



08.08.24

Flooded Area Assessment using KMeans clustering technique

By Penny (Qian) Yin



Story

Context

- Flooding -> damage, such as infrastructure destruction and water system pollution;
- The New England Region - challenges from coastal and inland flooding;
- Crucial to quantify the impacts of flood events.

Problem

- Precisely assessing the extent of flood can be challenging.

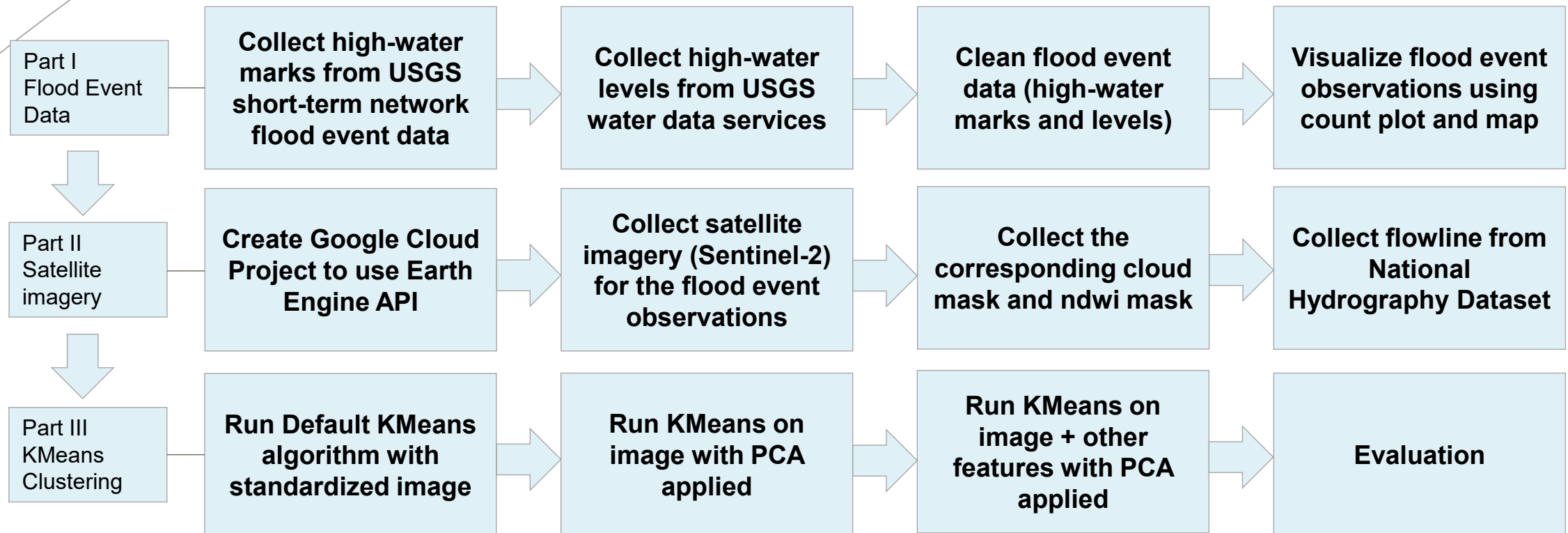
Proposal

- Satellite images – straightforward
- Develop an approach to help automatically identify the flooded area



* Image from [The Great Vermont Flood of July 2023](#)

