Fresh Perspectives

An Oceanographer’s View of Machine Learning in Great Lakes Science

Dani Jones
Associate Research Scientist (Research Faculty)
Cooperative Institute for Great Lakes Research (CIGLR)
University of Michigan
dannes@umich.edu

UNIVERSITY OF MICHIGAN

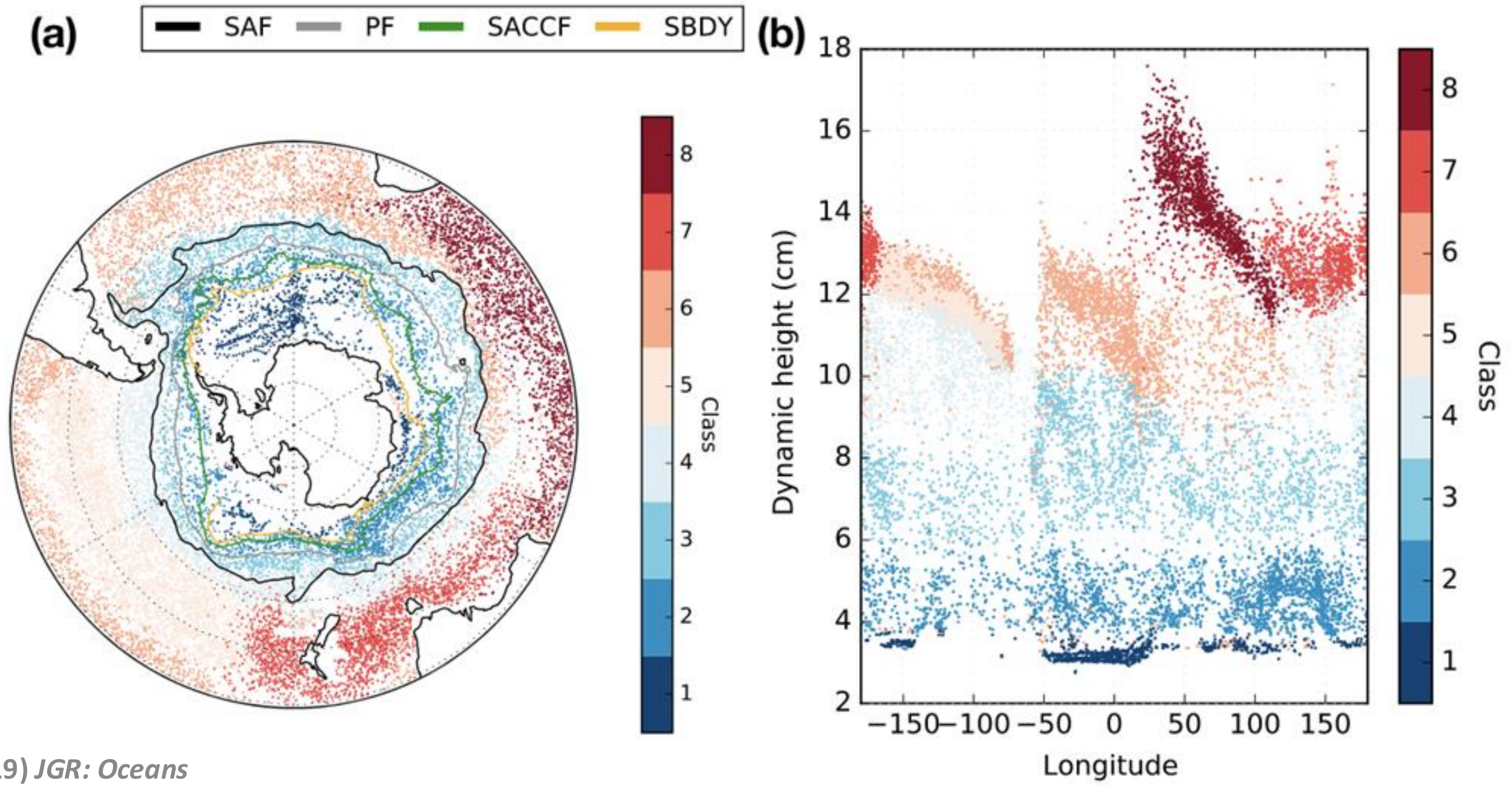
CIGLR Cooperative Institute for Great Lakes Research

How an oceanographer ended up at a Great Lakes research institute

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

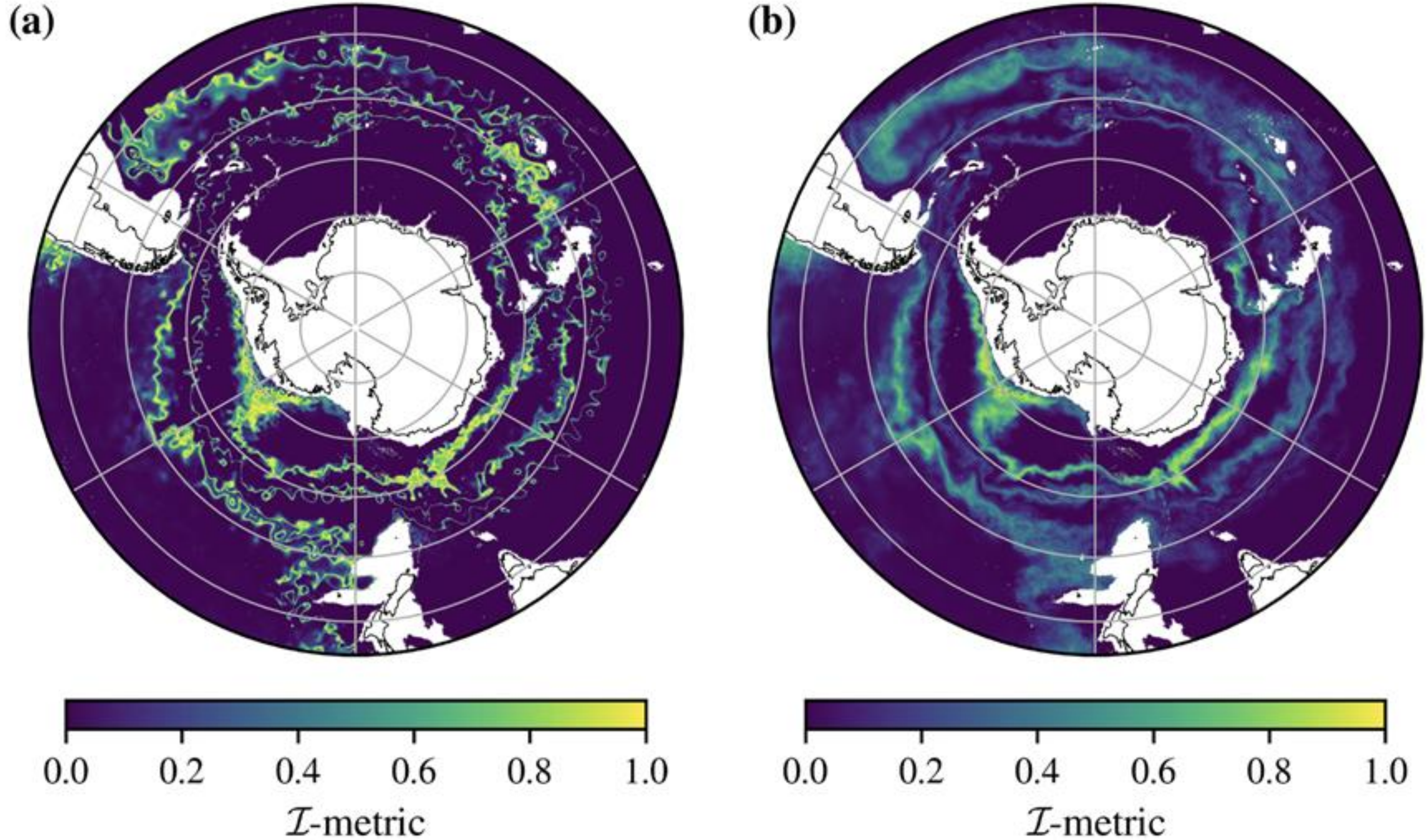
Cooperative Institute for Great Lakes Research
CIGLR
Great Lakes Science for Society

While at BAS: exploring machine learning



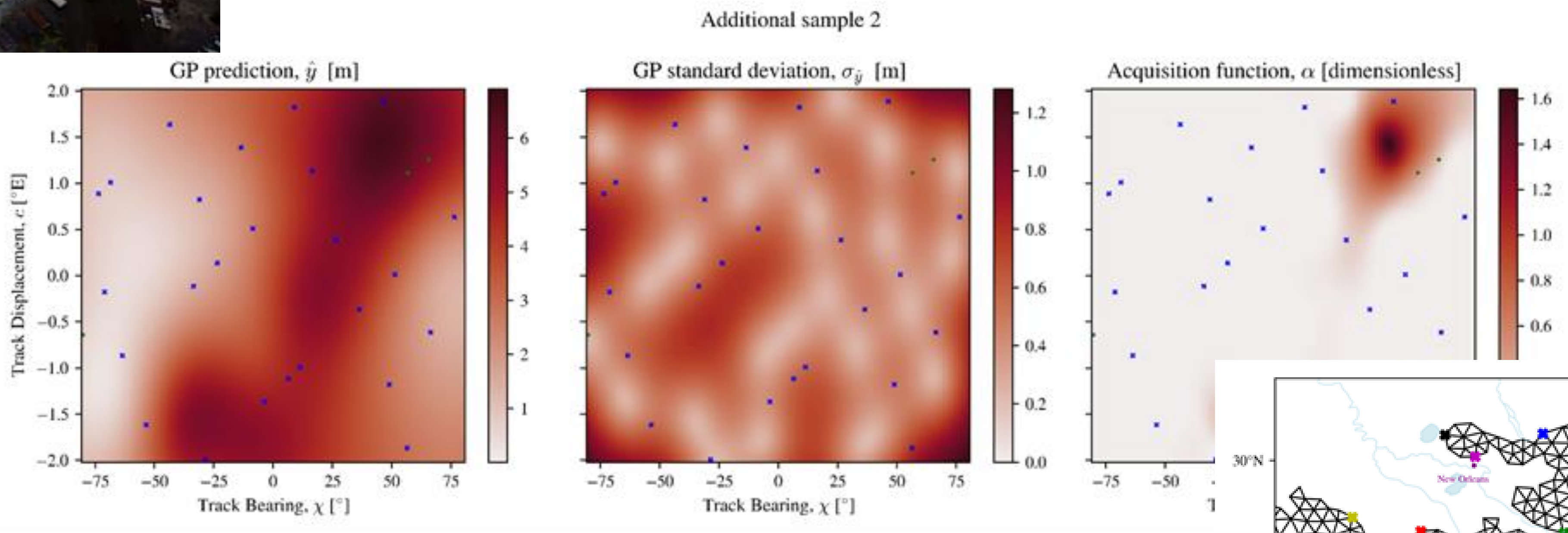
Jones et al. (2019) *JGR: Oceans*

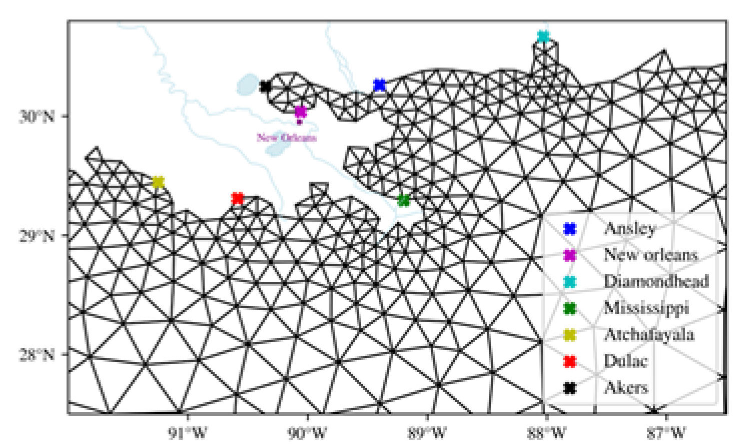
While at BAS: exploring machine learning



