Comparison of Remotely Sensed and Volunteered Geographic Information for water reservoirs

Maria Moscholaki
November 2020

Supervisors:
Clara Garcia-Sanchez
Balázs Dukai

Gennadii Donchyts
Christine Rogers

TU Delft

Deltares
Enabling Delta Life

Outline

- Introduction
- Methodology
- Implementation and Results
- Sensitivity Analysis
- Conclusions and Future Work

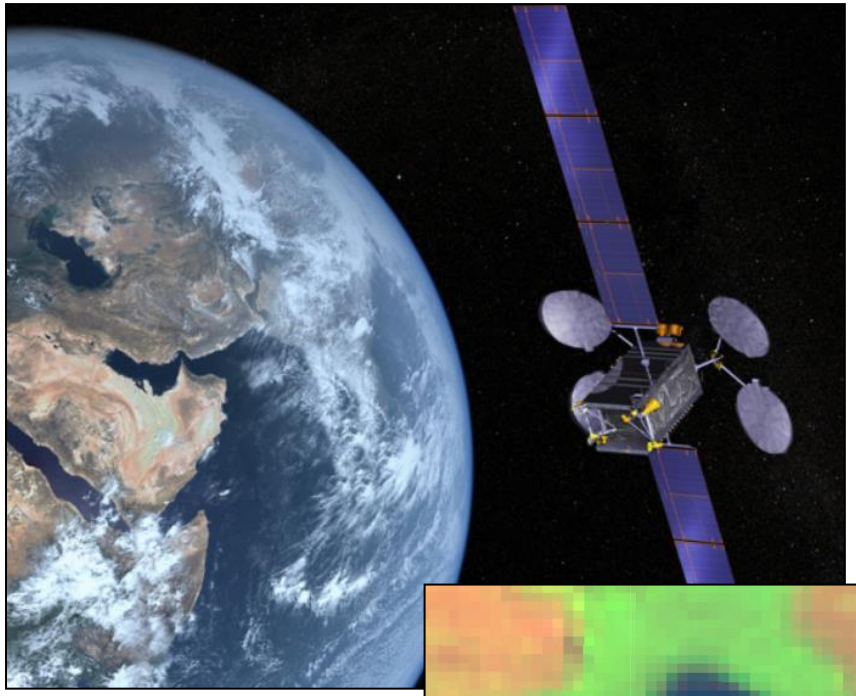
Outline

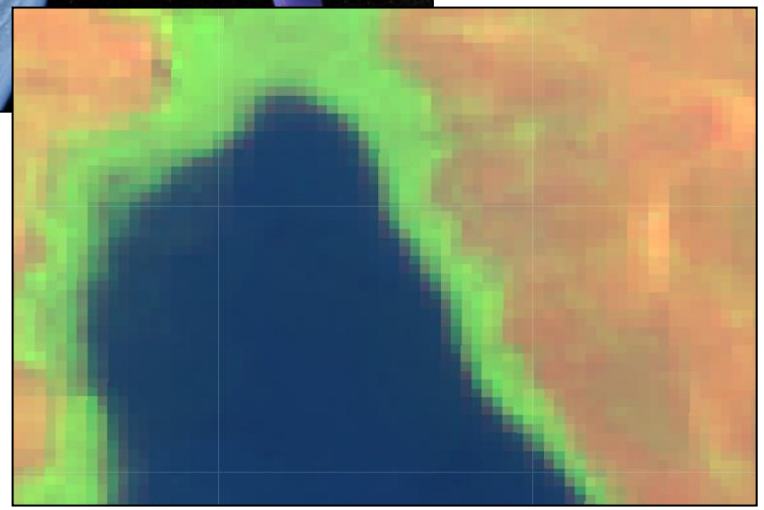
- Introduction
- Methodology
- Implementation and Results
- Sensitivity Analysis
- Conclusions and Future Work

Introduction
Motivation

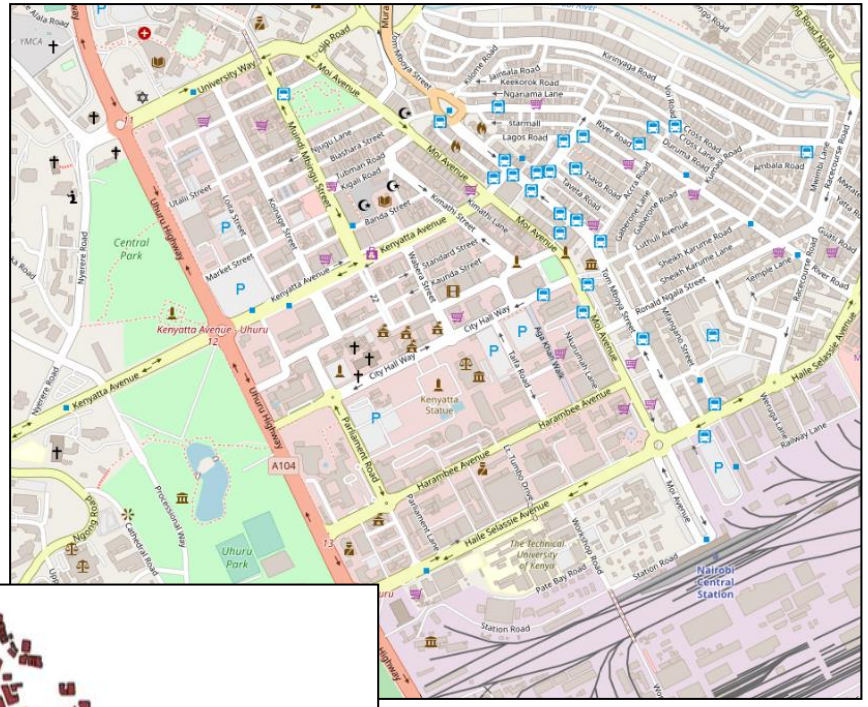
2 Types of Formats

Raster





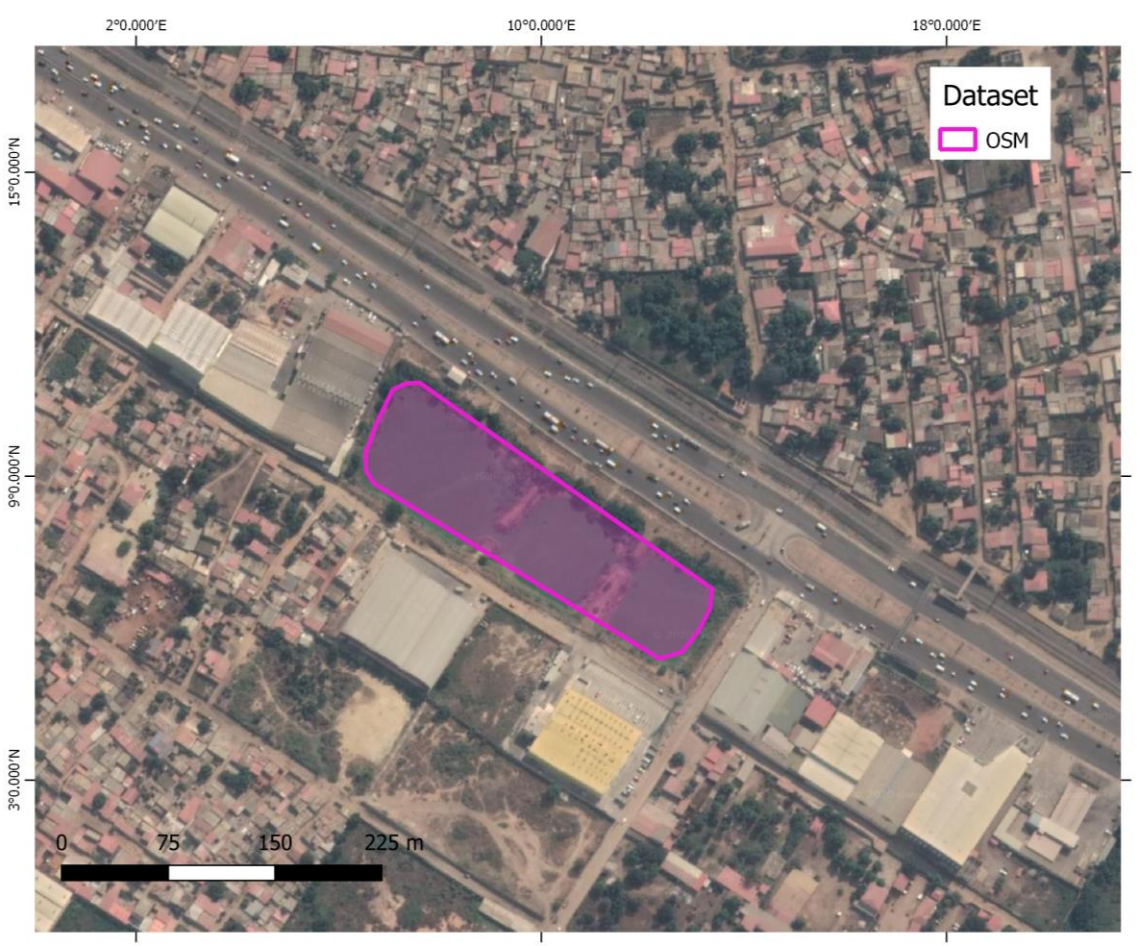
Vector



Introduction
Problem Statement

Data Dis-similarities

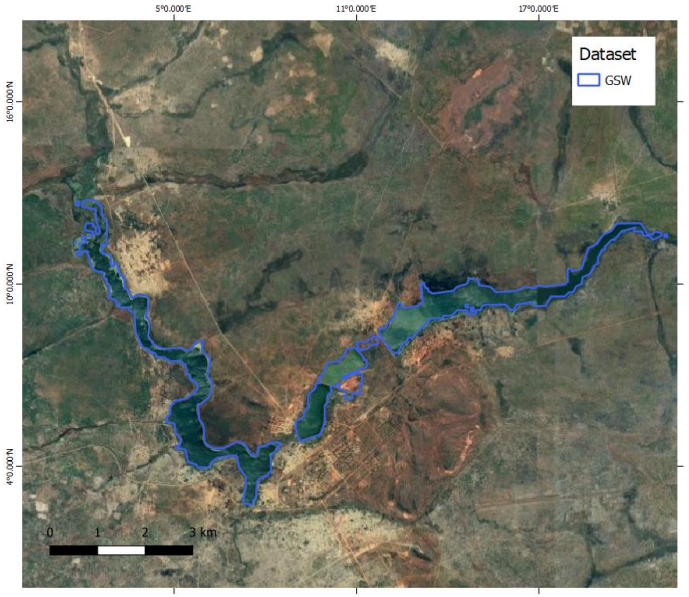
Data Dis-similarities

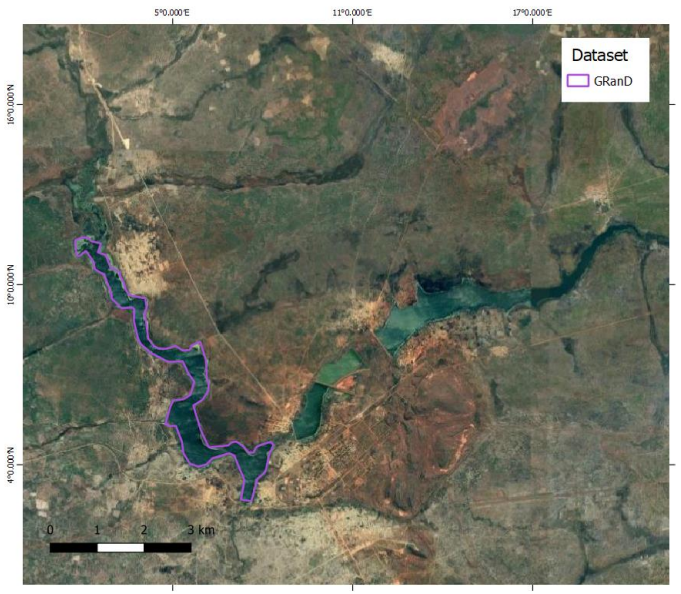


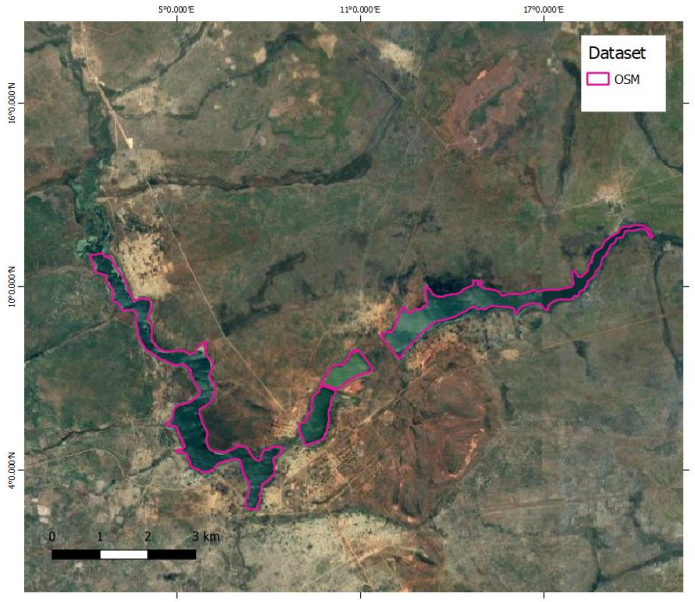
Data Dis-similarities

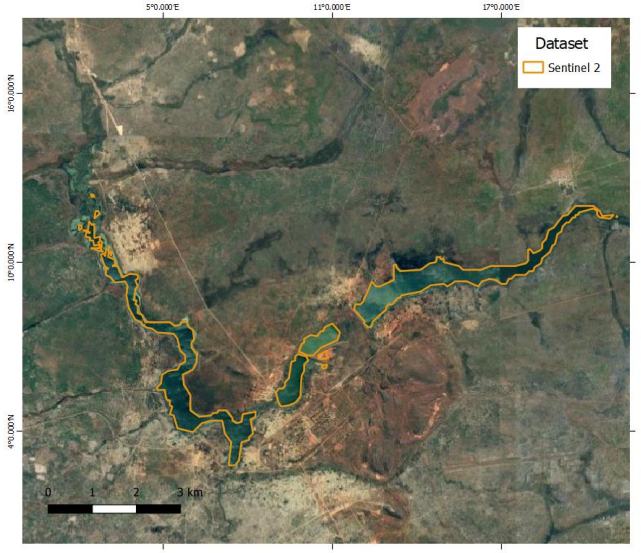
Introduction
Problem Statement

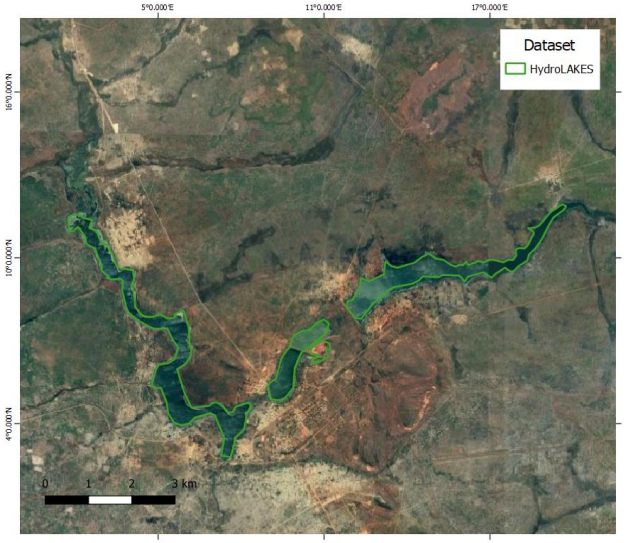
Data Dis-similarities











Research questions

What are the spatial differences between Earth Observation based and Volunteered Geographic Information for water reservoirs and how can they be addressed in an automated way at a large scale?

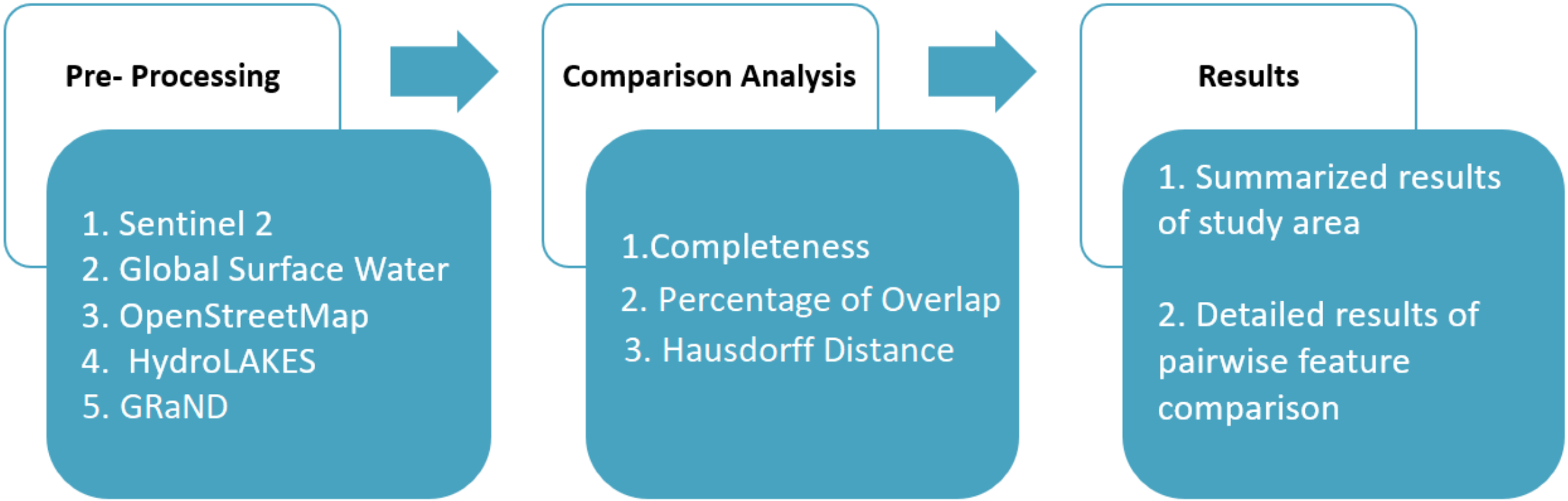
- *What are the differences in terms of spatial coverage?*
- *What are the differences in terms of positional accuracy?*

Outline

- Introduction
- Methodology
- Implementation and Results
- Sensitivity Analysis
- Conclusions and Future Work

Methodology

Flowchart



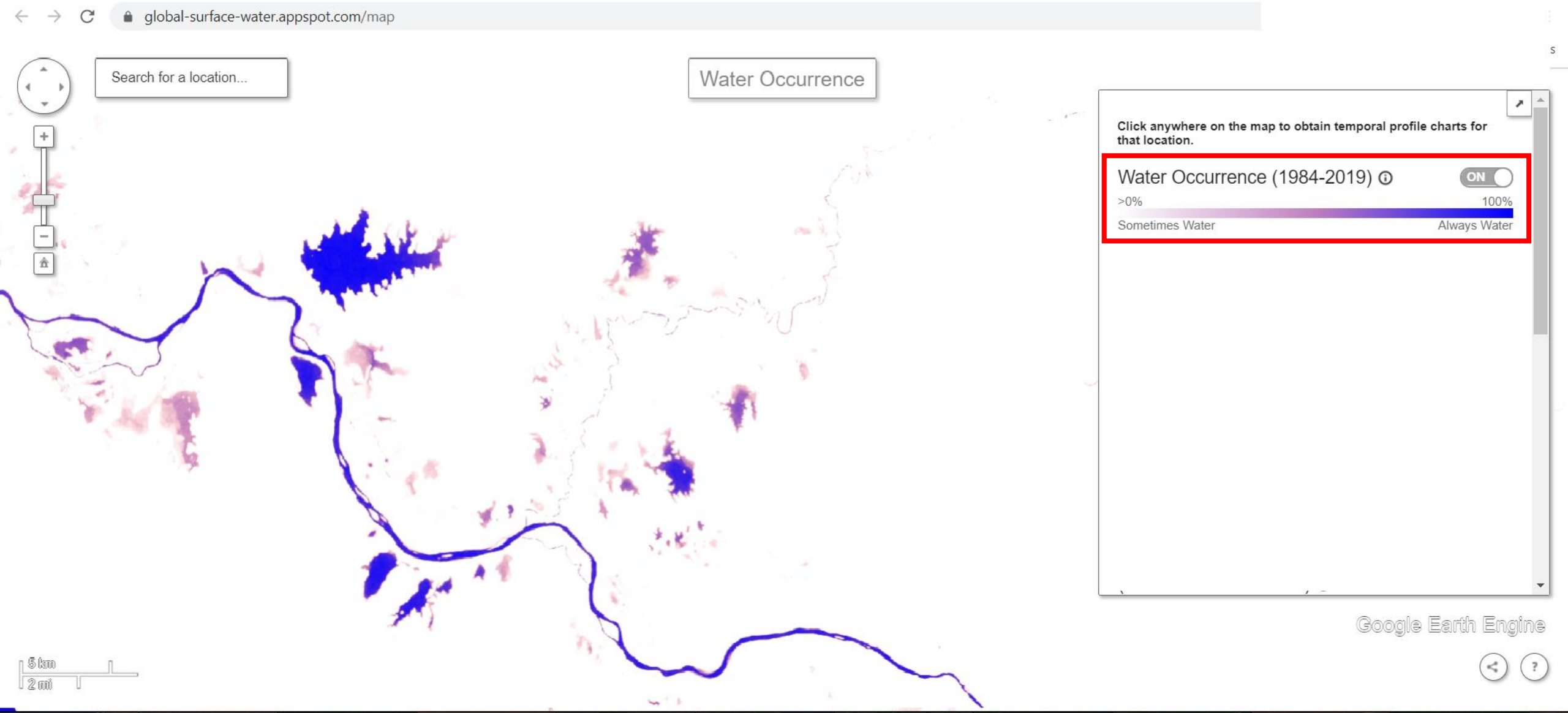
Methodology

Pre-Processing: Satellite Imagery

Seasonal and Annual changes in surface water dynamics

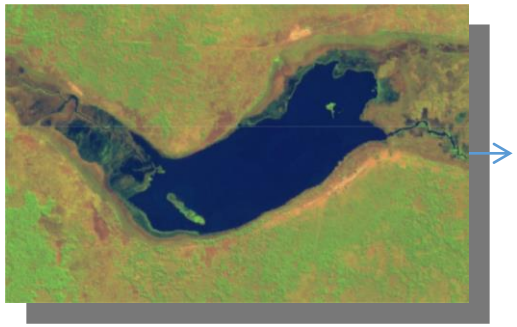
Methodology

Pre-Processing: Global Surface Water (GSW)



Methodology

Pre-Processing: Global Surface Water (GSW)



Water Occurrence satellite image

Repeat for every grid tile

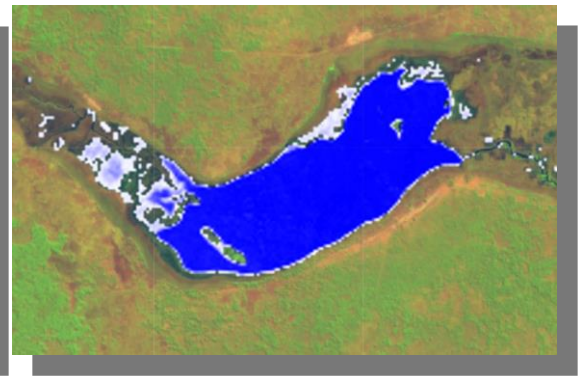
Bicubic Interpolation

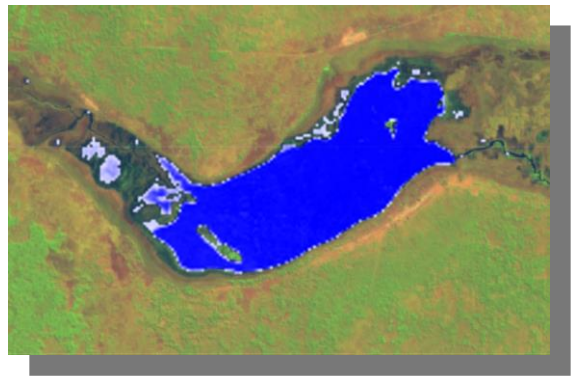
Threshold

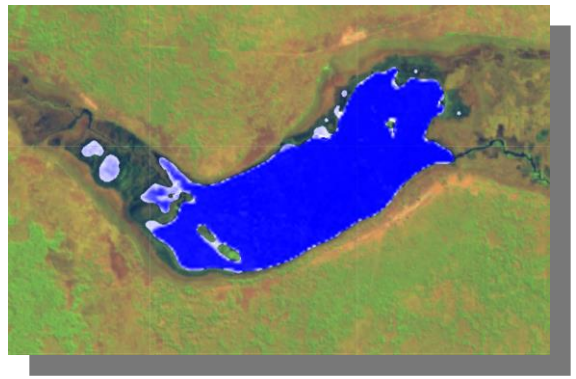
Closing morphological Operation

Boundary Simplification

Surface Water Vector

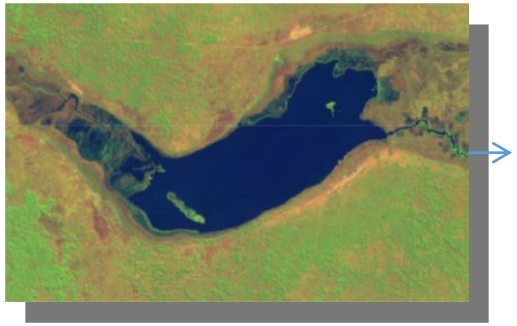






Methodology

Pre-Processing: Global Surface Water (GSW)