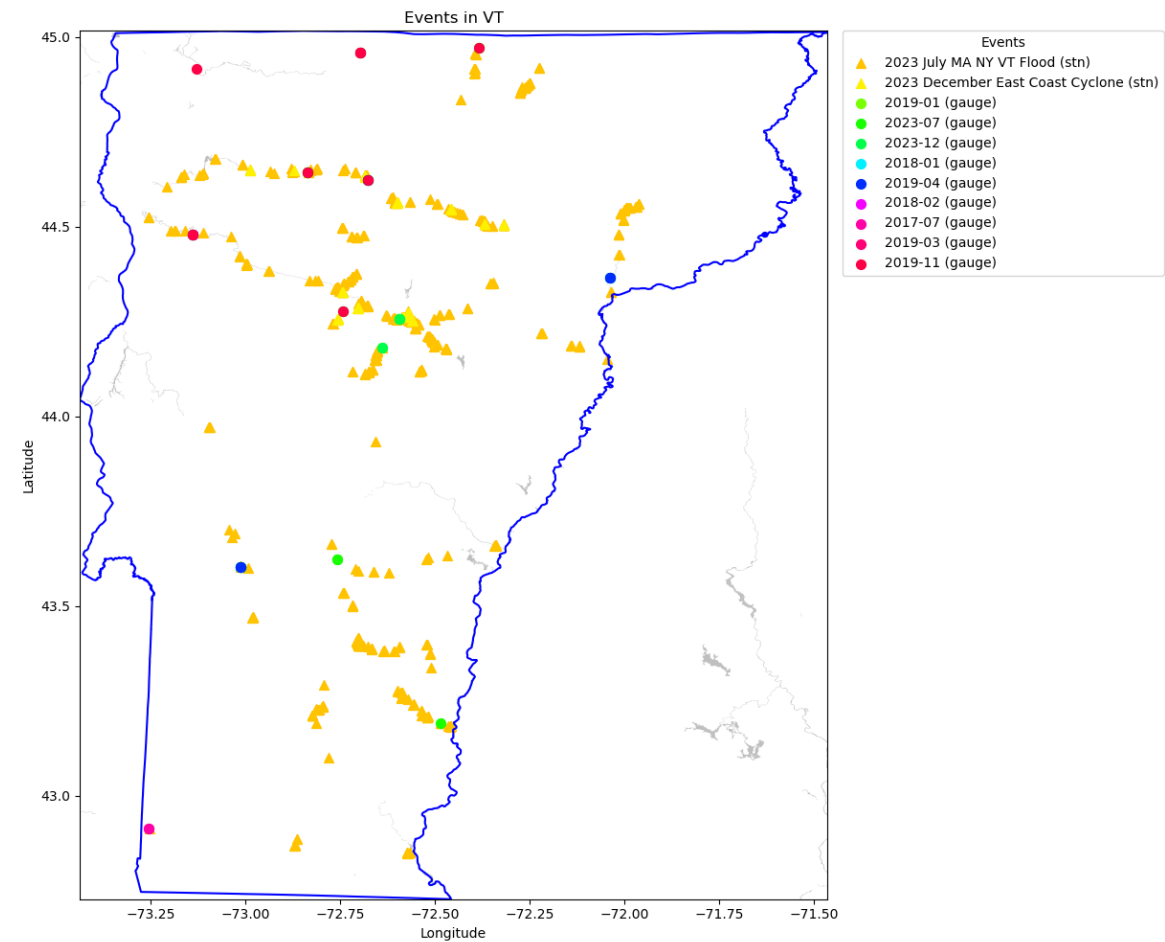
Workflow



- USGS: United States Geological Survey
- [High-water marks](#): Visual clues of peak stream height or surveying equipment
- High-water level: Real-time water level above [moderate flood stage](#) value for specified gauges from USGS Water Data Services (e.g., <https://water.noaa.gov/gauges/ASTM1>)

