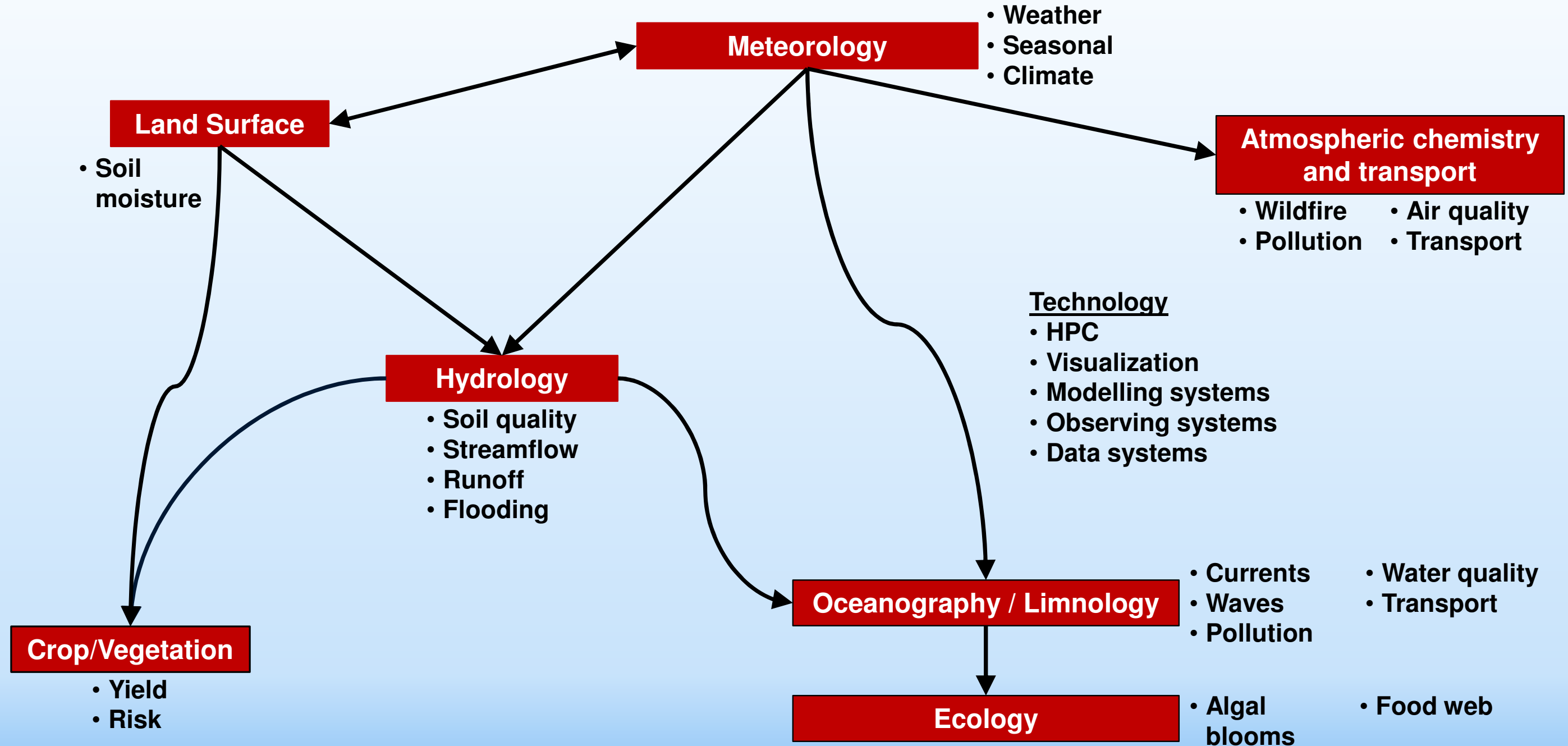


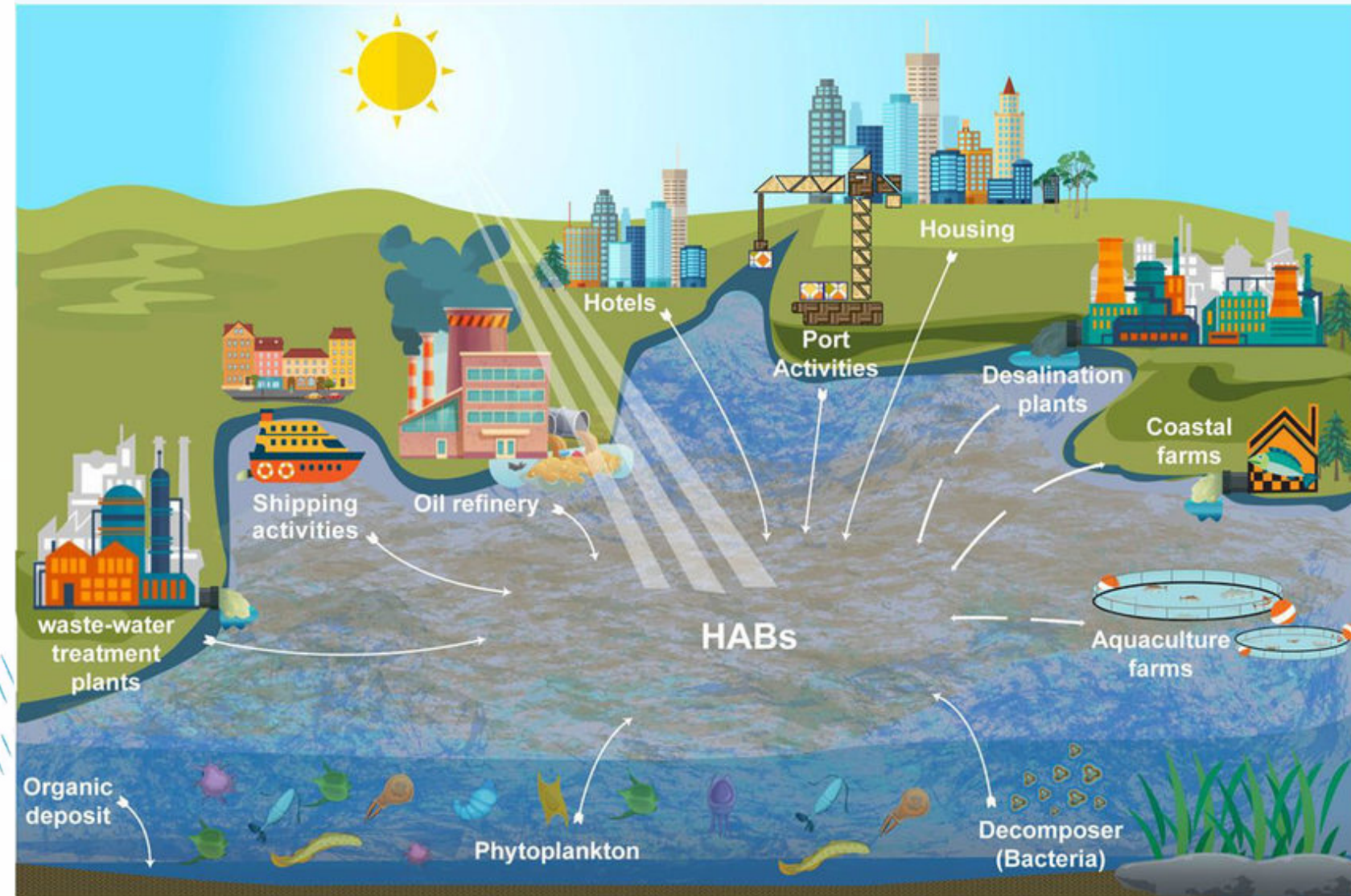
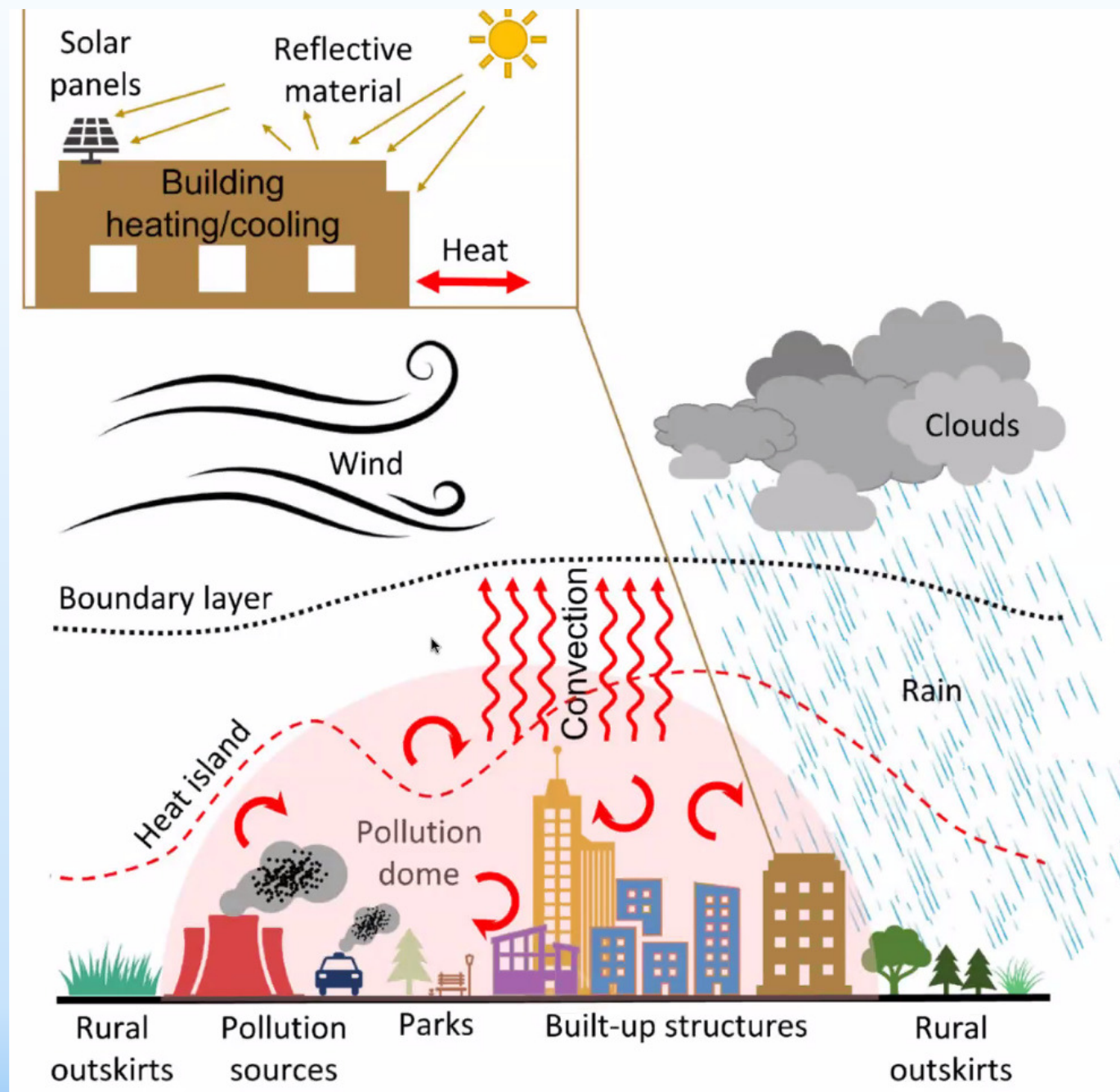
Furthering Research in Environmental Sciences on a Warming Planet Leveraging Advances in Open Data and Technology