Water Occurrence satellite image

Repeat for every grid tile

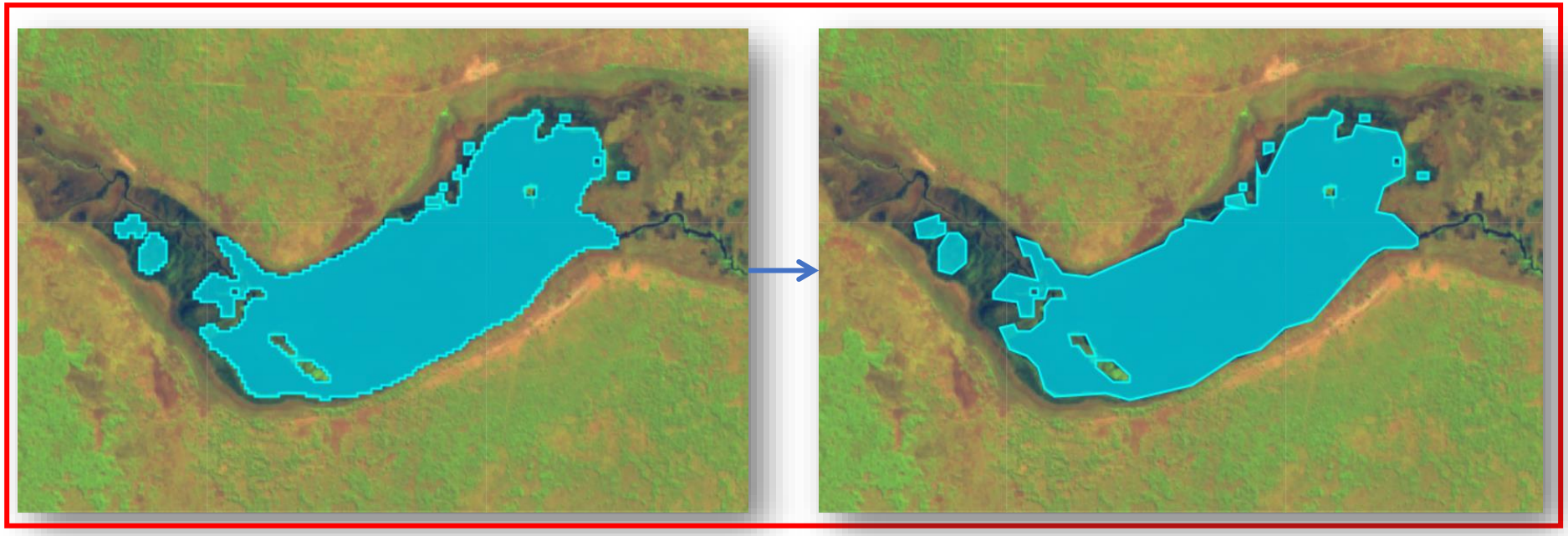
Bicubic Interpolation

Threshold

Closing morphological Operation

Boundary Simplification

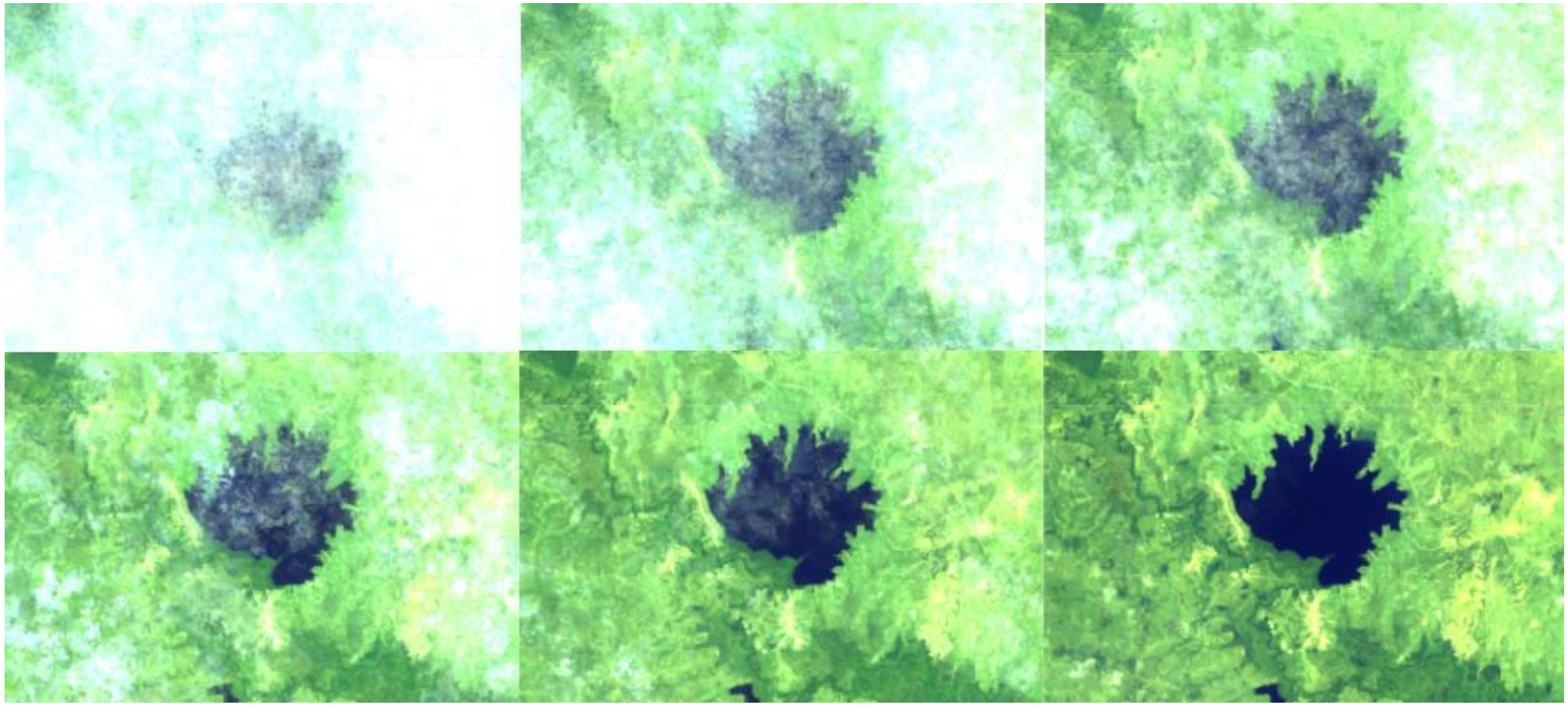
Surface Water Vector



Methodology

Pre-Processing: Sentinel 2

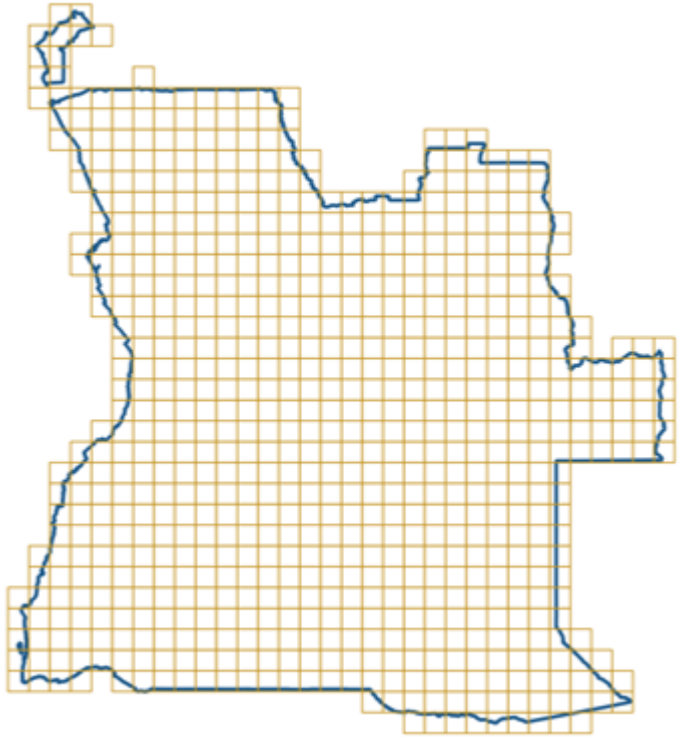
Acquisition Period 2016 -2019

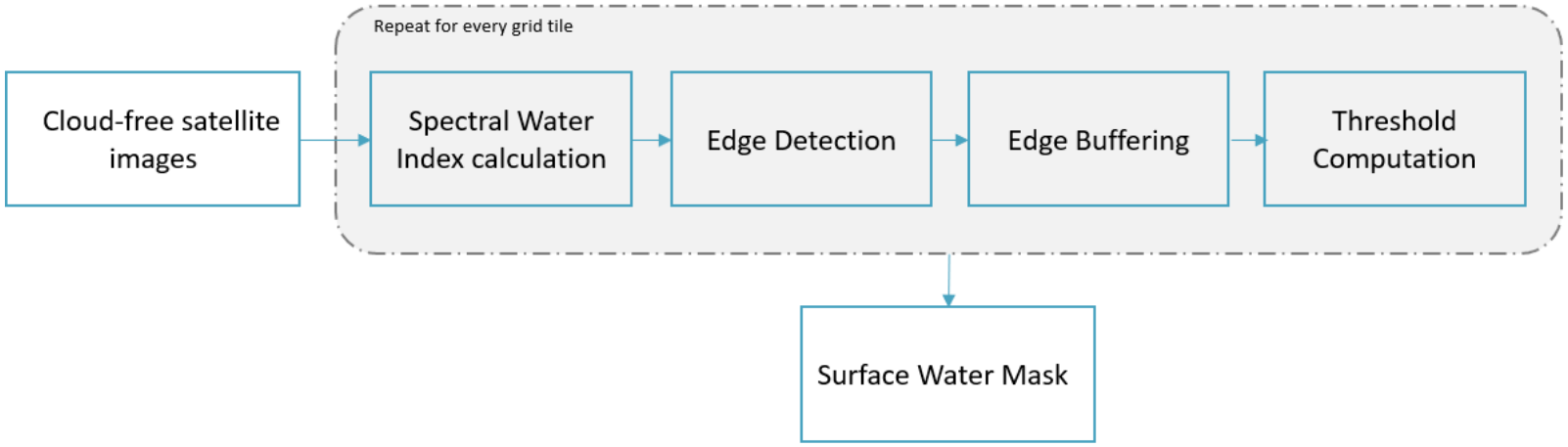


Methodology

Pre-Processing: Sentinel 2

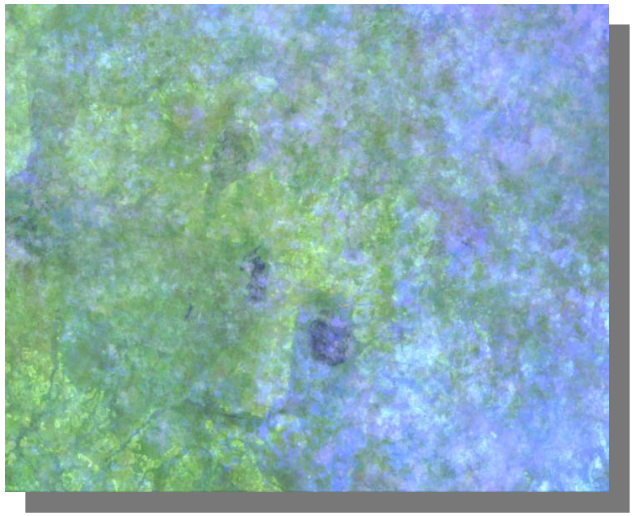
1km x 1km Grid over Study Area





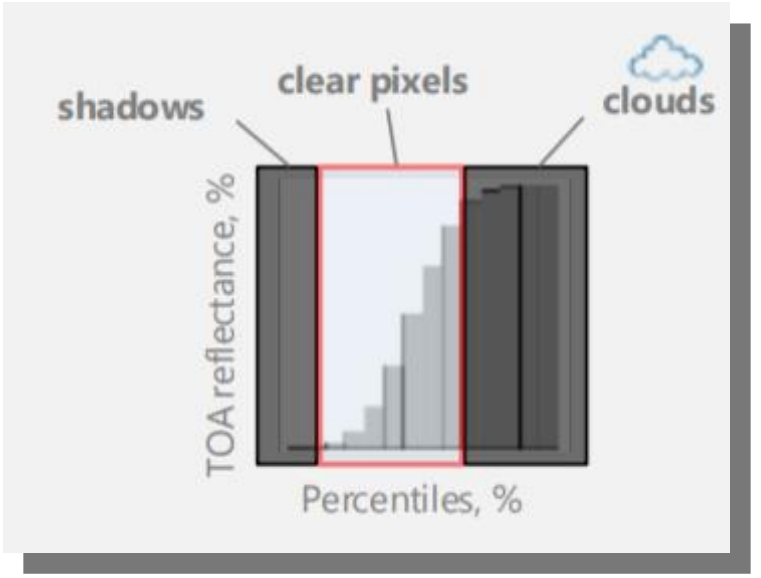
Methodology

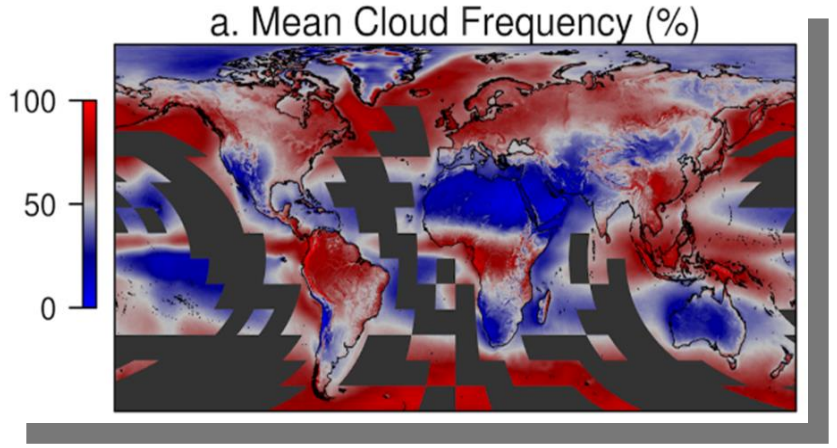
Pre-Processing: Sentinel 2



Sort Images depending on their cloud score

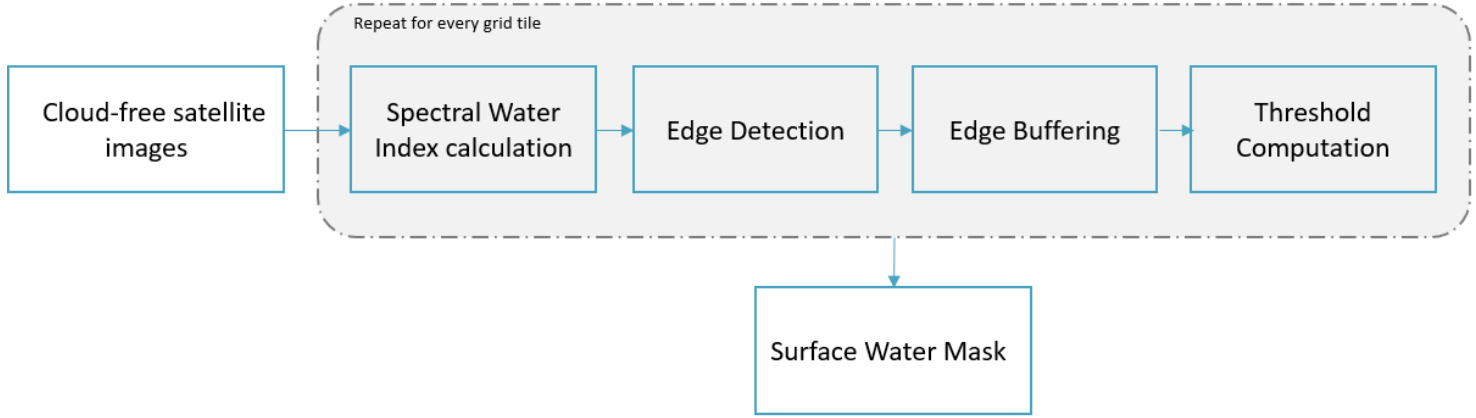
Selection of least cloudy images based on MODIS annual cloud frequency

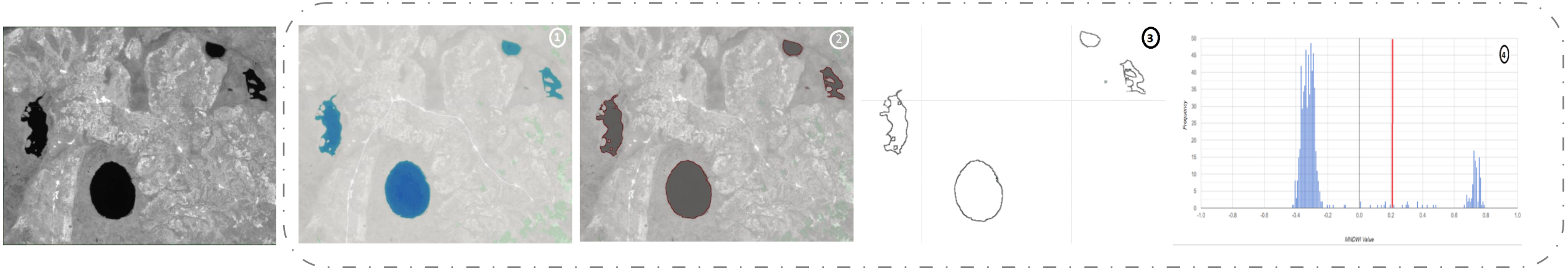


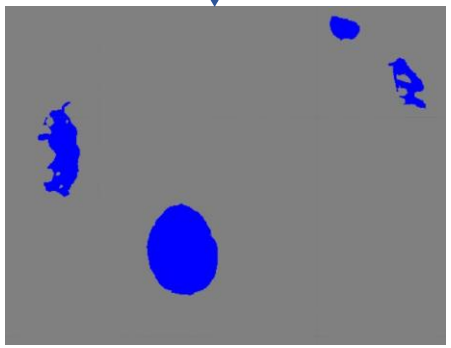


Methodology

Pre-Processing: Sentinel 2

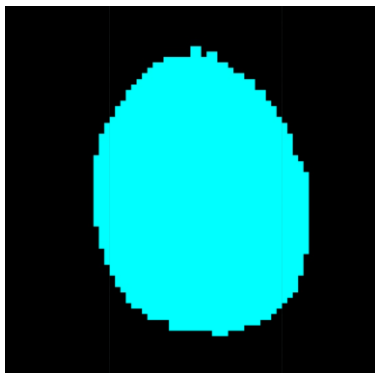


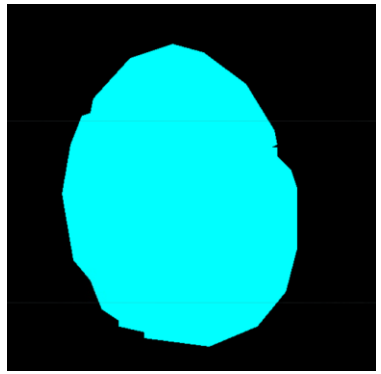




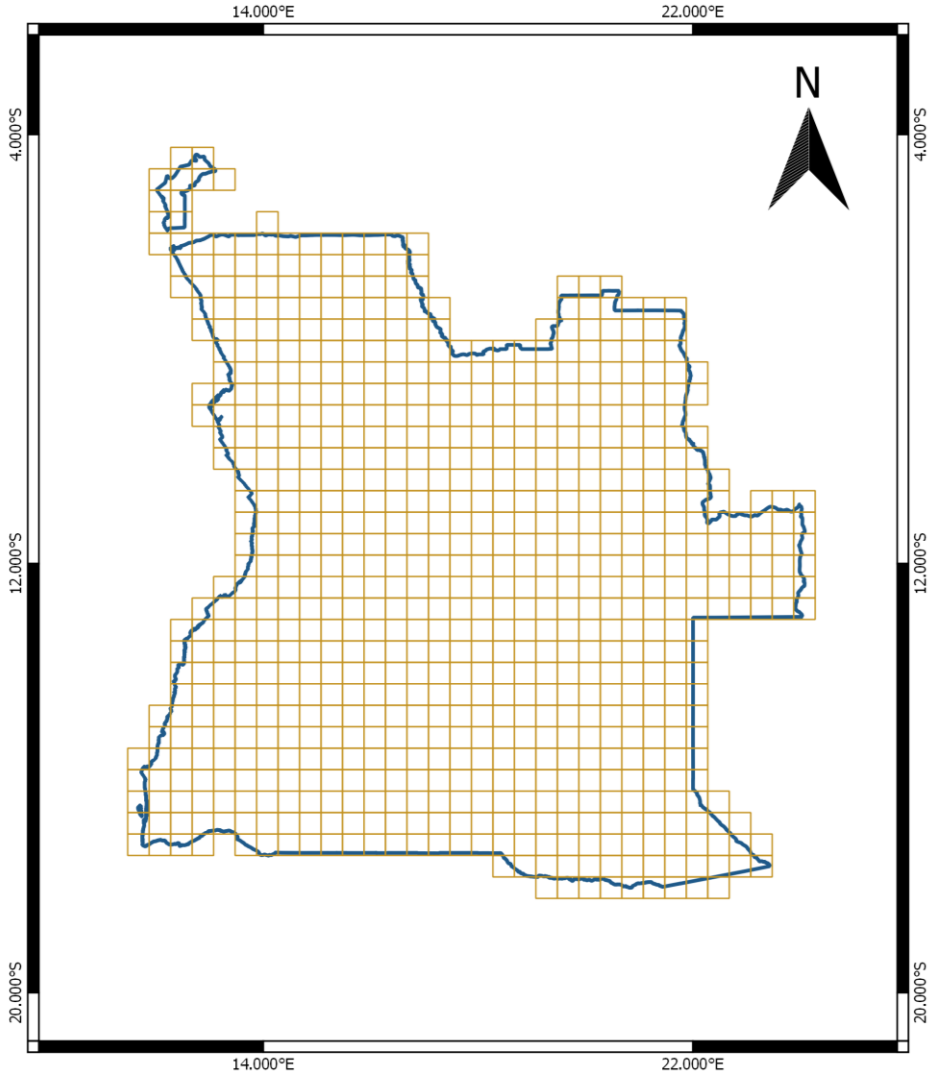
Vector Format

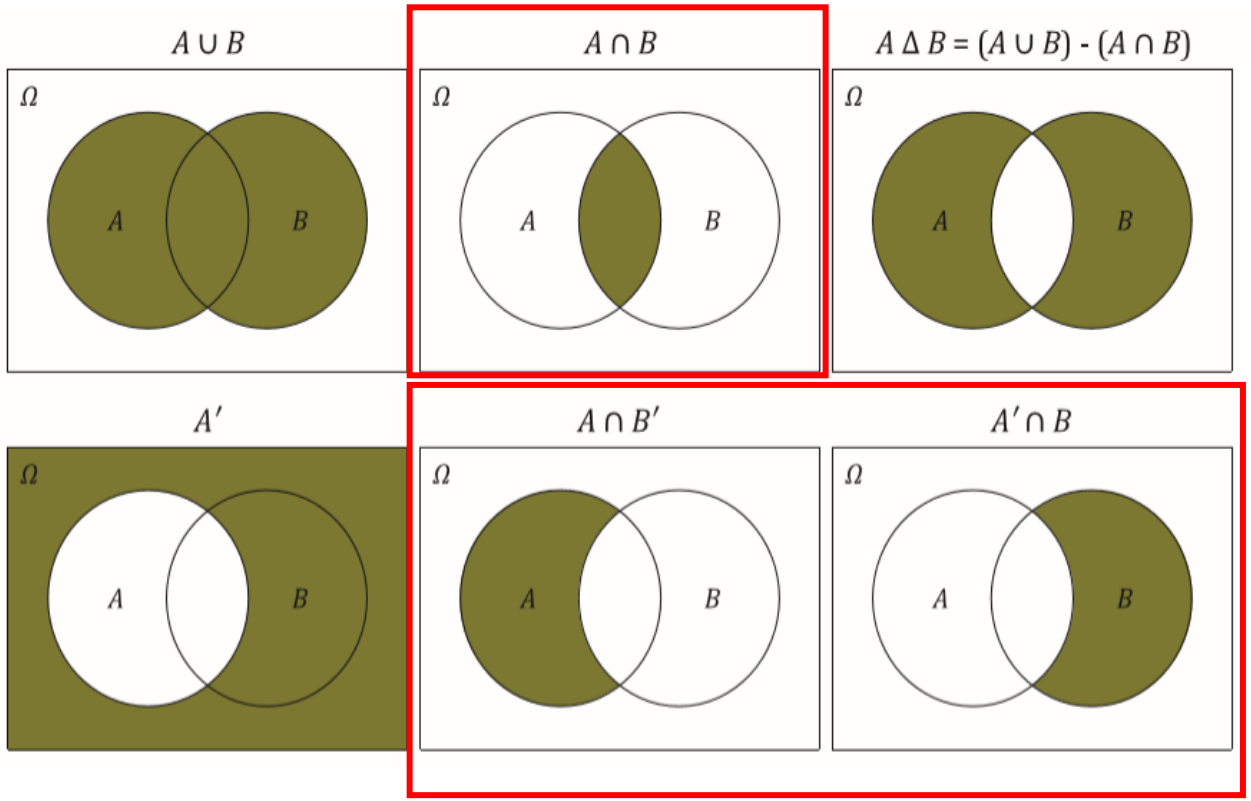
Simplification



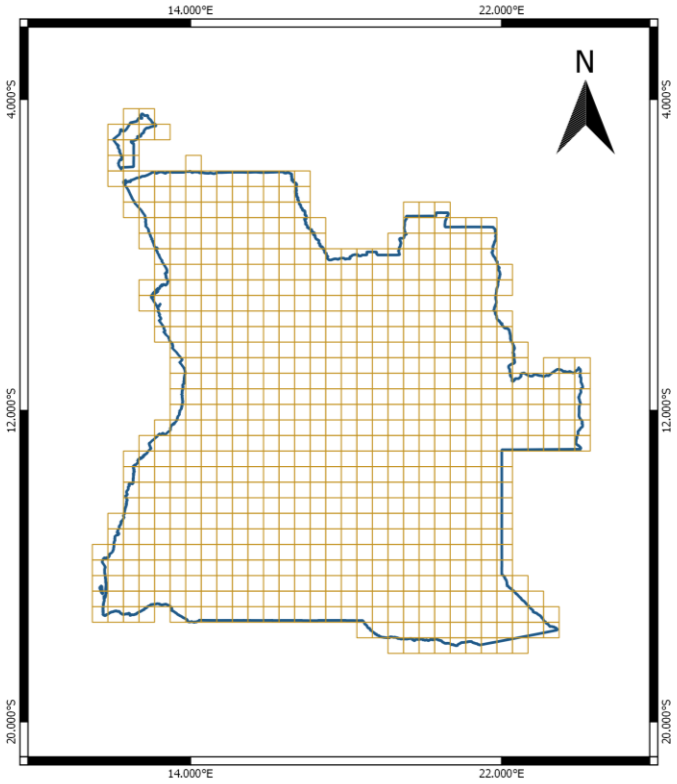


Completeness quality metric: Indicates water area completeness





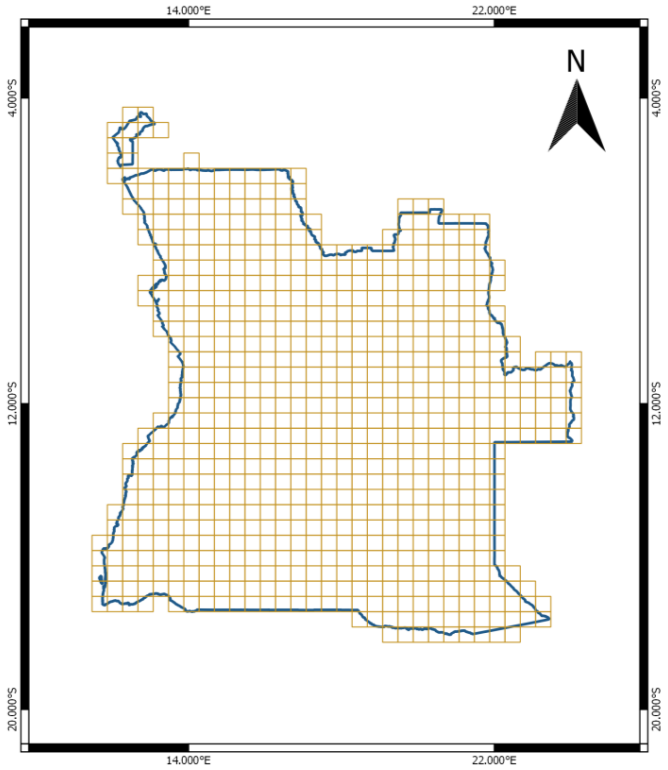
Methodology
Completeness



Per cell:

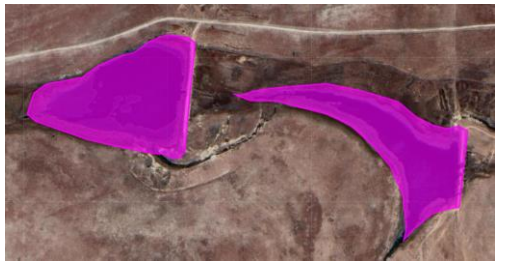
Methodology

Completeness

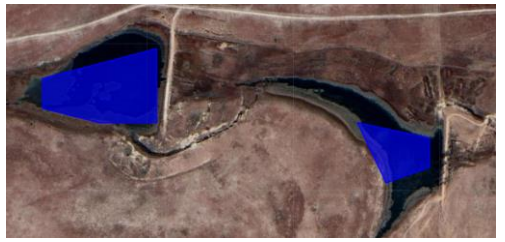


Per cell:

1. Total water area Dataset 1



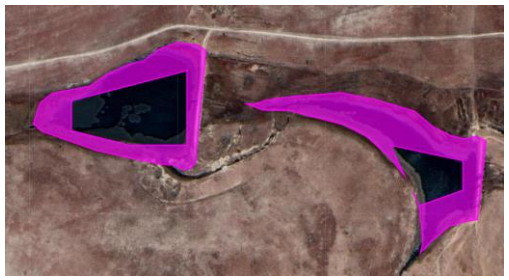
2. Total water area Dataset 2



3. Intersection Datasets 1&2

4. Missing water area Dataset 1

5. Missing water area Dataset 2

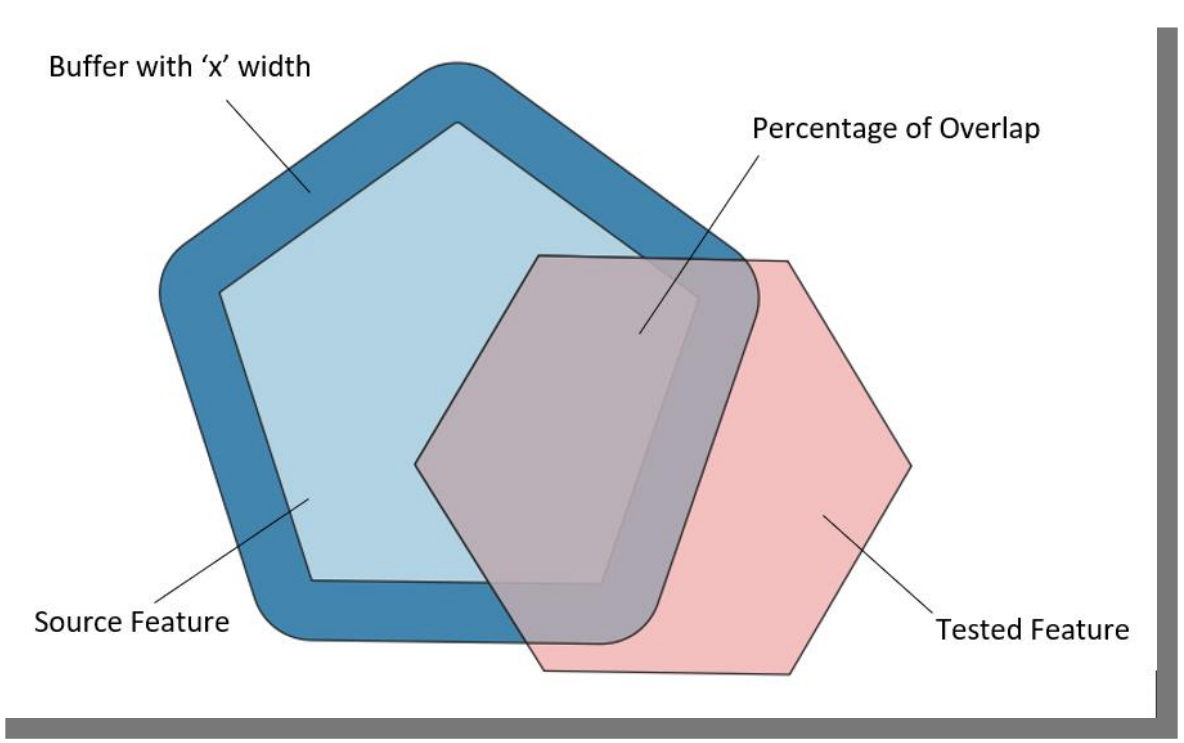


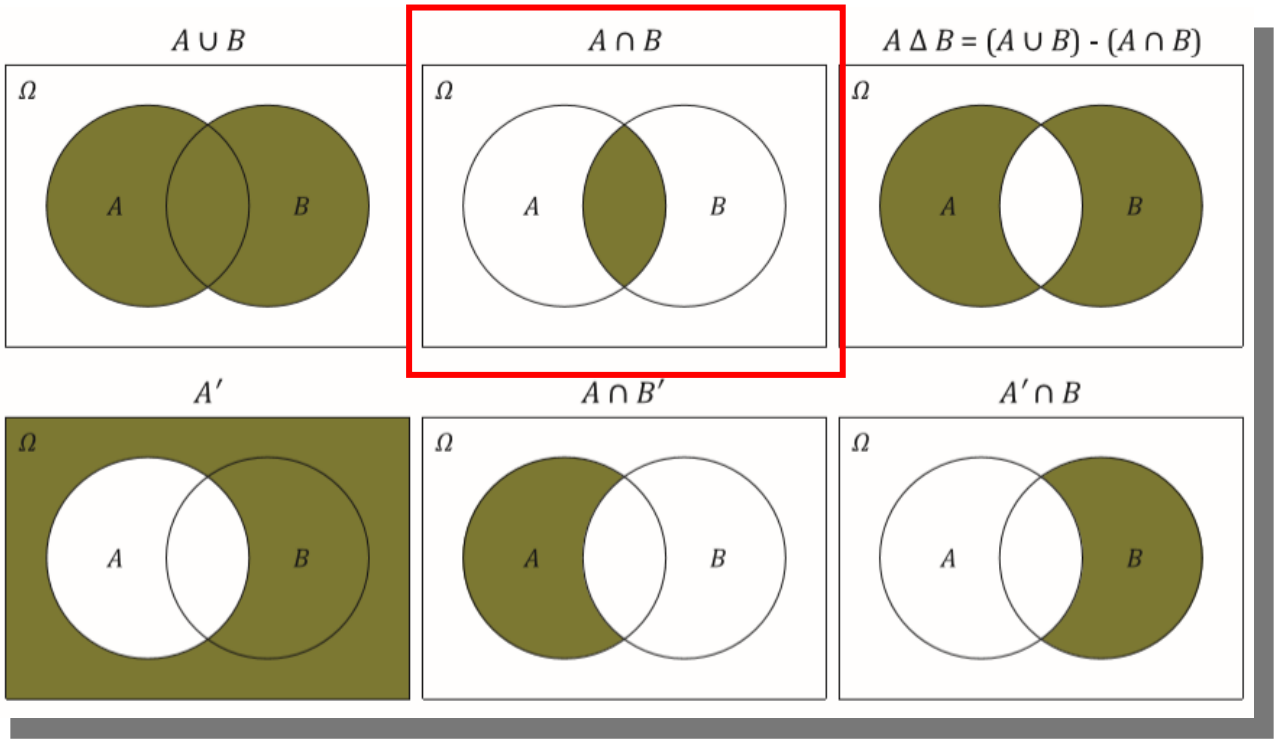
Methodology

Goodchild's Percentage of overlap

Percentage of Overlap quality metric:

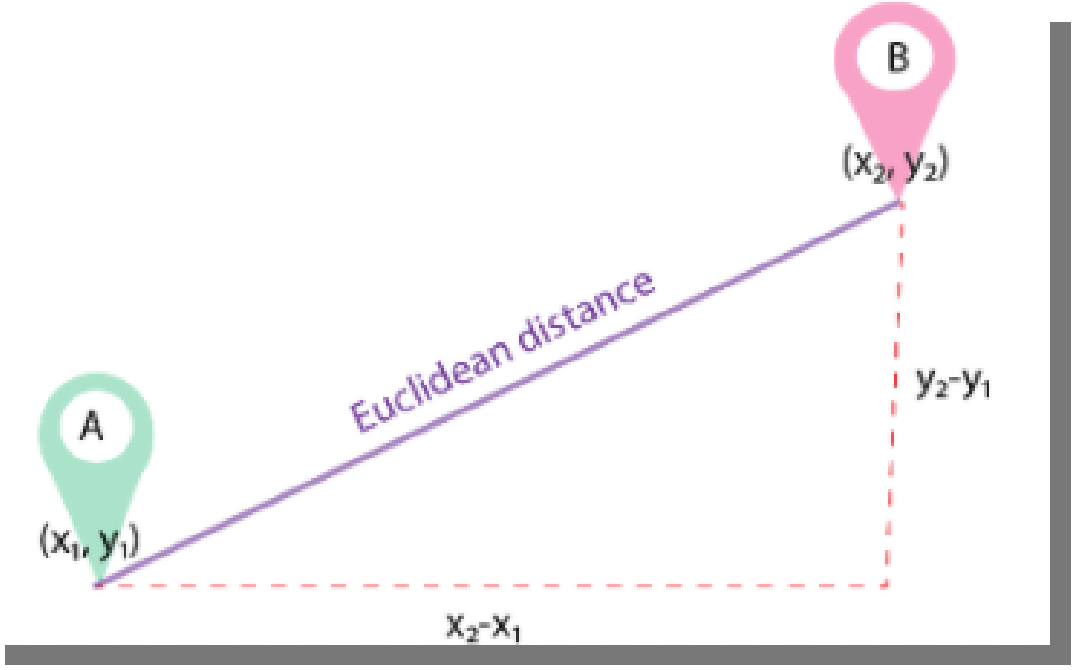
1. Feature Based Comparison
2. Indicates spatial offset between two features



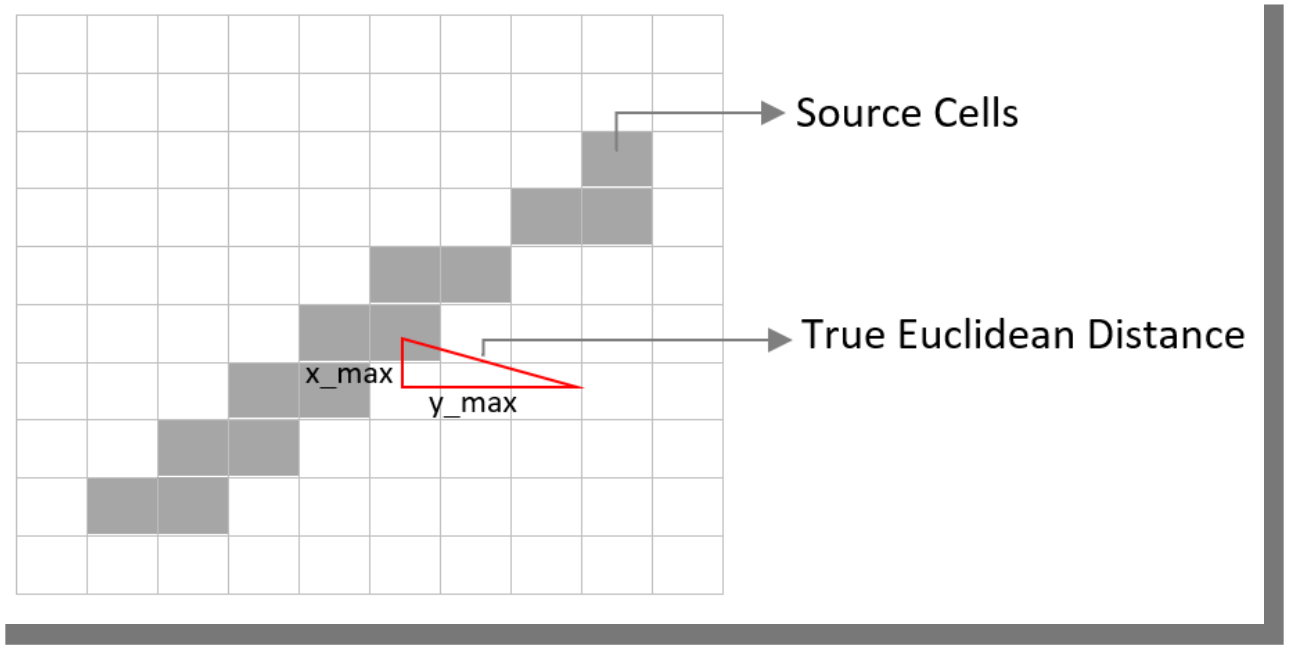


Point Based Comparison

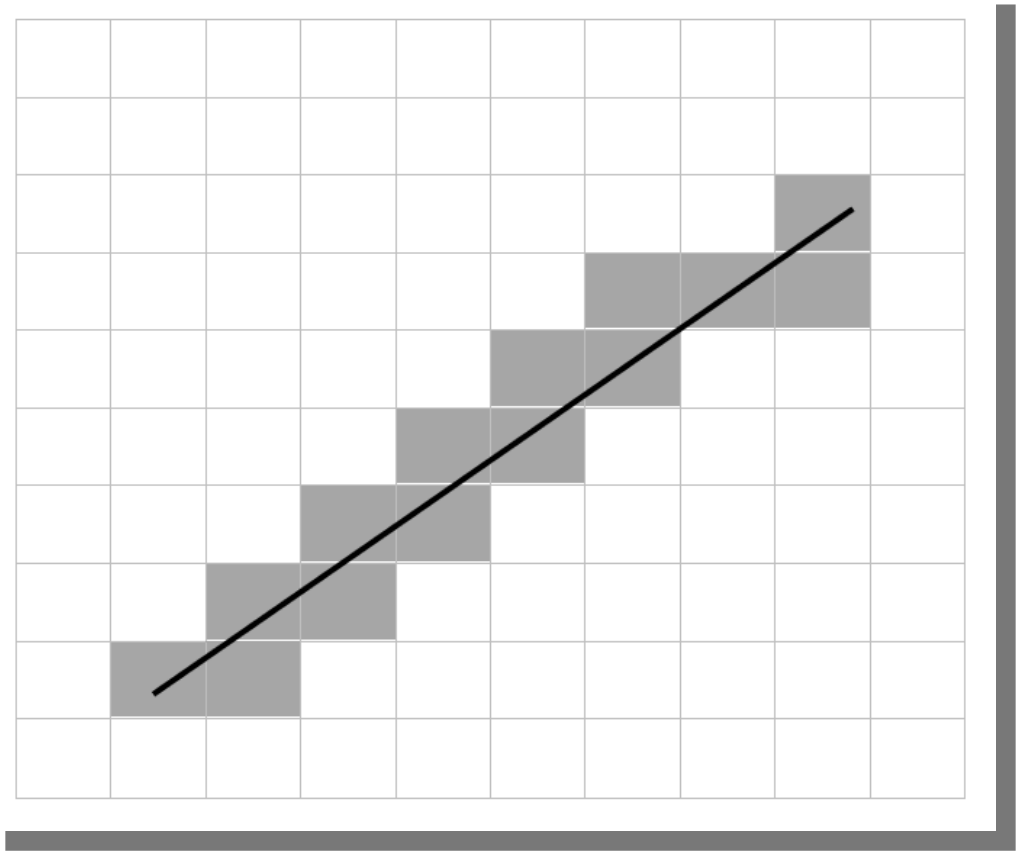
Euclidean Distance for vectors based on: Points