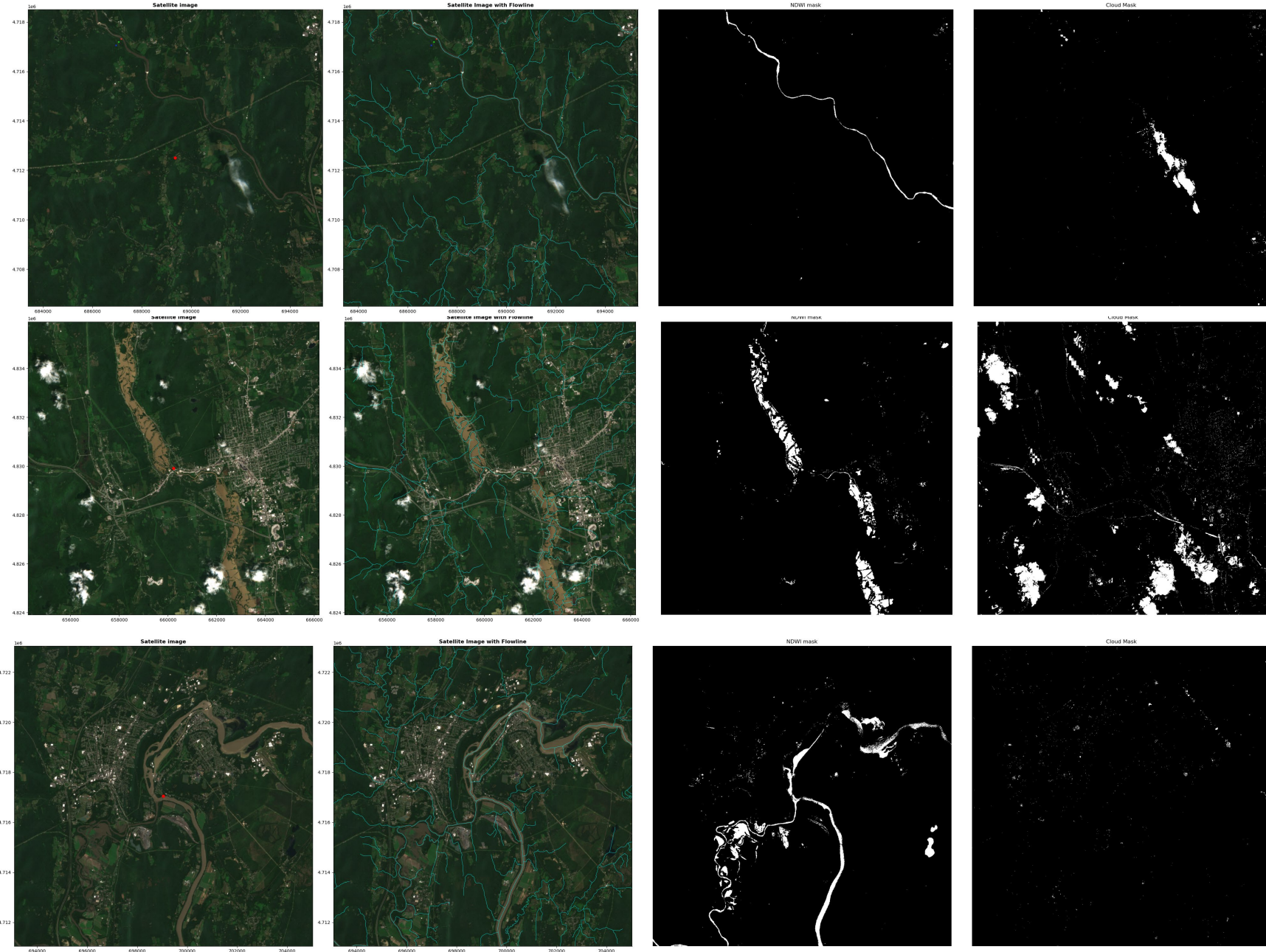
Part I Flood Event

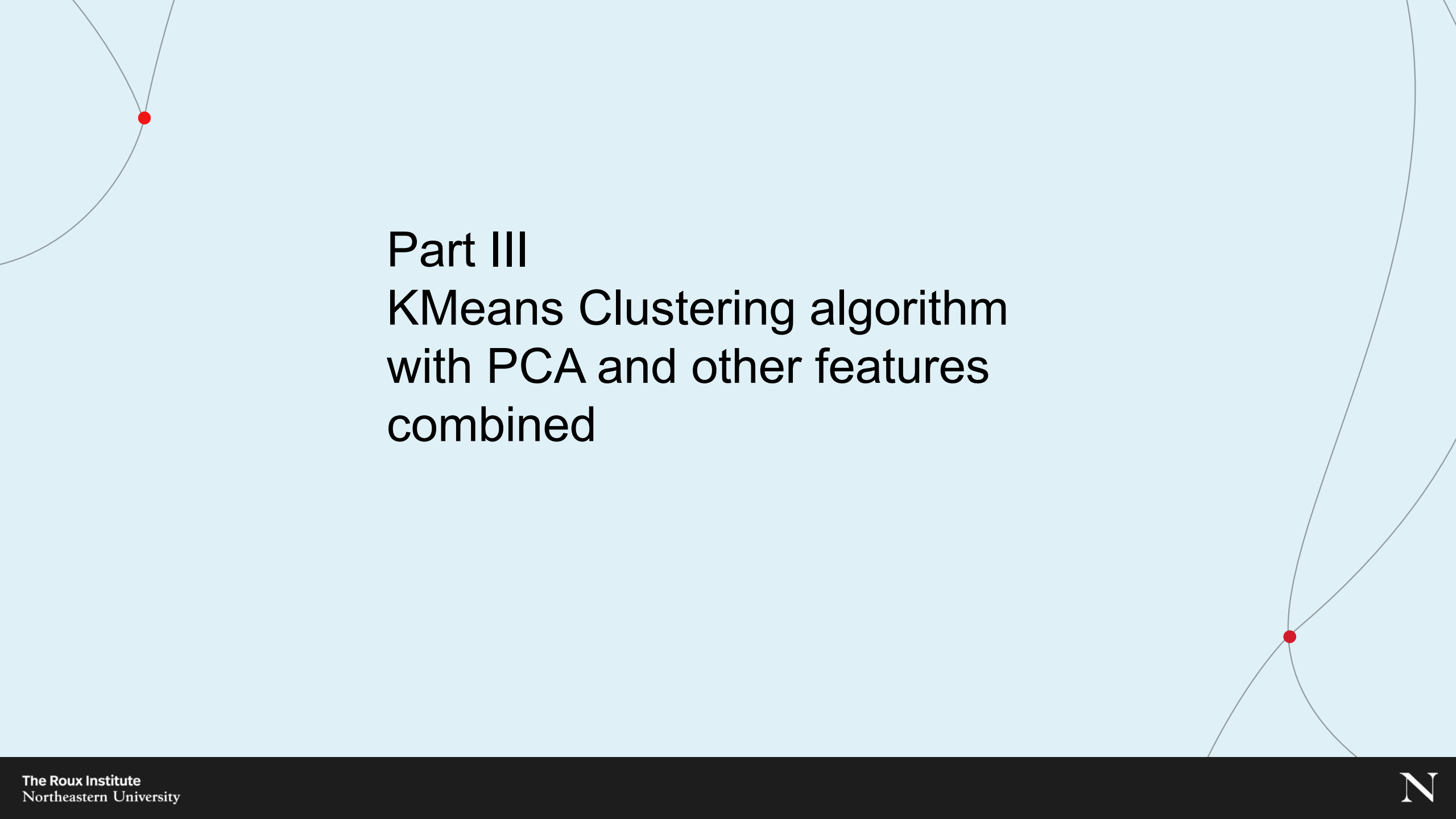
	High-water marks	High-water levels
Original Shape	53 x 3502	11 x 8326
Modified Shape	7 x 889	9 x 218
Top 3 Flood events	<ul style="list-style-type: none">2023 July MA NY VT Flood : 641 instances2018 March Extratropical Cyclone : 115 instances2018 January Extratropical Cyclone : 81instance	<ul style="list-style-type: none">2023-12 : 642023-07 : 272024-01 : 18
Top 3 States	<ul style="list-style-type: none">VT : 590MA : 262CT : 15	<ul style="list-style-type: none">CT : 57VT : 56ME :40
Note	Based on the counts, the majority of collected images will be from 2023-07.	



Part II Satellite Image 2023-07

1. [Sentinel -2](#) satellite image:
25 images during flood
events (8 notable)
2. Flowline from [National
Hydrography Dataset](#)
3. Normalized Difference
Water Index (NDWI):
threshold set to -0.1
4. Cloud Mask based on
[s2cloudless](#): urban area
and high reflectance





