Spatiotemporal Machine Learning: A Couple of Examples in Hydrology

Zhe Jiang (zhe.jiang@ufl.edu)

Dept. of Computer & Info. Sci. & Eng.

Center of Coastal Solutions

University of Florida

www.jiangteam.org

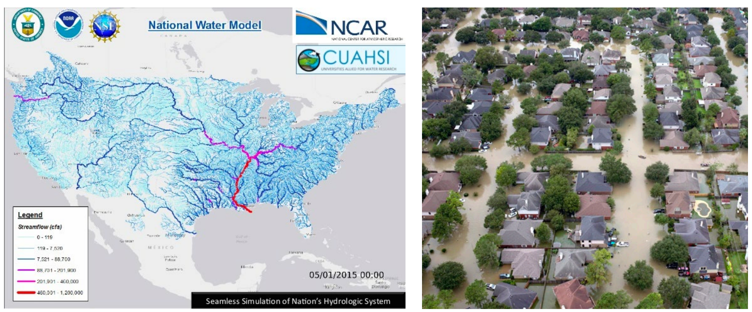
UNIVERSITY of FLORIDA

Outline

- Overview of Spatiotemporal Machine Learning
- Example 1: Terrain-aware flood mapping
- Example 2: National Hydrography Dataset refinement
- Summary

Societal Needs

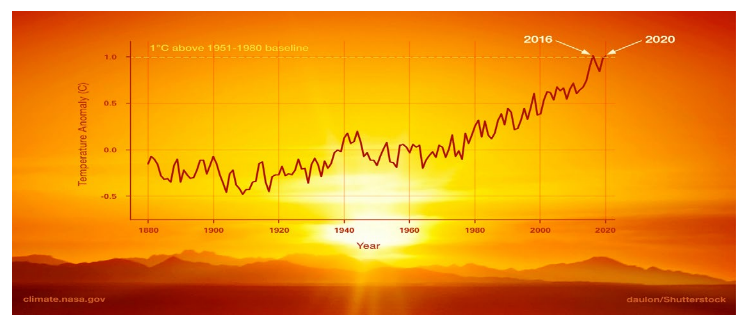
- The revolution of Artificial Intelligence and Machine Learning
  - Computer vision, natural language processing, games
  - Driven by big data, computational hardware, models/algorithms
- Can AI achieve the same level of success in Geo-domains?



National water resource management
(Source: NOAA, NBC news)

Agriculture and food security
(Source: USDA, foodbusinessnews)