Thomas et al. (2021) *Ocean Science*

While at BAS: exploring machine learning





Thomas et al., *in prep.*

CIGLR was expanding at the time

Research Institute

Directorate

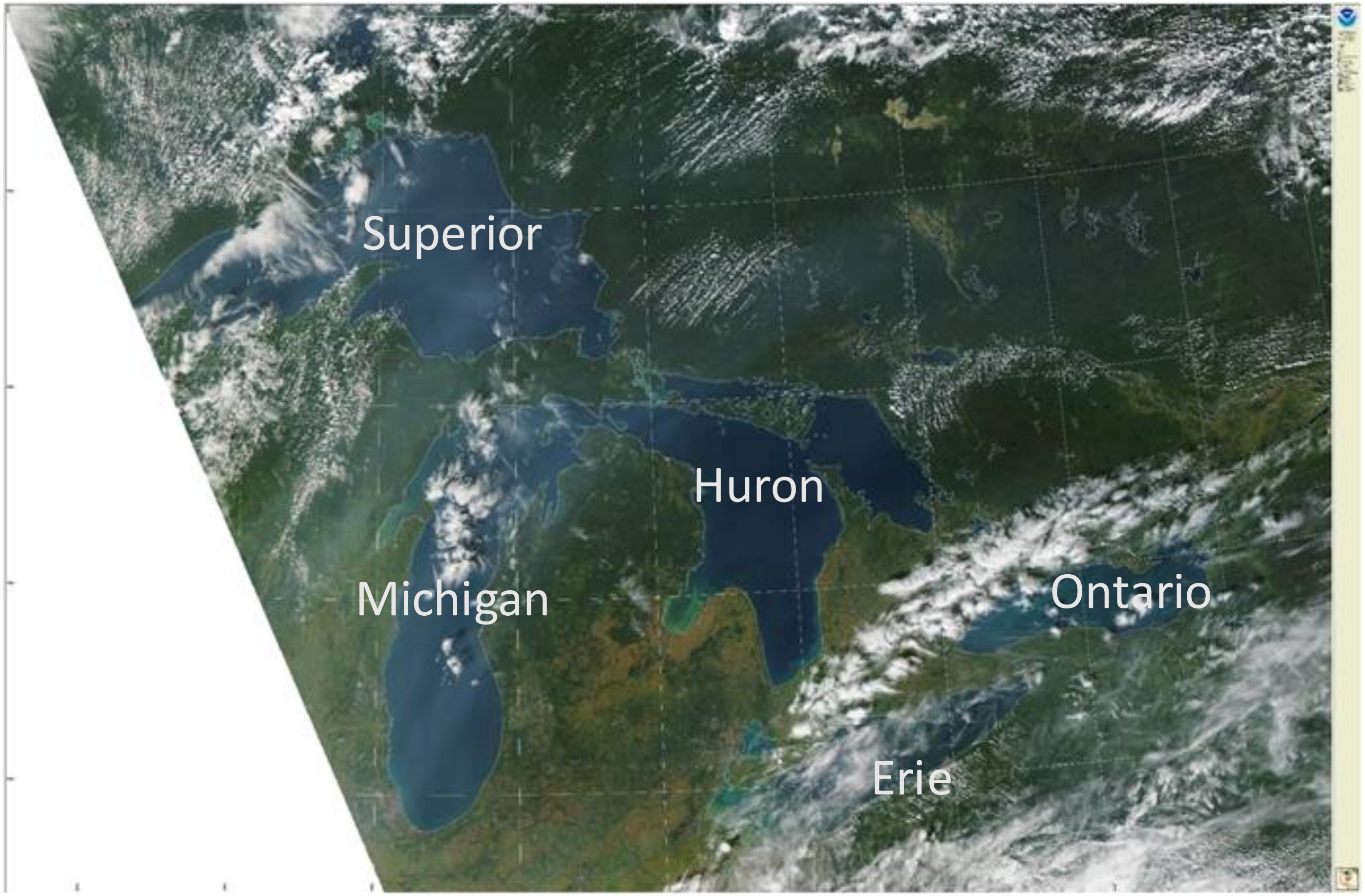
Research Engagement Team

Regional Consortium

CIGLR Summit: AI for the Great Lakes

July 2024, Dana Building, University of Michigan

The Laurentian Great Lakes are (kinda) big



Source: MODIS true color image from 17 September 2024

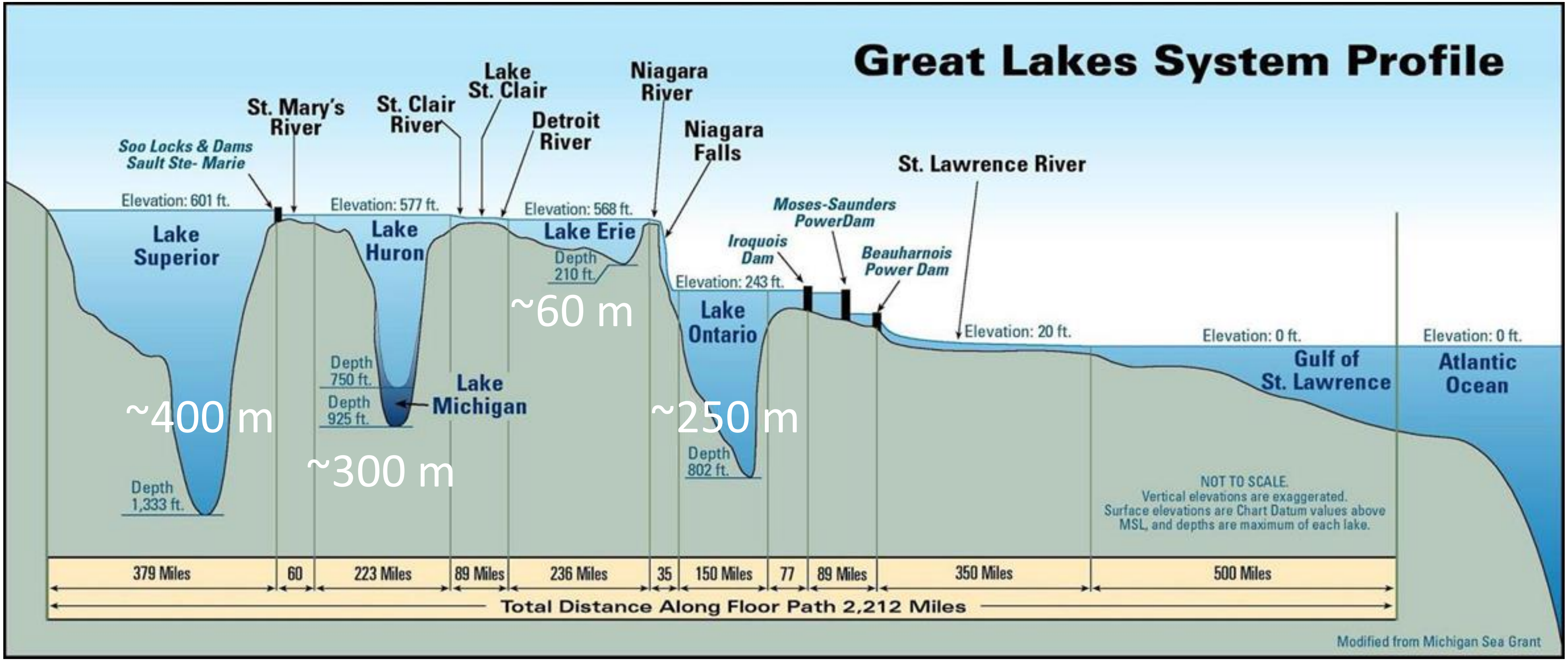
- Largest group of freshwater lakes in the world, containing:
  - ~80% of North America's surface freshwater
  - ~20% of the world's surface freshwater
- ~18,000 km of coastline
- Enough freshwater to submerge the continental US in ~3 m of water

The Laurentian Great Lakes are (kinda) big

	Flow rate [Sv]	Time needed at that flow rate to drain the Great Lakes
St. Lawrence River	0.01	~70 years
Alaska Coastal Current	0.9	~10 months
Gulf Stream	30	~8 days
Antarctic Circumpolar Current (ACC)	170	~35 hours