Part III

KMeans Clustering algorithm with PCA and other features combined

Default

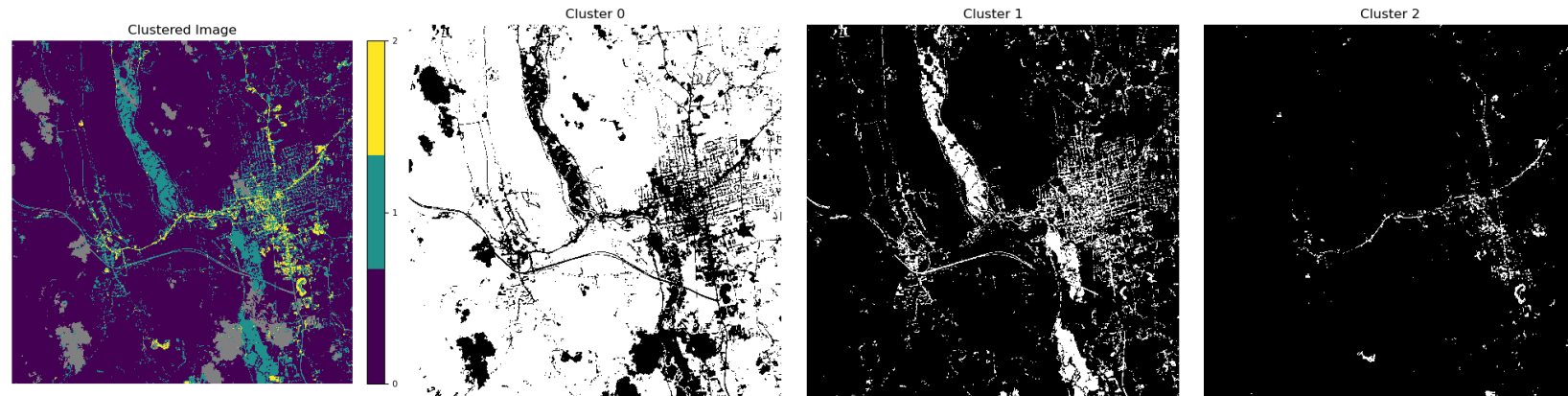


Image -> PCA

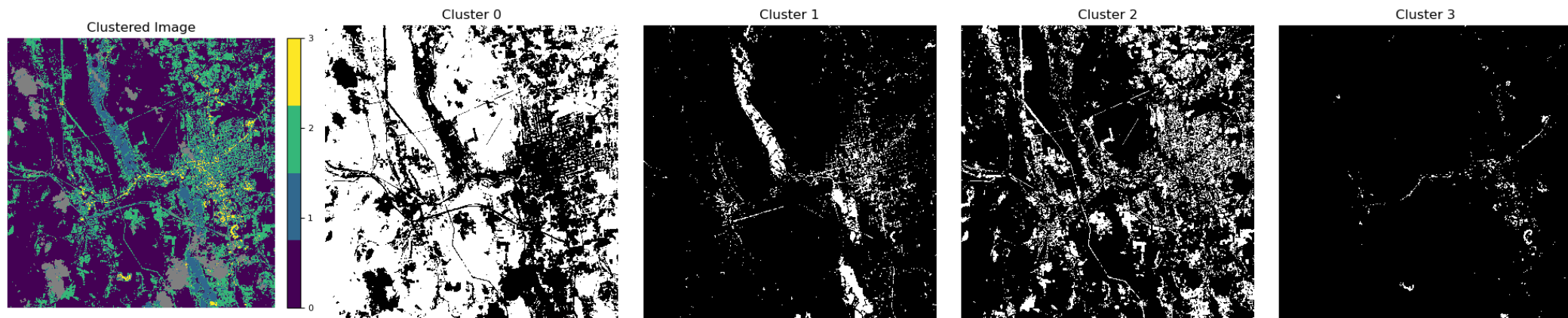
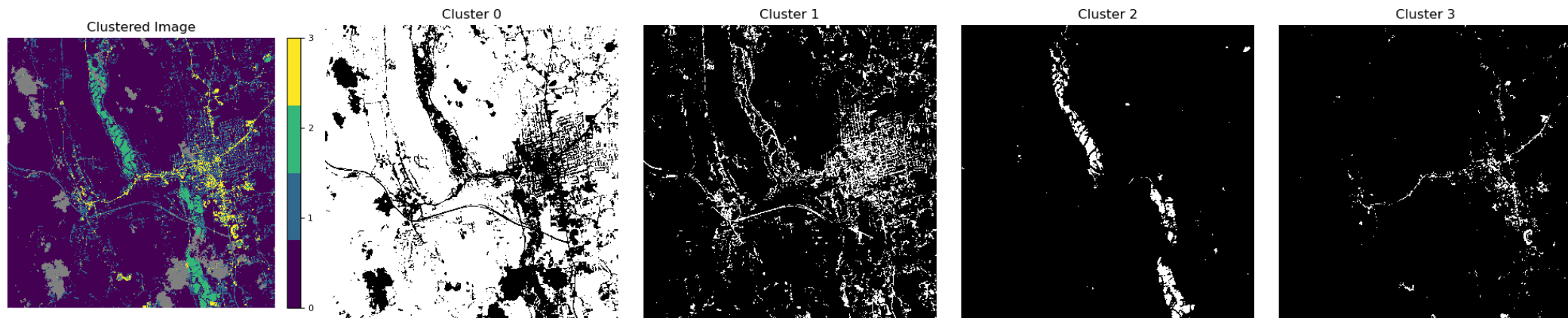
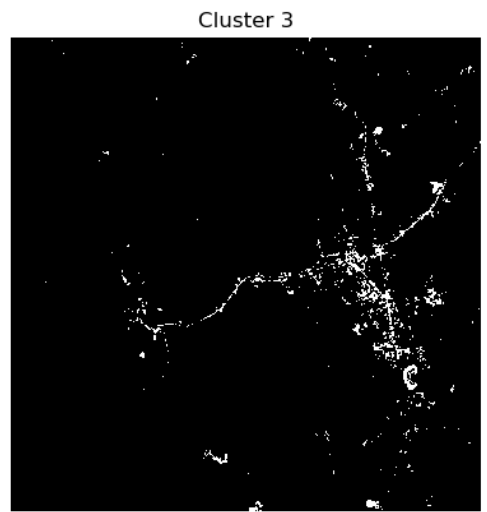
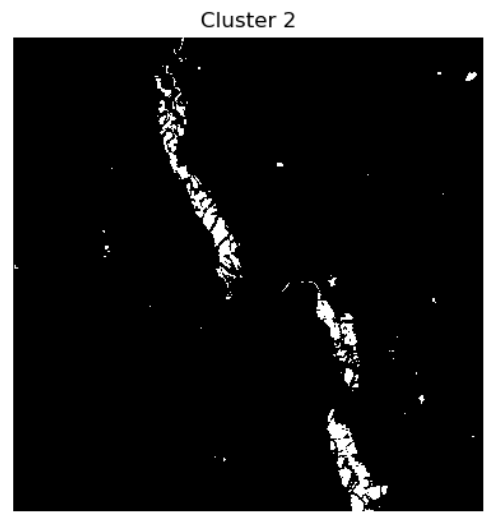
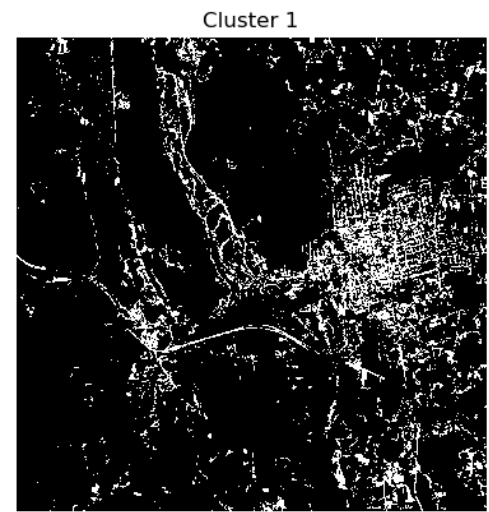
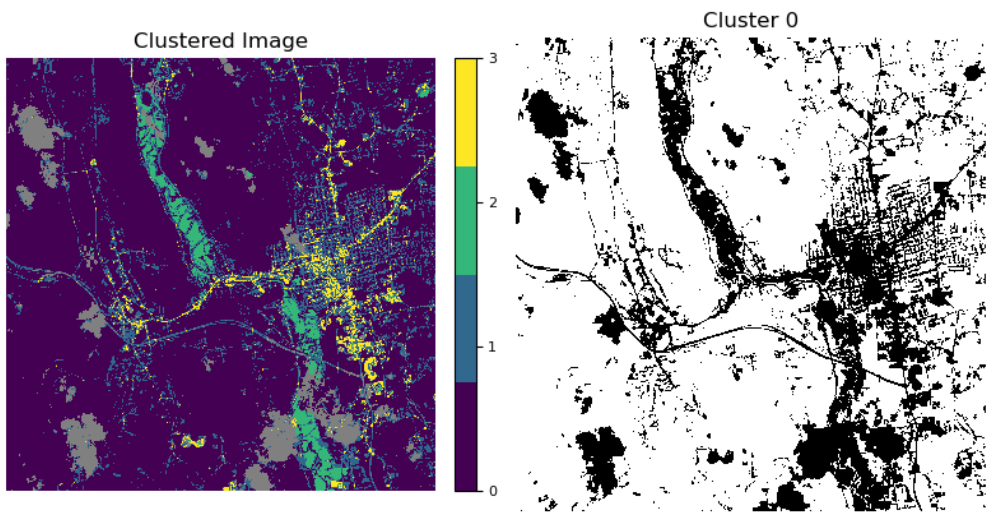
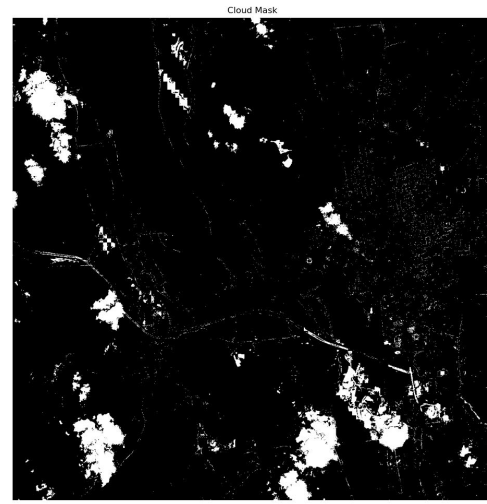