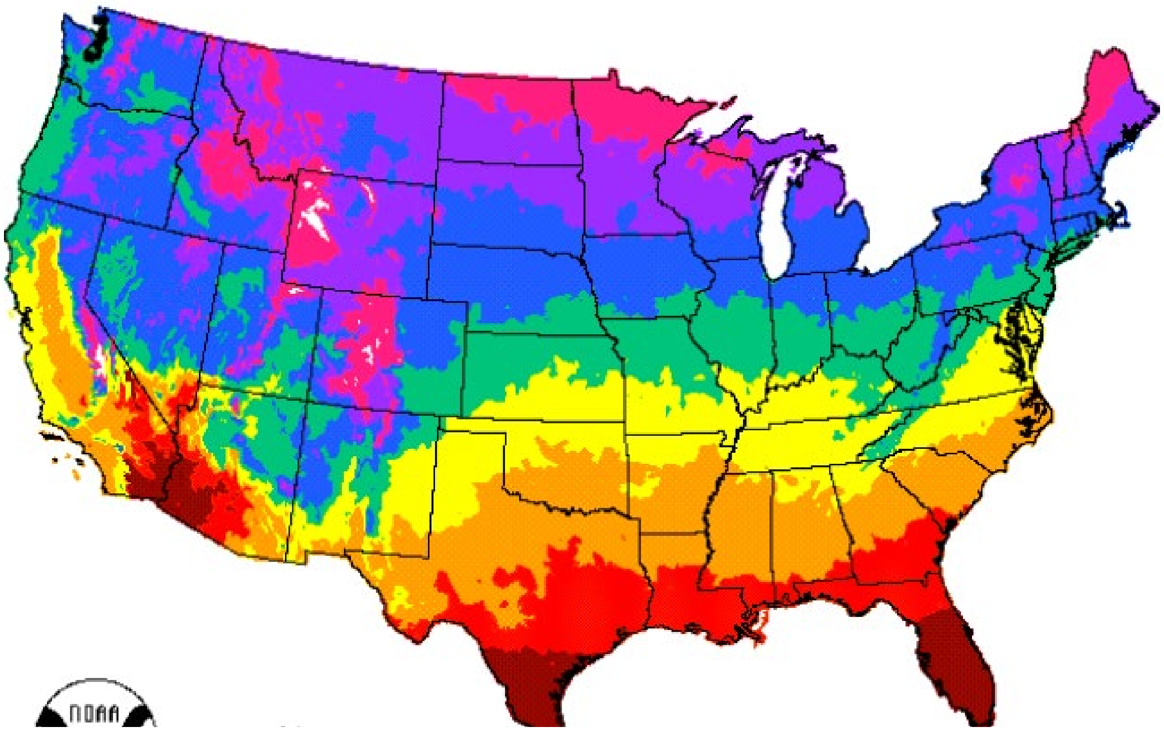
Coastal hazards (storm surge, algal bloom)
(Source: apnews, NOAA)



Mitigation and adaptation to climate change
(Source: NASA)

What are The Unique Challenges of Spatiotemporal ML?

- Spatio-temporal auto-correlation, teleconnections, heterogeneity

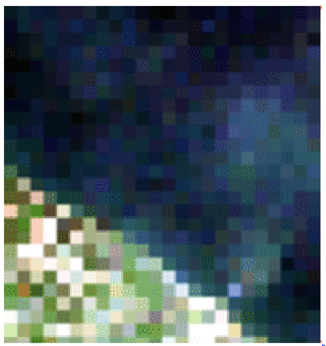


Temperature in the U.S. (NOAA)

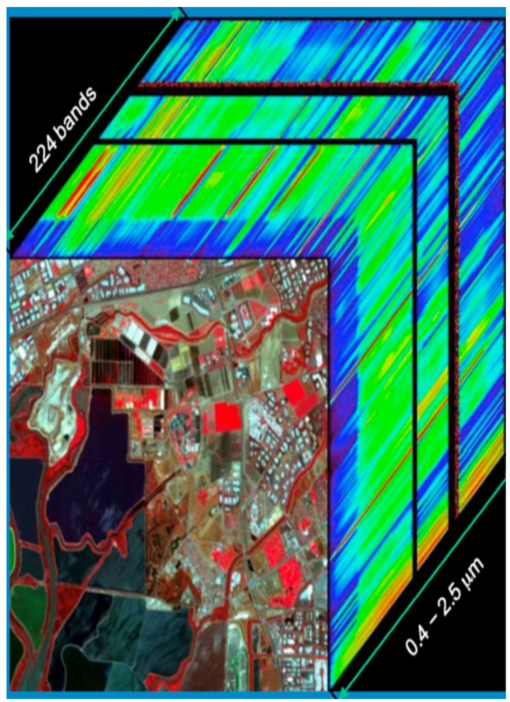
What are The Unique Challenges of Spatiotemporal ML?