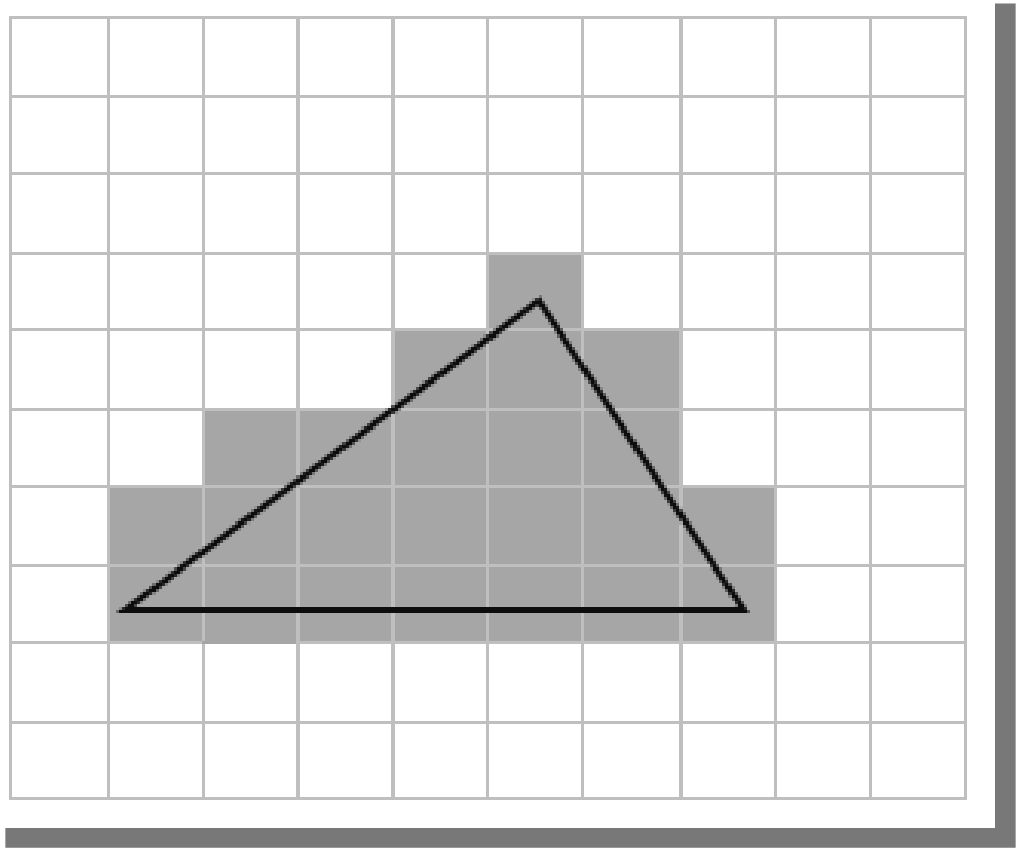


Euclidean Distance for rasters based on: Cells



Rasterization of Line and Polygonal Primitives





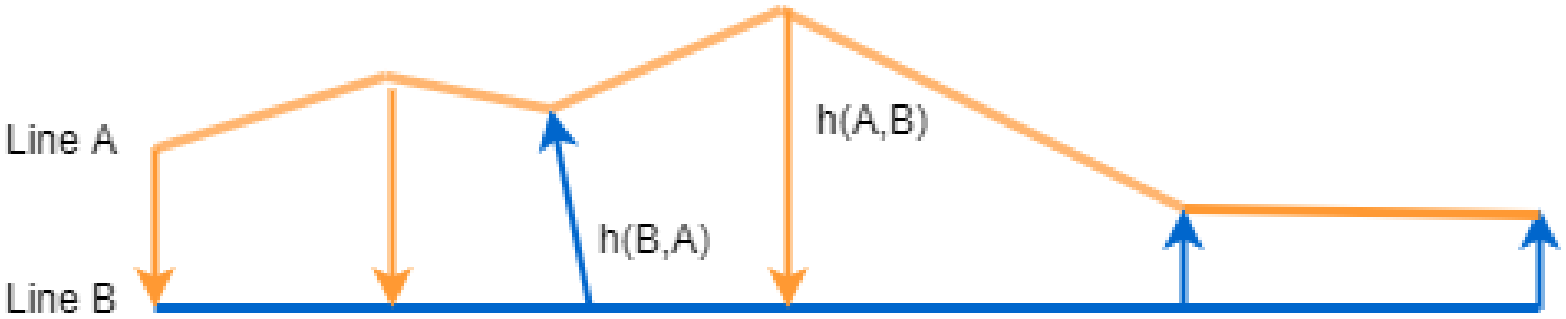
Euclidean Distance Map (EDM) of Rasterized Linear Feature

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	1	1	0	0	0
0	0	0	0	1	1	0	0	0	0
0	0	0	1	1	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

6.40	5.66	5.00	4.48	4.12	3.61	2.83	2.24	2.00	2.24
5.66	5.0	4.24	3.61	3.16	2.83	2.24	1.41	1.00	1.41
5.00	4.24	3.61	2.83	2.24	2.00	1.41	1.00	0.00	1.00
4.24	3.61	2.83	2.24	1.41	1.00	1.00	0.00	0.00	1.00
3.61	2.83	2.24	1.41	1.0	0.00	0.00	1.00	1.00	1.41
2.83	2.24	1.41	1.00	0.00	0.00	1.00	1.41	2.00	2.24
2.24	1.41	1.00	0.00	0.00	1.00	1.41	2.24	2.83	3.16
1.41	1.00	0.00	0.00	1.00	1.41	2.24	2.83	3.61	4.12
1.00	0.00	0.00	1.00	1.41	2.24	2.83	3.61	4.24	5.00
1.41	1.00	1.0	1.41	2.24	2.83	3.61	4.24	5.00	5.66

Methodology
Hausdorff Distance

Hausdorff Distance: Indicates maximum deviation and shape similarity between two features



Bidirectional Distances:

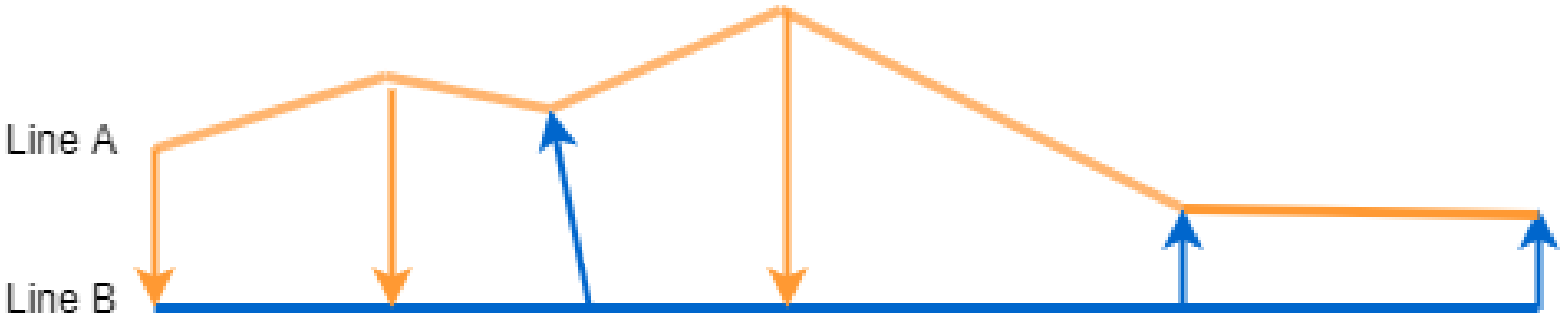
$$h(A,B) = max\left\{min\ d(a,b)\right\}$$

$$h(B,A) = max\left\{min\ d(b,a)\right\}$$

Method 1

$$H(A,B) = max\left\{h(A,B), h(B,A)\right\}$$

Hausdorff Distance



Bidirectional Distances

$$h(A,B) = avg\{min(a,b)\}$$

$$h(B,A) = avg\{min(b,a)\}$$

Method 2:

$$H(A,B) = h(A,B) + h(B,A)$$

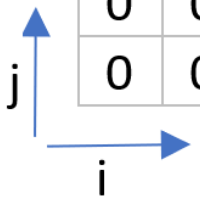
Methodology

Hausdorff Distance

Object A

Polygon

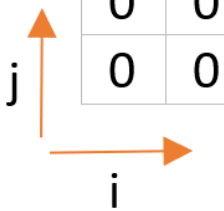
0	0	0	0	0	0	0
0	1	1	1	0	0	0
0	1	1	1	0	0	0
0	1	1	1	1	0	0
0	0	0	1	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0



Polygon

Object B

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	1	1	1	0
0	0	0	1	1	1	0
0	0	0	1	1	1	0
0	0	0	0	0	0	0



Euclidean Distances from B to A

1.4	1.0	1.0	1.0	1.4	2.2	3.2
1.0	0.0	0.0	0.0	1.0	2.0	2.8
1.0	0.0	0.0	0.0	1.0	1.4	2.2
1.0	0.0	0.0	0.0	0.0	1.0	2.0
1.4	1.0	1.0	0.0	0.0	1.0	2.0
2.2	2.0	1.4	1.0	1.0	1.4	2.2
3.2	2.8	2.2	2.0	2.0	2.2	2.8

Intersection of EDM of A with object B

1.4	1.0	1.0	1.0	1.4	2.2	3.2
1.0	0.0	0.0	0.0	1.0	2.0	2.8
1.0	0.0	0.0	0.0	1.0	1.4	2.2
1.0	0.0	0.0	0.0	0.0	1.0	2.0
1.4	1.0	1.0	0.0	0.0	1.0	2.0
2.2	2.0	1.4	1.0	1.0	1.4	2.2
3.2	2.8	2.2	2.0	2.0	2.2	2.8

Euclidean Distance values and maximum value from B to A

Euclidean Distances from A to B

3.6	2.8	2.2	2.0	2.4	2.8	3.2
3.2	2.2	1.4	1.0	1.4	2.0	2.2
3.0	2.0	1.0	0.0	1.0	1.0	1.4
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.2	2.2	1.4	1.0	1.0	1.0	1.4

j

i

Intersection of EDM of B with object A

3.6	2.8	2.2	2.0	2.4	2.8	3.2
3.2	2.2	1.4	1.0	1.4	2.0	2.2
3.0	2.0	1.0	0.0	1.0	1.0	1.4
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.0	2.0	1.0	0.0	0.0	0.0	1.0
3.2	2.2	1.4	1.0	1.0	1.0	1.4

j

i

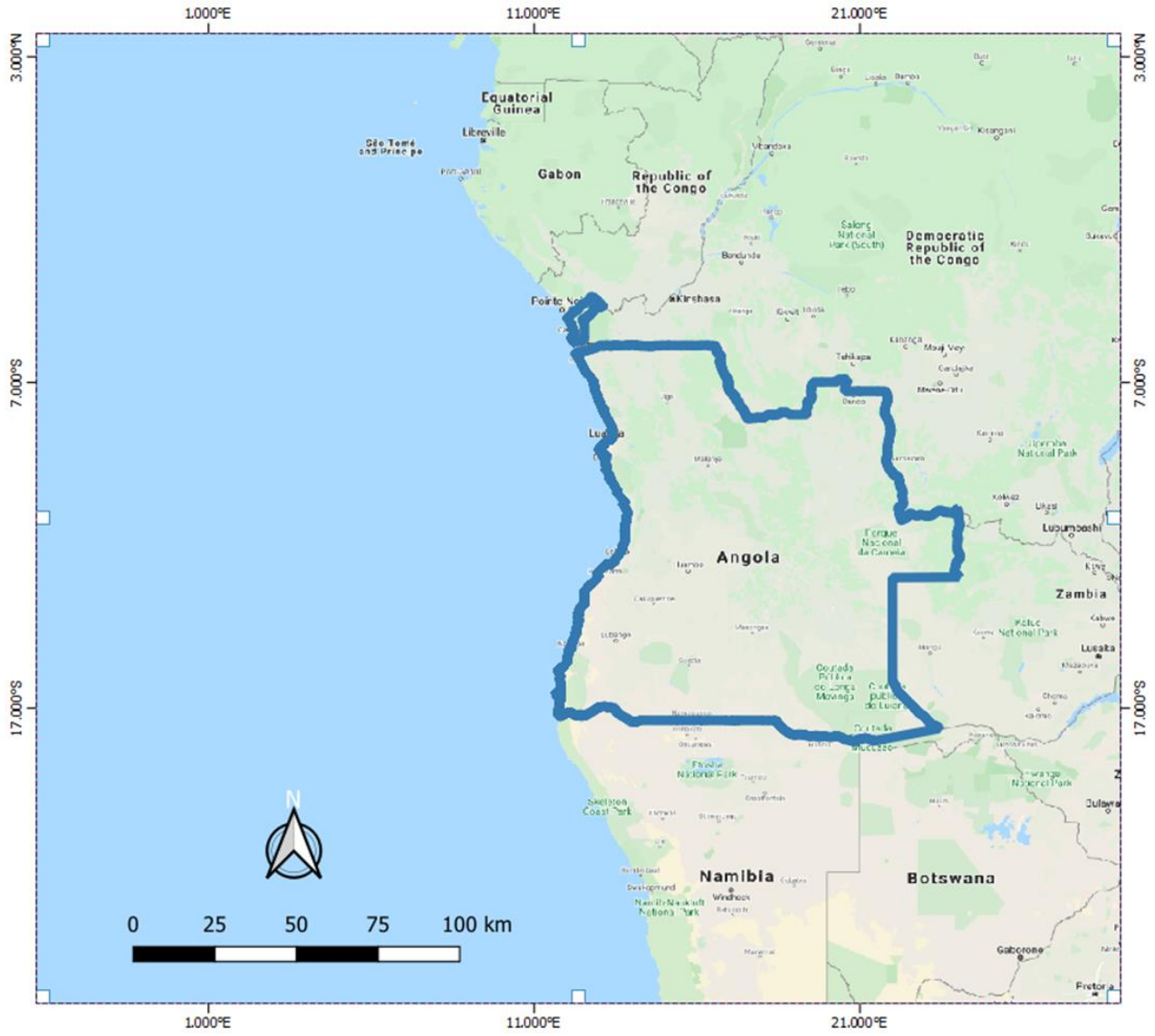
Euclidean Distance values and maximum value from A to B

Outline

- Introduction
- Methodology
- Implementation and Results
- Sensitivity Analysis
- Conclusions and Future Work

Implementation

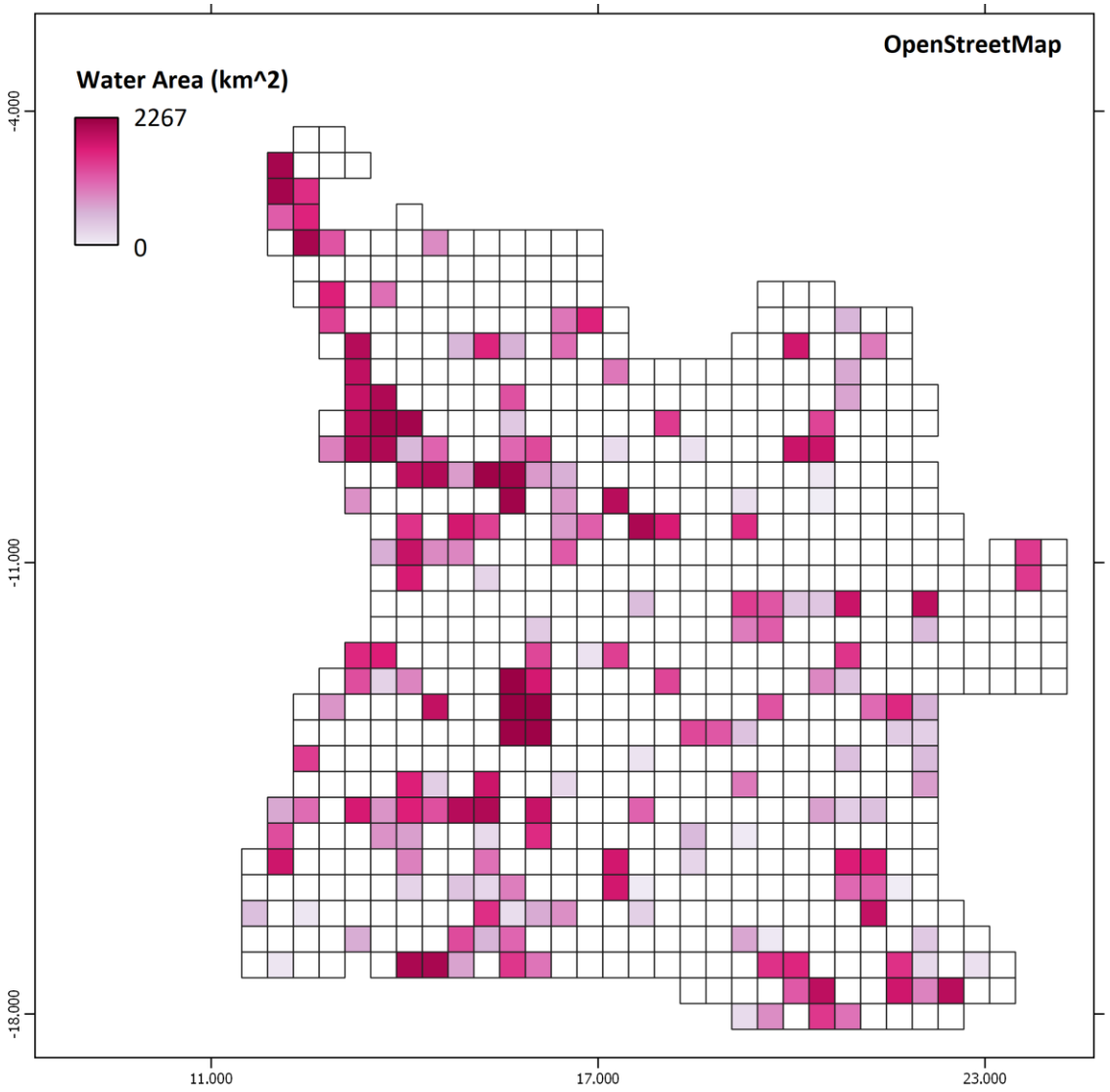
Study area

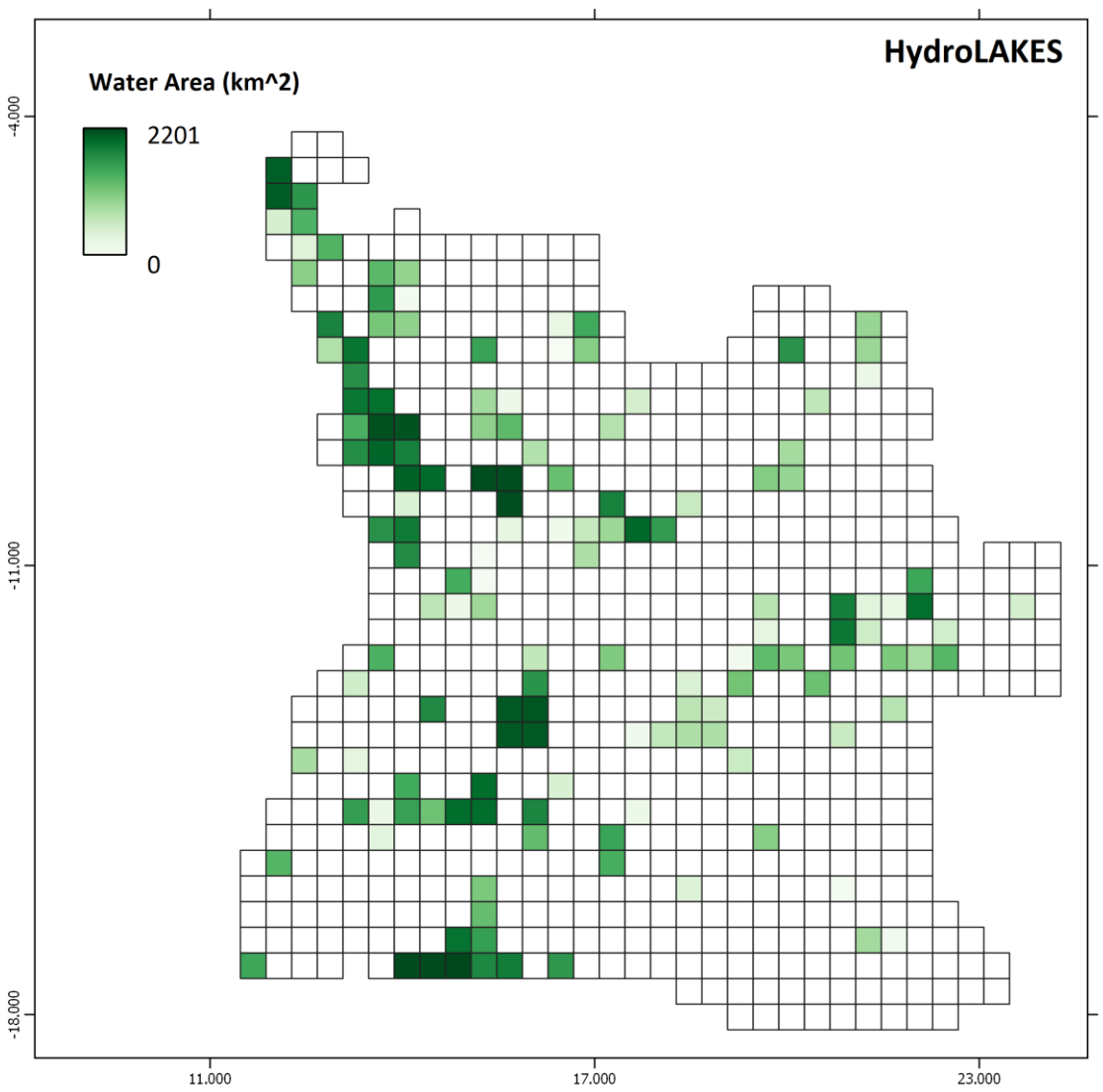


Implementation& Results

Completeness

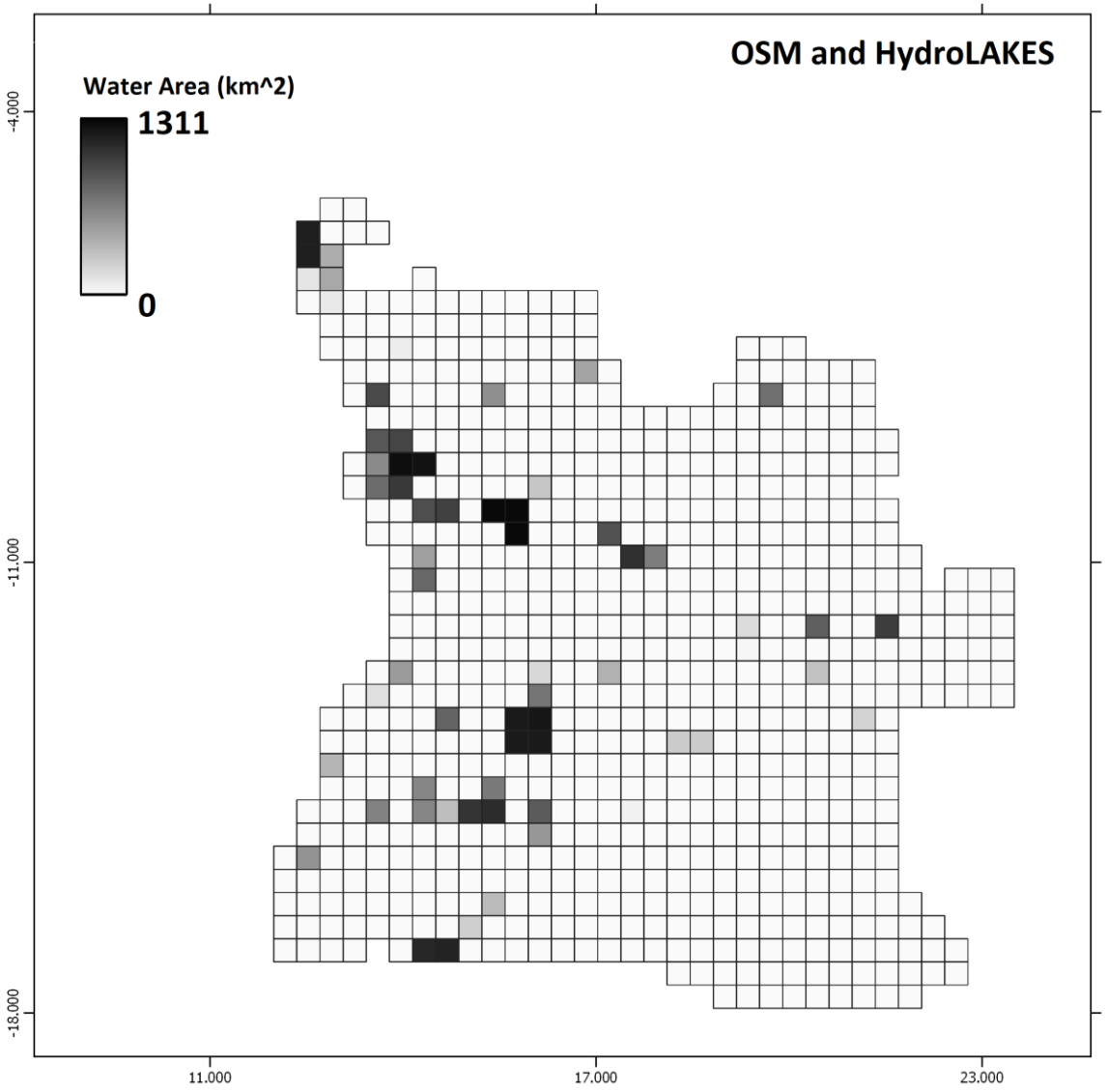
Thematic Differences between OSM and HydroLAKES (water /no water)





Implementation& Results

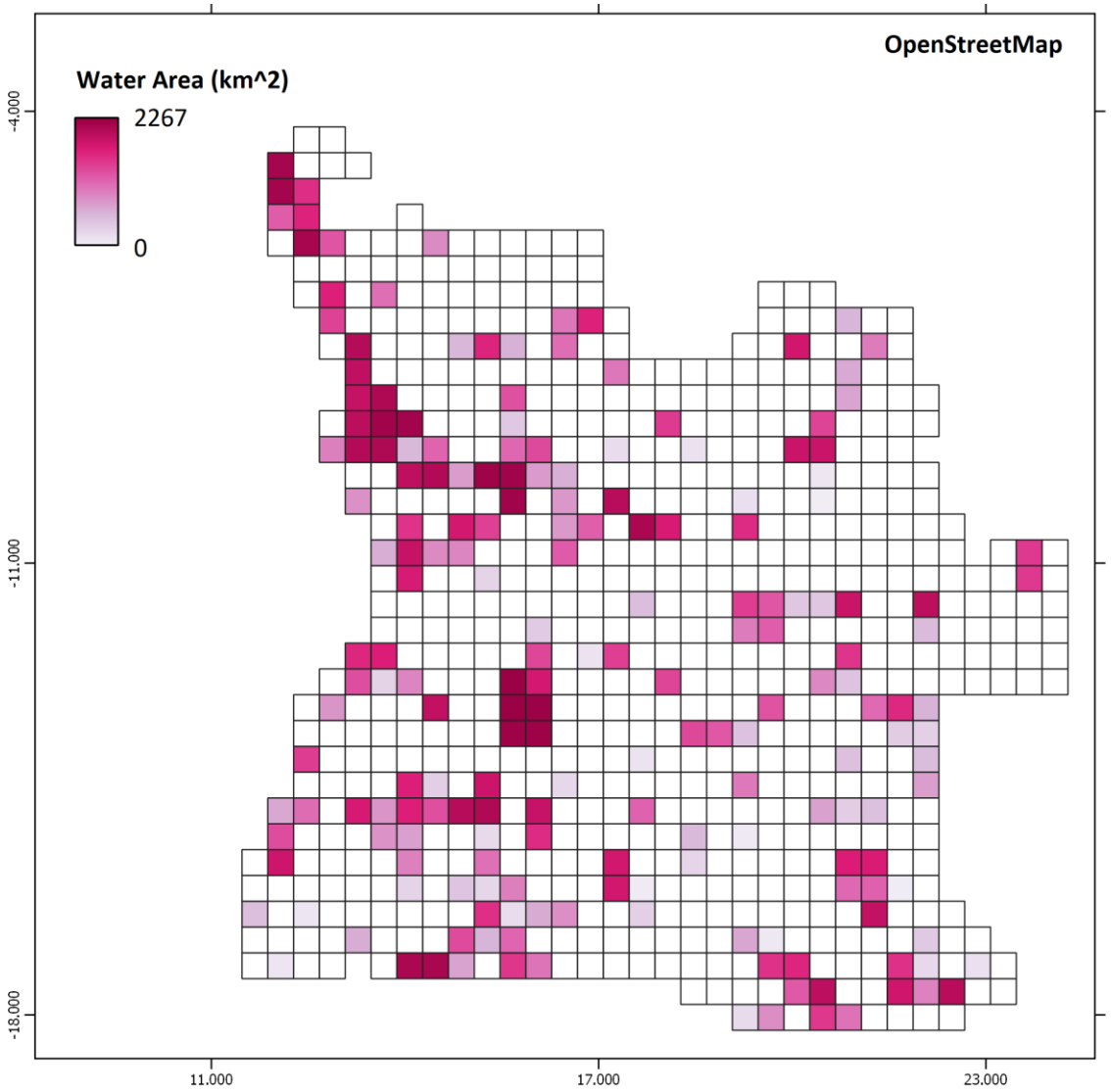
Completeness

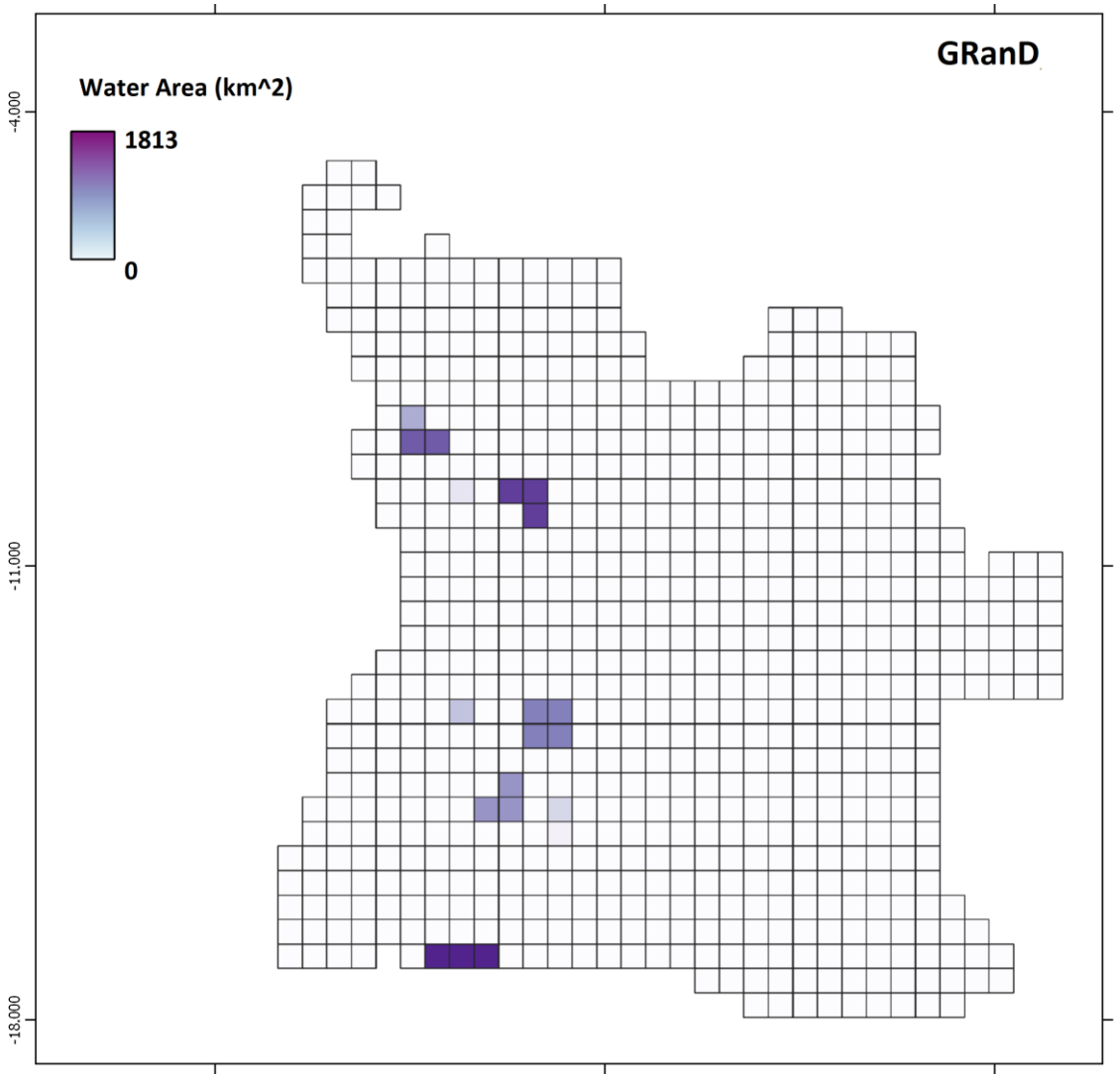


Overlap between OSM and HydroLAKES:

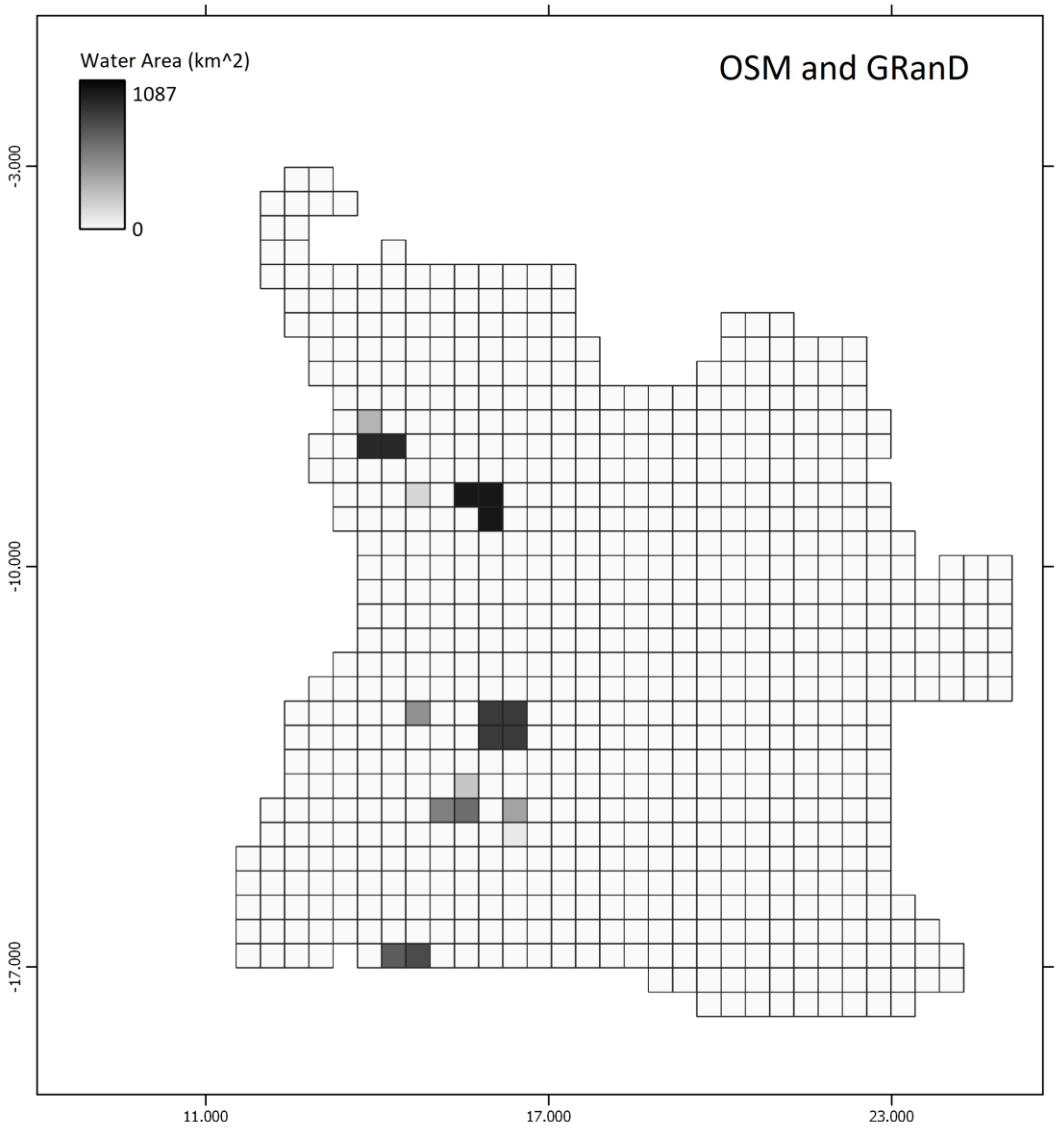
Dataset	Area (km^2)	Ratio (%)
Total Water	5437	100
OSM	2267	42
HydroLAKES	2201	41
OSM and HydroLAKES	1311	24
OSM, no HydroLAKES	956	18
HydroLAKES, no OSM	3171	58