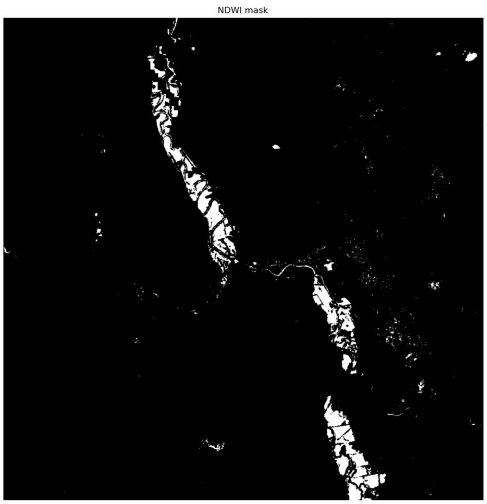
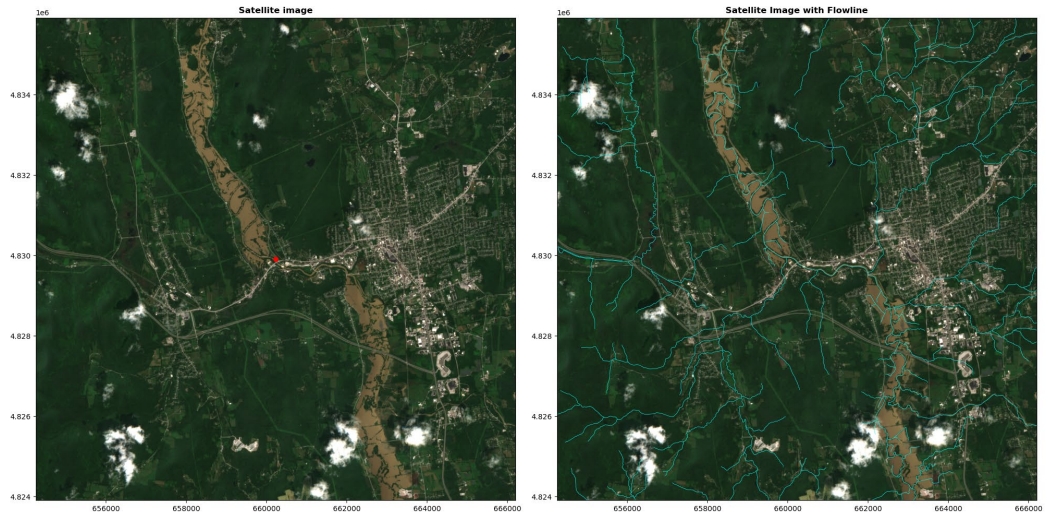


Image with
NDWI feature ->
PCA



SUMMARY - Optimal KMeans Clustering Result

- Reduced noise





THANK YOU!

Discussion

- Satellite imagery can be a straightforward way to check the flooding situation in a specific area; however; not all flood events can be caught.
- Combining the satellite imagery and NDWI (Normalized Difference Water Index) helps the clustering algorithm.

Future work

- Flowline data from NHD includes all flowing water features -> noise.
- At the current stage, the evaluation part focuses on visual inspection -> quantitative evaluation.
- Ensure the reproducibility.
- Algorithm improvement: K-means clustering process is being killed (the process is running out of memory)

Acknowledgements

- Prof. Bogden (Supervisor)
- Sam from USGS (Stakeholder)