- Spatio-temporal auto-correlation, teleconnections, heterogeneity
- Multiple spatial, temporal, spectral resolutions
- Diverse noise, missing data and gaps

MODIS

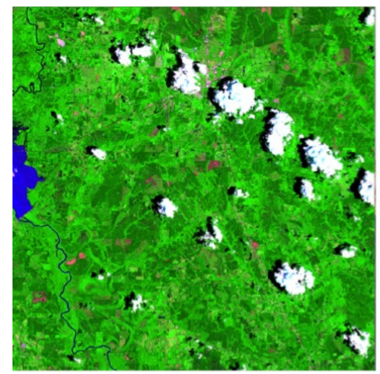


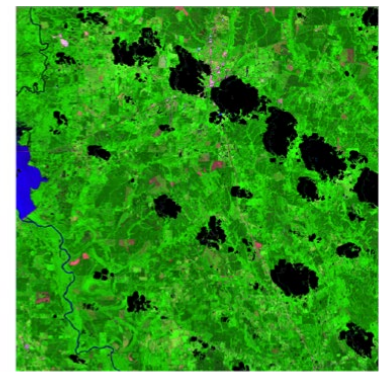
Landsat 8



Hyperspectral imagery

Landsat with cloud mask





Noisy radar image

Gap in Landsat 7

(Image source: USGS, ESA, routledge.com, Concha et al., Shermeyer et al.)

What are The Unique Challenges of Spatiotemporal ML?

- Spatio-temporal auto-correlation, teleconnections, heterogeneity
- Multiple spatial, temporal, spectral resolutions
- Diverse noise, missing data and gaps
- Domain physics and constraints (e.g., terrain, topography)

(Image source: amerikaplus.nl)

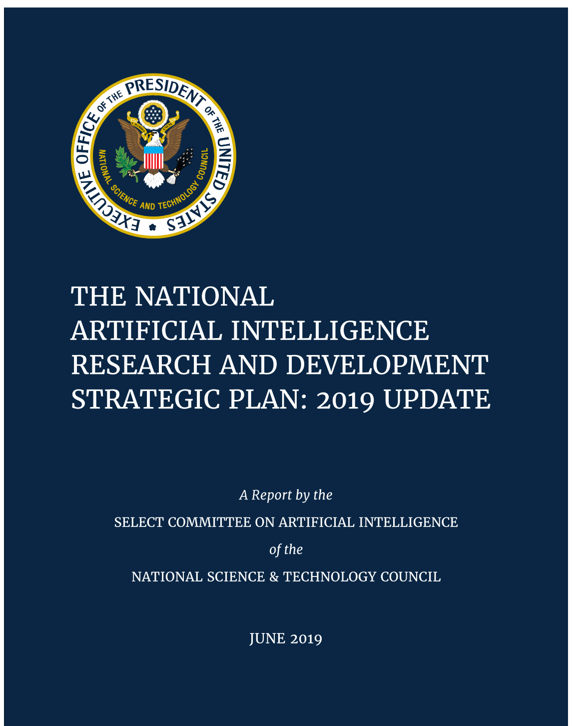
What are The Unique Challenges of Spatiotemporal ML?

- Spatio-temporal auto-correlation, teleconnections, heterogeneity
- Multiple spatial, temporal, spectral resolutions
- Diverse noise, missing data and gaps
- Domain physics and constraints (e.g., terrain, topography)
- Paucity of ground truth

USGS field crew during a Colorado flood
(Source: USGS)

What are The Unique Challenges of Spatiotemporal ML?

- Spatio-temporal auto-correlation, teleconnections, heterogeneity
- Multiple spatial, temporal, spectral resolutions
- Diverse noise, missing data and gaps
- Domain physics and constraints (e.g., terrain, topography)
- Paucity of ground truth