Thematic Differences between OSM and GRaND (water/no water)





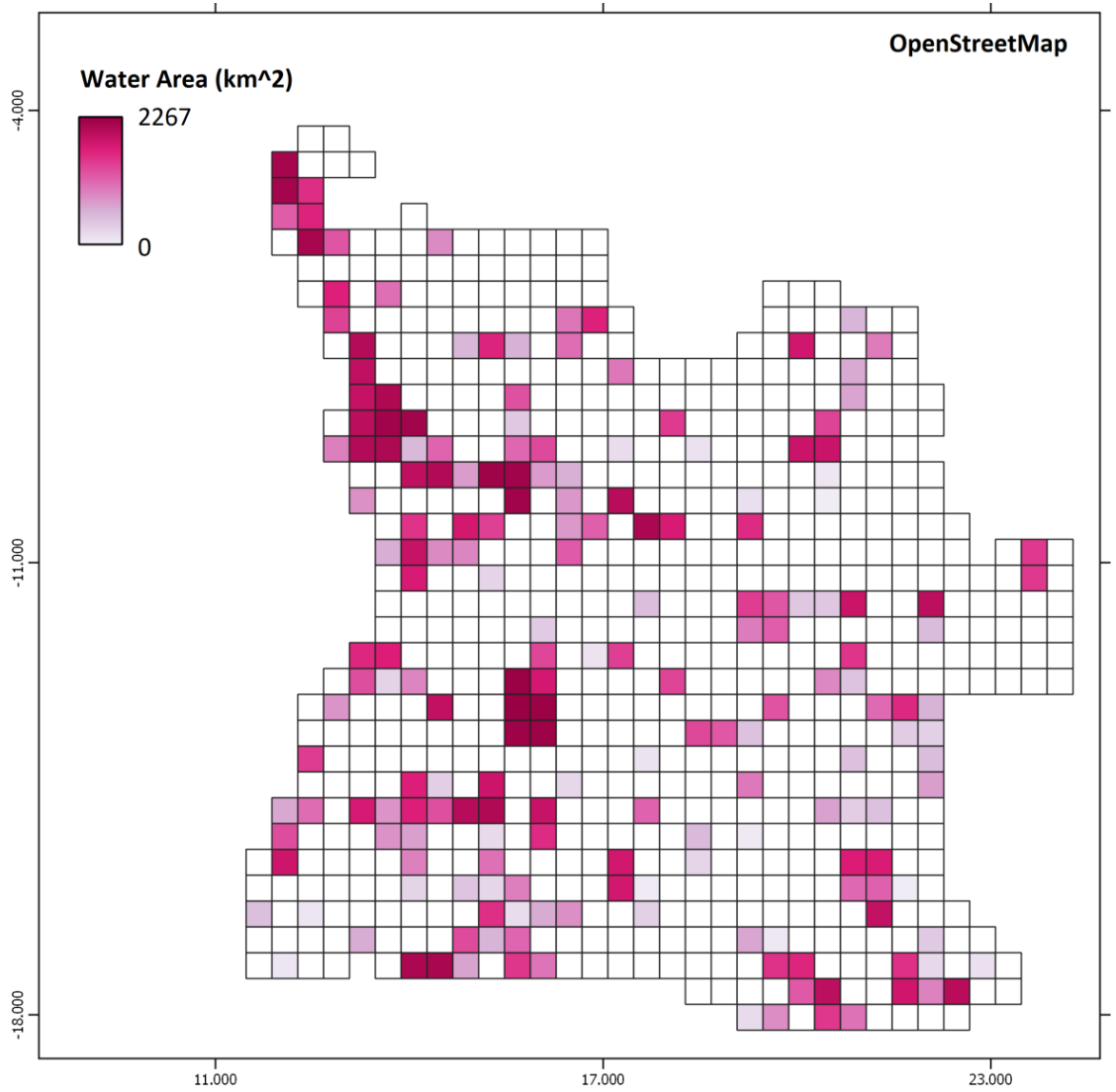
Implementation
Completeness

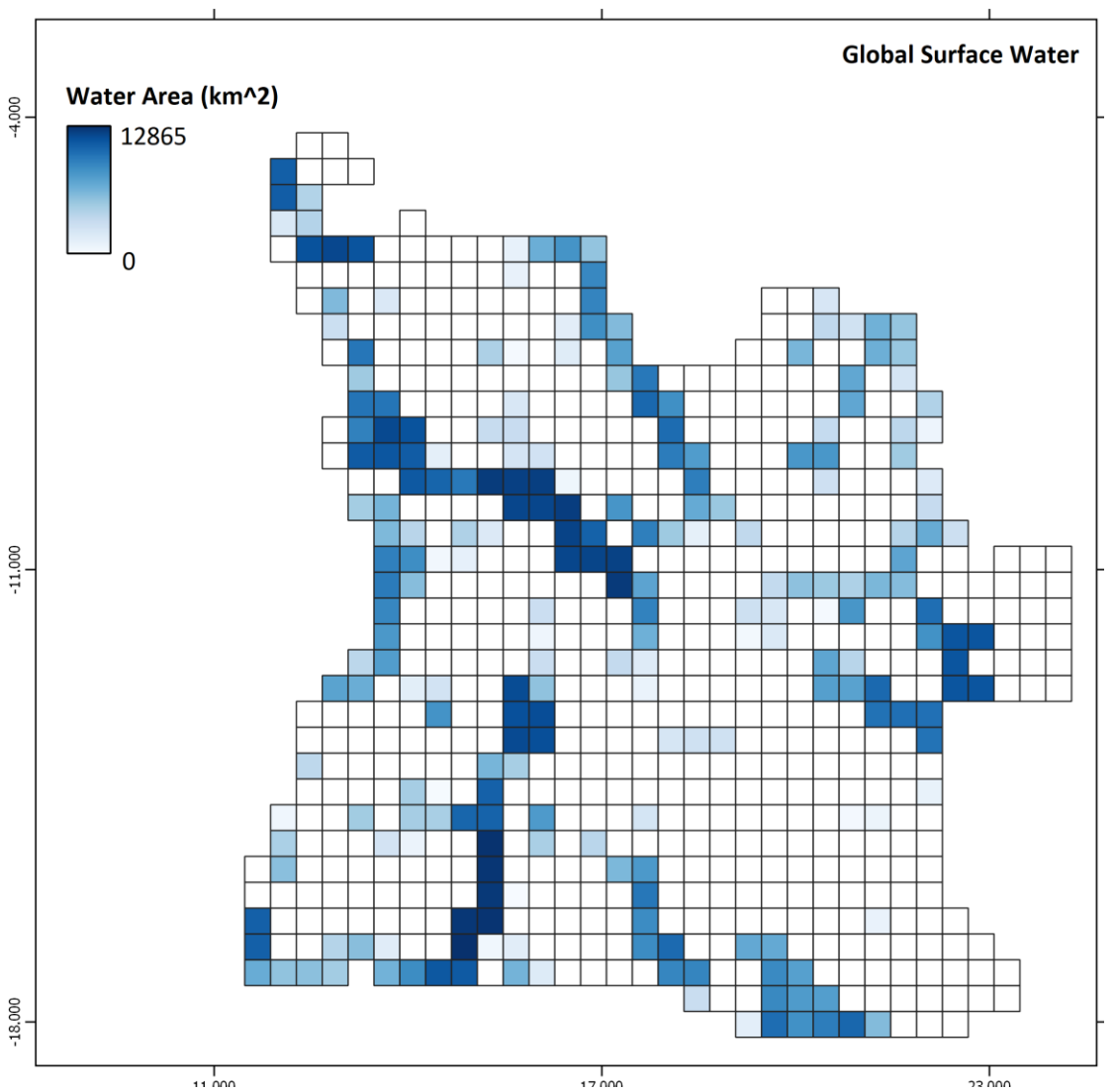


Overlap between OSM and GRaND:

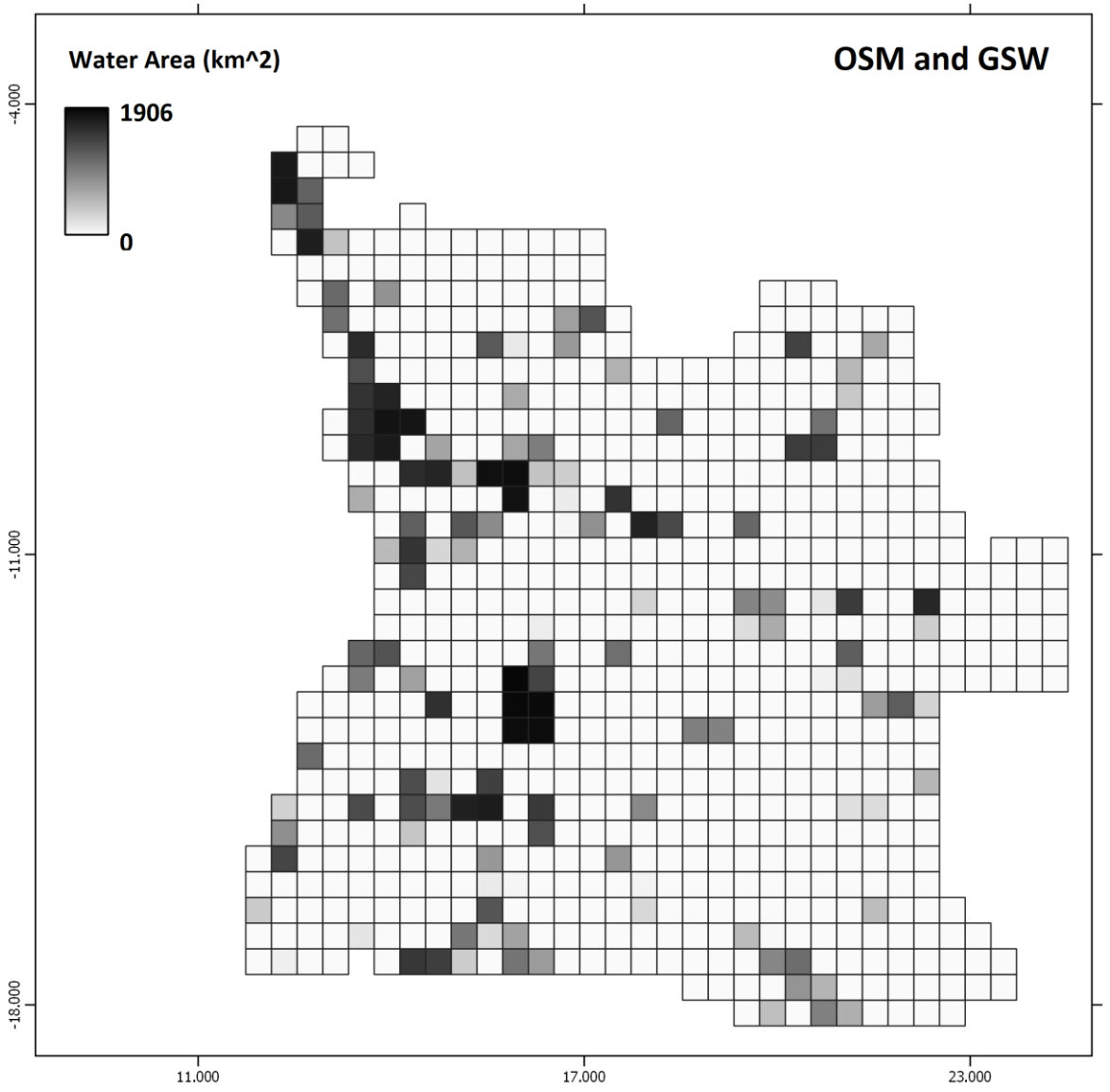
Dataset	Area (km^2)	Ratio (%)
Total Water	5030	100
OSM	2267	45
GRanD	1813	36
OSM and GRaND	1087	22
OSM, no GRaND	1180	24
GRaND, no OSM	2763	54

Thematic Differences between OSM and Global Surface Water (water/no water)





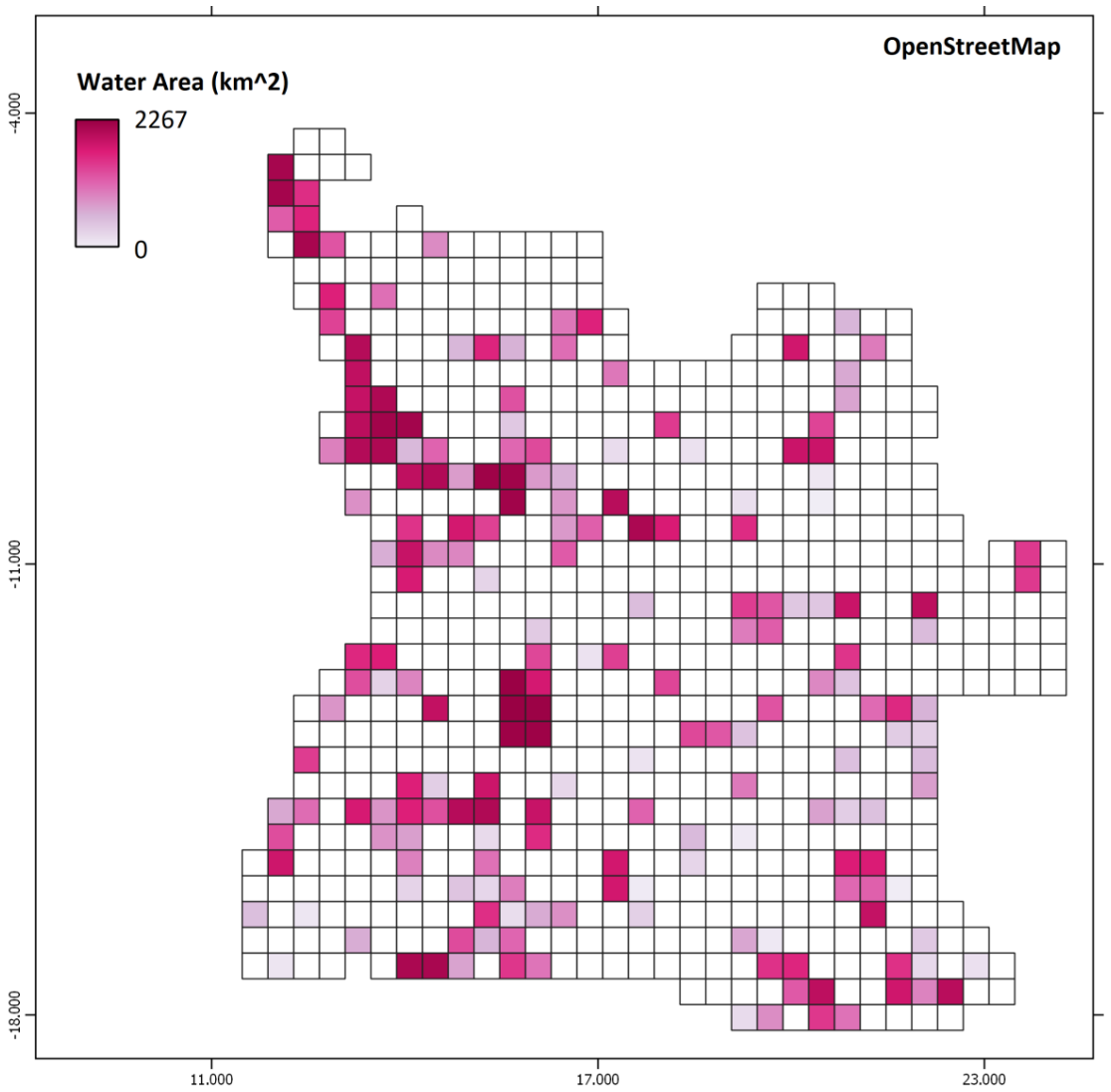
Implementation
Completeness

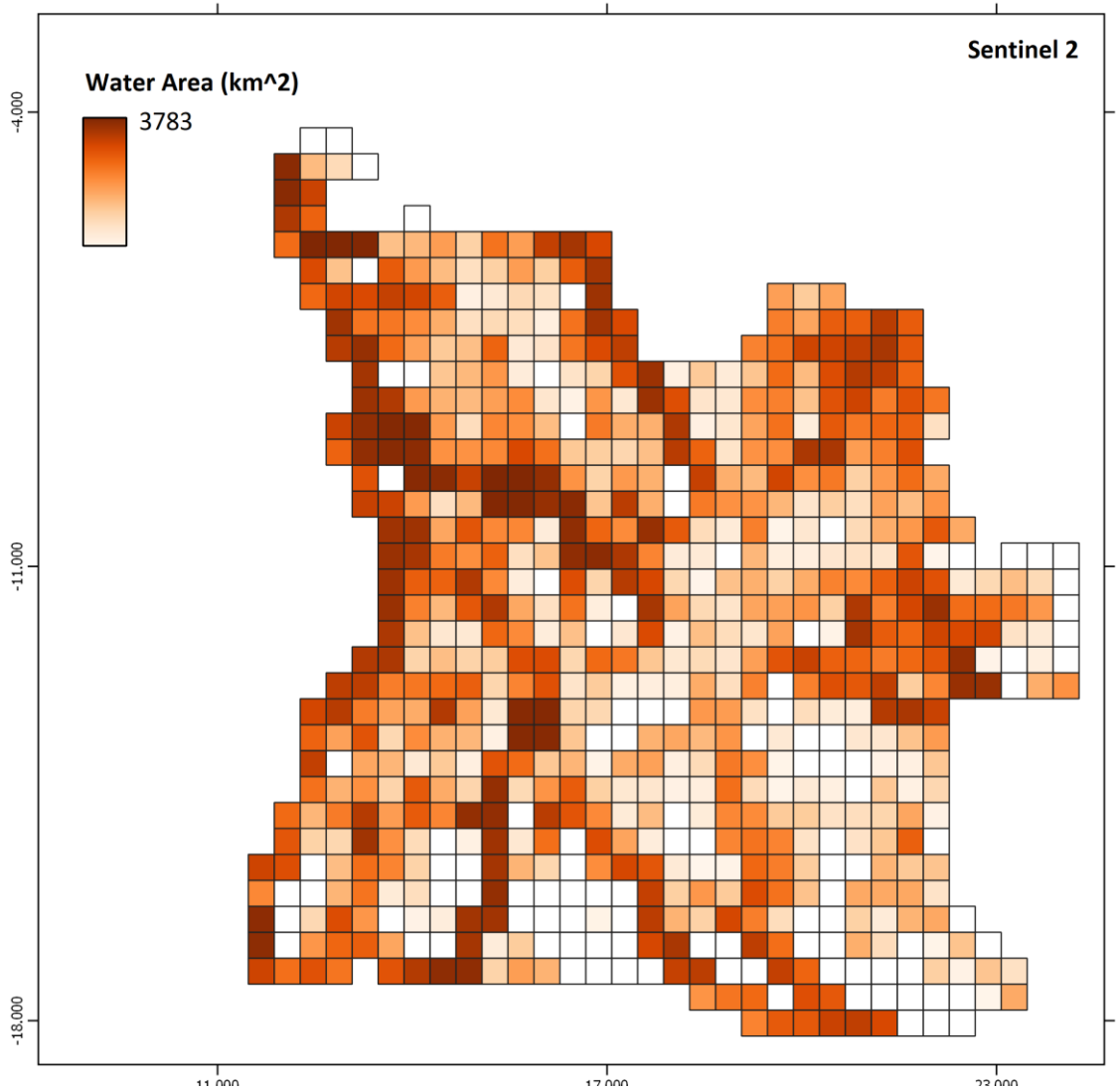


Overlap between OSM and GSW:

Dataset	Area (km^2)	Ratio (%)
Total Water	21759	100
OSM	2267	10
GSW	12865	59
OSM and GSW	1906	9
OSM, no GSW	8894	41
GSW, no OSM	10959	50

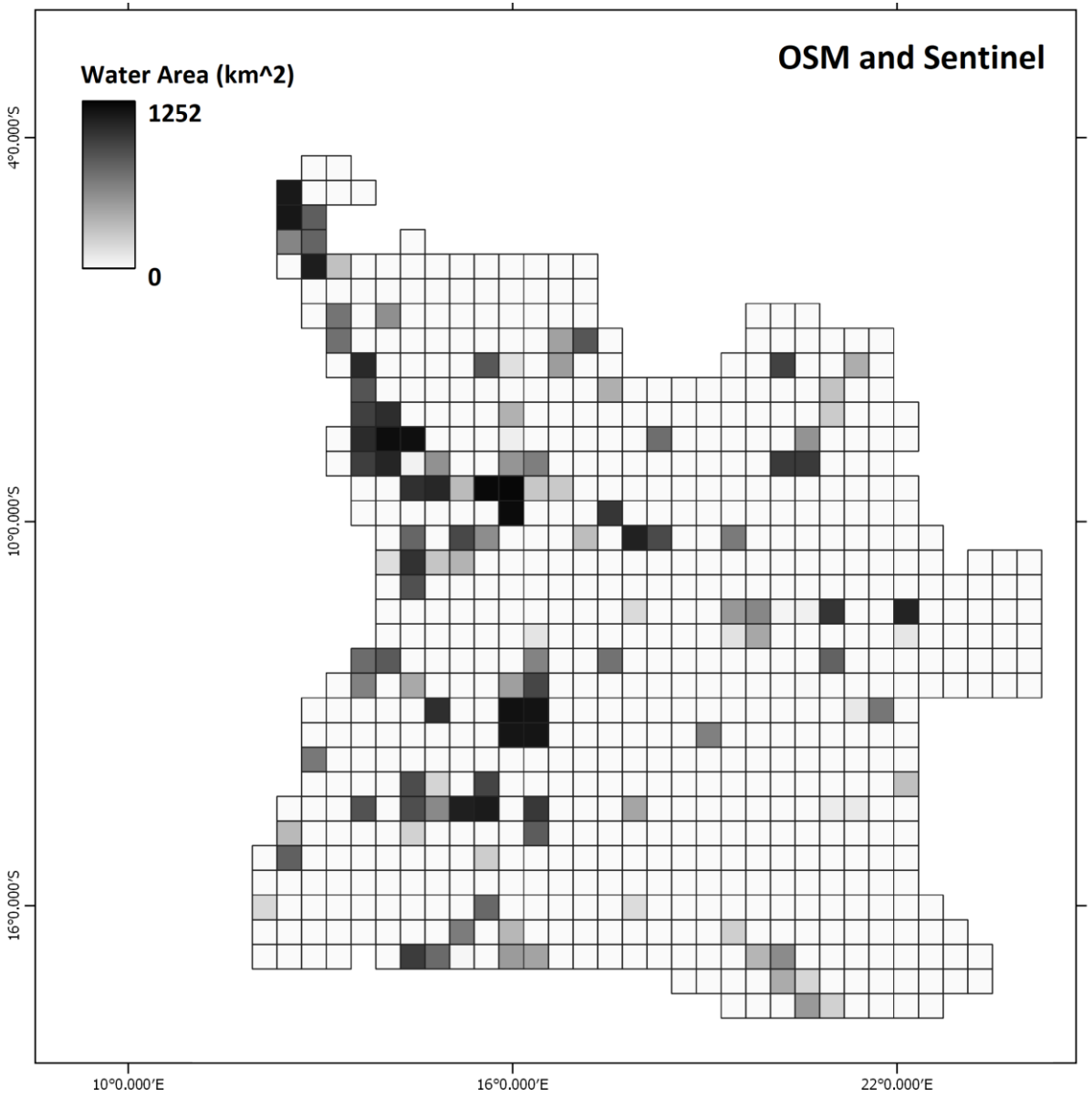
Thematic Differences between OSM and Sentinel 2





Implementation & Results

Completeness



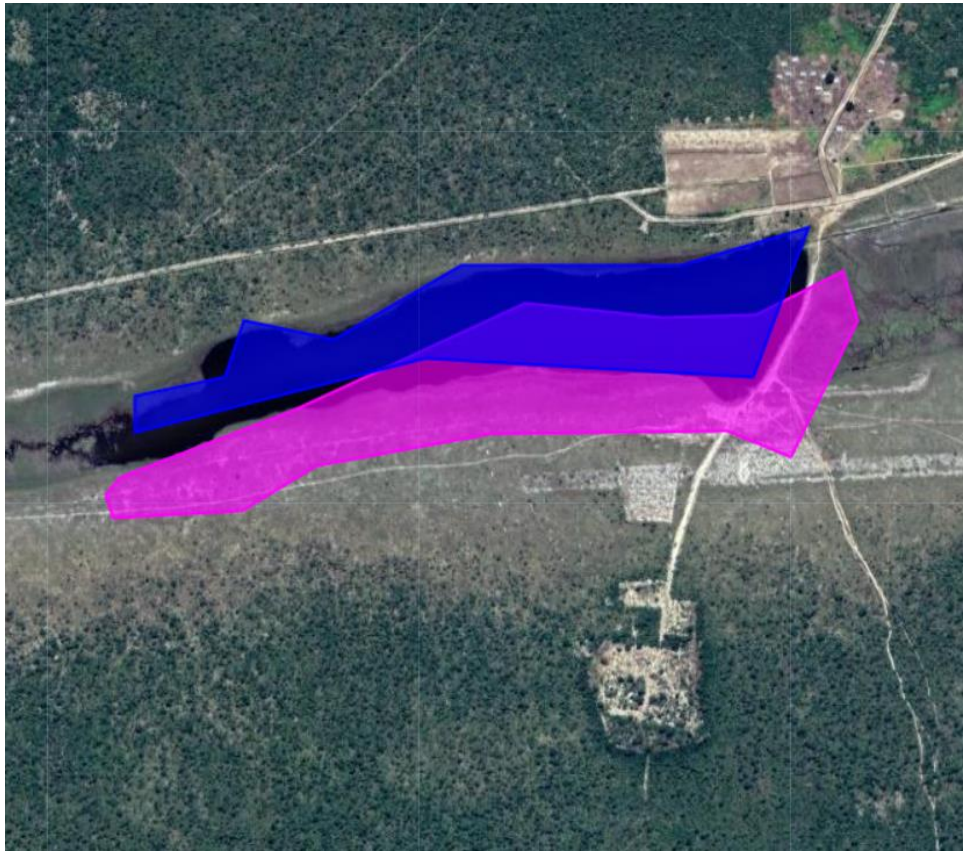
Overlap between OSM and Sentinel 2:

Dataset	Area (km^2)	Ratio (%)
Total Water	6933	100
OSM	2267	33
Sentinel 2	3783	55
OSM and Sentinel 2	1252	19
OSM, no Sentinel 2	3151	45
Sentinel 2, no OSM	2530	36