Image from “Draining the Great Lakes”, *Nat. Geo.*

The Laurentian Great Lakes are (kinda) big



The Laurentian Great Lakes are regionally significant

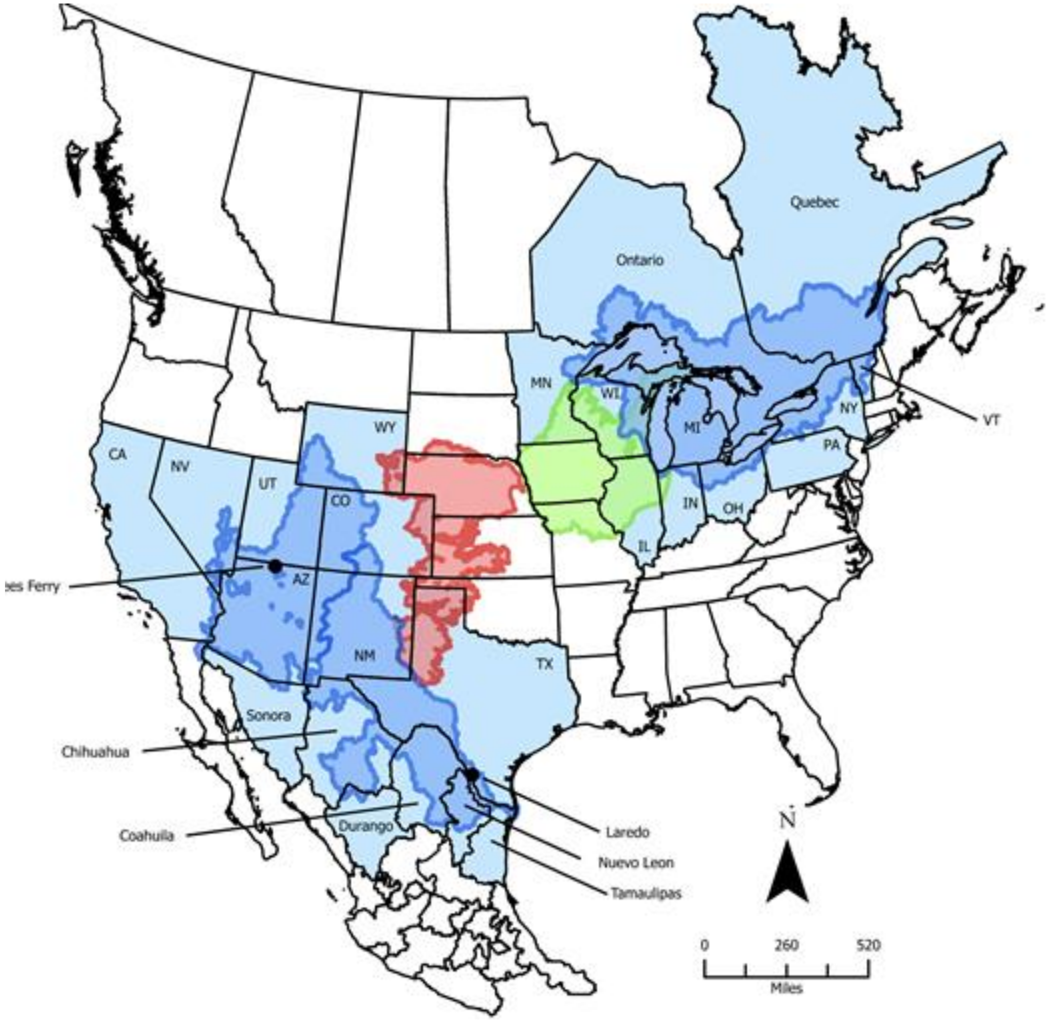
Source: GreatLakesNow.org

- They support many communities: **~34 million people** in the US and Canada rely on the Great Lakes for water, food, commerce, hydropower, and recreation

- Impact weather patterns and regional climate

- Host considerable **biodiversity** (more than 3,500 species of plants and animals)

The Laurentian Great Lakes are regionally significant



Gronewold et al., (2024), *Nature Communications*

The Great Lakes serve as a key example of:

- **cross-border water management agreements**
- understanding and adapting to **regional climate change**

The Great Lakes span a wide range of conditions

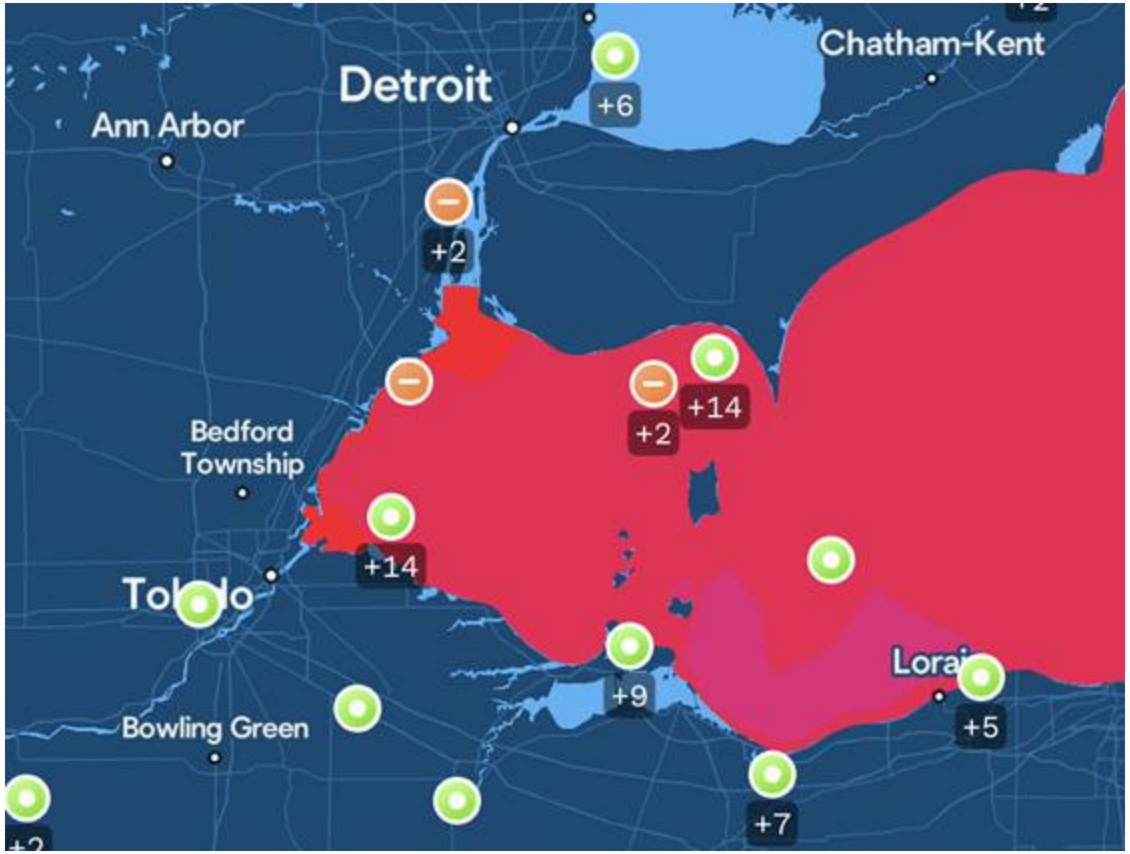
Lake Superior	Lake Ontario
Deep, sea-like	Shallow, highly variable
Relatively unregulated	Heavily regulated
Colder inland climate	Warmer, maritime-influenced climate

Why should oceanographers care about the Great Lakes?

They are a **natural laboratory** for studying:

- the role of air-water fluxes in setting heat and volume budgets,
- the interplay between variability in stratification, ice cover, circulation, and solar radiation and the resulting impacts on:
  - biogeochemical cycling (e.g., nitrogen and phosphorus cycling, important for understanding harmful algal blooms),
  - microbial communities, wider ecosystems, and invasive species
- exchange flows (e.g., the Straits of Mackinac)

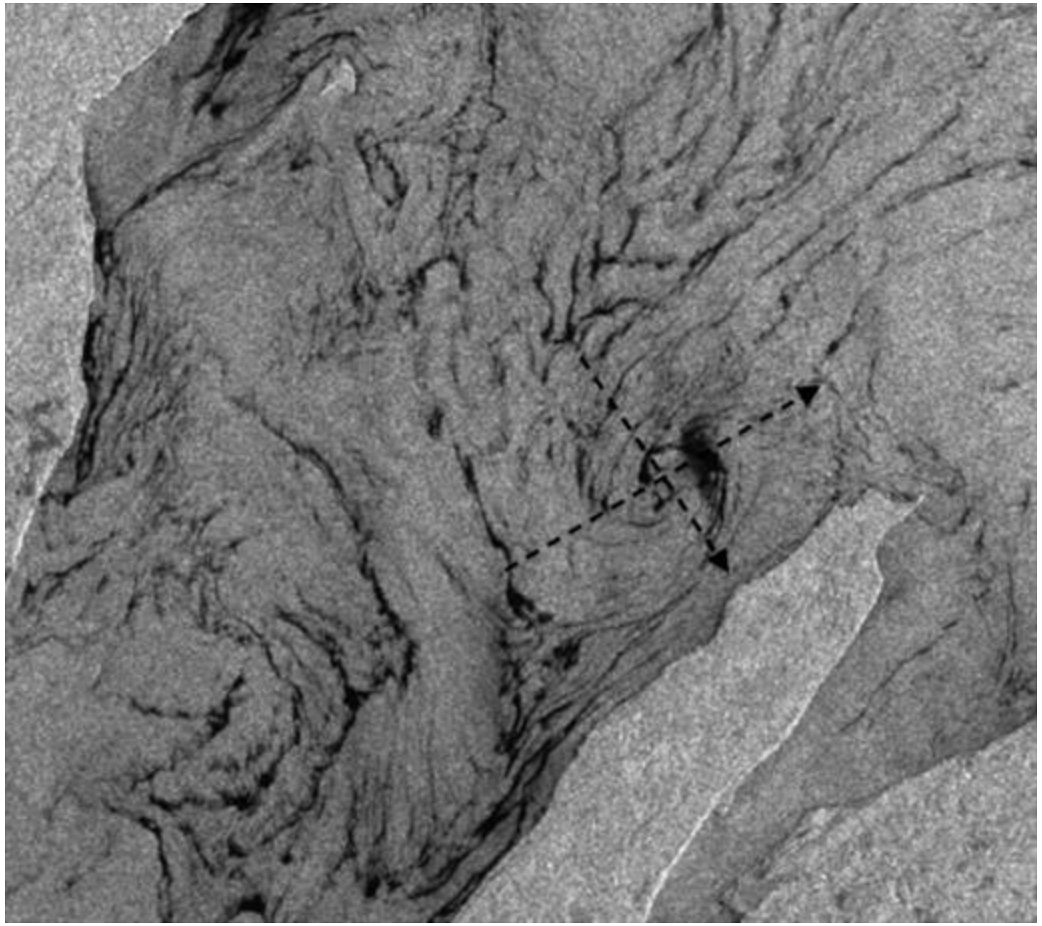
Why should oceanographers care about the Great Lakes?



- Some regions are among the **best monitored marine environments** in the world
- **Winter is relatively understudied**; knowledge gained here could translate to polar environments
- Typically **much easier to get to** and operate in than much of the open ocean!