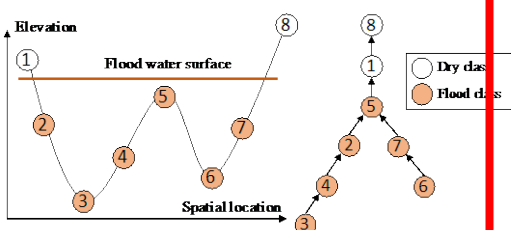


AI for spatial and spatiotemporal data listed as **open research question.**

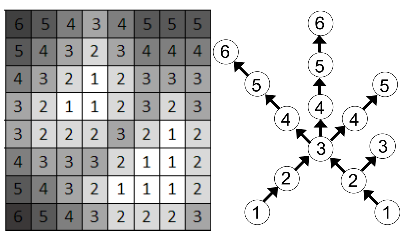
> Many AI applications are interdisciplinary in nature and make use of heterogeneous data. Further investigation of multimodality machine learning is needed to enable knowledge discovery from a wide variety of different types of data (e.g., discrete, continuous, text, spatial, temporal, spatio-temporal, graphs). AI investigators must determine the amount of data needed for training and to properly

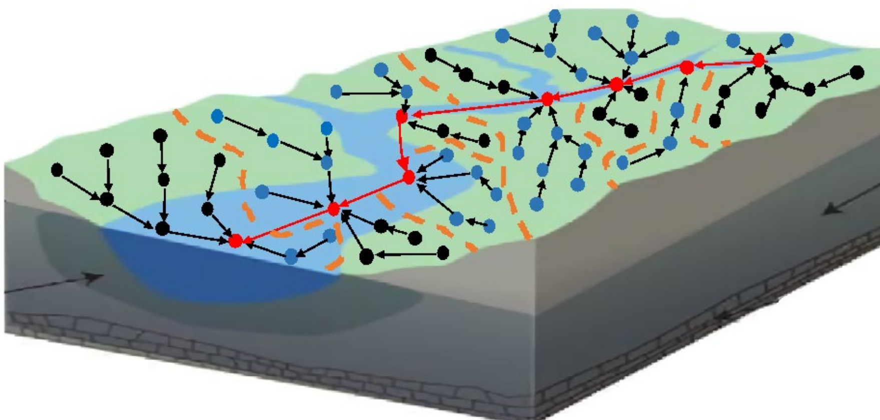
Topography-aware Flood Mapping from Satellite Imagery

- Spatial structure representation of terrains
  - Spatial graph (reverse tree): partial order class dependency
  - Focus on **tree** (not Directed Acyclic Graph) for simplicity and efficiency



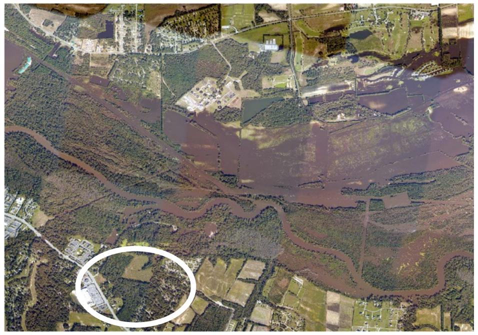
(a) Reverse tree (KDD18)

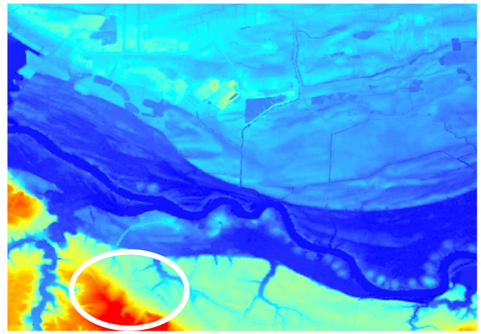

(b) Contour tree (KDD19)

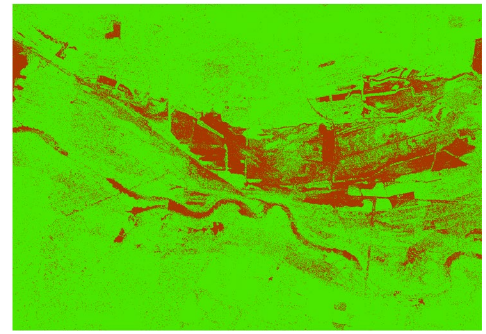

(c) Region-wise flow trees (ongoing with federal agencies)