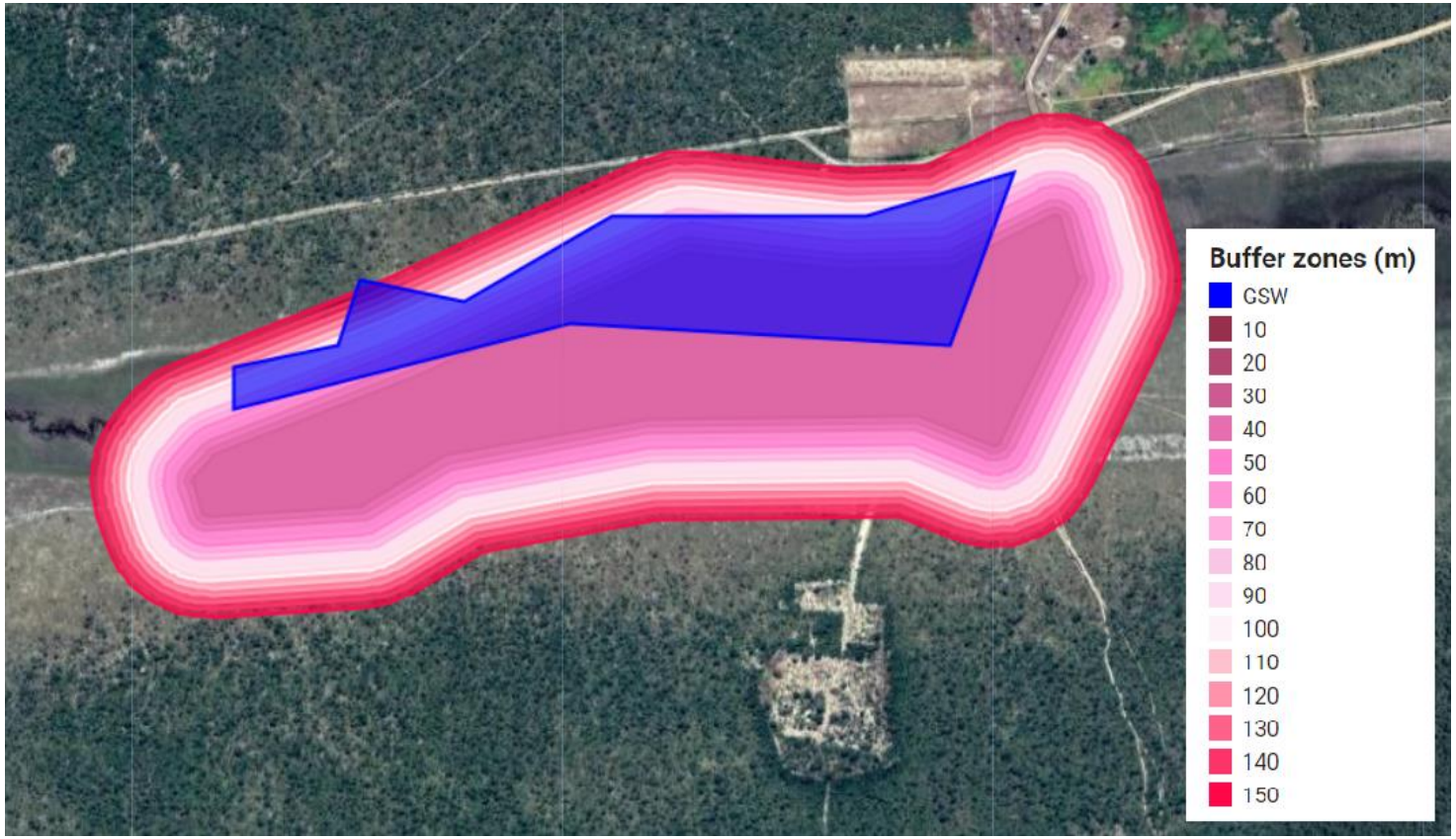
Implementation & Results

Goodchild's Percentage of Overlap

Features OSM and GSW



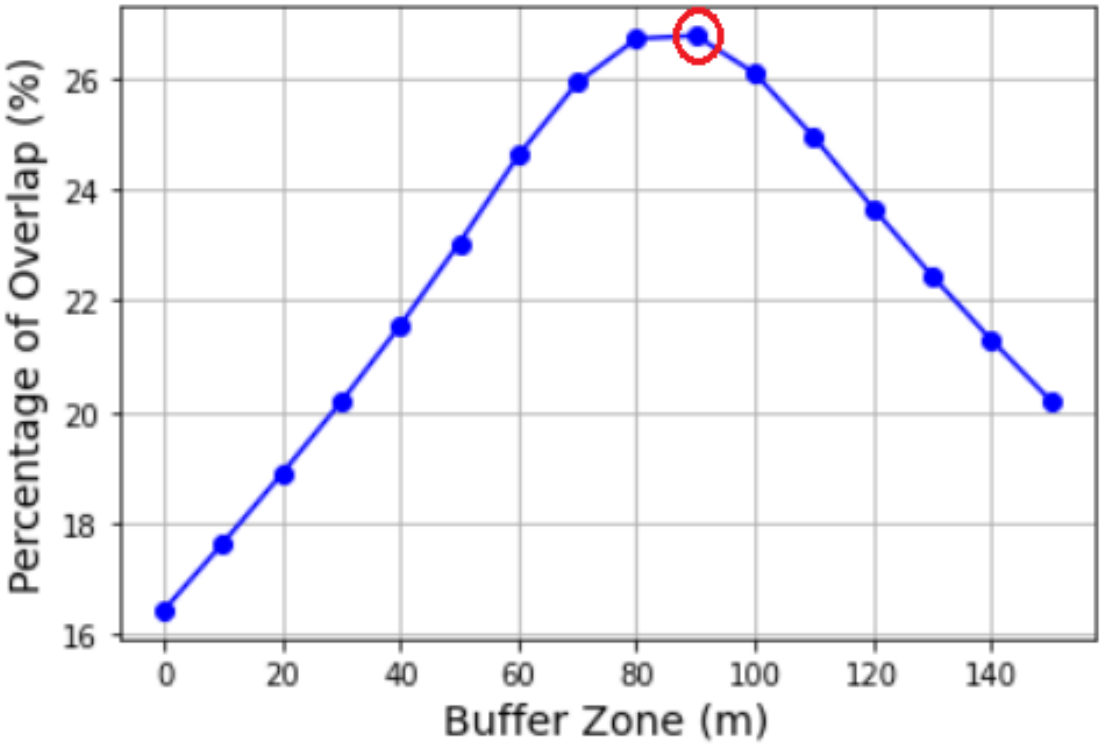
Increasing Buffer zones [0,150] m



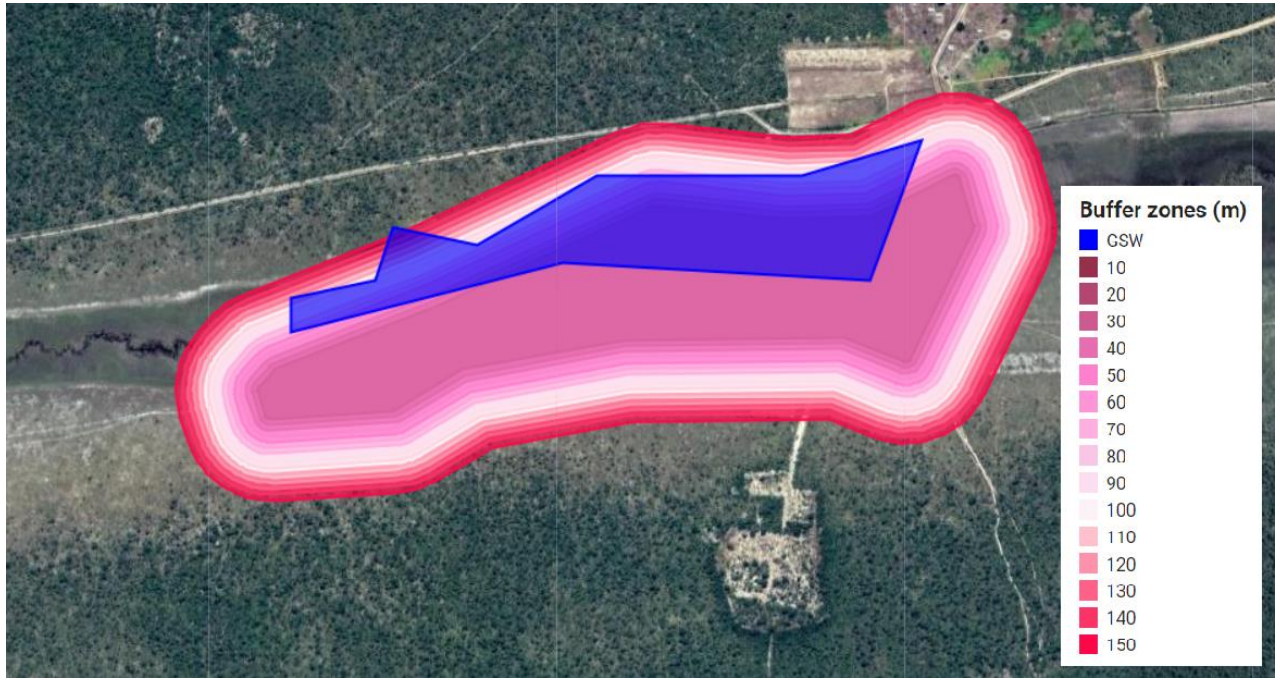
Implementation & Results

Goodchild's Percentage of Overlap

Distance between OSM and GSW feature



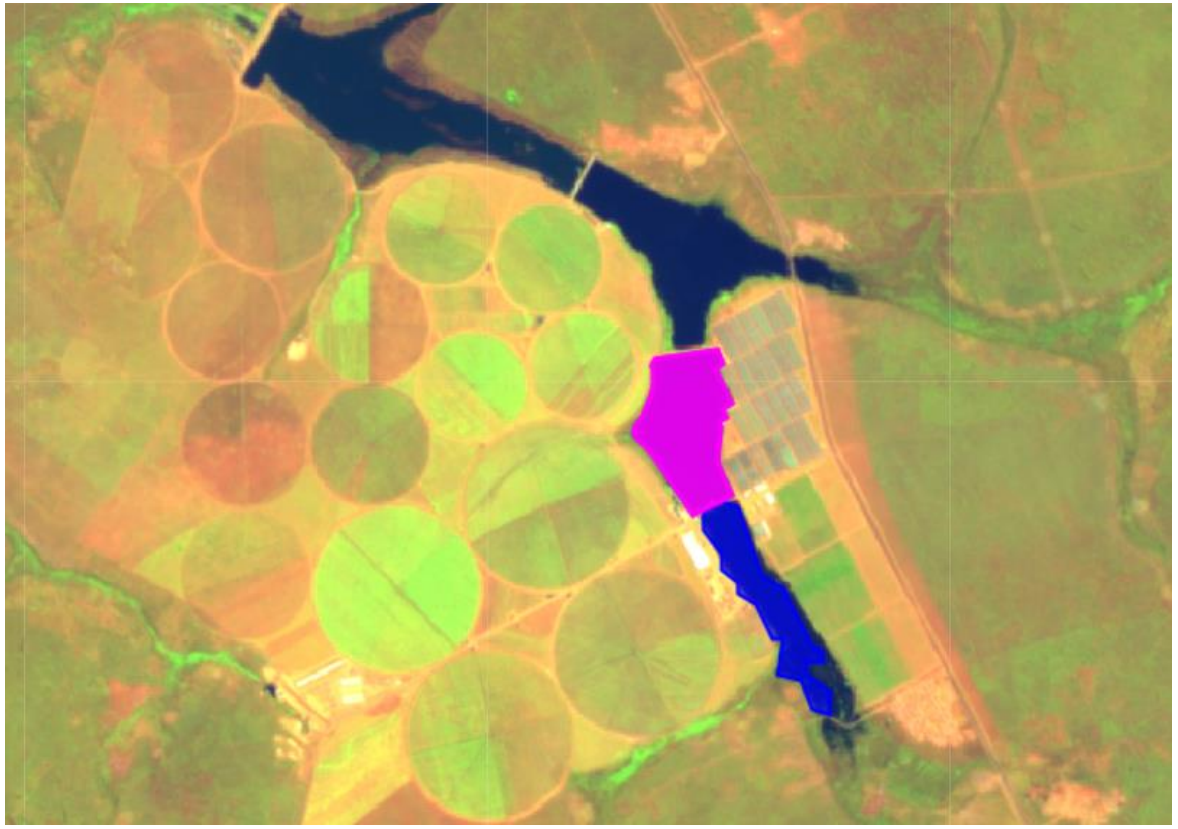
Peak of histogram



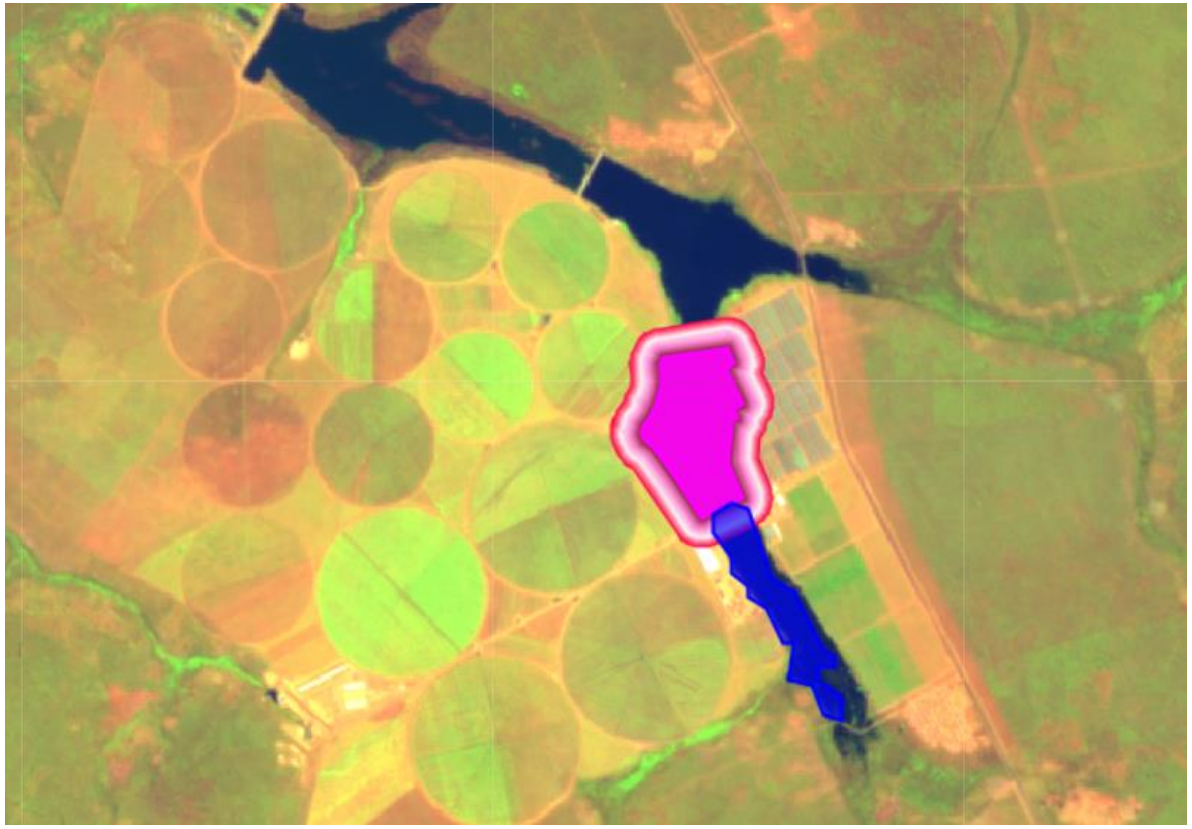
Implementation & Results

Goodchild's Percentage of Overlap

Features OSM and GSW



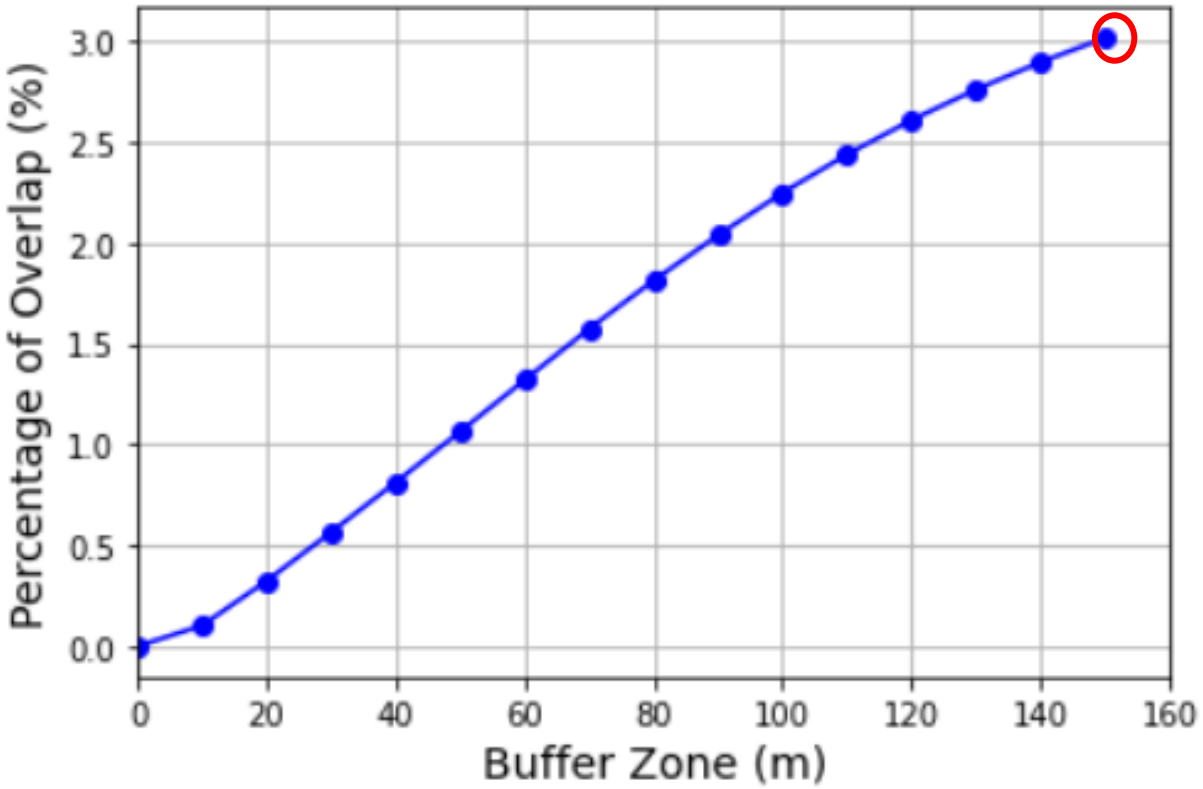
Increasing Buffer zones [0,150] m

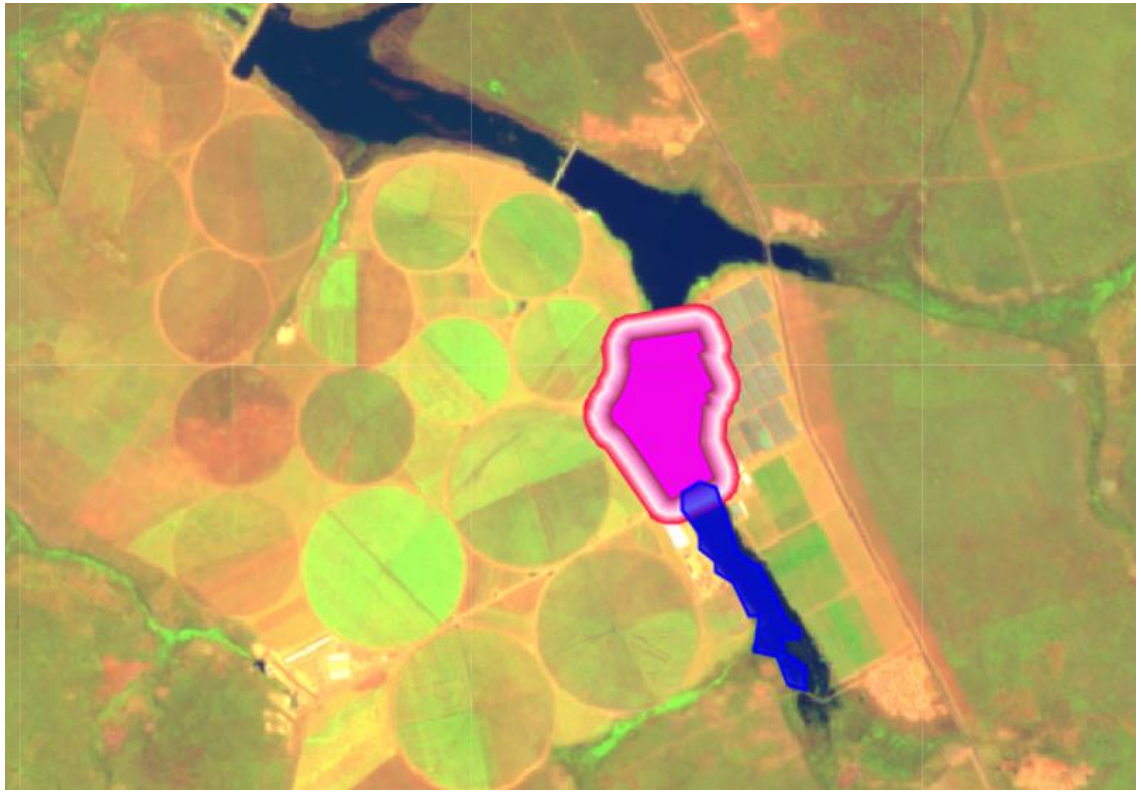


Implementation & Results

Goodchild's Percentage of Overlap

Distance between OSM and GSW feature

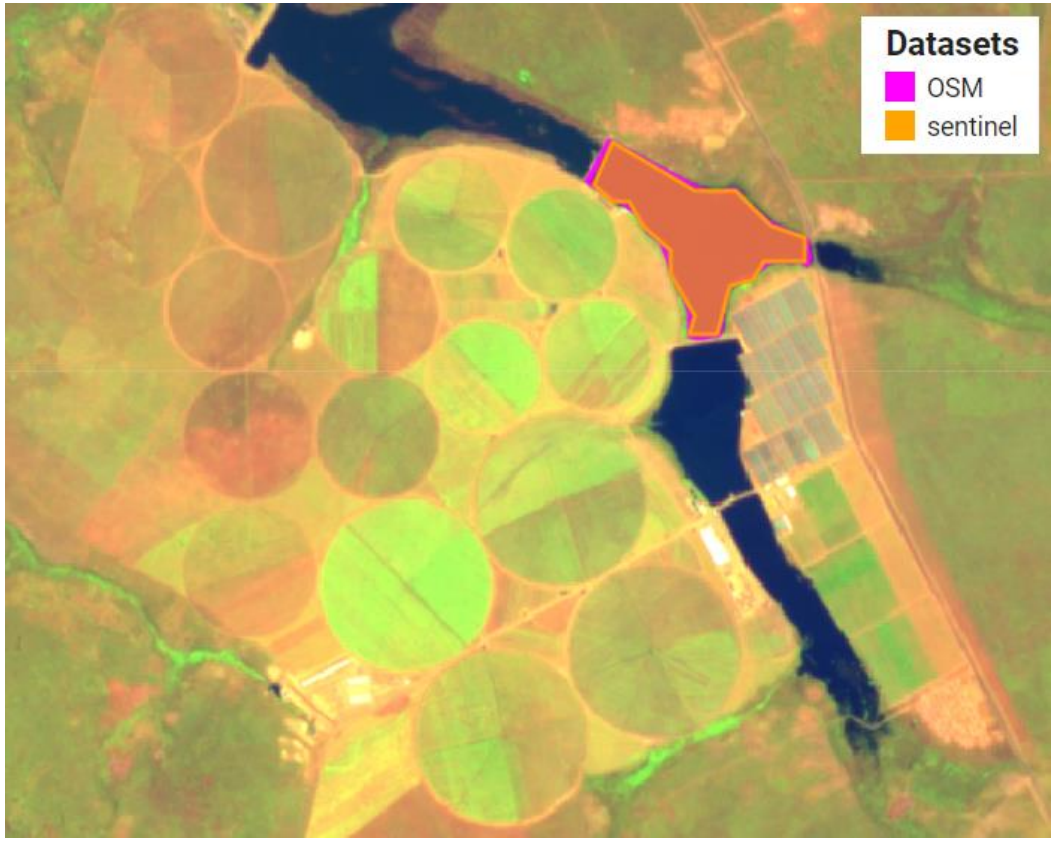




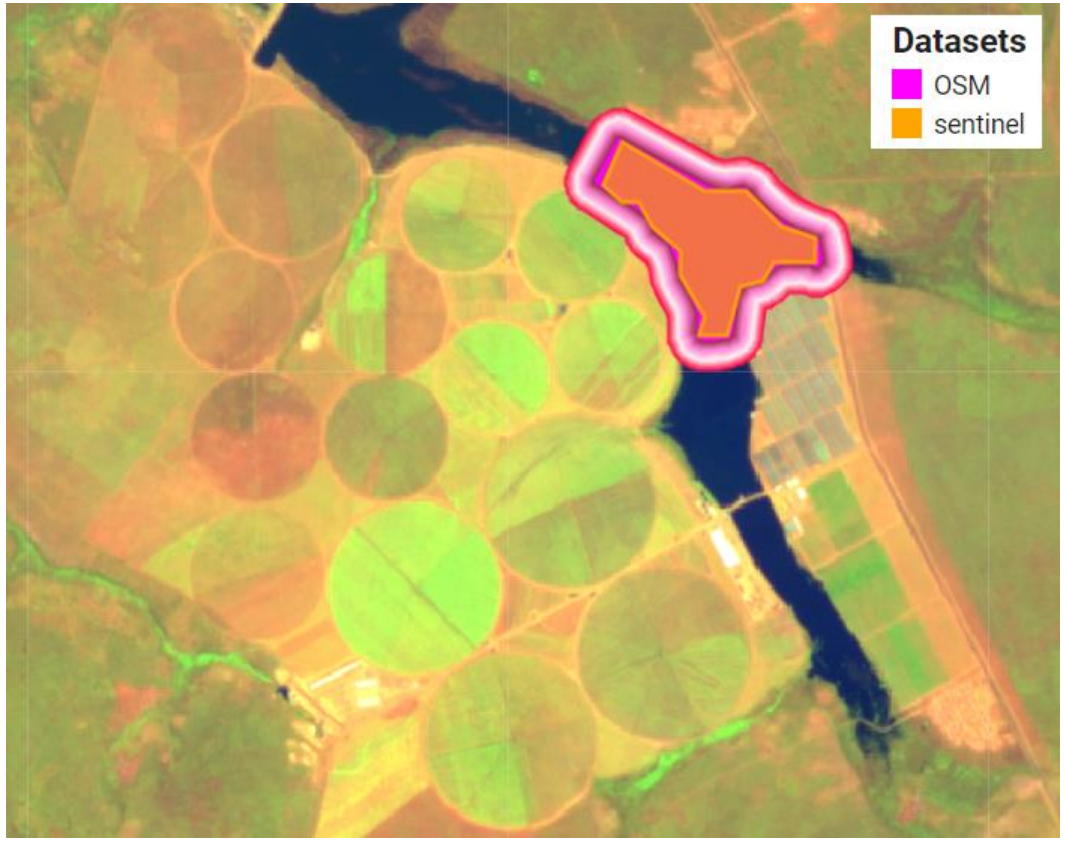
Implementation & Results

Goodchild's Percentage of Overlap

Features OSM and Sentinel 2



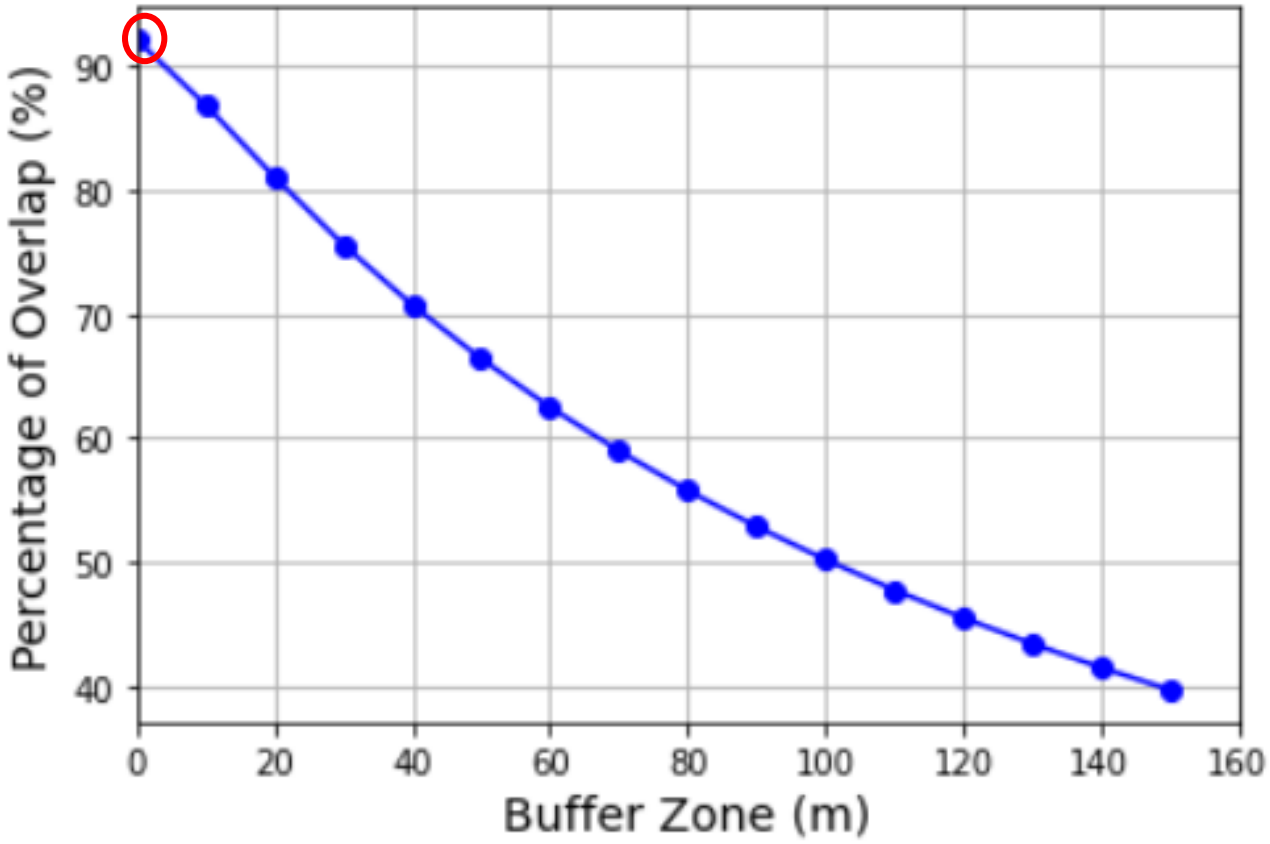
Increasing Buffer zones [0,150] m

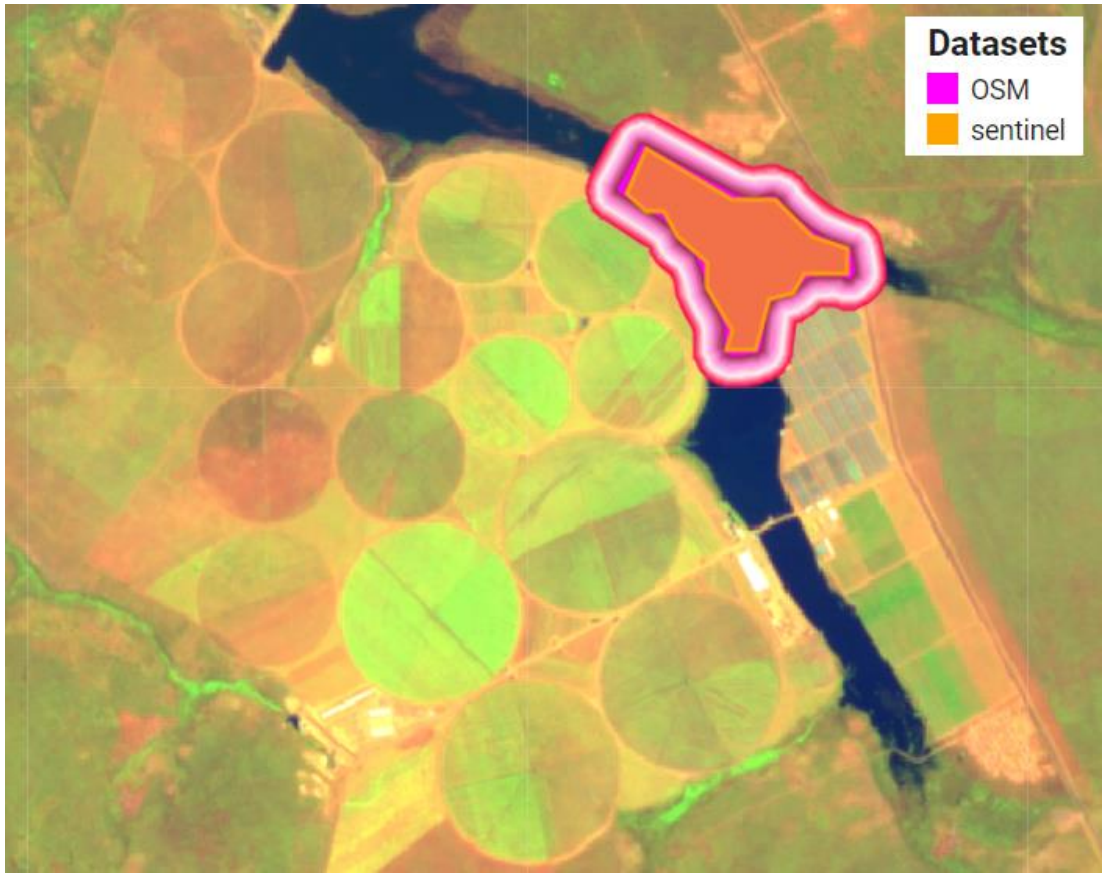


Implementation & Results

Goodchild's Percentage of Overlap

Distance between OSM and Sentinel 2 feature

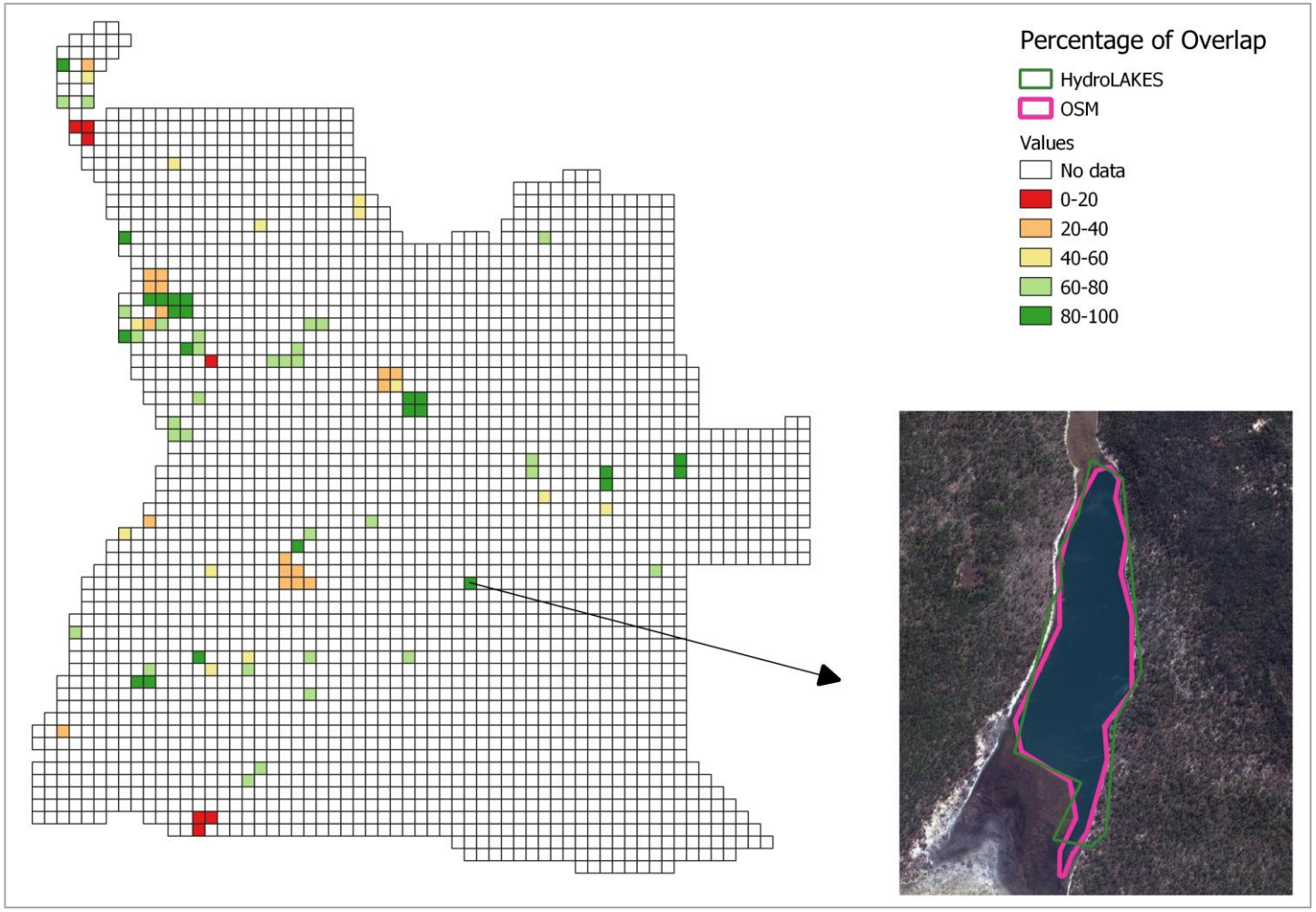




Implementation & Results

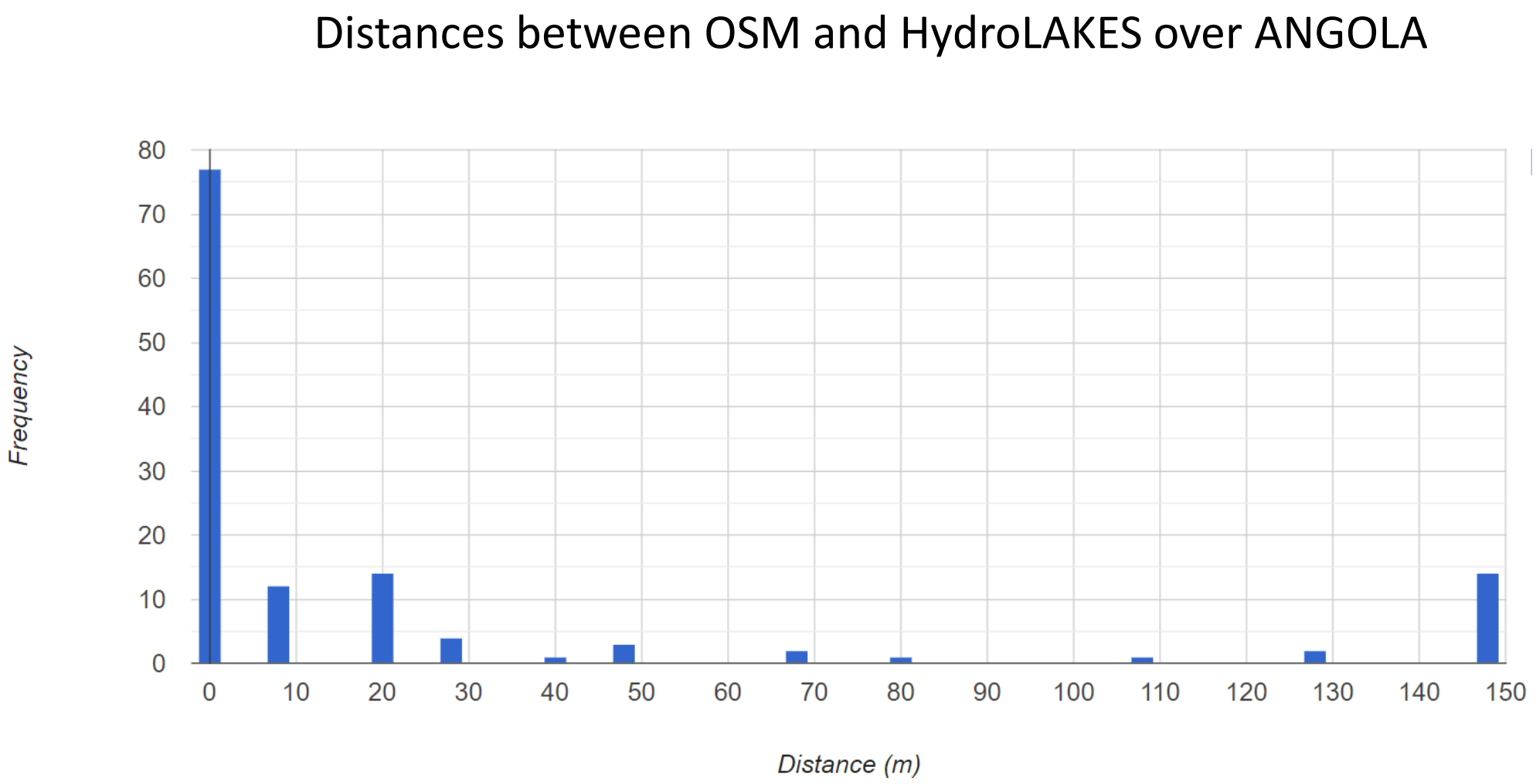
Goodchild's Percentage of Overlap

Percentage of Overlap
between OSM and HydroLAKES



Implementation & Results

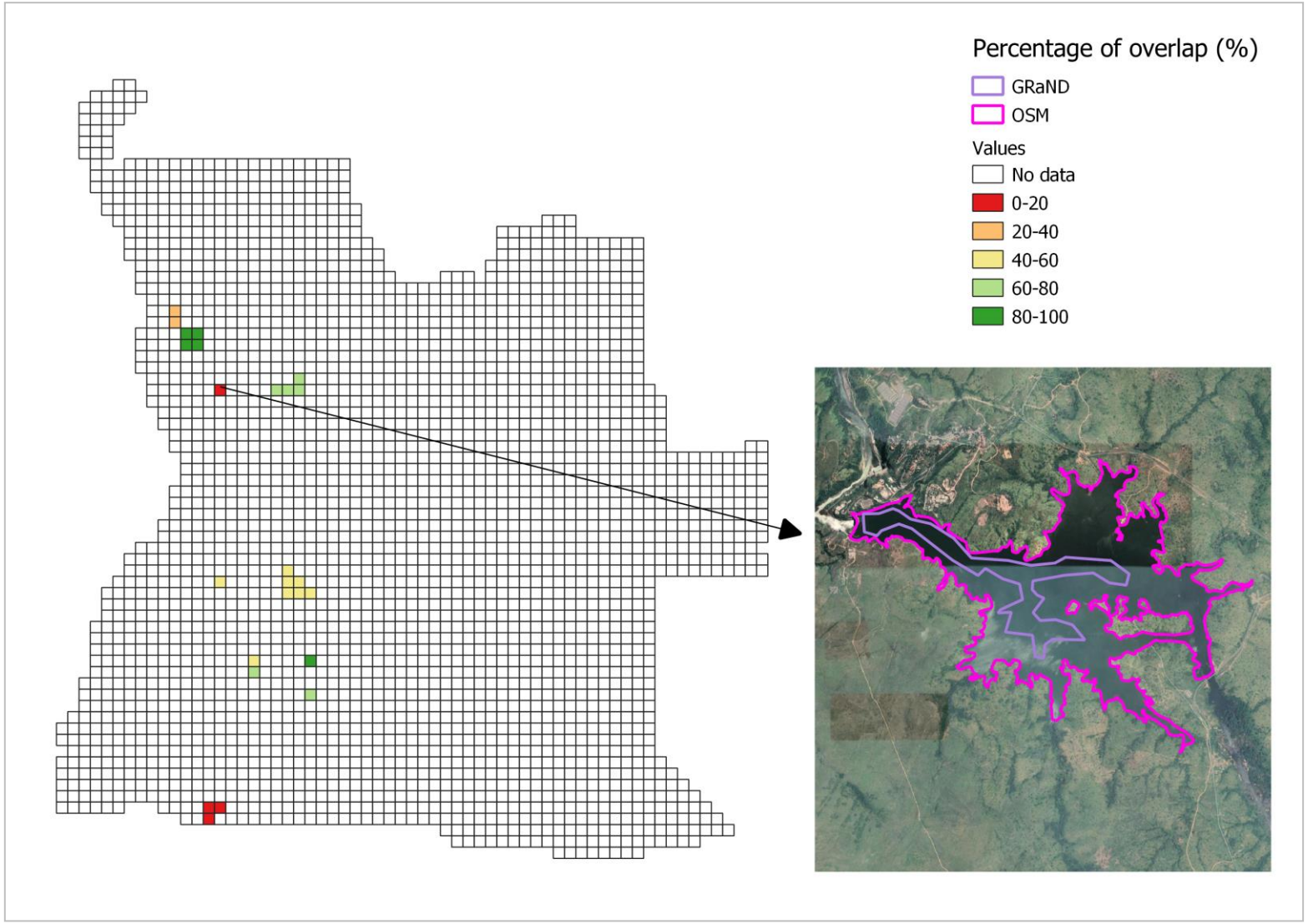
Goodchild's Percentage of Overlap



Implementation & Results

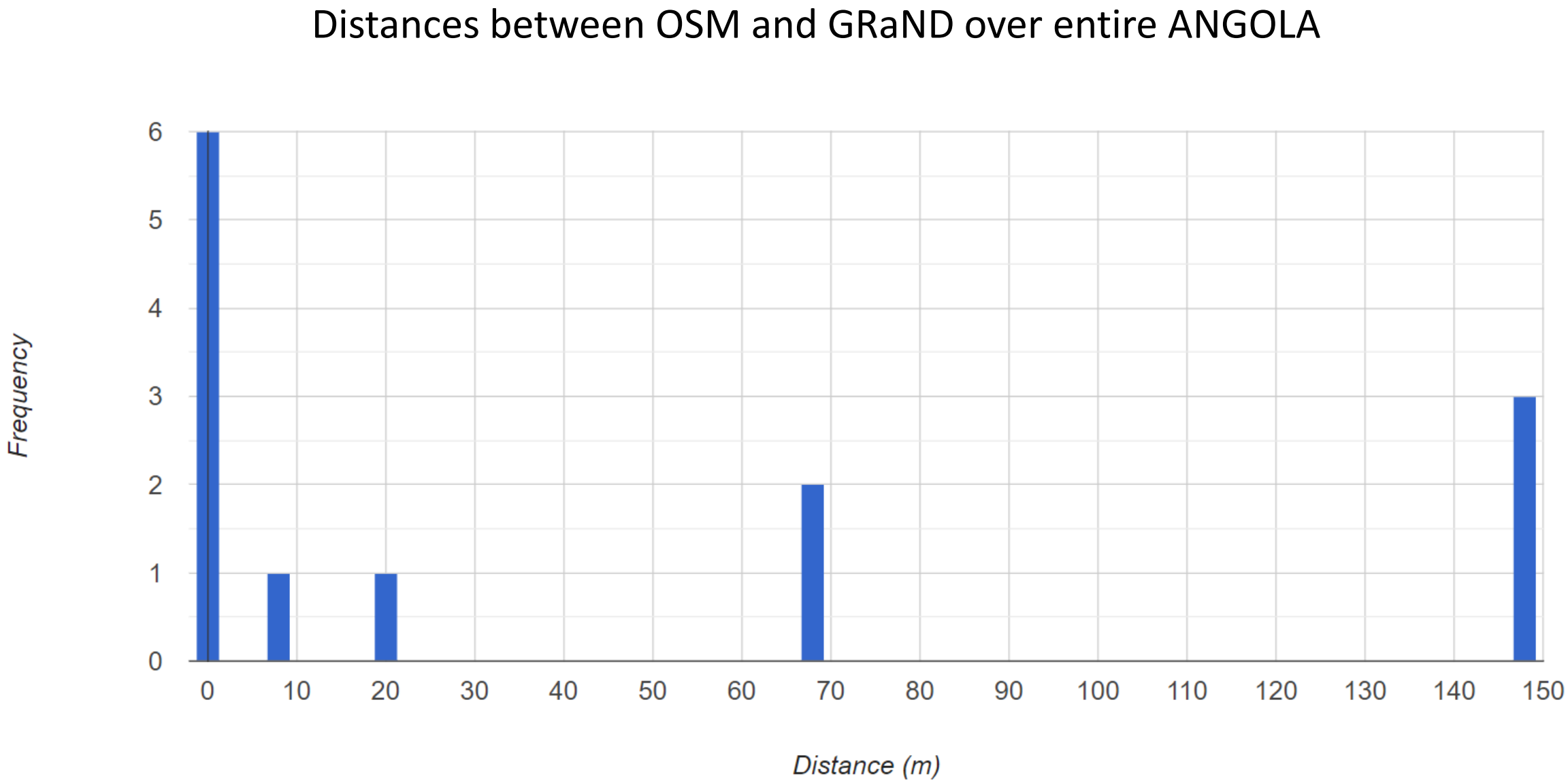
Goodchild's Percentage of Overlap

Percentage of Overlap between OSM and GRaND:



Implementation & Results

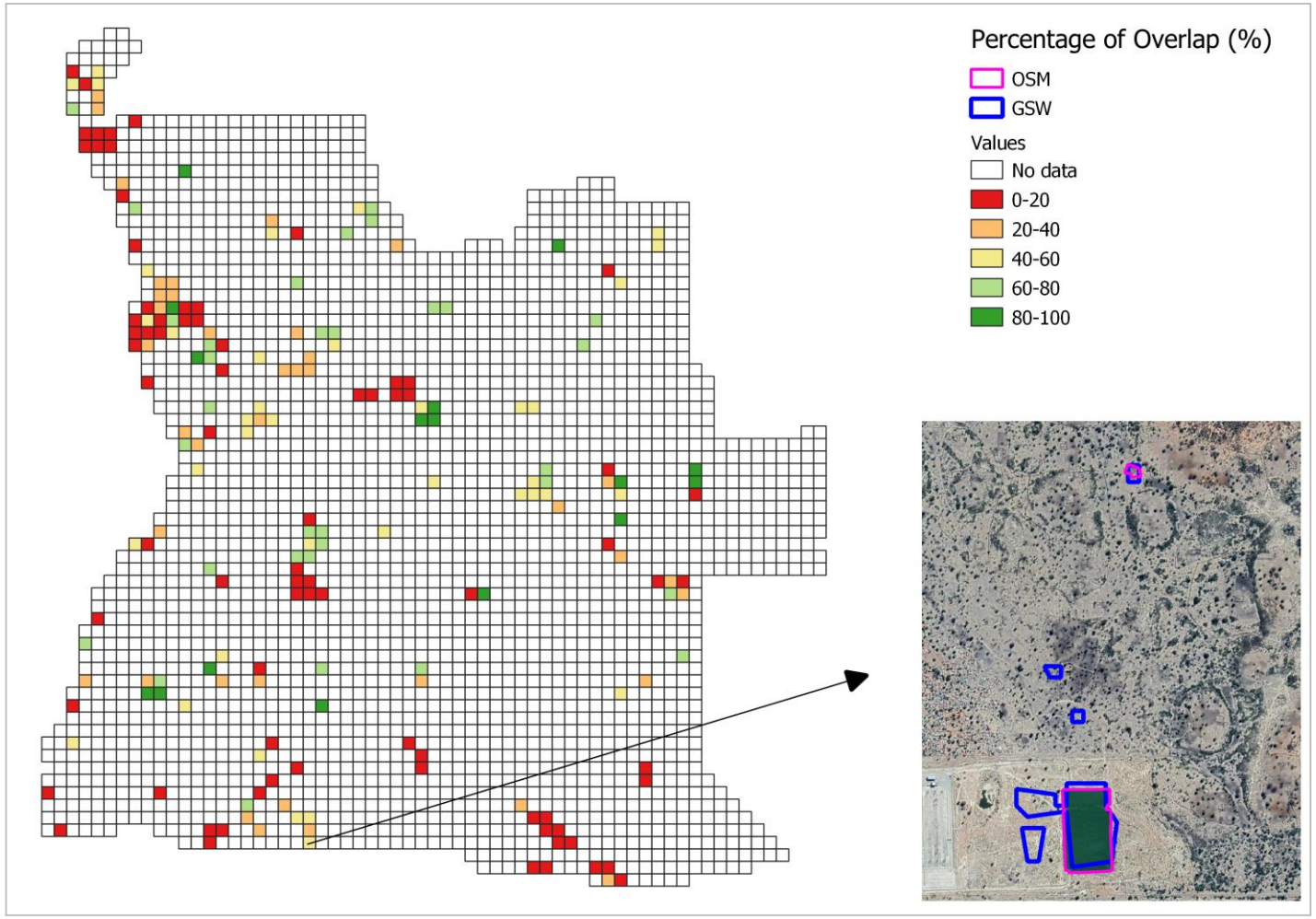
Goodchild's Percentage of Overlap



Implementation & Results

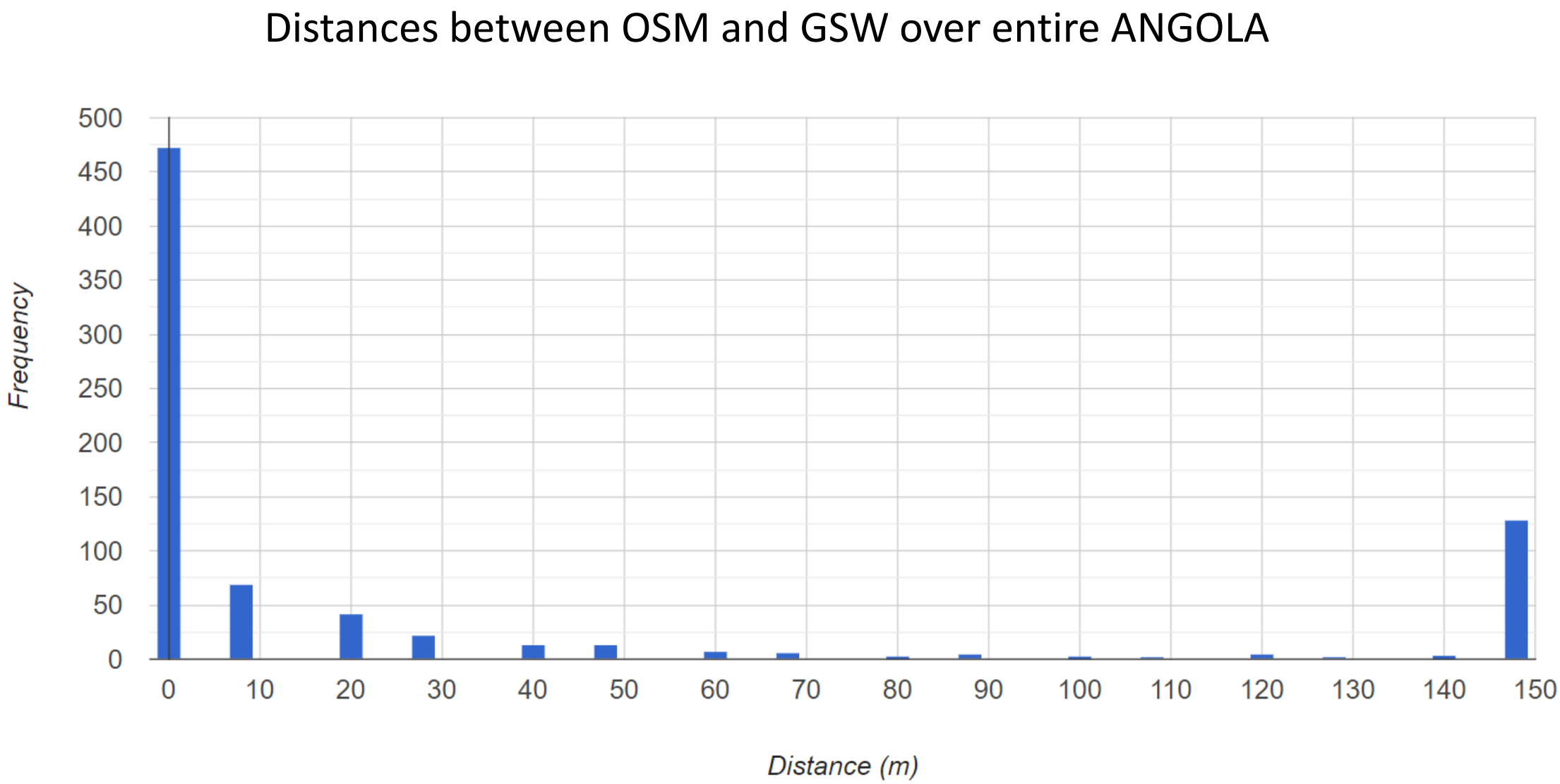
Goodchild's Percentage of Overlap

Percentage of Overlap between OSM and GSW



Implementation

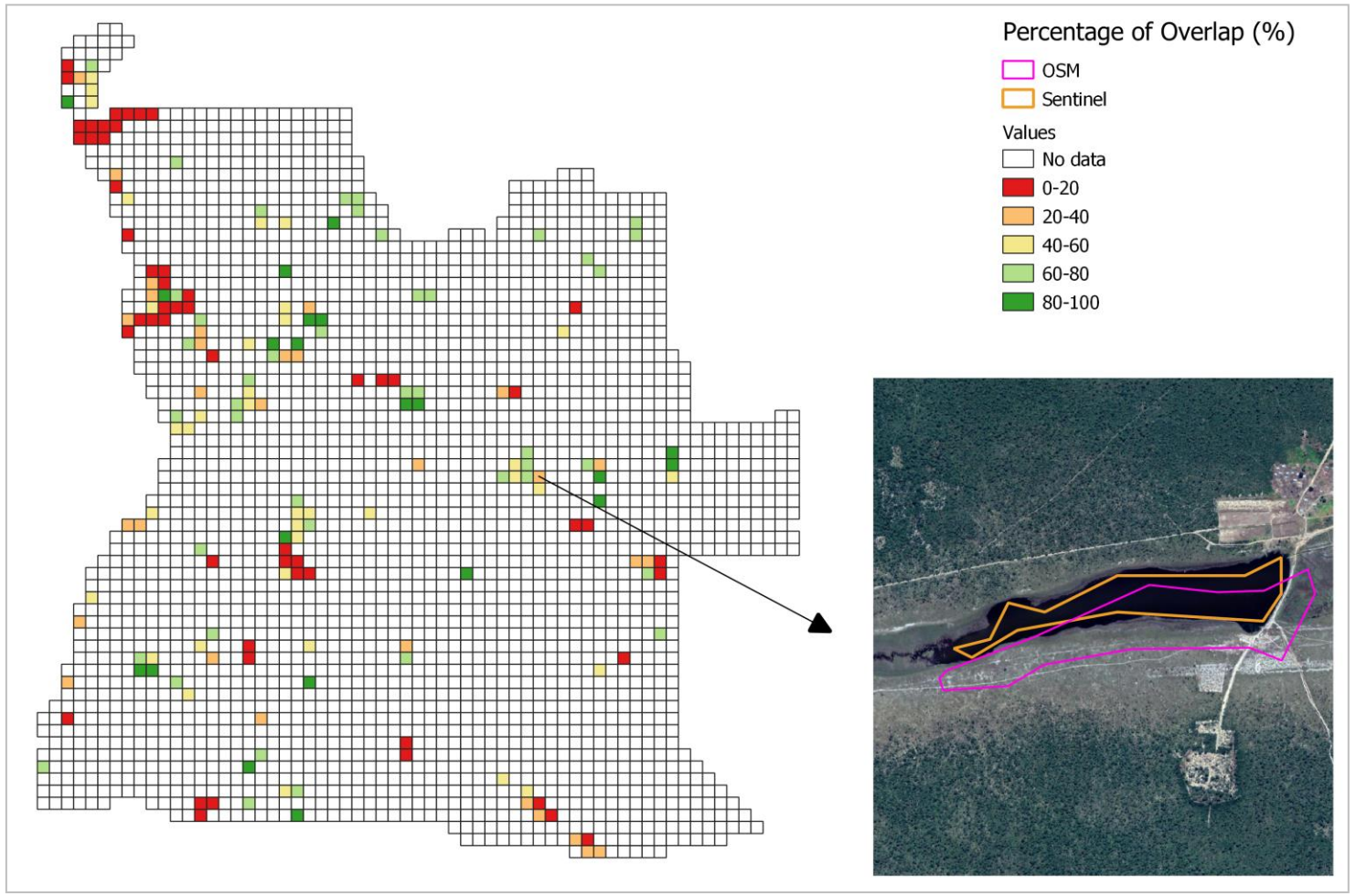
Positional Accuracy: Goodchild's Method



Implementation & Results

Goodchild's Percentage of Overlap

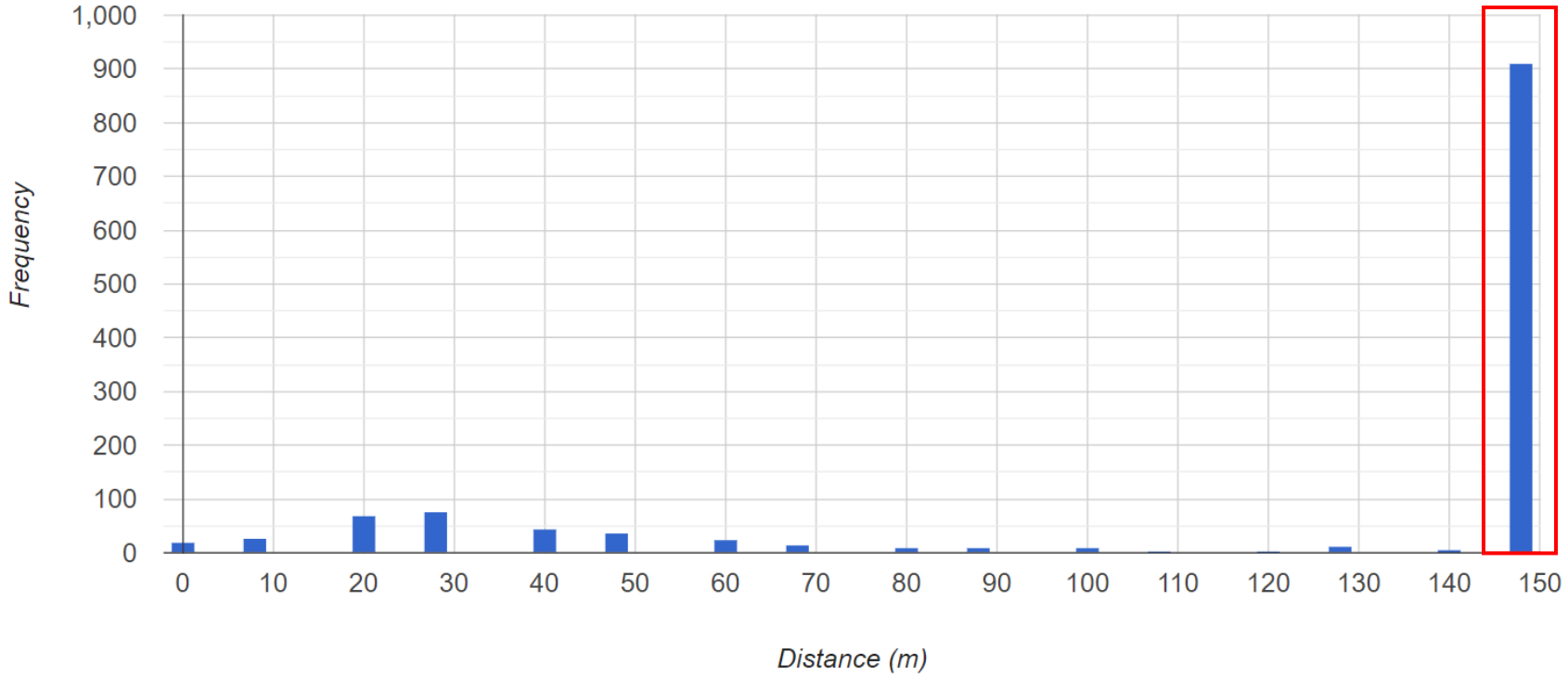
Percentage of Overlap between OSM and Sentinel 2



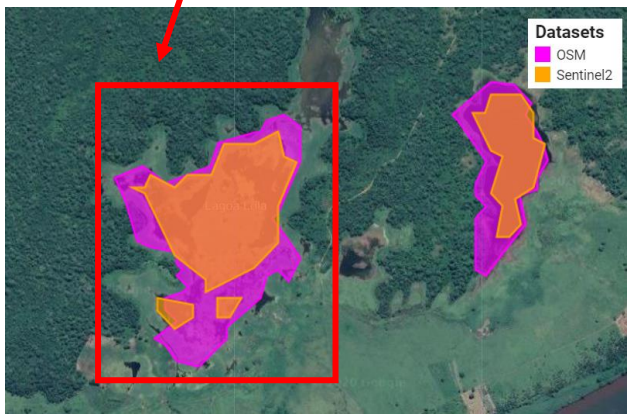
Implementation

Positional Accuracy: Goodchild's Method

Distances between OSM and Sentinel 2 over entire ANGOLA

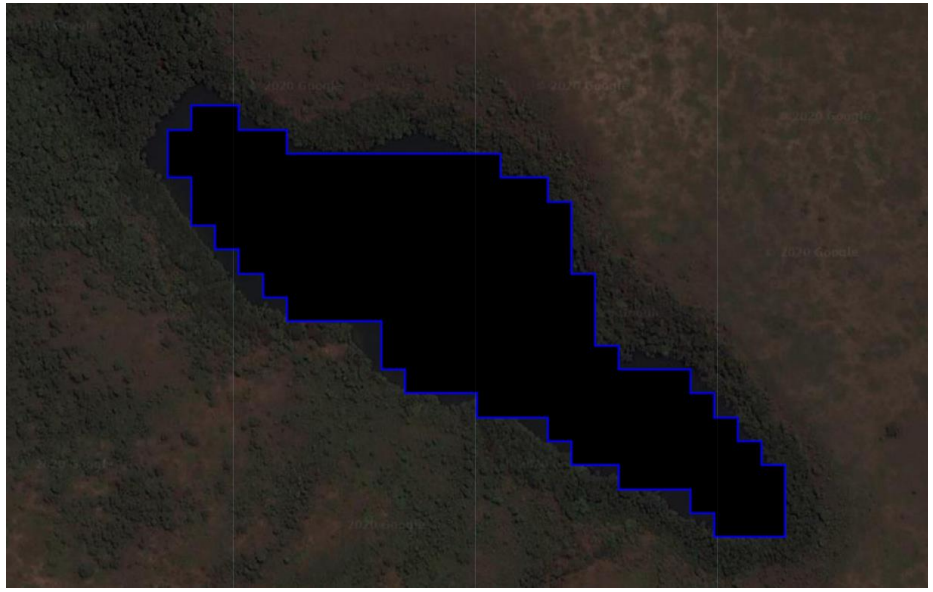


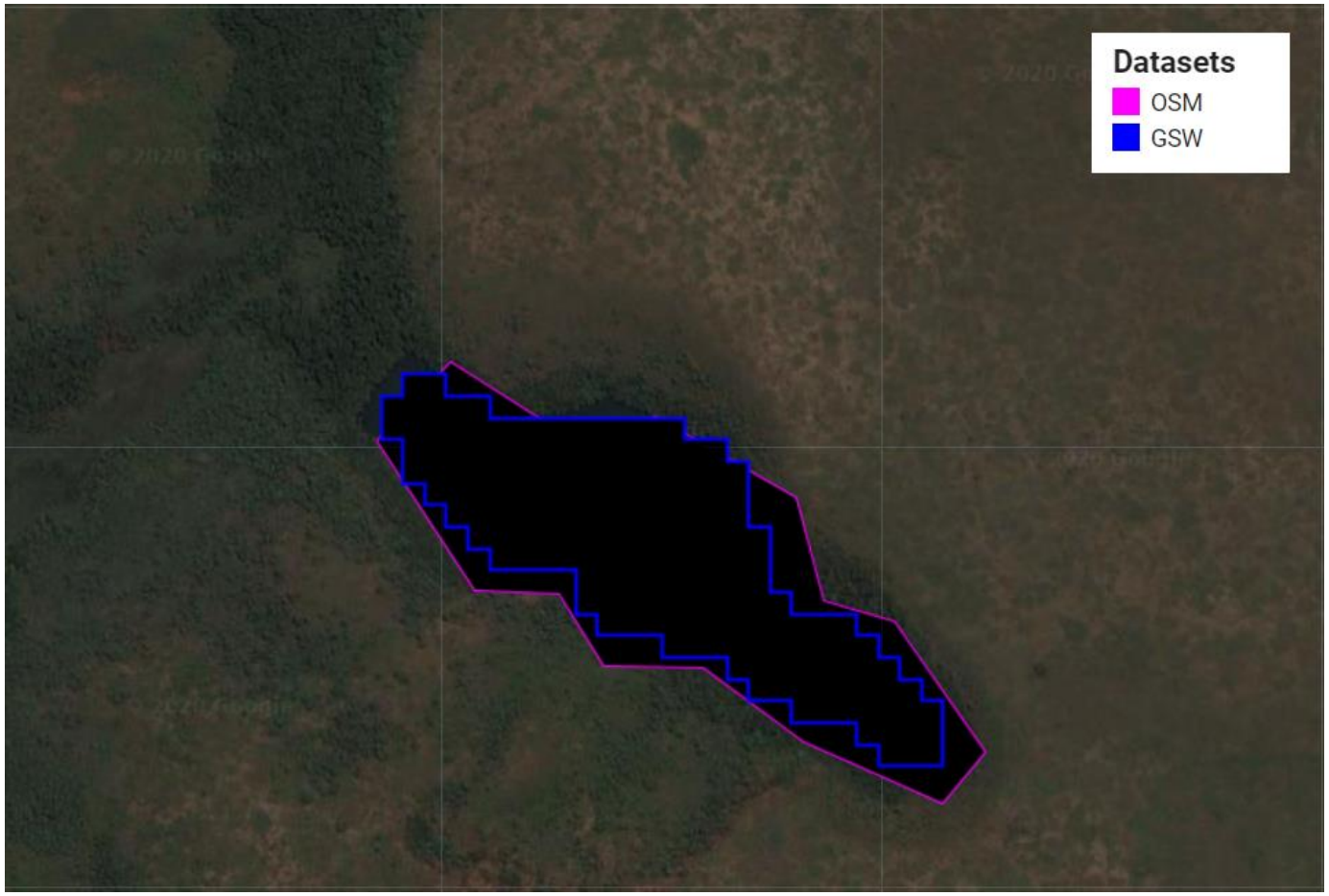
Clusters of Sentinel 2 water polygons



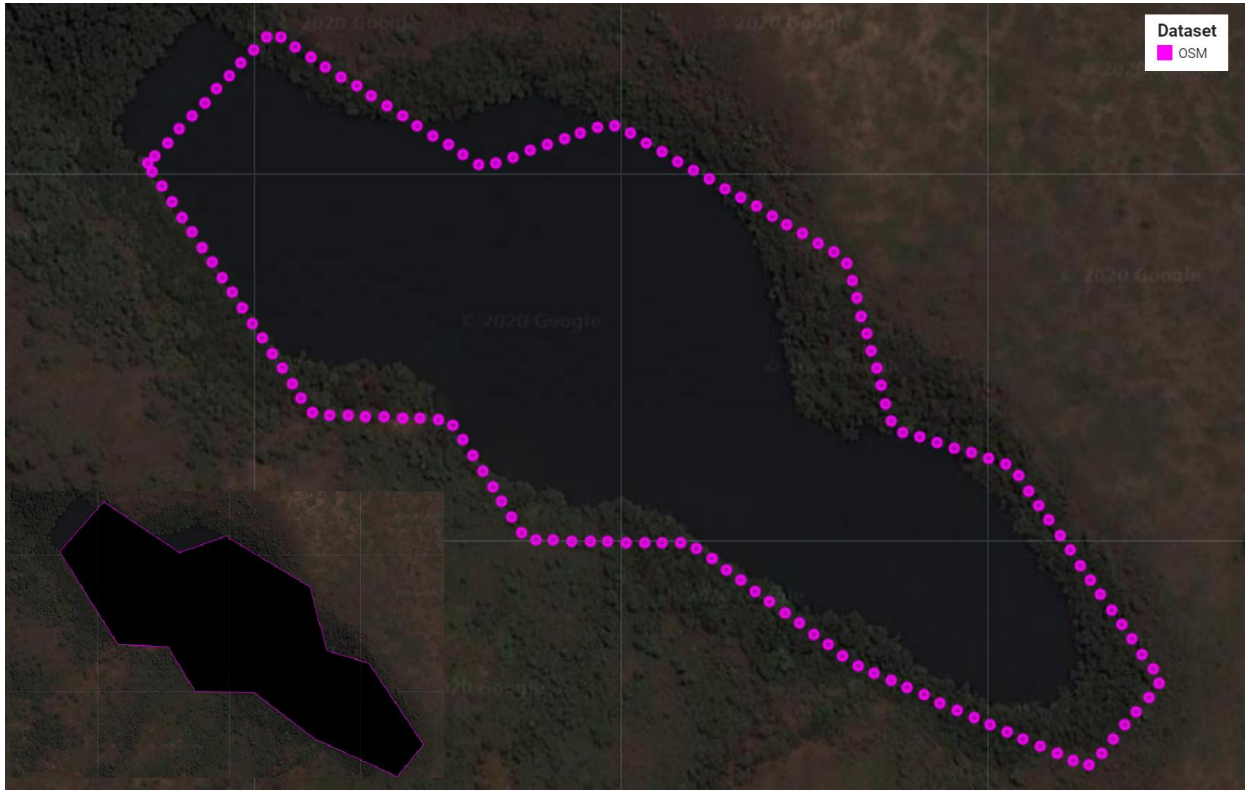
Implementation & Results

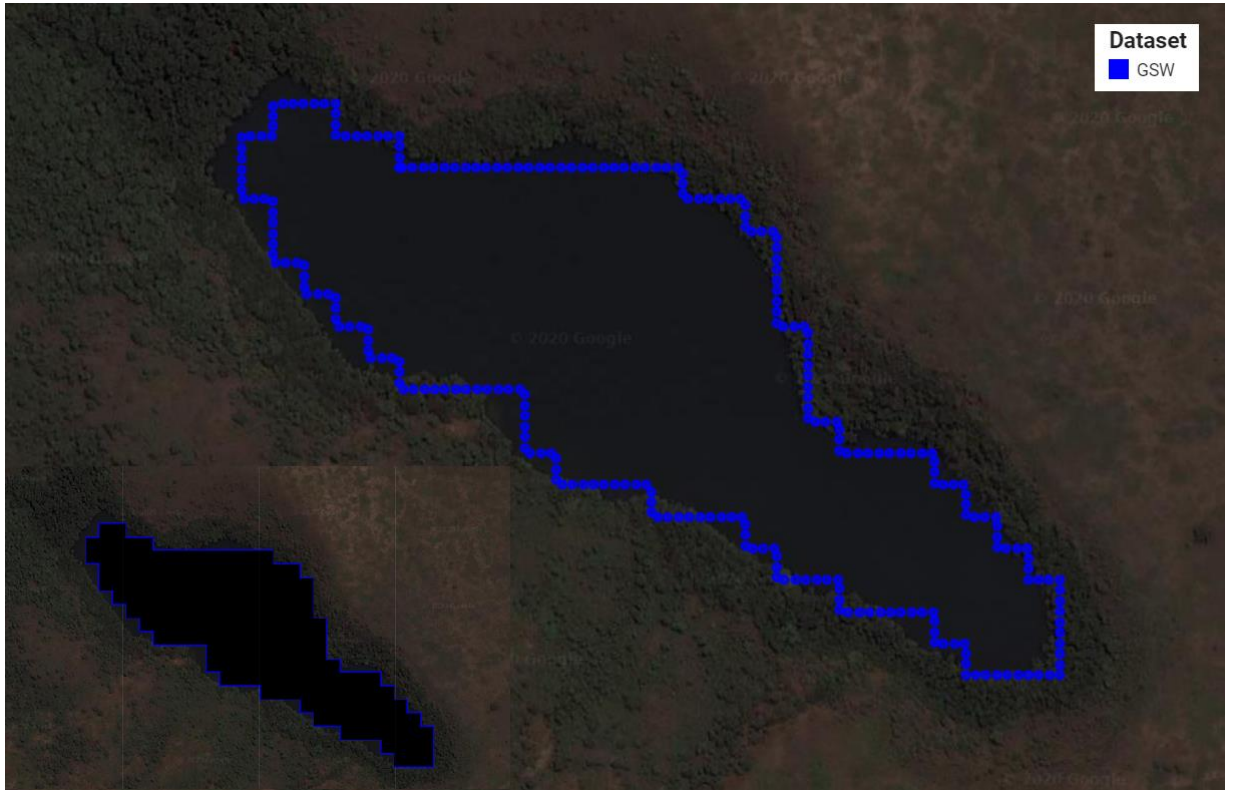
Hausdorff Distance



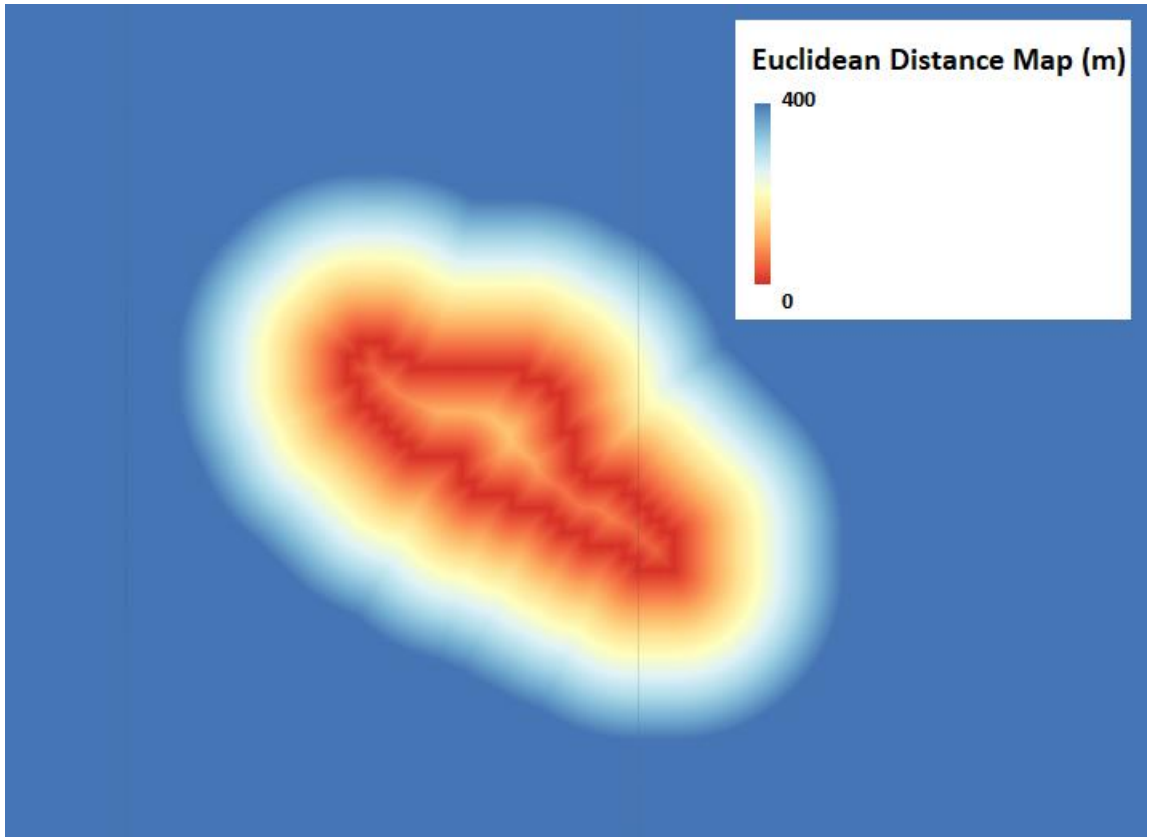


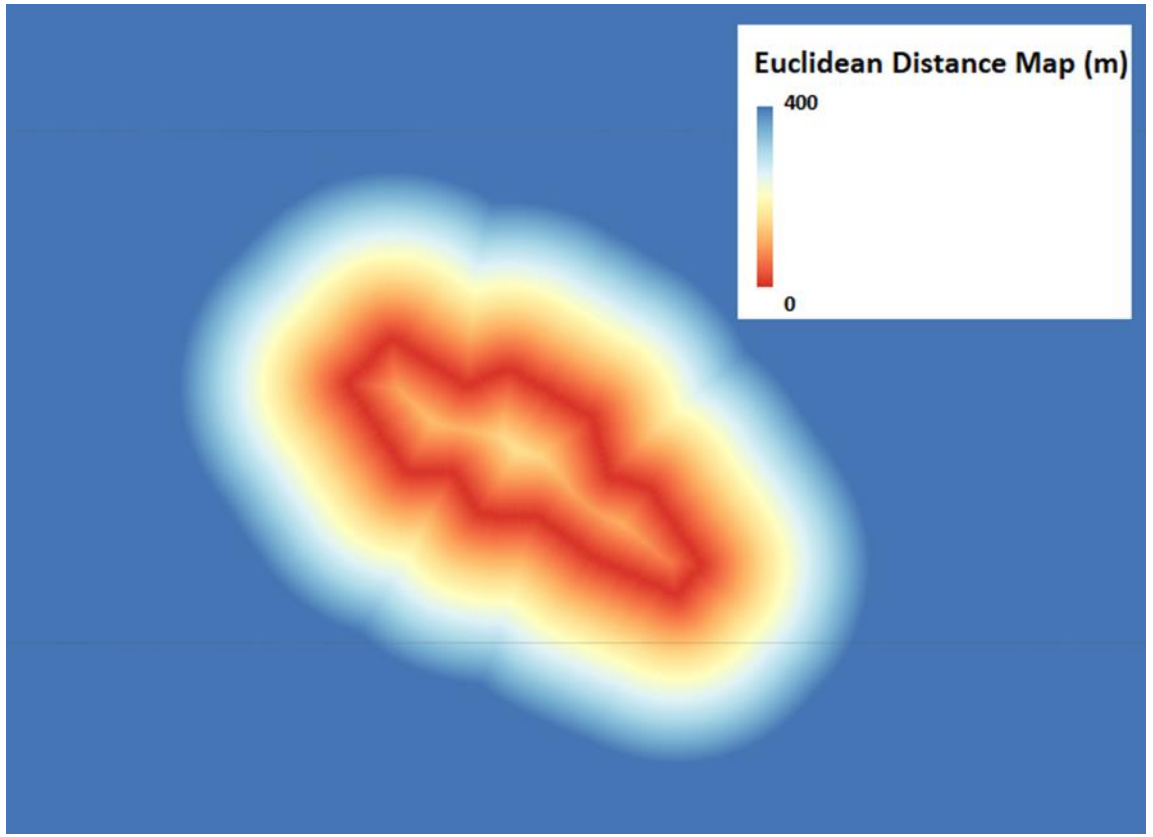
Sampling points on OSM and GSW features





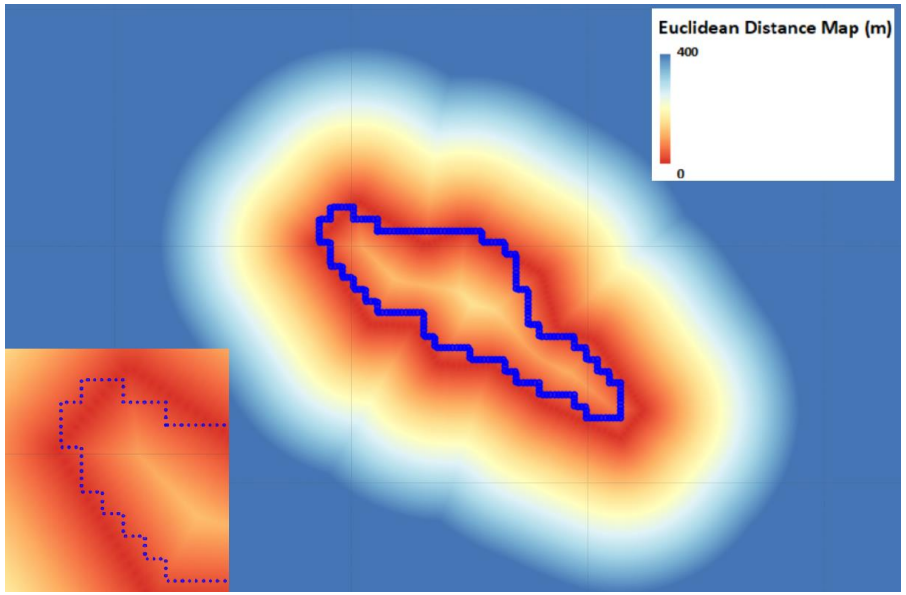
Euclidean Distance Maps of OSM and GSW features