Source: seagull.glos.org

The Great Lakes as inland freshwater seas



McKinney et al., (2012), *J. Great Lakes Res.*

- On the largest scales, Great Lakes circulation has a Rossby number comparable to that of basin-scale ocean gyres
- Coastal upwelling
- Mesoscale and submesoscale phenomena
- Distinct nearshore and open water dynamical regimes

The Great Lakes as inland freshwater seas

- Tilted seasonal thermoclines and associated geostrophic flows (baroclinic “coastal jets” during summer stratification)
- Prominent seasonal cycles in convective mixing and sea ice

Source: NOAA GLERL

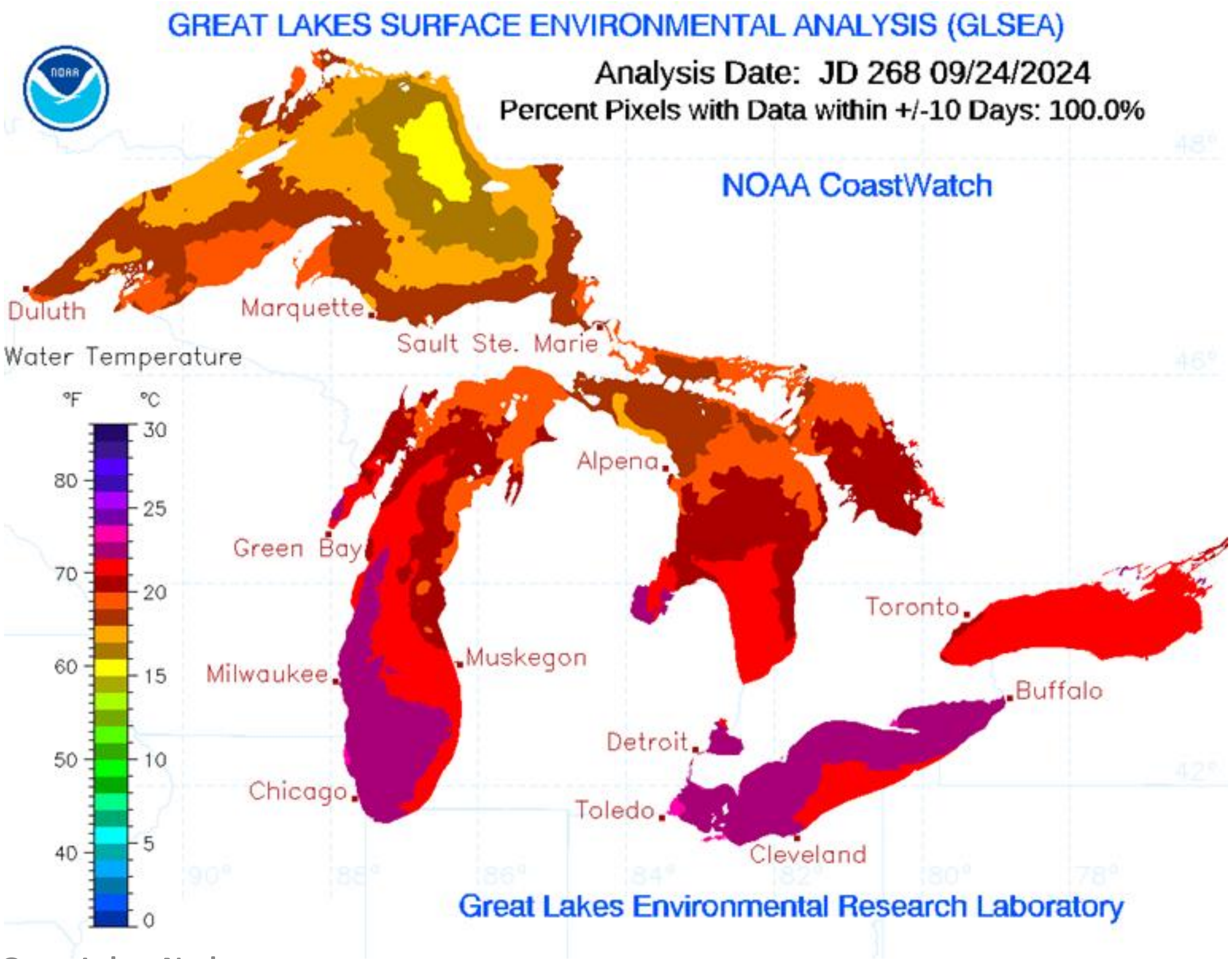
Of course, there are some differences…

Baroclinic Rossby radius of deformation is about an order of magnitude smaller than that of the midlatitude ocean (weaker stratification from the **lack of salinity gradients)**