Tree construction based on
H. Carr et al. 2003, with customization

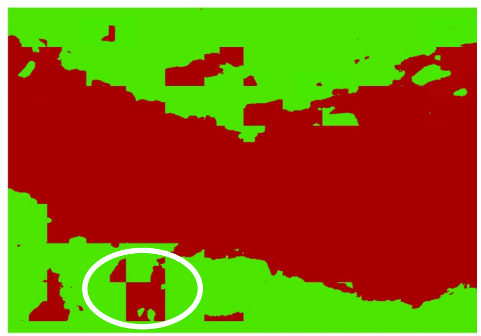
Result Visualization

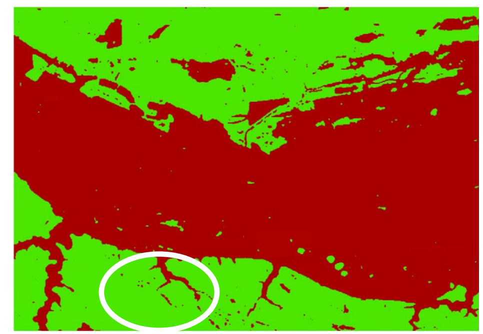

High-resolution aerial image (Hurricane Matthew, 2016)


Digital elevation model


Random forest


Gradient boosted model


Deep learning (U-Net)


Hidden Markov tree

Analysis:

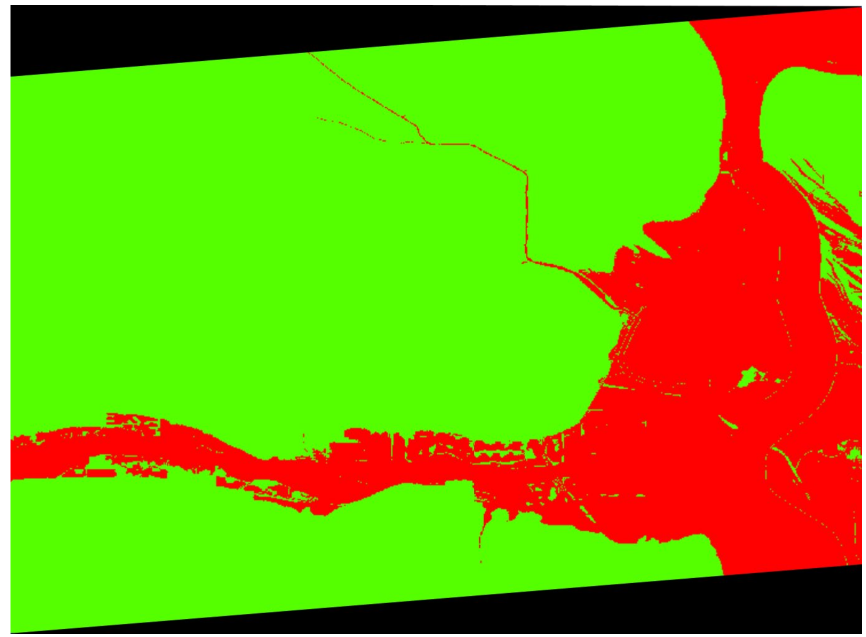
- Non-spatial classifier performed poorly (significant false negatives)
- U-Net performed better but still confused in highly vegetated areas
- HMT performed the best