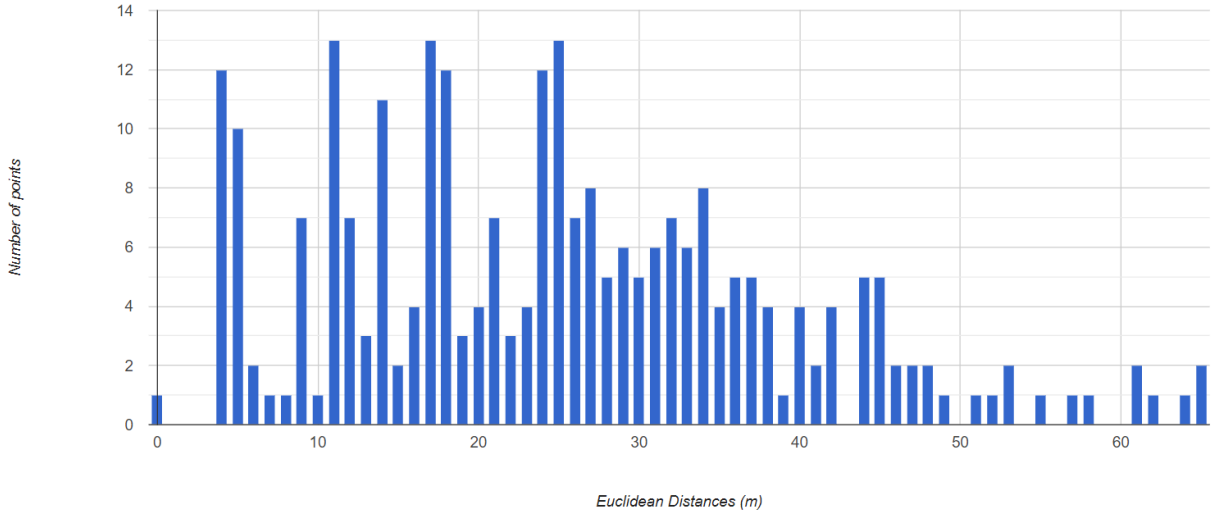


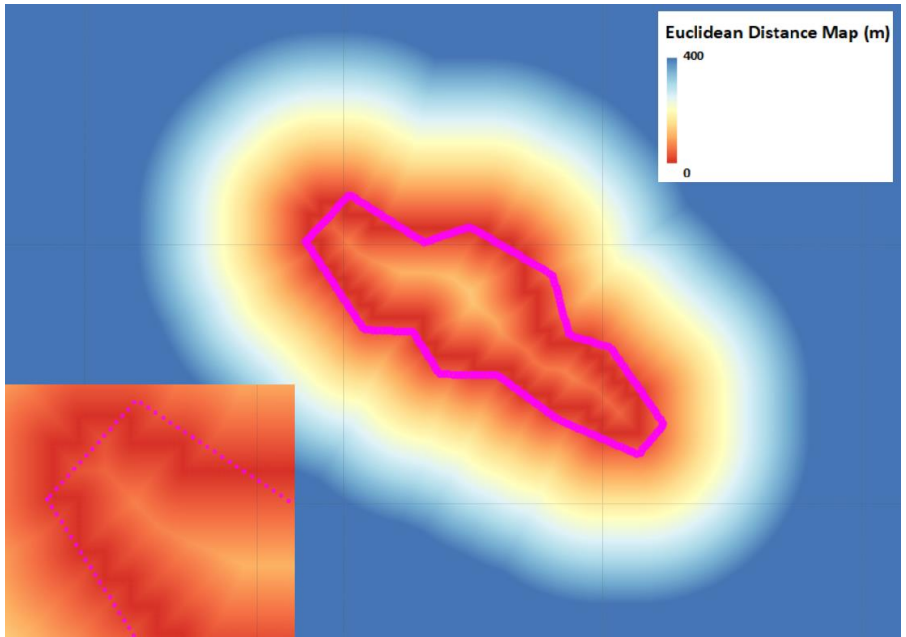


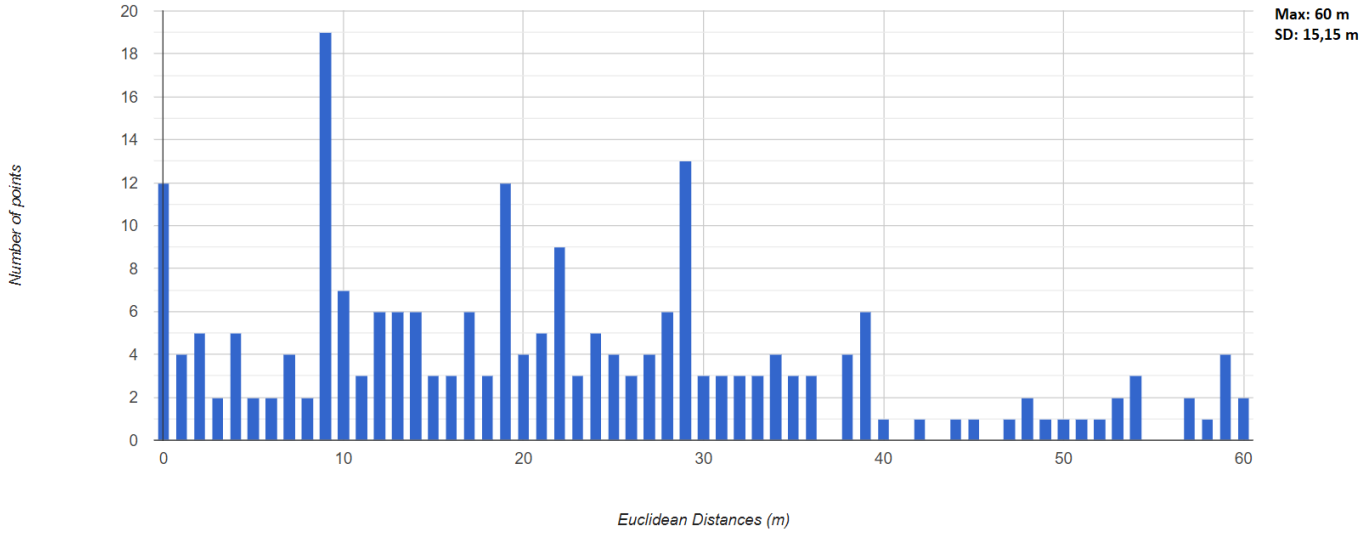
Implementation & Results

Hausdorff Distance





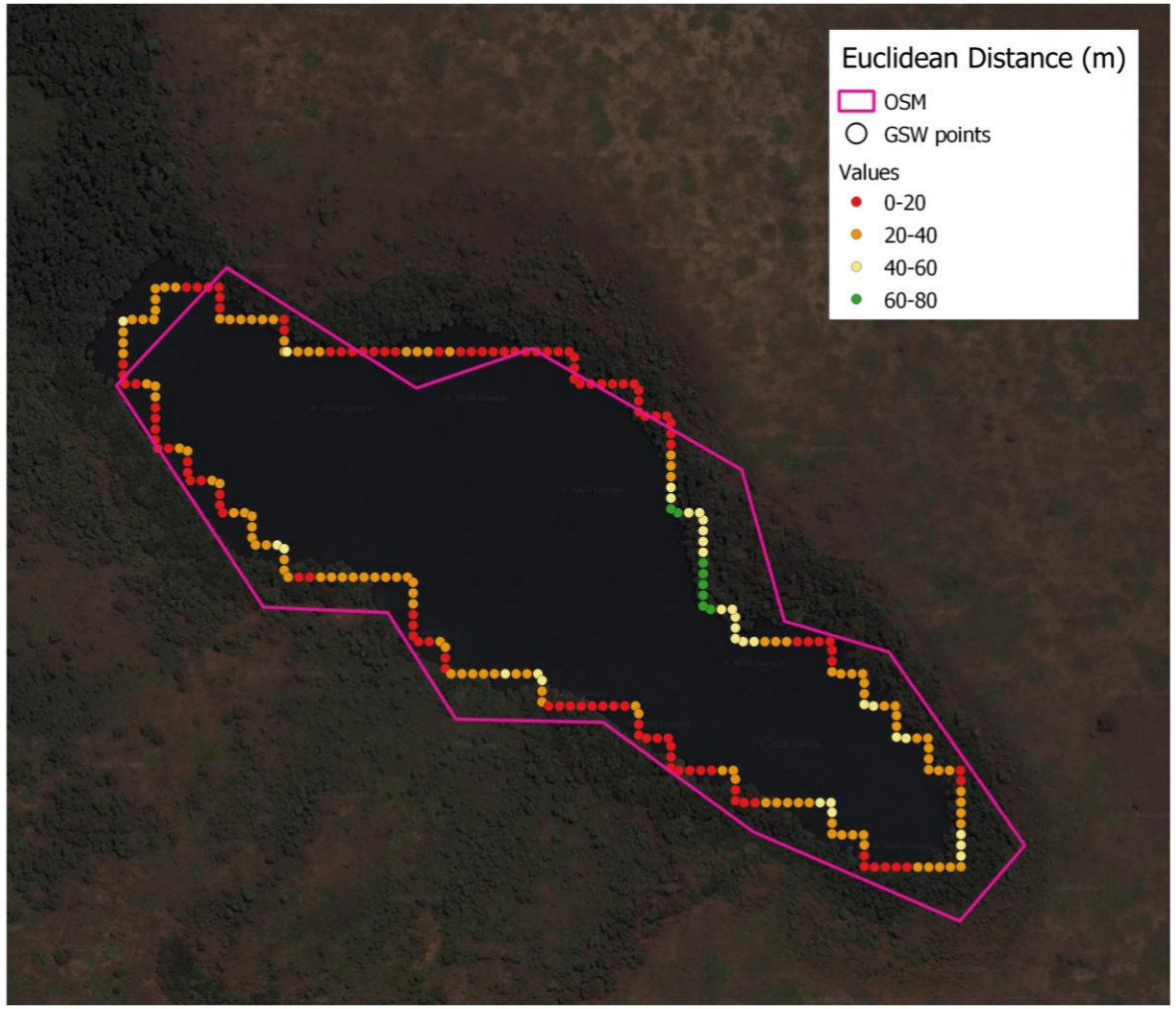




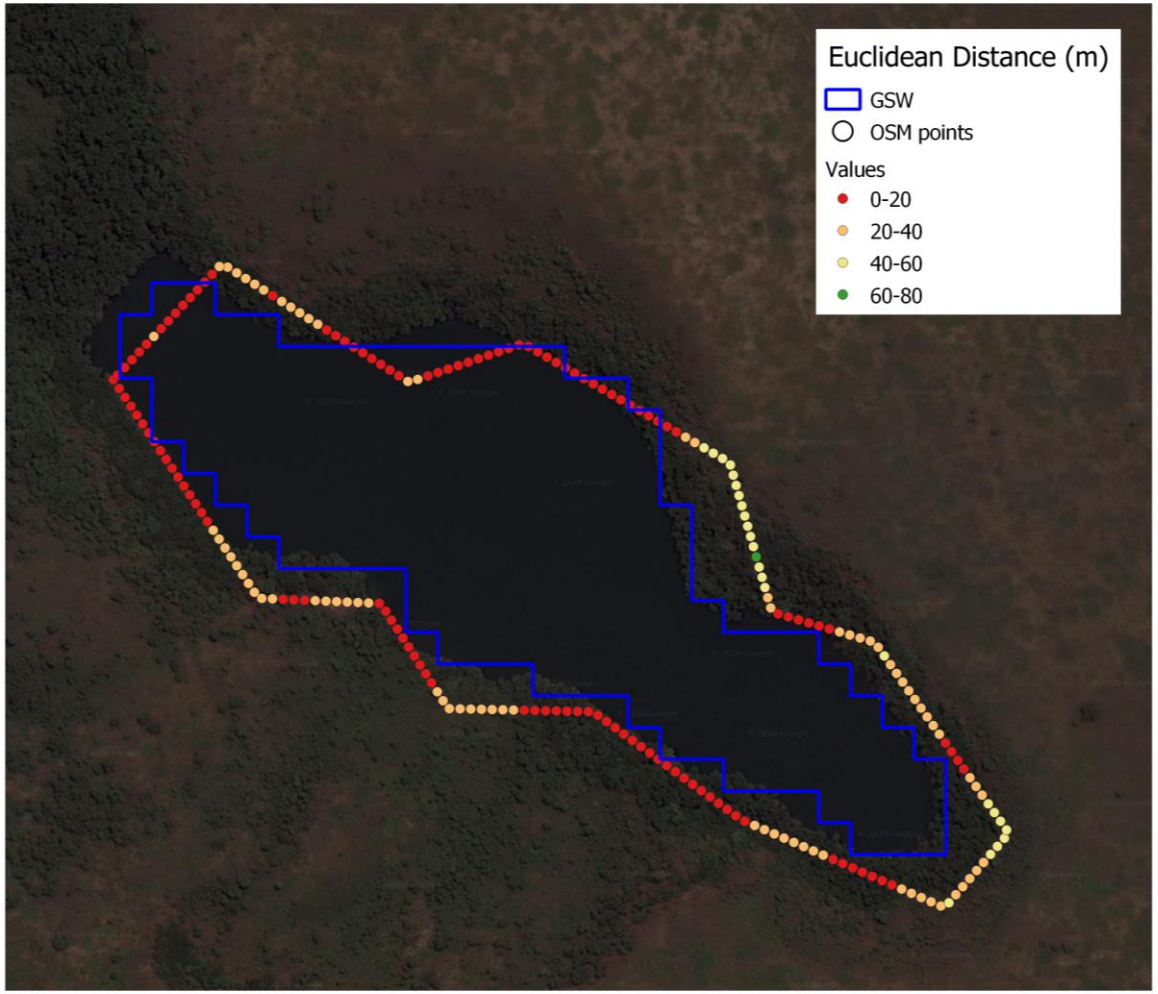
Implementation & Results

Hausdorff Distance

Euclidean Distances of GSW to OSM

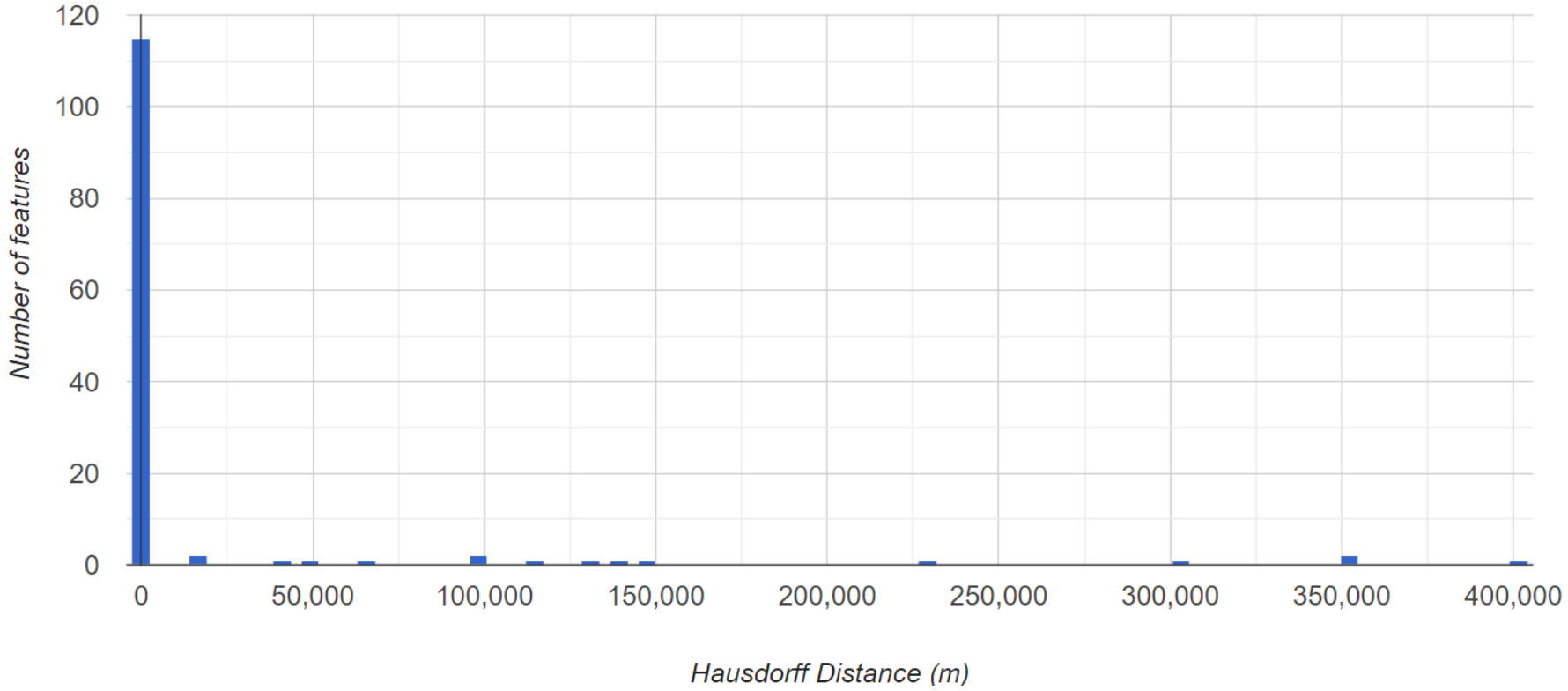


Euclidean Distances of OSM to GSW



Implementation & Results

Hausdorff Distance

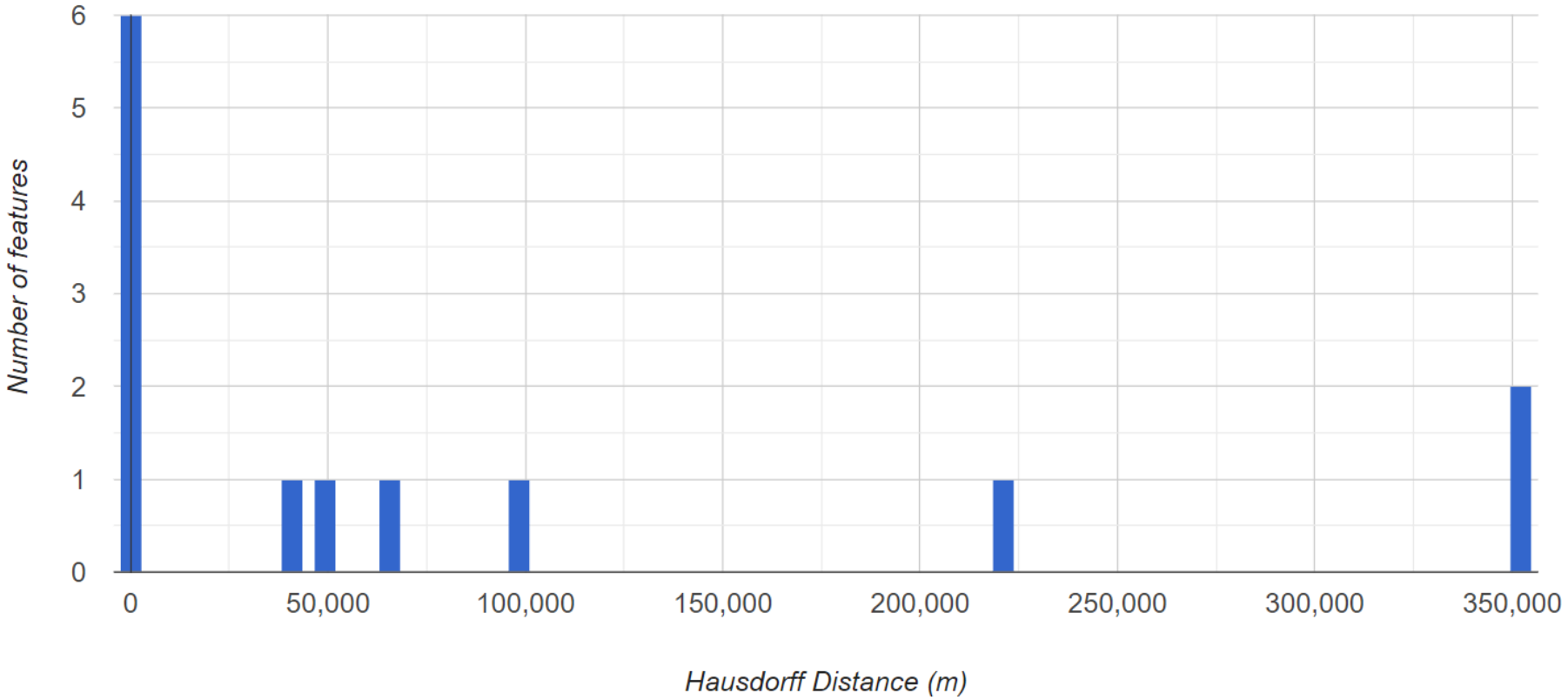


Hausdorff Distance between OSM and HydroLAKES:

Mean: 20017.22 m
Median: 206.19 m

Implementation & Results

Hausdorff Distance

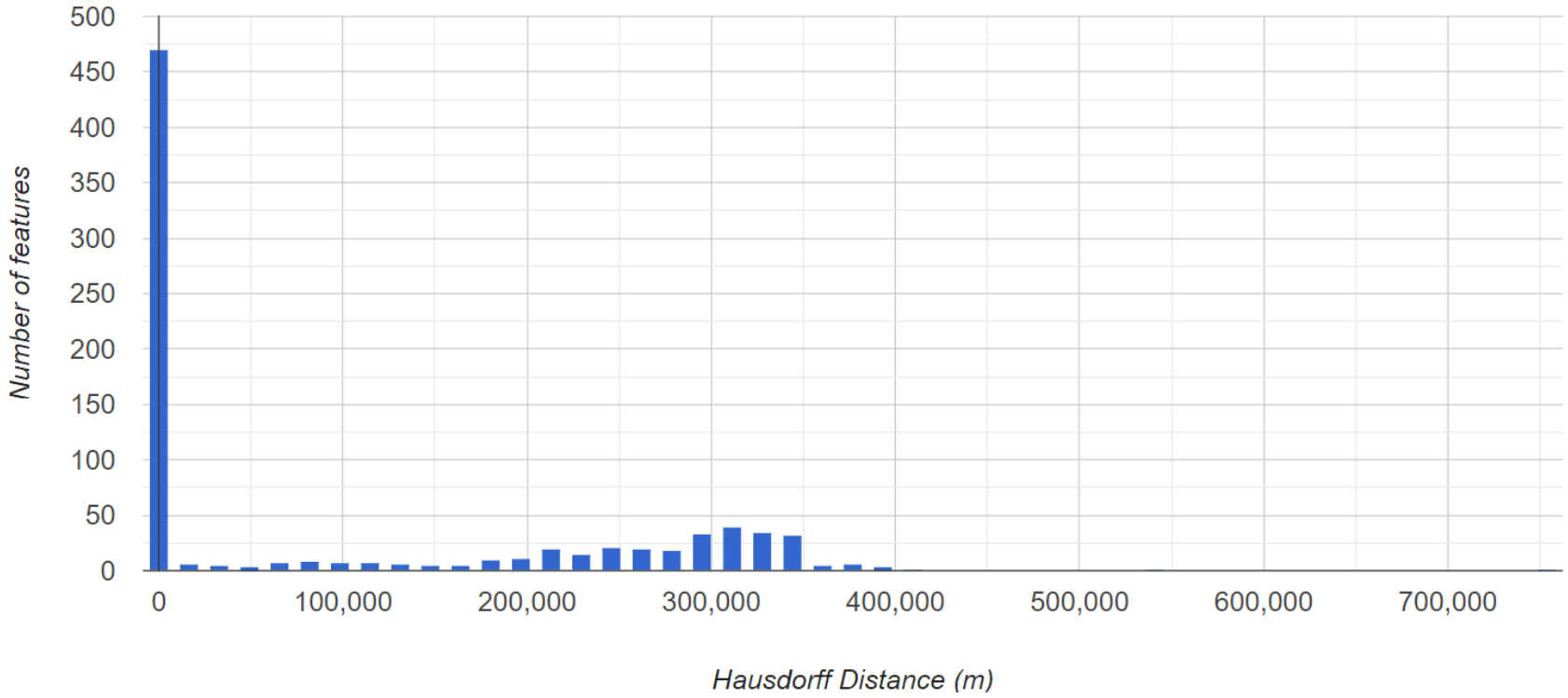


Hausdorff Distance between OSM and GRaND:

Mean: 93129.97 m
Median: 41761.69 m

Implementation & Results

Hausdorff Distance

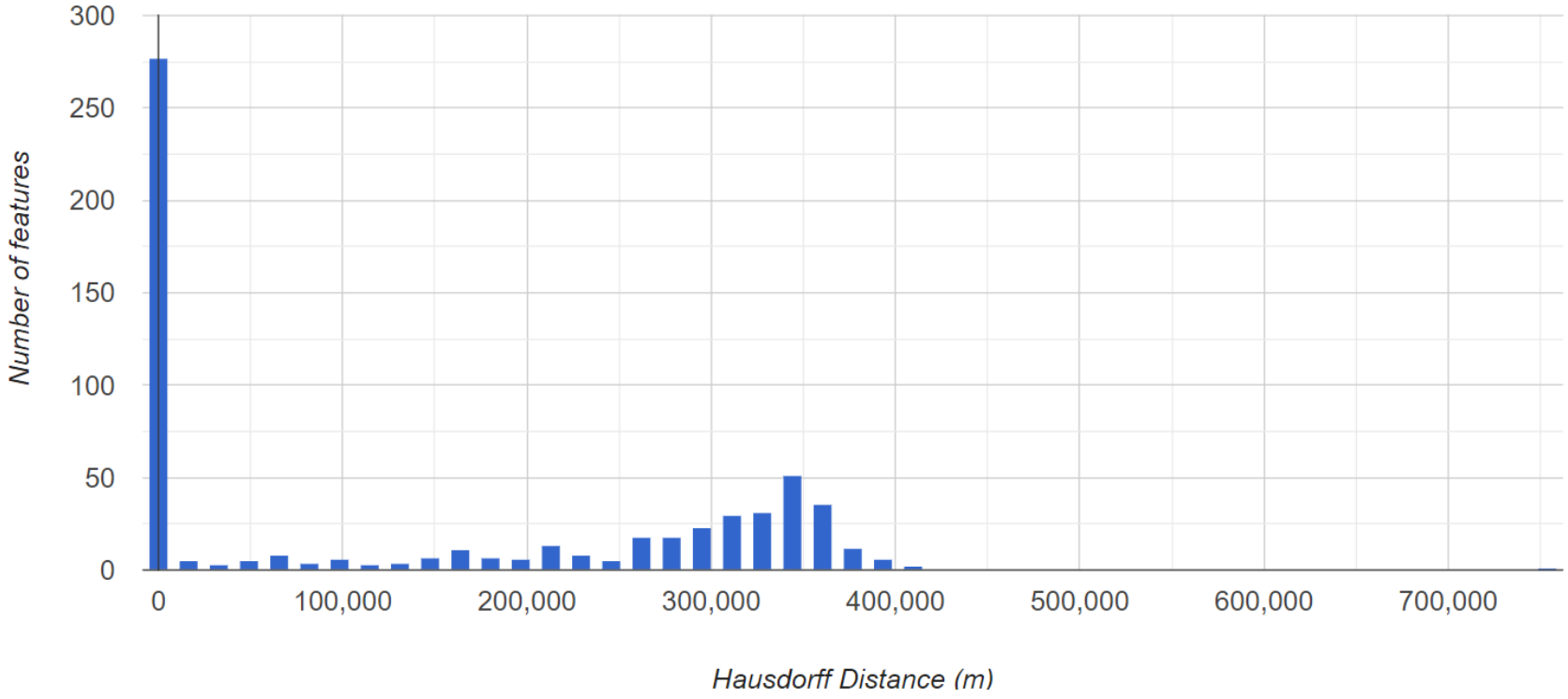


Hausdorff Distance between OSM and GSW :

Mean: 109133.59 m
Median: 236.07 m

Implementation & Results

Hausdorff Distance



Hausdorff Distance between OSM and subset of Sentinel 2 :