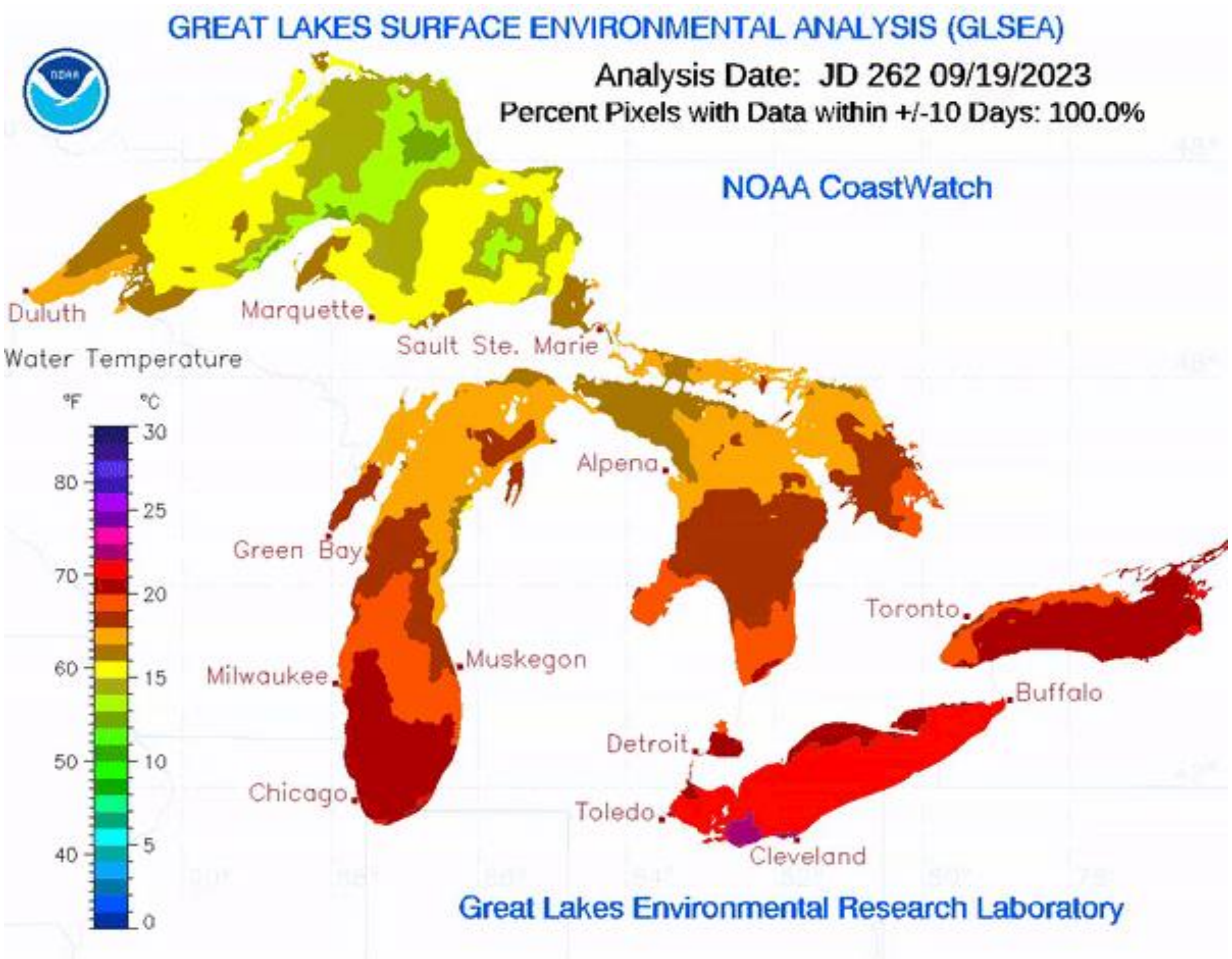
No appreciable tidal mixing

No open boundary conditions to consider

Credit: Morton (2020), *EOS*

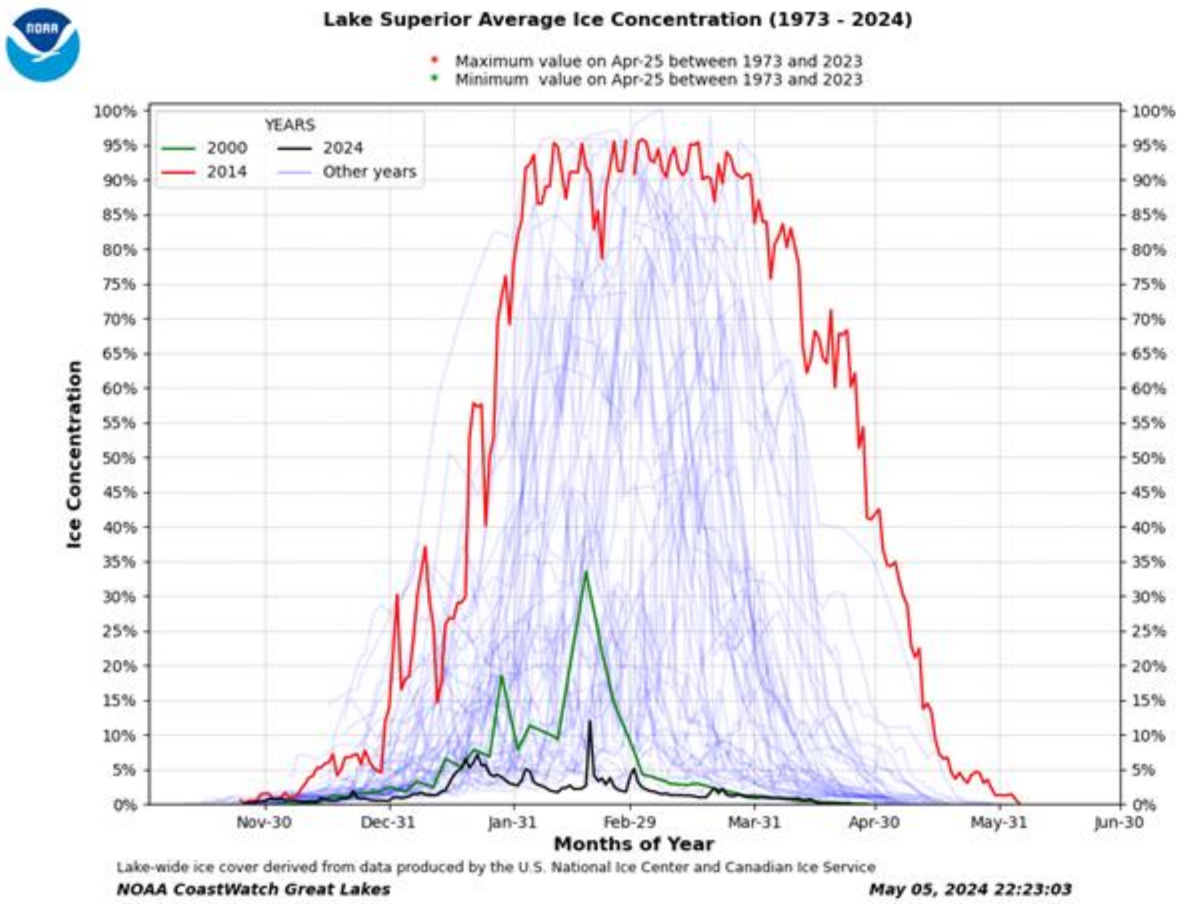


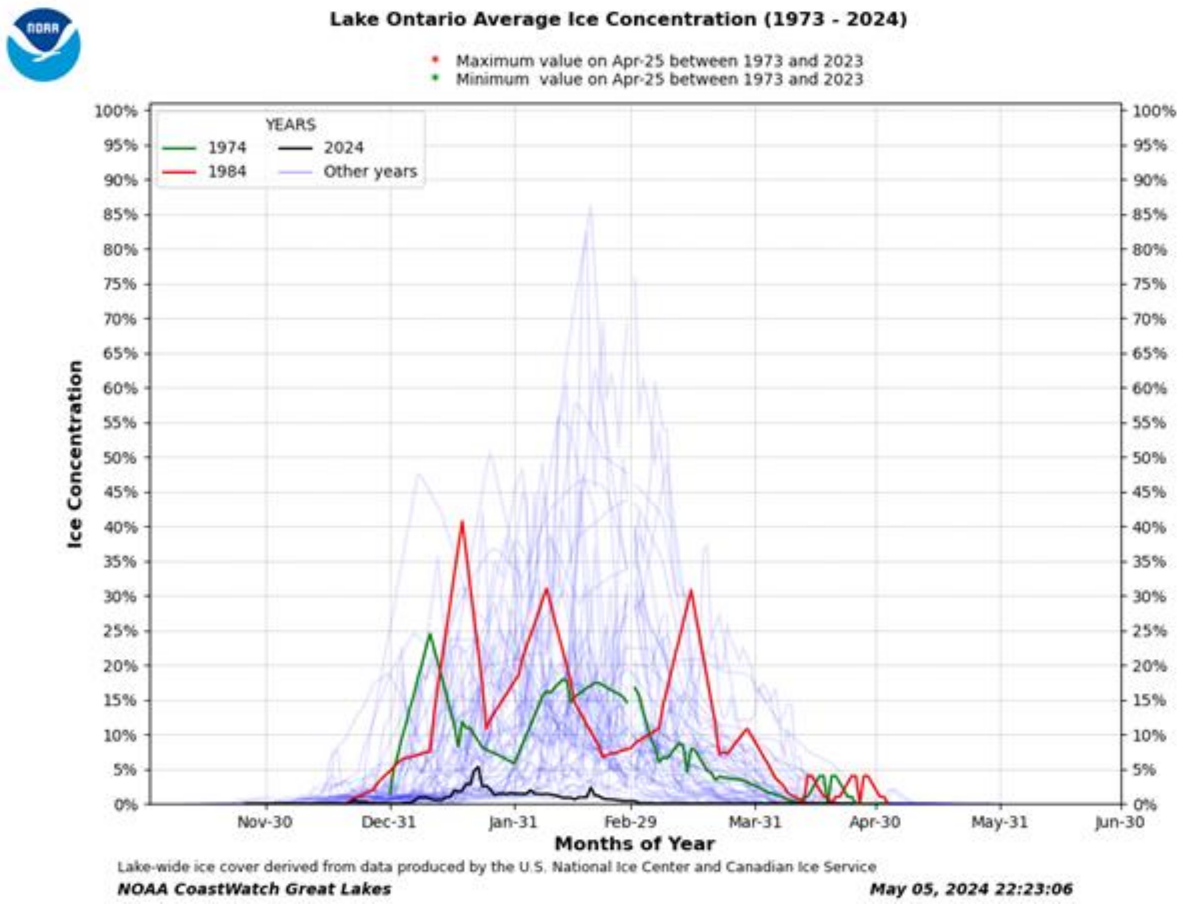
Source: NOAA CoastWatch Great Lakes Node



Source: NOAA CoastWatch Great Lakes Node

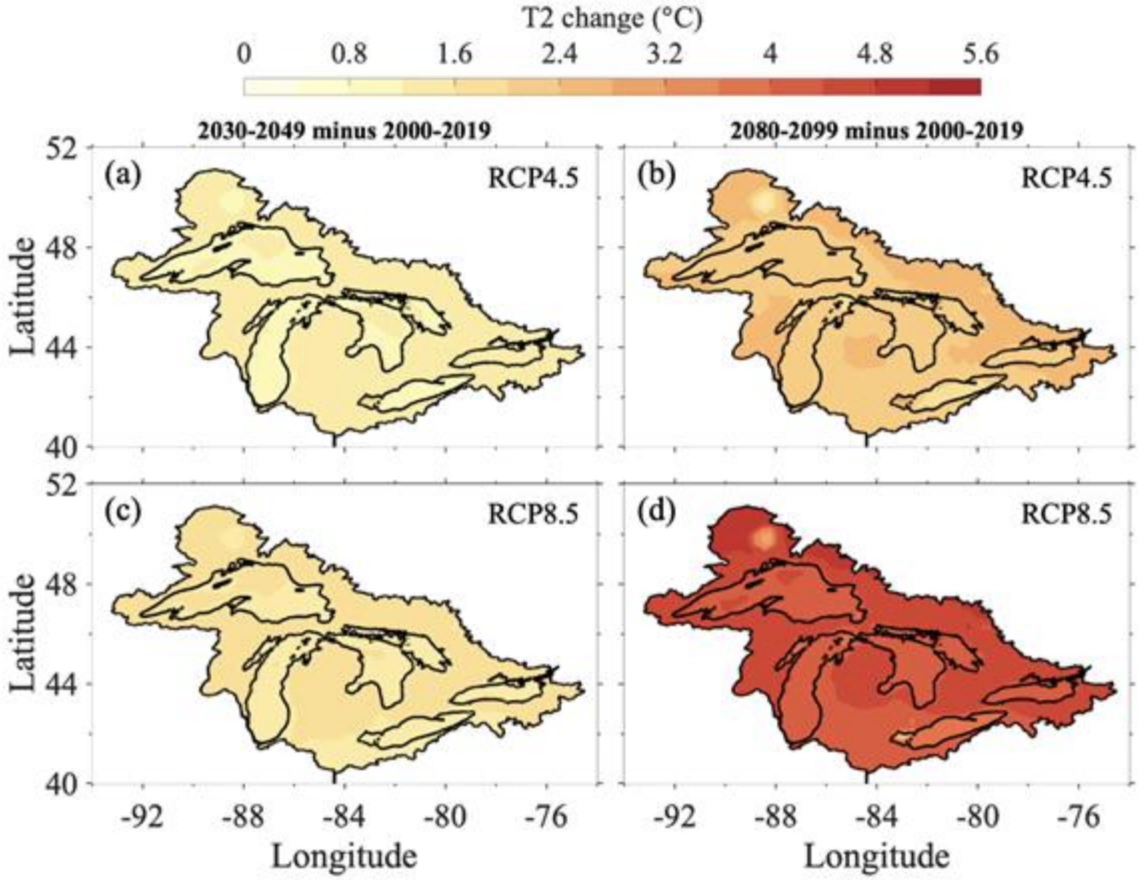
Ice cover has considerable interannual variability





Source: NOAA CoastWatch Great Lakes Node

A region under threat

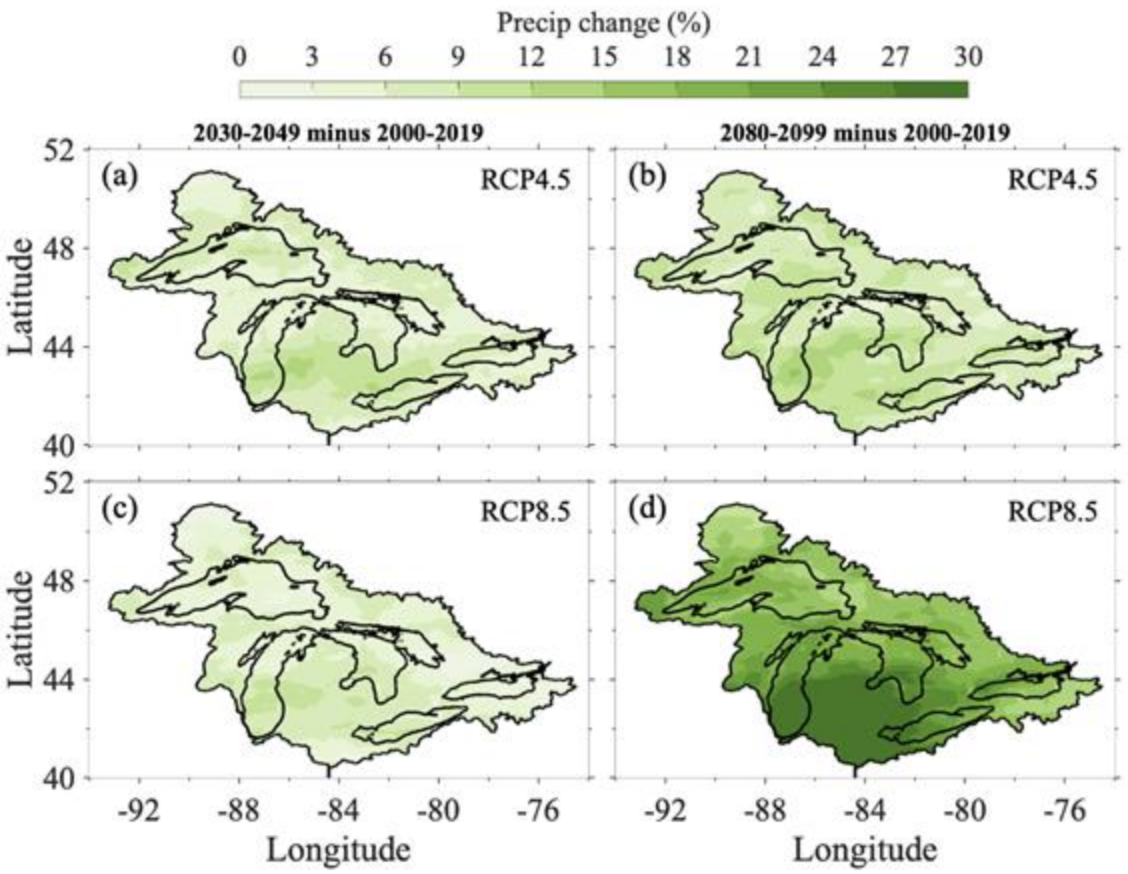


Xue et al. (2022), *Geosci. Model Dev.*

Climate change is projected to increase precipitation, air temperatures, and evaporation

The overall impact on water levels remains uncertain: will they be higher, lower, and/or more variable?