Note: permanent water can be further removed to show flooded area

HMT in A Real-World Riverine Flood

- Collaboration with USGS, NGA, NOAA NWC

Worldview image in **Omaha, Nebraska**, 2019



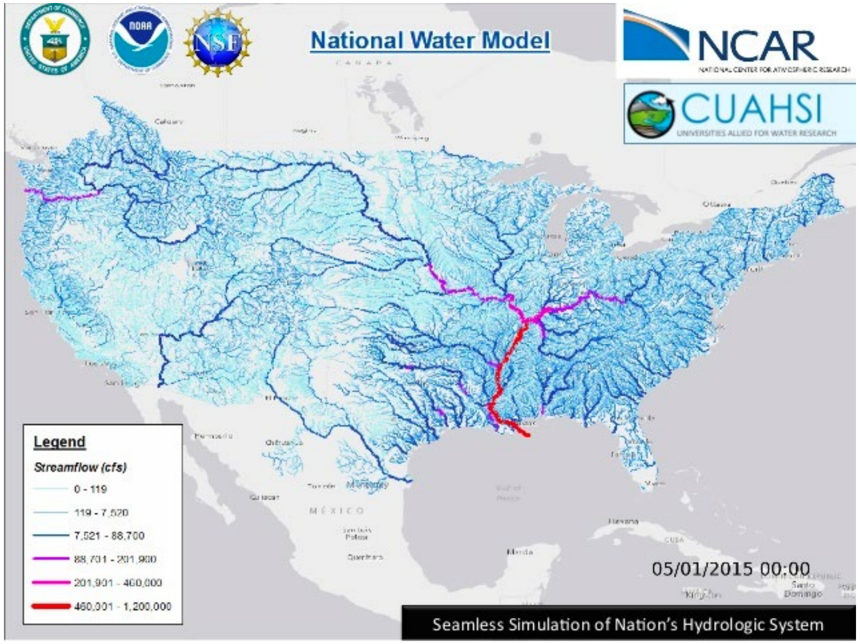
Deep neural network results

After using Hidden Markov trees

Analysis: integrating domain knowledge (topography constraint) enhances machine learning performance

National Hydrography Dataset Refinement

- National Hydrography Dataset (NHD)
  - Widely used for surface water body features
  - With high-resolution remote sensing data, USGS is refining NHD to a higher resolution




Rivers


Creeks

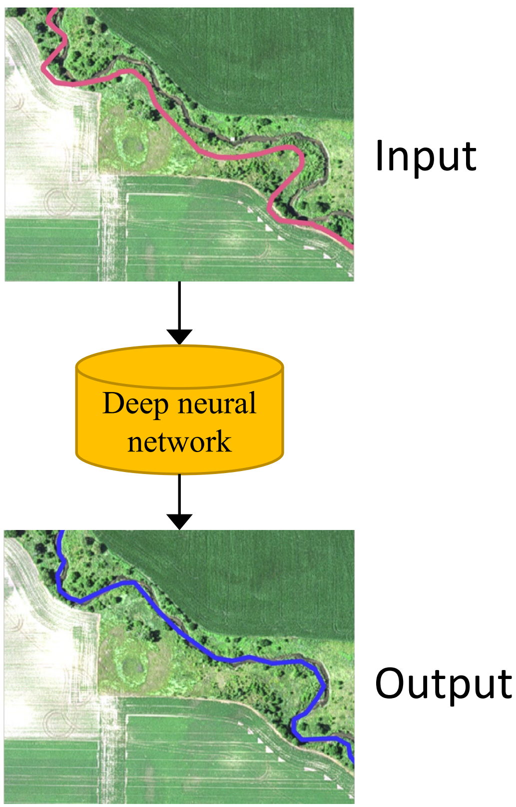

Lakes


Ponds

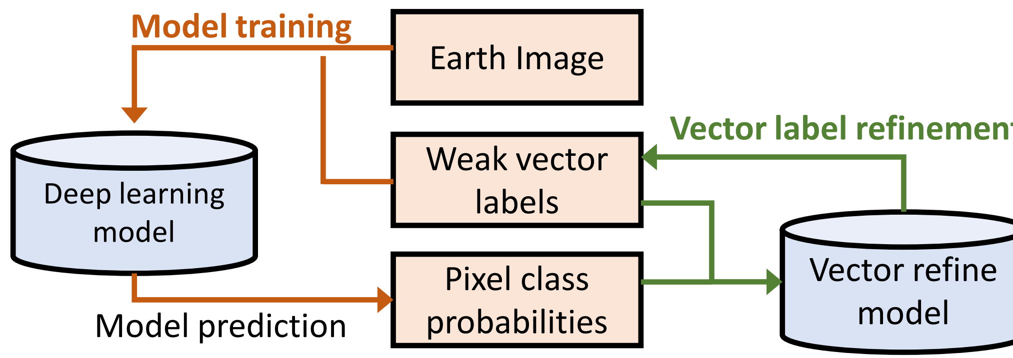
National Hydrography Dataset (NHD) refinement

