for year 2015**
- Black noise represent the points with net negative progress (as per classifier) over 5 years
- Noise is spread uniformly, but has fallen wrt to the previous plot
- Counts:
    - Positives: 1278962
    - Negatives: 344959



Counts for positives & negatives

S.No	Spatial Smoothing Approach		Positives	Negatives	Comments
1.	Neighbour count	Threshold = 2	1173105	450816	Noise zones
2.		Threshold = 2 (Cloud)	1239672	384249	
3.		Threshold = 3	1197145	426776	
4.		Threshold = 3 (Cloud)	1261377	362544	
5.	Built-up Score	With 2015 images	1232998	390923	Uniform
6.		Without 2015 images	1278962	344959	

Further Plan of Action

1. Analysis and approximation on the wrongly classified points
2. Account for cloud cover to improve the classification
3. Use the above approximation for finding correlation with disease spread

Thank you