A region under threat



Climate change projections for the Great Lakes have been crude

The Great Lakes are often not represented at all, or they are represented by a simple 1-D model

This approximation has impacts on regional climate projections

Xue et al. (2022), *Geosci. Model Dev.*

Could machine learning (ML) and artificial intelligence (AI) transform Great Lakes science, restoration, and management?

A "digital twin" approach would move the field away from watershed-specific studies to a unified, region-wide approach

The Great Lakes can serve as a unique test bed for the application of ML/AI to large, open-water systems and the associated ecosystems

ML/AI "emulators" could make it easier for stakeholders to engage with process-based and risk evaluation models

What is artificial intelligence (AI) and machine learning (ML)?

AI: Machines / algorithms that can perform tasks that normally require human intelligence, such as learning, reasoning, using and comprehending language, decision making, etc.

ML: Subset of AI focused on data-driven learning and prediction (has some overlap with statistics)

Great Lakes science and management is just starting to use ML, and we're in the very early stages of conceptualizing how we might use AI

Cooperative Institute for Research in the Atmosphere (CIRA) ML Philosophy

Using machine learning to inform life-and-death decision making, e.g., to make decisions on which regions to evacuate, is very different from using machine learning for standard computer science applications.

In particular, in order for machine learning approaches to be accepted into operational use for weather forecasting, they must be as robust and interpretable as possible.

Connecting Models and Observations

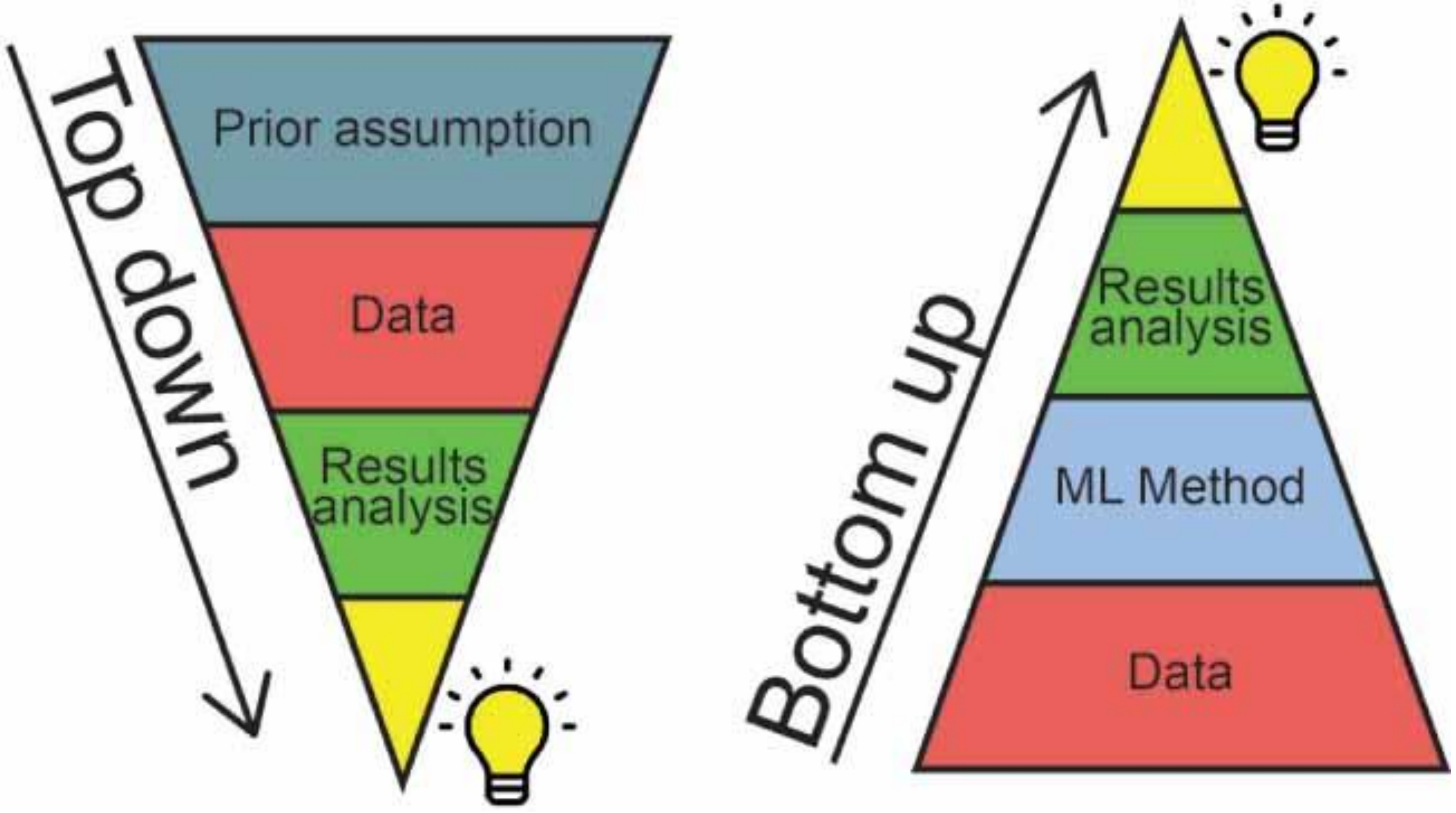
Cooperative Institute for Research in the Atmosphere (CIRA) ML Philosophy

The best way to achieve robustness and interpretability is to make ML models **as simple as possible**, using any of the following approaches:

- Integrating expert knowledge (e.g. feature engineering)
- Simplifying neural networks where possible (e.g. ablation studies)
- Providing well-calibrated uncertainty estimates (use a variety of metrics)
- Applying Explainable AI (XAI) methods

CIRA
Connecting Models and Observations

ML/AI can be used as a "hypothesis generation engine"



Sonnewald et al. (2023)

The Great Lakes are a unique test bed for ML/AI applications in environmental sciences

Extensive coastlines (relevant to freshwater and marine coastal systems globally)

High density of observations per square kilometer in some regions (e.g. western Lake Erie)

Closed boundaries

Wide variety of physical and regulatory conditions

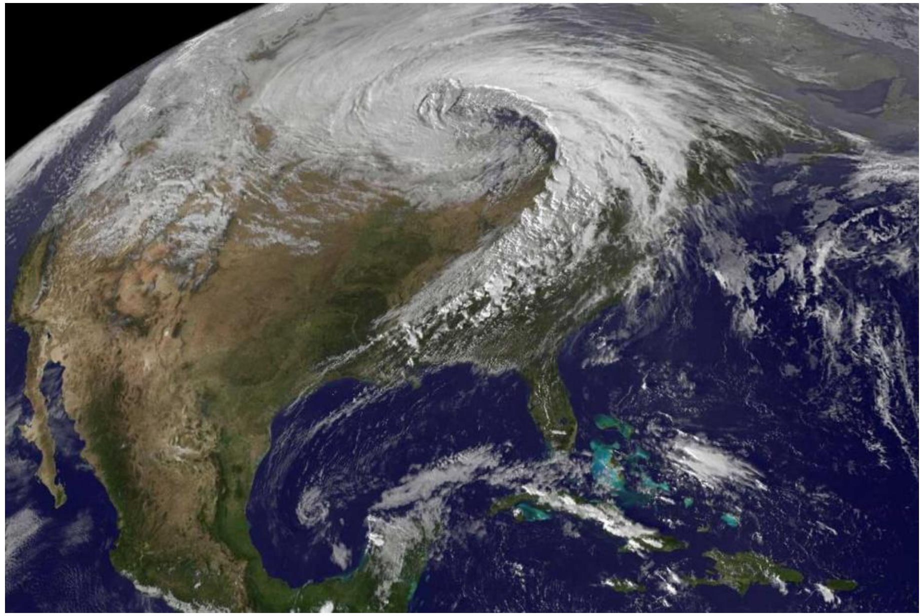
Tightly integrated into regional ecosystems and communities

Credit: Morton (2020), *EOS*

Weather as driving Great Lakes physics

Mid-latitude cyclones produce:

- winds that
  - drive lake currents and sediment transport
  - alter wave conditions that contribute to coastal erosion
  - lead to anomalous evaporation and air-water heat flux
- precipitation that
  - affects water levels (directly and via runoff)
- changes in air temperature that affect ice conditions



Can ML improve weather predictions (generally and for the Great Lakes?)