Weakly Supervised Learning for National Hydrography Dataset (NHD) Refinement

Problem Definition



Approach: Weakly supervised spatial learning

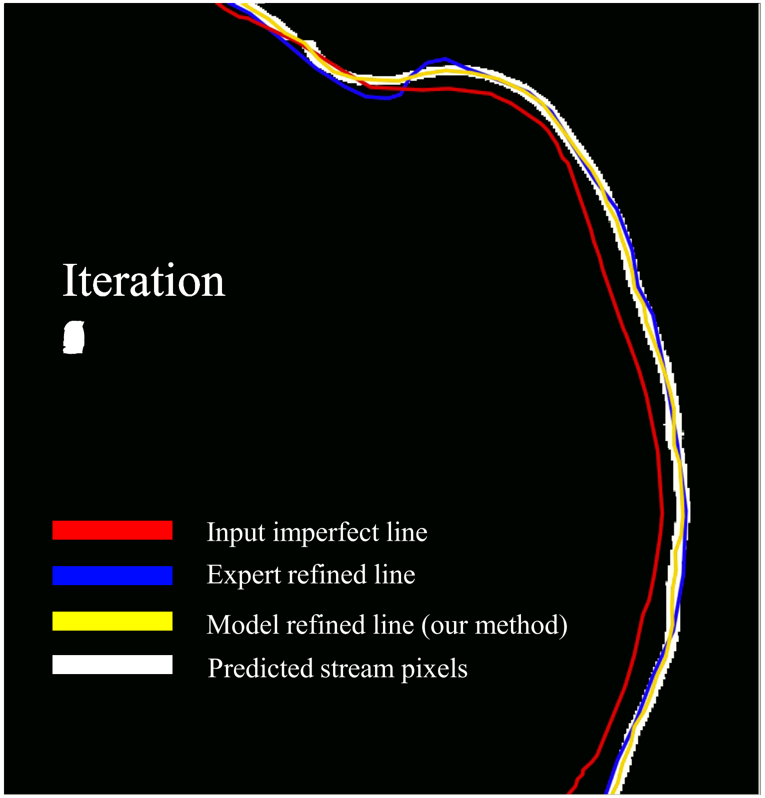


Evaluation

Method	Class	Confusion Matrix		Precision	Recall	F1 score
U-Net	Non-stream	9818176	79801	0.99	0.99	0.99
	Stream	79672	57551	**0.39**	**0.44**	**0.42**
U-Net with self-training	Non-stream	9747792	150185	1.00	0.98	0.99
	Stream	48750	88473	**0.37**	**0.64**	**0.47**
Our Method	Non-stream	9867813	30164	0.99	0.99	0.99
	Stream	57614	79609	**0.71**	**0.58**	**0.64**

Evaluation: A Case Study and Interpretation

- Initially:
  - Stream predictions are wide
  - Due to imperfect polyline labels
- After iterations:
  - Stream predictions narrower
  - Refined line converges



The proposed method successfully refines vector labels with registration errors

Conclusion and Future Work

- Spatiotemporal machine learning is important but technically challenging
- Addressing the challenges motivate new ML research

- **Future Directions**
  - Integrating AI and physics (process-based model, numerical model)
  - Fusing multiple data sources (remote sensor, in-situ sensors, “social” sensor or citizen science, simulation data)
  - Looking for collaborations with domain scientists!