Mean: 152535.25 m

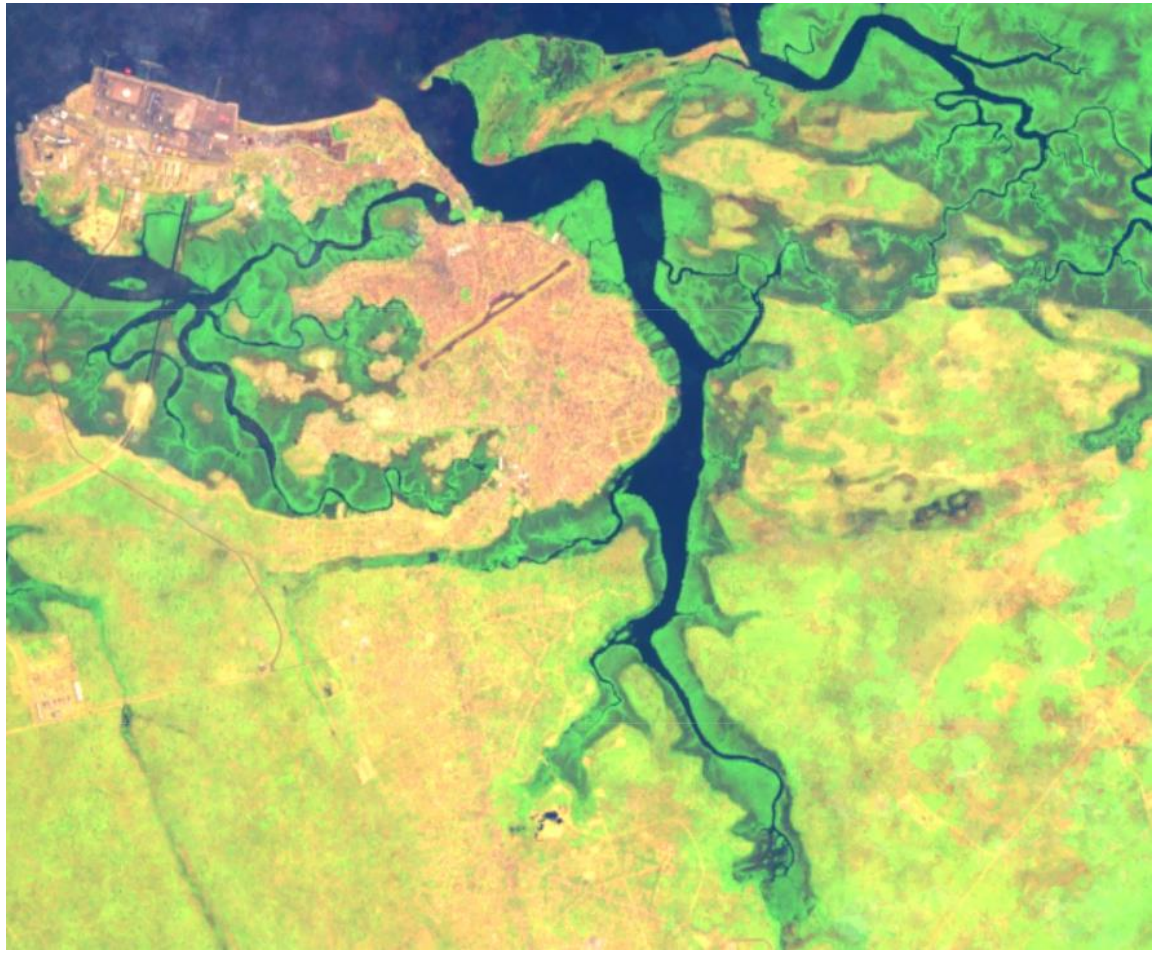
Median: 92301.64 m

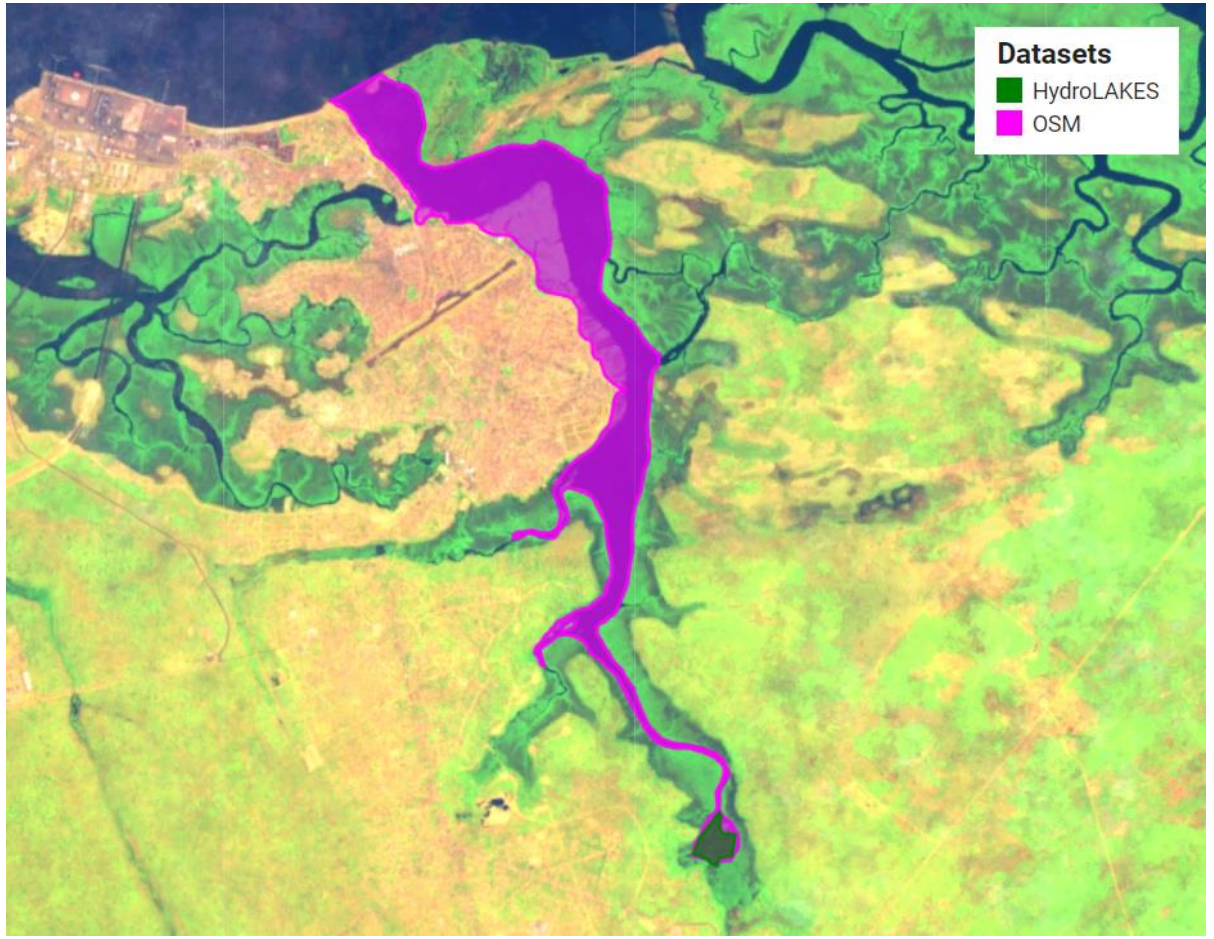
Outline

- Introduction
- Methodology
- Implementation and Results
- Sensitivity Analysis
- Conclusions and Future Work

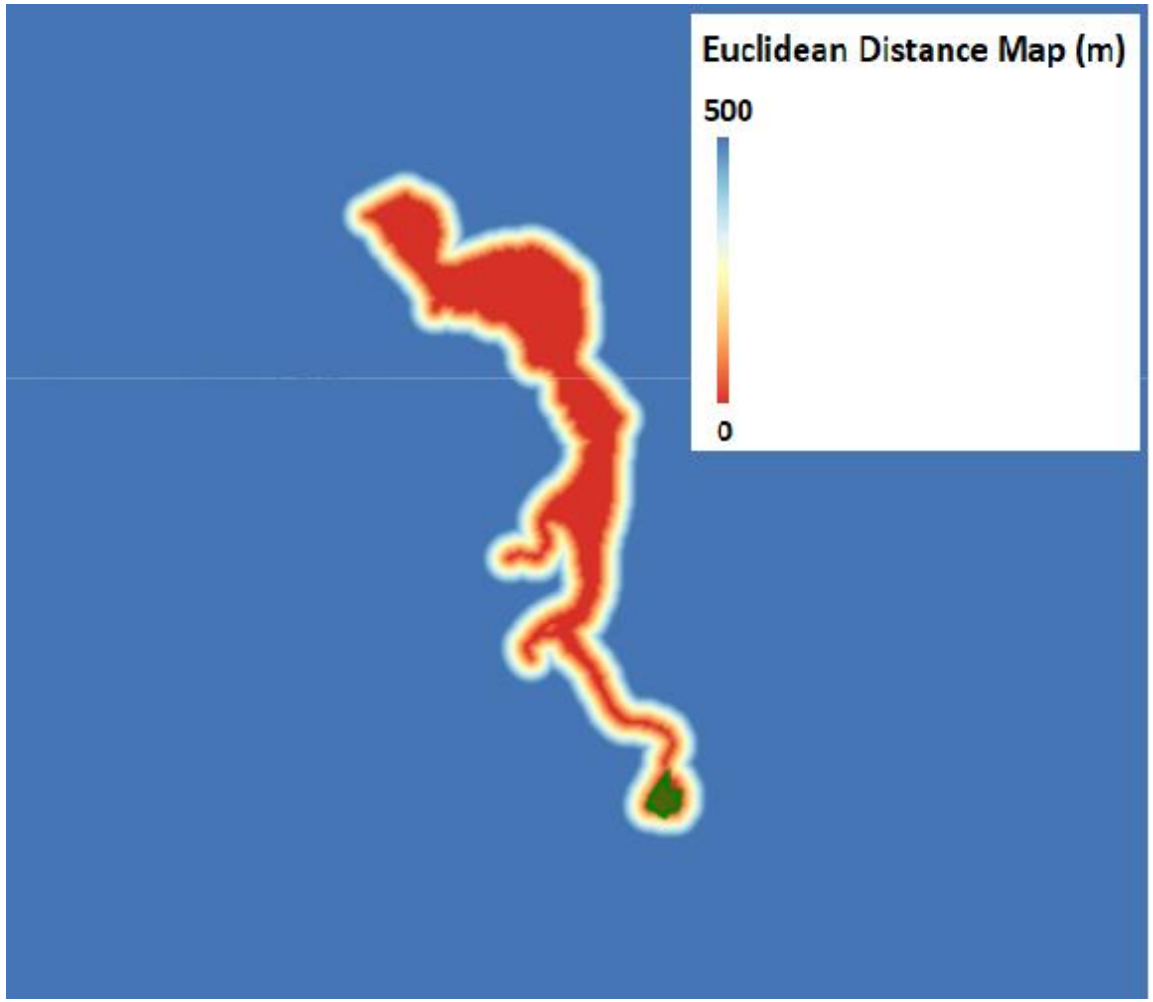
Sensitivity Analysis

Experiments





EDM of OSM feature without (left) and with (right) point sampling



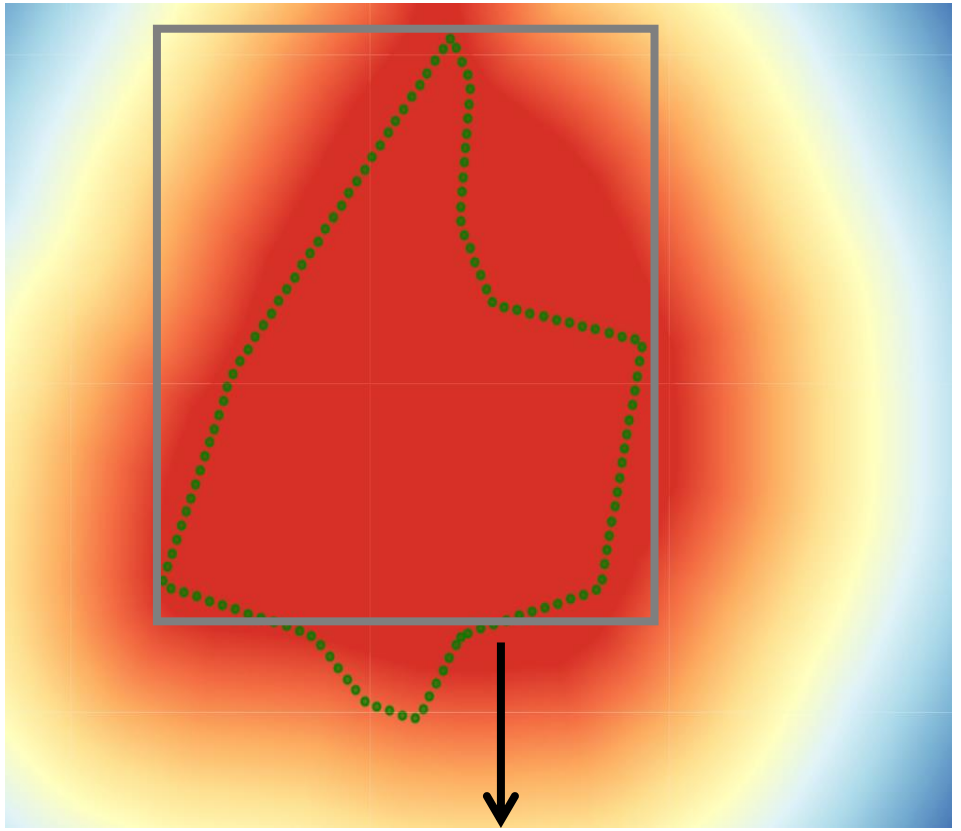
Sensitivity Analysis

Experiments

Euclidean Distances from HydroLAKES to OSM feature

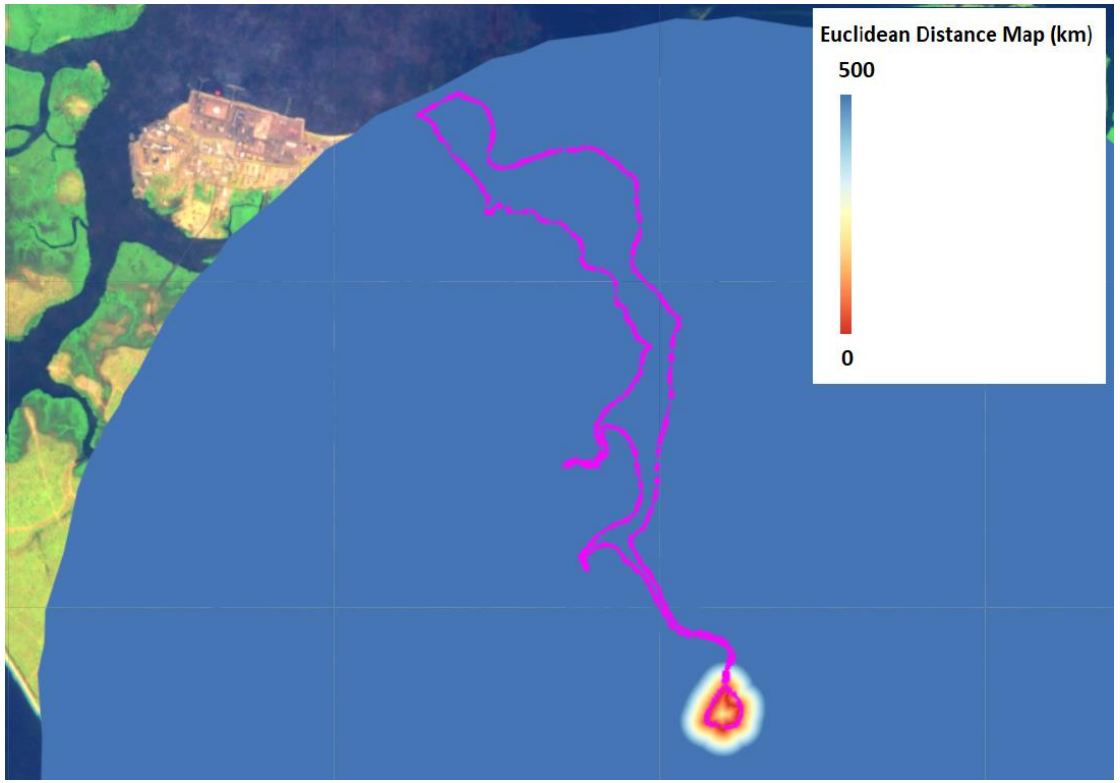
	Euclidean Distance (m)	
Statistical Unit	**Without Sampling**	**With Sampling**
Max	50.67	152.64
Mean	3.57	47.04
Median	0	39.80
Mode	0	43.65
Standard Deviation	11.55	31.80

Without sampling, all HydroLAKES points inside of the OSM red zone appear as zero distance!!



Points with falsely estimated zero distance values

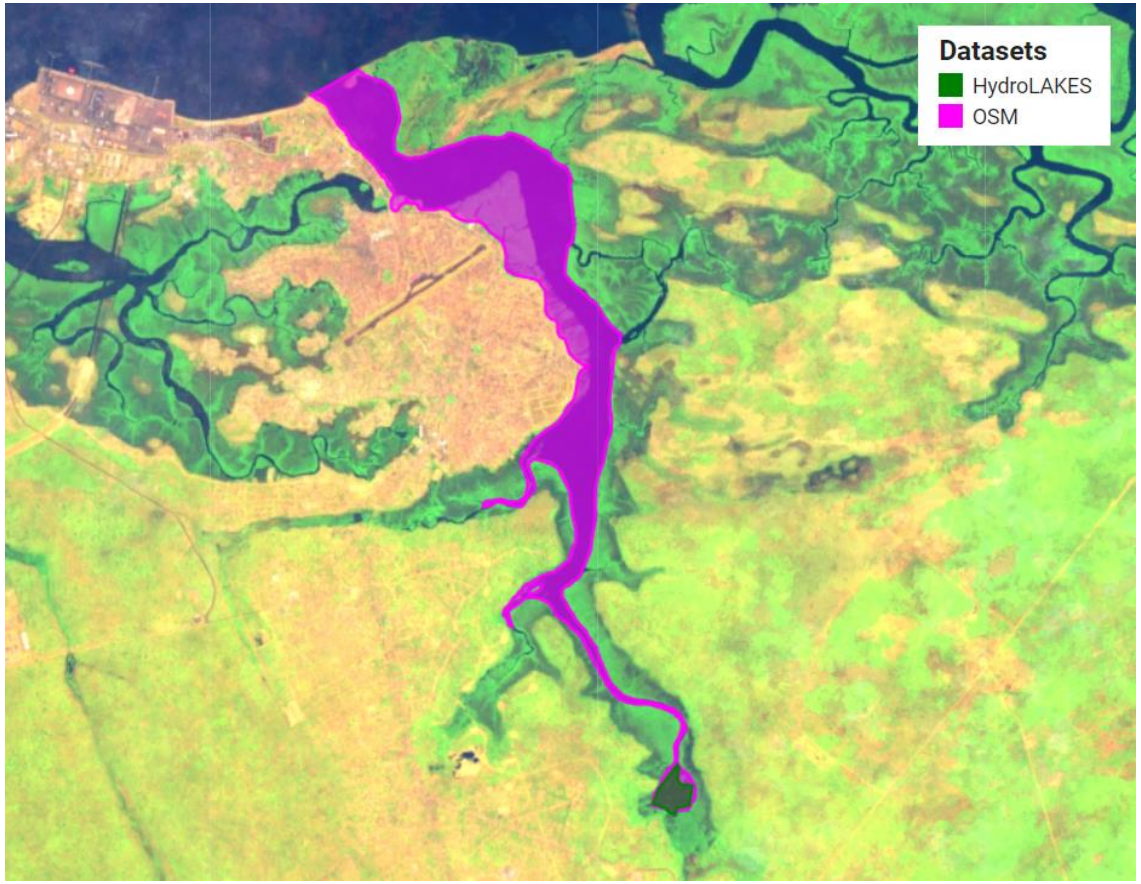
Euclidean Distances from OSM to HydroLAKES feature



Statistical Tool	Euclidean Distance (km)
Max	152.64
Mean	47.04
Median	39.80
Mode	43.65
Standard Deviation	31.80

Sensitivity Analysis

Experiments



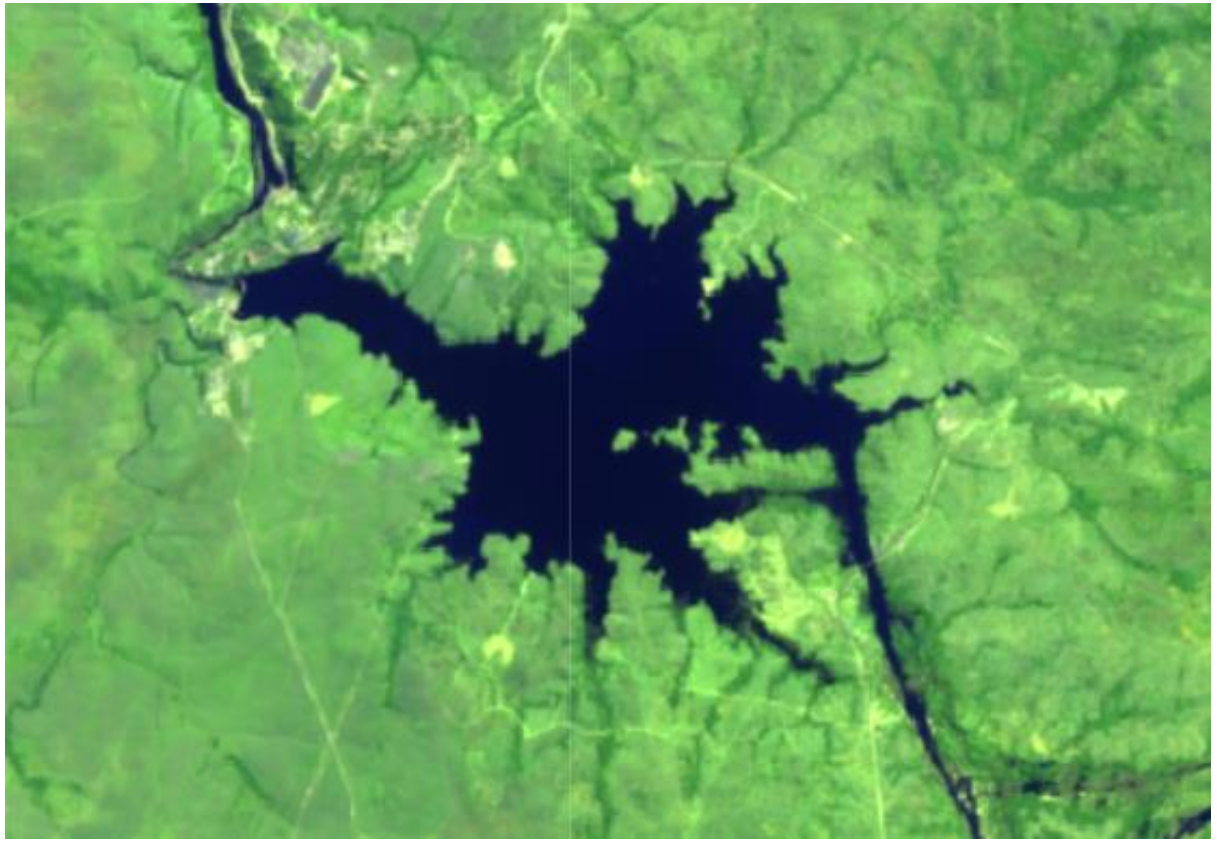
Without Point Sampling:
Hausdorff Distance = 360.06 km

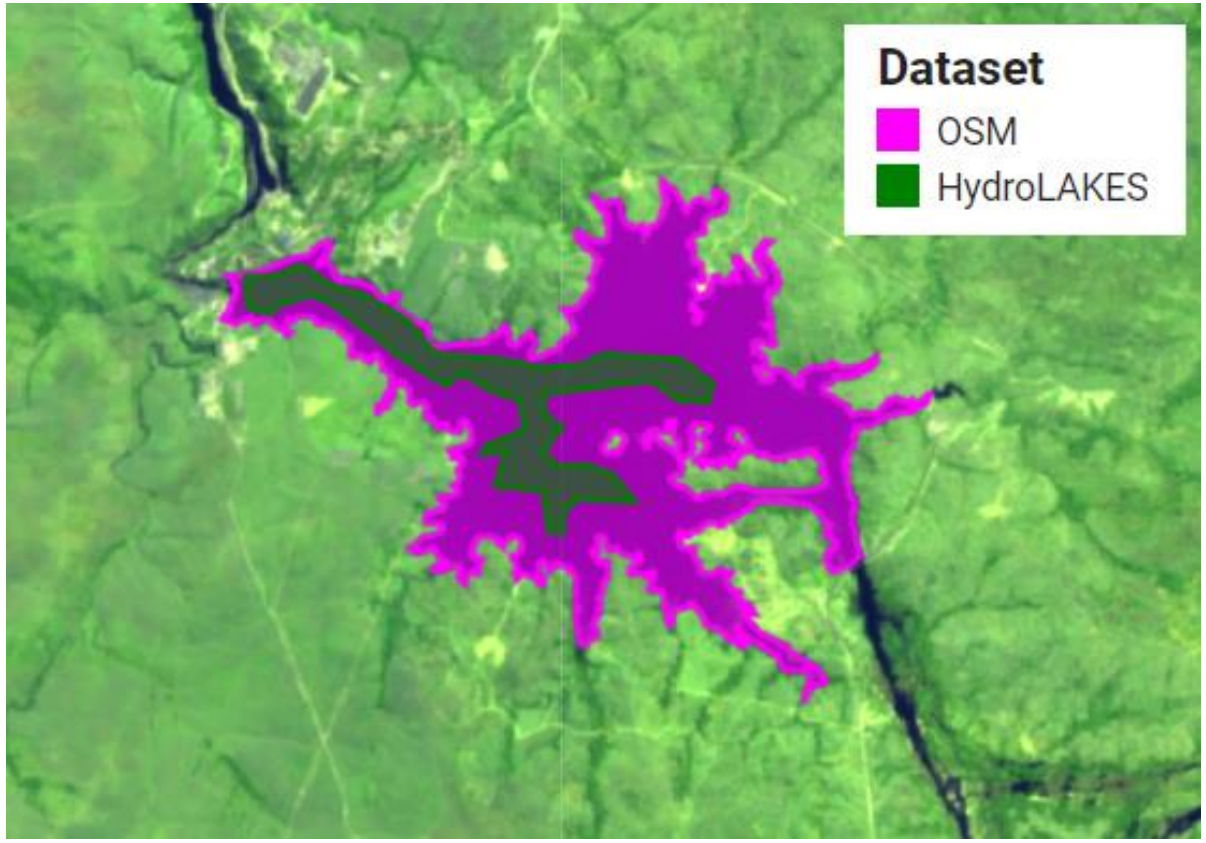
With Point Sampling:
Hausdorff Distance = 360.02 km

Difference : 40 m

Sensitivity Analysis

Experiments

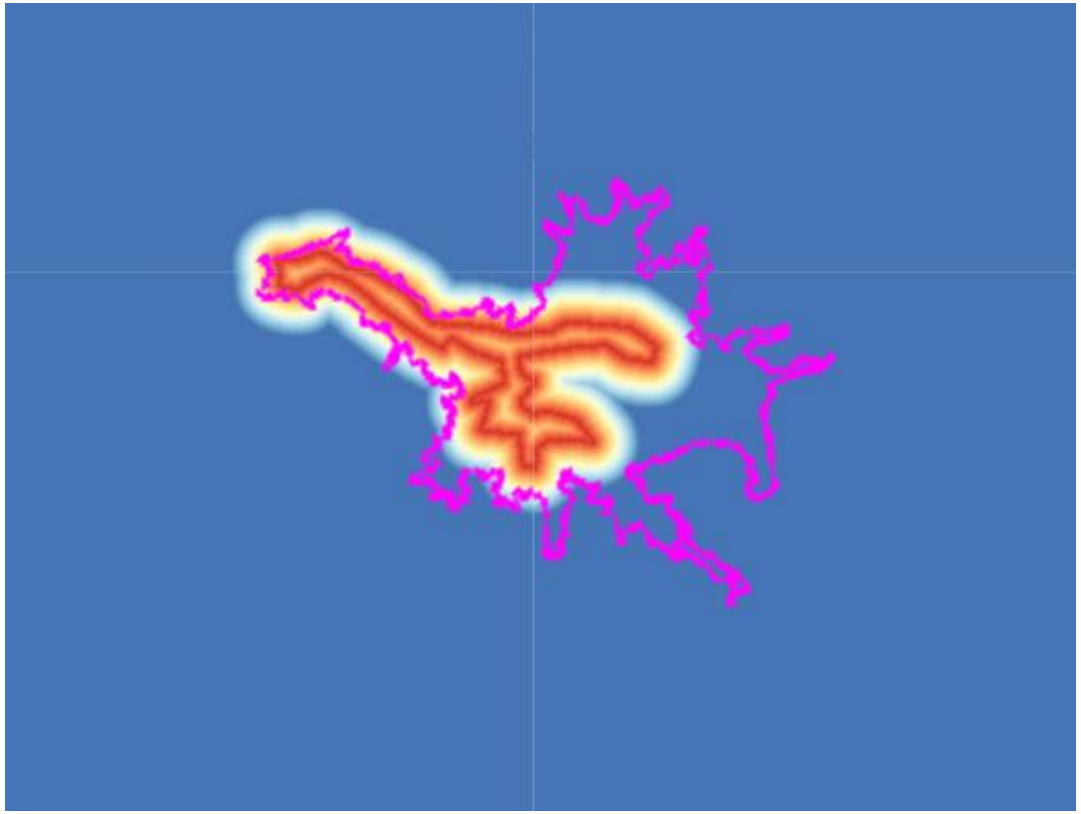




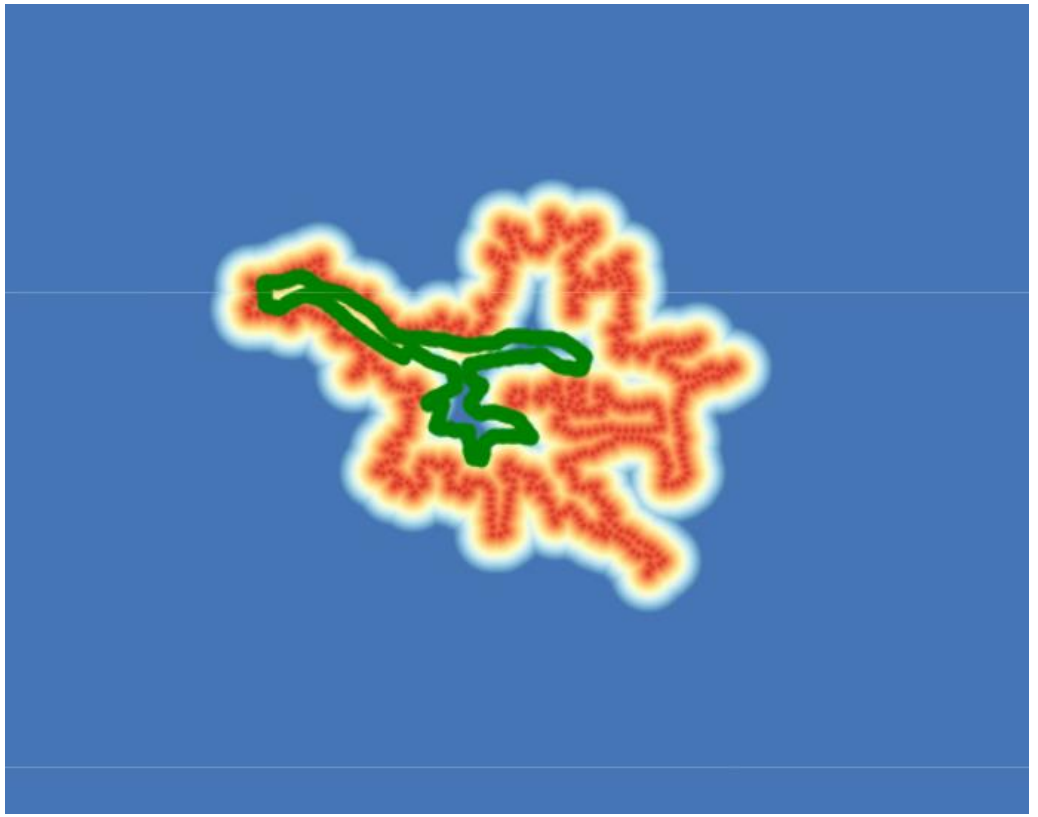
Sensitivity Analysis

Experiments

Euclidean Distances from
OSM to HydroLAKES feature



Euclidean Distances from
HydroLAKES to OSM feature



Sensitivity Analysis

Experiments

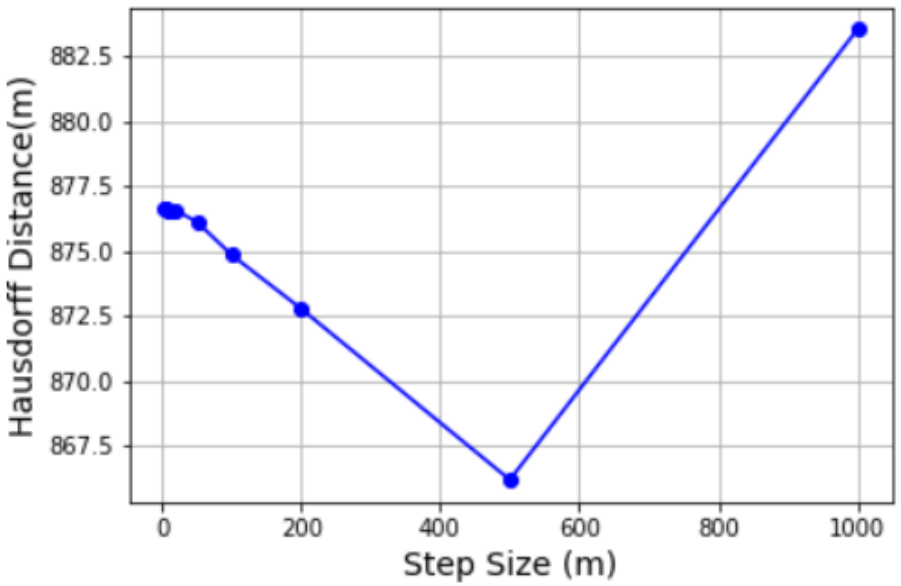
Euclidean Distances between OSM and HydroLAKES feature for various step sizes

	Euclidean Distances (m)	
Step Sizes (m)	**Mean ED (OSM to HydroLAKES)**	**Mean ED (HydroLAKES to OSM)**
3	663.95	212.72
5	663.94	212.69
8	663.86	212.68
10	663.88	212.65
20	663.90	212.65
50	663.73	212.38
100	663.17	211.67
200	661.30	211.47
500	655.64	210.60
1000	657.46	226.05

Sensitivity Analysis

Experiments

Step Sizes (m)	Amount of Points	Hausdorff Distance (m)	Relative Difference (%)
3	14.357	876.67	-
5	8615	876.63	- 0.004
8	5385	876.54	- 0.010
10	4308	876.53	- 0.001
20	2154	876.55	+ 0.002
50	862	876.11	- 0.050
100	432	874.84	+ 0.145
200	217	872.77	- 0.237
500	88	866.20	- 0.753
1000	45	883.51	+ 1.998



Conclusions

Research Question 1:

What are the differences in terms of spatial coverage?

- In terms of common surface water area:

1.	OSM and HydroLAKES	24%
2.	OSM and GRaND	22%
3.	OSM and GSW	9%
4.	OSM and Sentinel 2	19%

Conclusions

Research Question: 2

What are the differences in terms of positional accuracy?

- *Goodchild's Percentge of overlap: Within 10 m distance from OSM:*
  1. *57% of HydroLAKES*
  2. *46 % GRaND*
  3. *15 % GSW*
  4. *1.7 % Sentinel 2*

- *Median Hausdorff Distance:*
  1. *OSM and HydroLAKES: 206.19 m*
  2. *OSM and GRaND: 41761.69 m*
  3. *OSM and GSW : 236.07 m*
  4. *OSM and Sentinel 2: 92301.64 m*

Conclusions

Main Research Question:

What are the spatial differences between Earth Observation based and Volunteered Geographic Information for water reservoirs and how can they be addressed in an automated way at a large scale?

- *Accuracy comparison of the datasets at a large scale is possible by exploiting the planetary-scale analysis capabilities of Google Earth Engine. With an optimization of the Hausdorff Distance algorithm, the computational speed could be increased, making a global scale analysis for all three quality metrics possible.*

Future Work

- Refinement of Sentinel 2 and GSW data, by improving surface water detection methods.
- Exclusion of river information from Sentinel 2 and GSW datasets.
- Merging of Sentinel 2 and GSW clusters of smaller water bodies, into single homologous water reservoirs.
- Classification into good and bad reservoir matches based on quality metrics. Creation of data fusion algorithm that combines the strengths of all water reservoir datasets to create one with better overall quality.
- Classification of water reservoirs into types of water features (e.g. lakes, agricultural water reservoirs, valley-dammed reservoirs etc.)

Thank you for your attention!

References:

1. https://www.mygrandforksnow.com/17177/columbia-river-treaty-community-discussions-highlighted-in-2019-report
2. Planetary-scale geospatial analysis with Google Earth engine, Gennadii Donchyts & Josh Friedman, Deltares, https://www.slideshare.net/Delft Software Days/dsd-int-2015-planetaryscale-geospatialanalysis-with-google-earth-engine-gennadii-donchyts-amp-josh-friedman-deltares
3. https://www.btlliners.com/can-irrigation-reservoirs-save-our-farms
4. http://floodlist.com/australia/insurers-can-get-better-responding-natural-disasters
5. https://spacenews.com/abs-orders-abs-2-satellite-space-systemsloral/
6. Allan Cain, Afonso Cupi Baptista, Community Management and the Demand for ‘Water for All’ in Angola’s Musseques.