- Effective forecasting relies on the ability to estimate the spread of possible results
- AI-based weather prediction (AIWP) models struggle to compete with physics-based models when it comes to the growth of estimates of forecast uncertainty
    - They lack the sensitive dependence on initial conditions characteristic of weather
- This highlights the importance of including physical constraints (e.g. hybrid models, post-processing of dynamical models) and other forms of domain expertise

Credit: Paul Roebber

Decision-relevant areas that can potentially be improved by ML/AI

Opportunity	Example
Weather forecasting	Hypothesis Generation
Land surface and hydrological process modeling	LSTM for streamflow forecasts, reconstructions
Urban hydrologic processes	More Efficient Data Processing and Analysis
Water balance forecasting	Tools for Decision Makers and Stakeholders
Beach monitoring	Microbial contamination

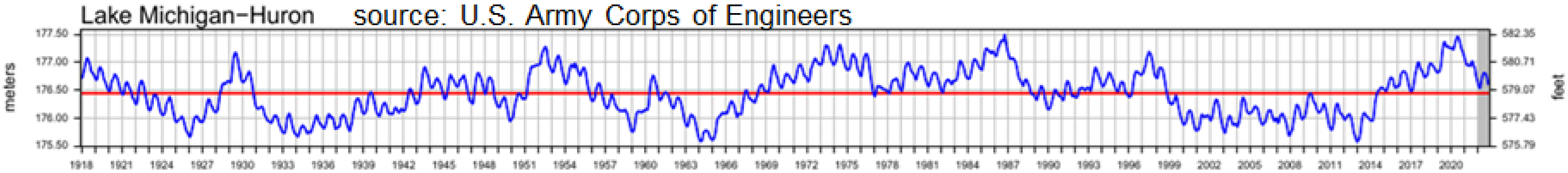
Opportunities for ML/AI in Environmental Science

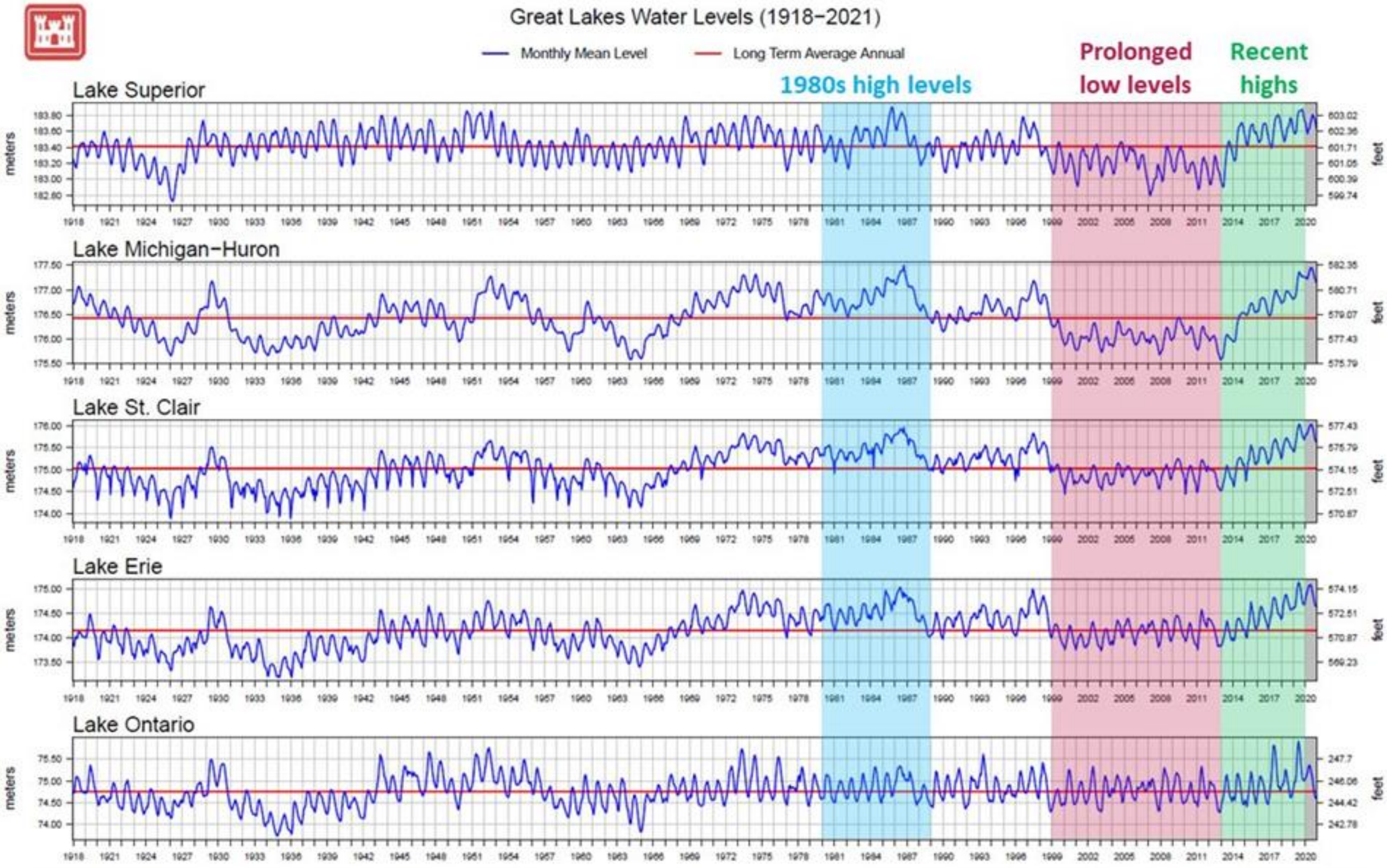
Opportunity	Example
Enhanced Data Analysis	Hypothesis Generation
Predictive Modeling	More Accurate, Efficient Forecasts
Automation of Routine Tasks	More Efficient Data Processing and Analysis
Real-time Monitoring and Decision Support	Tools for Decision Makers and Stakeholders

Risks of ML/AI in Environmental Science

Risk	Mitigation
Data Quality Issues and Bias	Robust Data Management
Low Interpretability / Transparency	Explainable AI
Ethical Concerns (e.g. Equitable Use)	Clear Ethical Guidelines, Open Platforms
Resource and Skill Gaps	Capacity Building
Over-reliance on Automated Systems	Collaboration and Governance

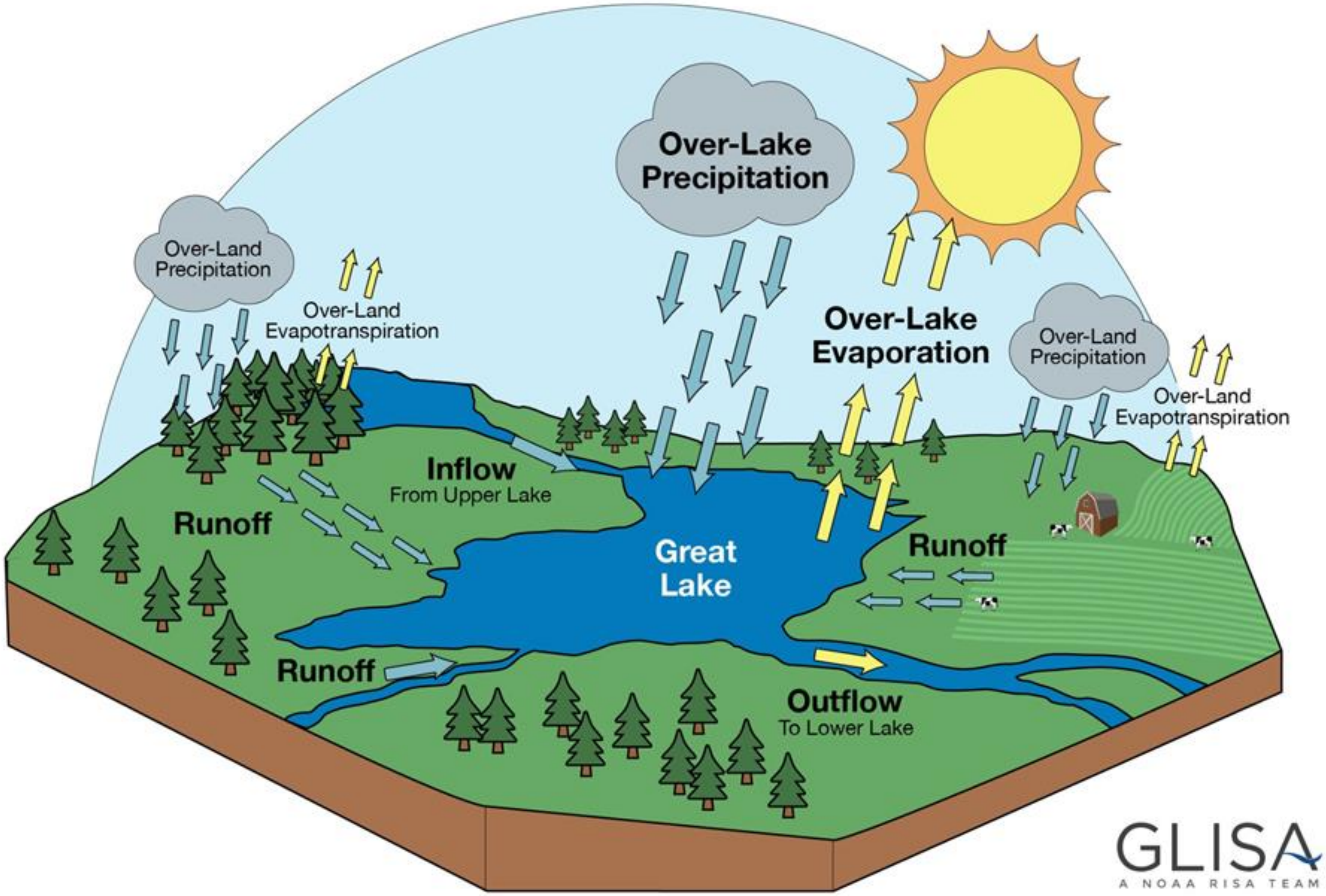
Communities and infrastructure along the coast of the Great Lakes are impacted by water level changes

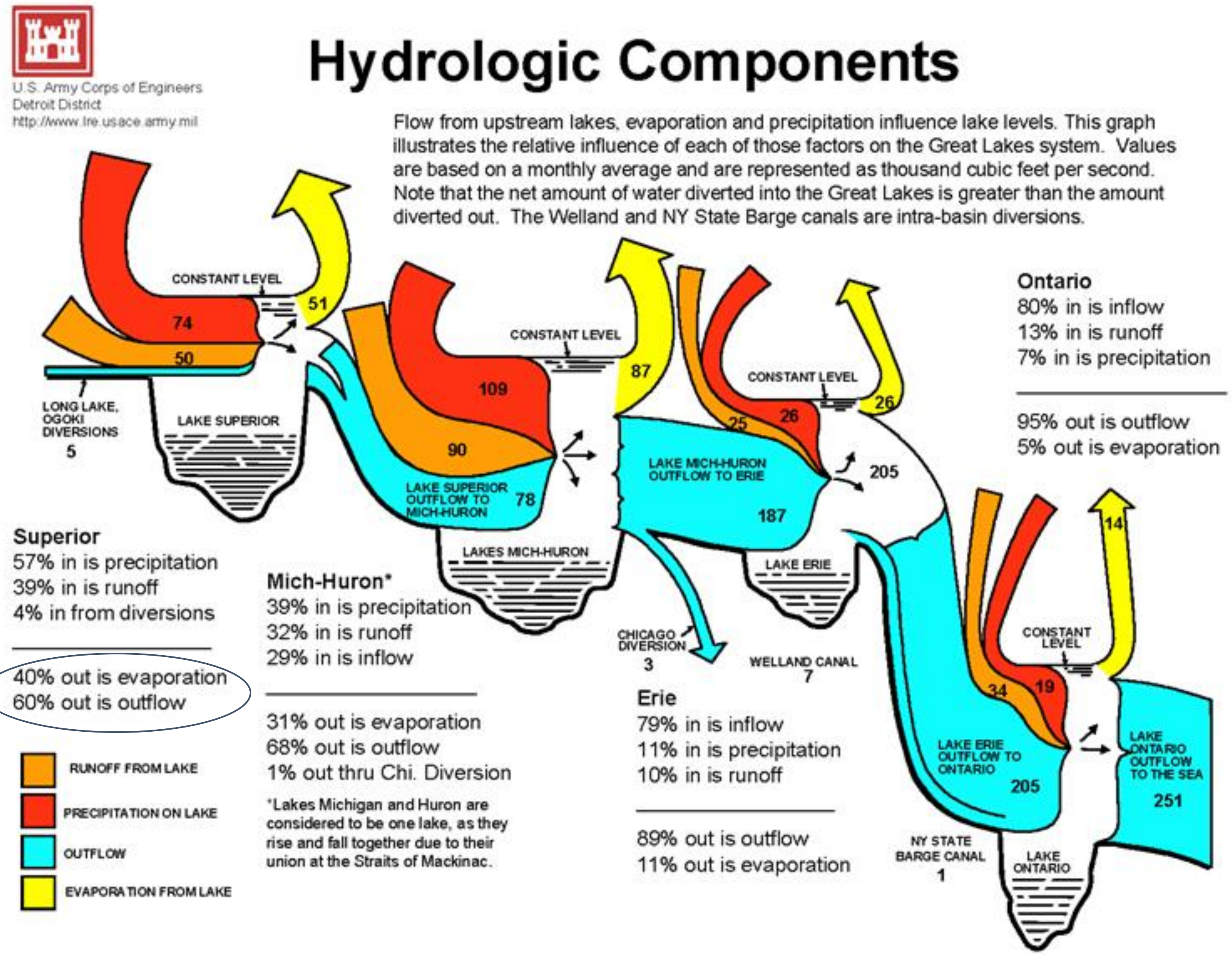




The monthly average levels are based on a network of water level gages located around the lakes.
Elevations are referenced to the International Great Lakes Datum (1985).

Water levels have been coordinated through 2019. Values highlighted in gray are provisional.


Over-Lake
Precipitation
Over-Land
Precipitation
Over-Land
Evapotranspiration
Over-Lake
Evaporation
Over-Land
Precipitation
Over-Land
Evapotranspiration
Inflow
From Upper Lake
Runoff
Great
Lake
Runoff
Runoff
Outflow
To Lower Lake
GLISA
A NOAA RISA TEAM



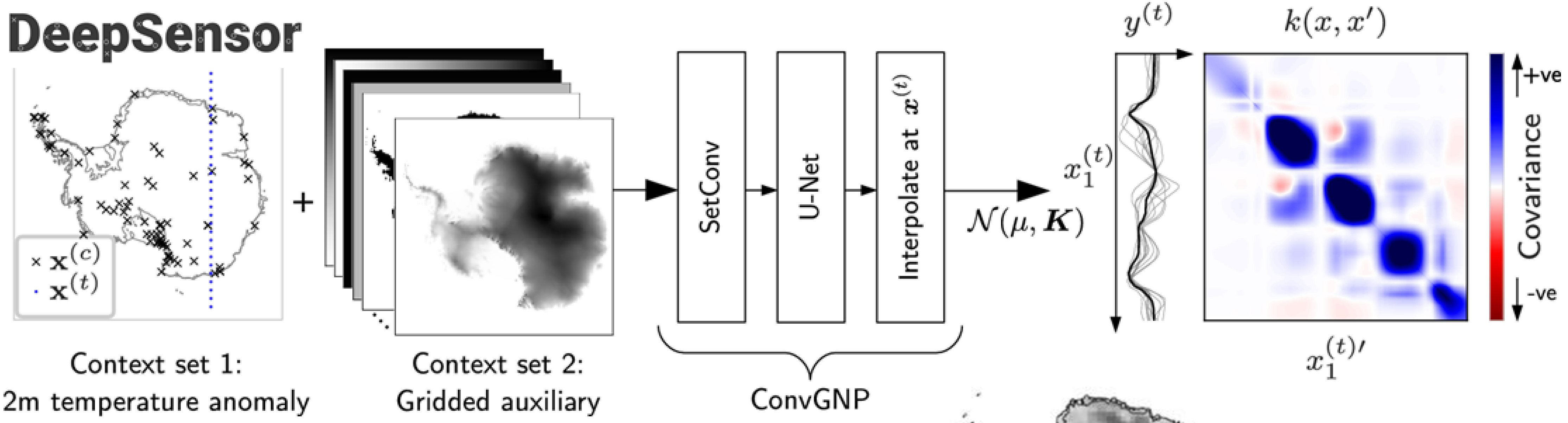
Source: US Army Corps of Engineers - Detroit District

CIGLR+GLERL are working to advance subseasonal-to-annual water level forecasts

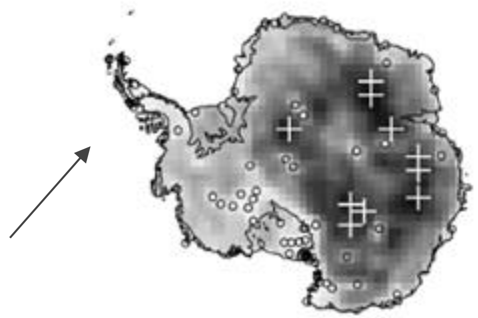
Water Level Forecast (1 month to 1 year)

- How does the global & regional climate affect water level forecasts for the Great Lakes? — Atmospheric Science
- What is the predictability of various hydroclimate factors for water level forecast? — Hydrology & Water Management
- How can we improve the reliability & accuracy of data-driven water level forecasts for operational applications? — Data Science
- What would make water level forecasts more useable, useful & efficient? — Research Engagement
- How do coastal managers use forecasts to mitigate risks? — Social Science

Gaussian neural processes and active learning



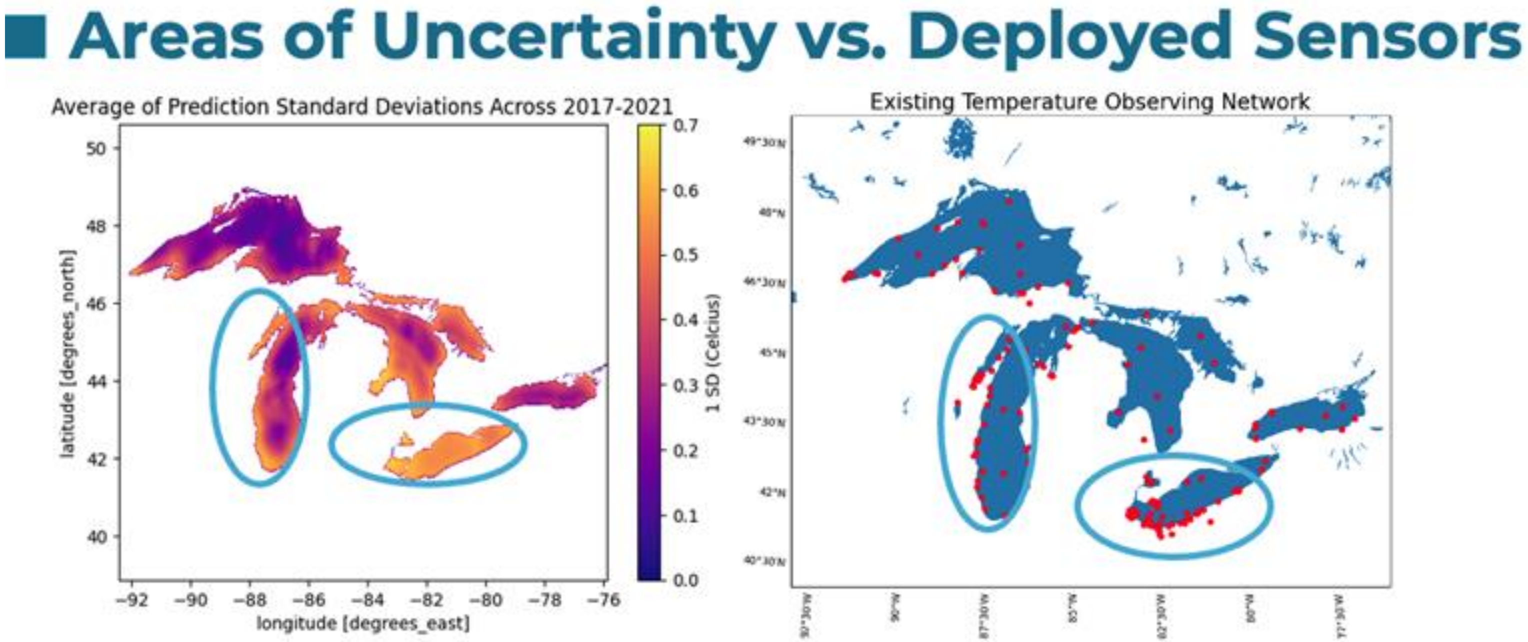
Active learning identifies areas where new sensors would best help reduce uncertainty (+)



Andersson et al., (2023), *Environmental Data Science*

Sensor placement for the Great Lakes

- Select **target variables** in collaboration with the Great Lakes Observing System (GLOS) organization
- "**Fuze together**" available observational and model data to fit the ConvNP model using DeepSensor
- Use active learning to evaluate the existing network and/or suggest **network expansions**
- Blue-sky extension: incorporate **human dimensions** via economic data, usage data



Credit: Erin Redding, CIGLR Summer Fellow

The wider context of ML/AI in env. sci.

- NOAA AI Strategic Plan (2021-2025)
- NOAA Modeling Strategy
- NOAA-GLERL Strategic Plan
- NOAA Center for AI (NCAI)
- NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)
- NASA-JPL Science Understanding through Data Science (SUDS)
- Pangeo, EarthMover
- Generative weather models (NVIDIA, Google DeepMind)

We are launching the “Great Lakes AI Lab”

- Perspective article for AIES
- Slack channel
- GitHub organization with collaborative repositories
- Partnership with UM MIDAS, NOAA

GREAT LAKES
AI LAB

We are launching the “Great Lakes AI Lab”

- Perspective article for AIES
- Slack channel
- GitHub organization with collaborative repositories
- Partnership with UM MIDAS, NOAA

Thank you!

GREAT LAKES
AI LAB