Lloyd Treinish
IBM Distinguished Engineer
Chief Scientist for Climate and Weather
Head, Atmospheric Science
IBM Thomas J. Watson Research Center
Yorktown Heights, NY, USA
`lloydt@us.ibm.com`

16 November 2023



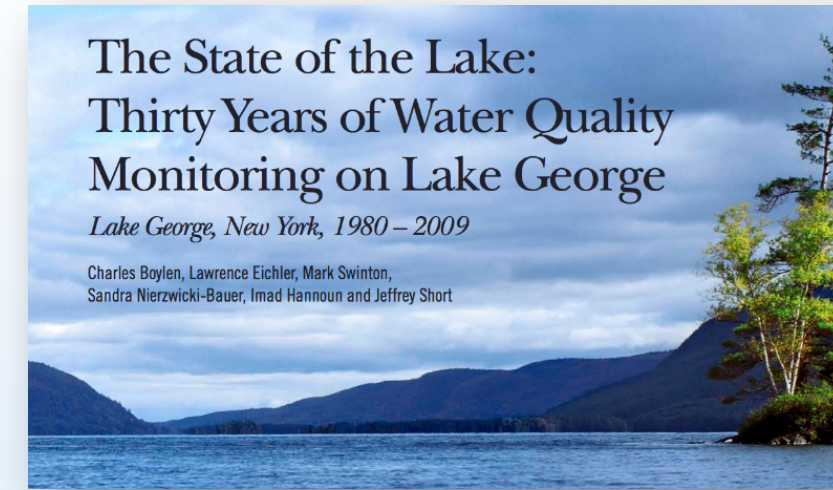
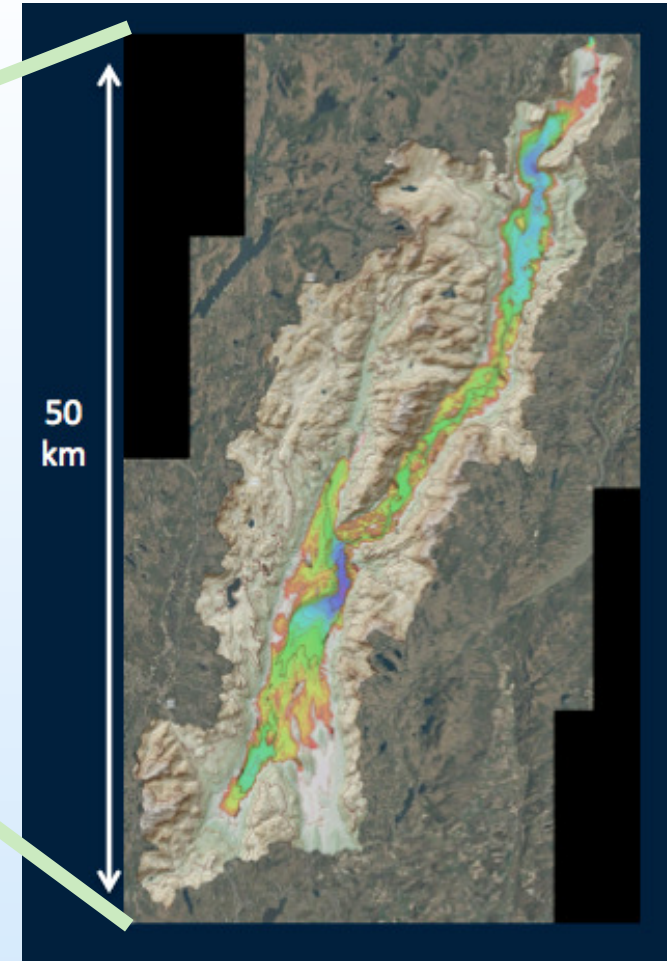
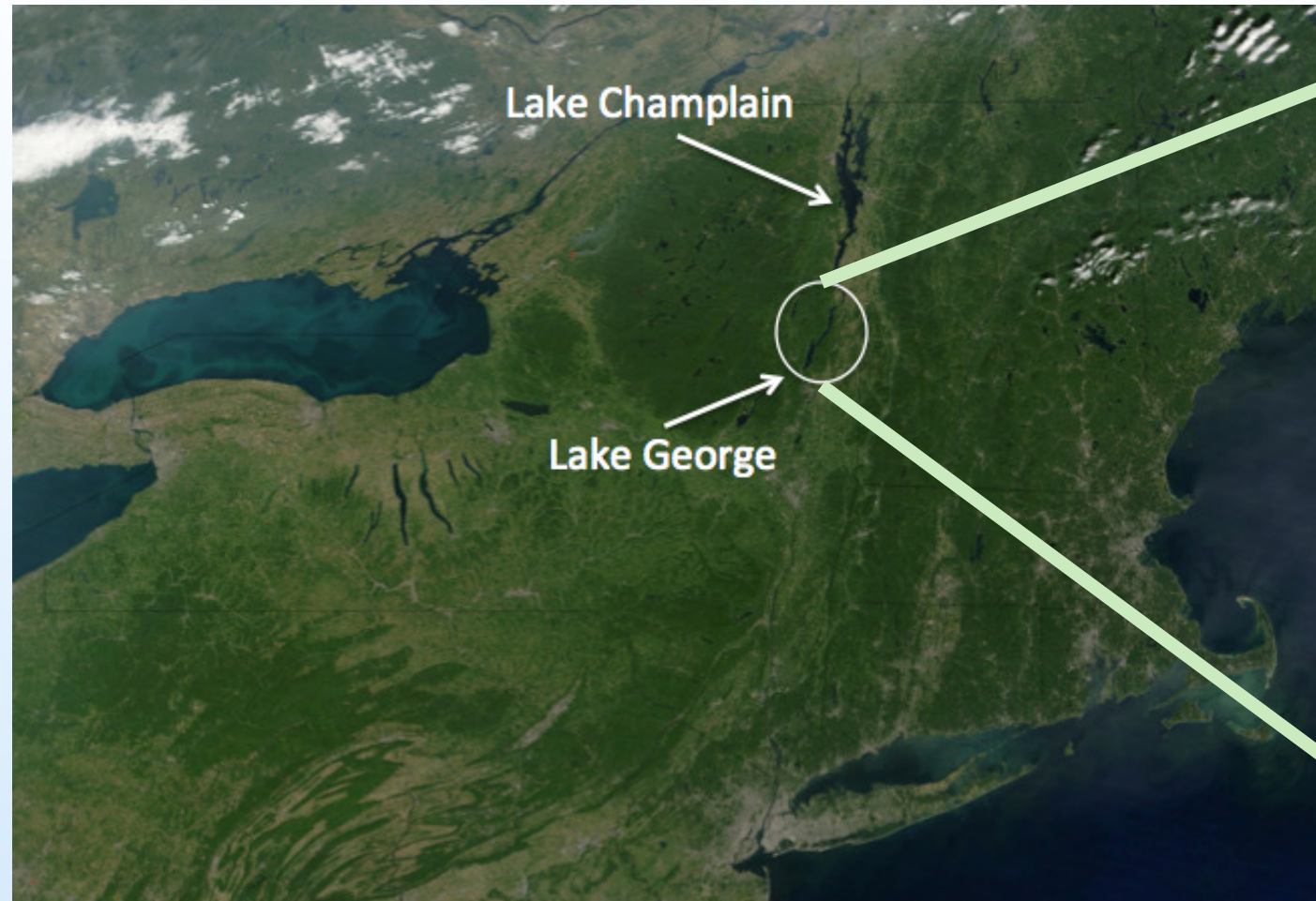
Urban Heat Islands, Heat Waves, Urban Ecosystems, Water Stress, Severe Weather, Climate Change, ...



- Environmental monitoring and prediction
- Adaptive modeling and visualization of the impact of climate change and urbanization
- Targeted weather forecasting for extreme events

Each topic will show the use of both open source software (environmental modeling and data management, visualization and analysis) and open data sets.

Environmental Monitoring and Prediction



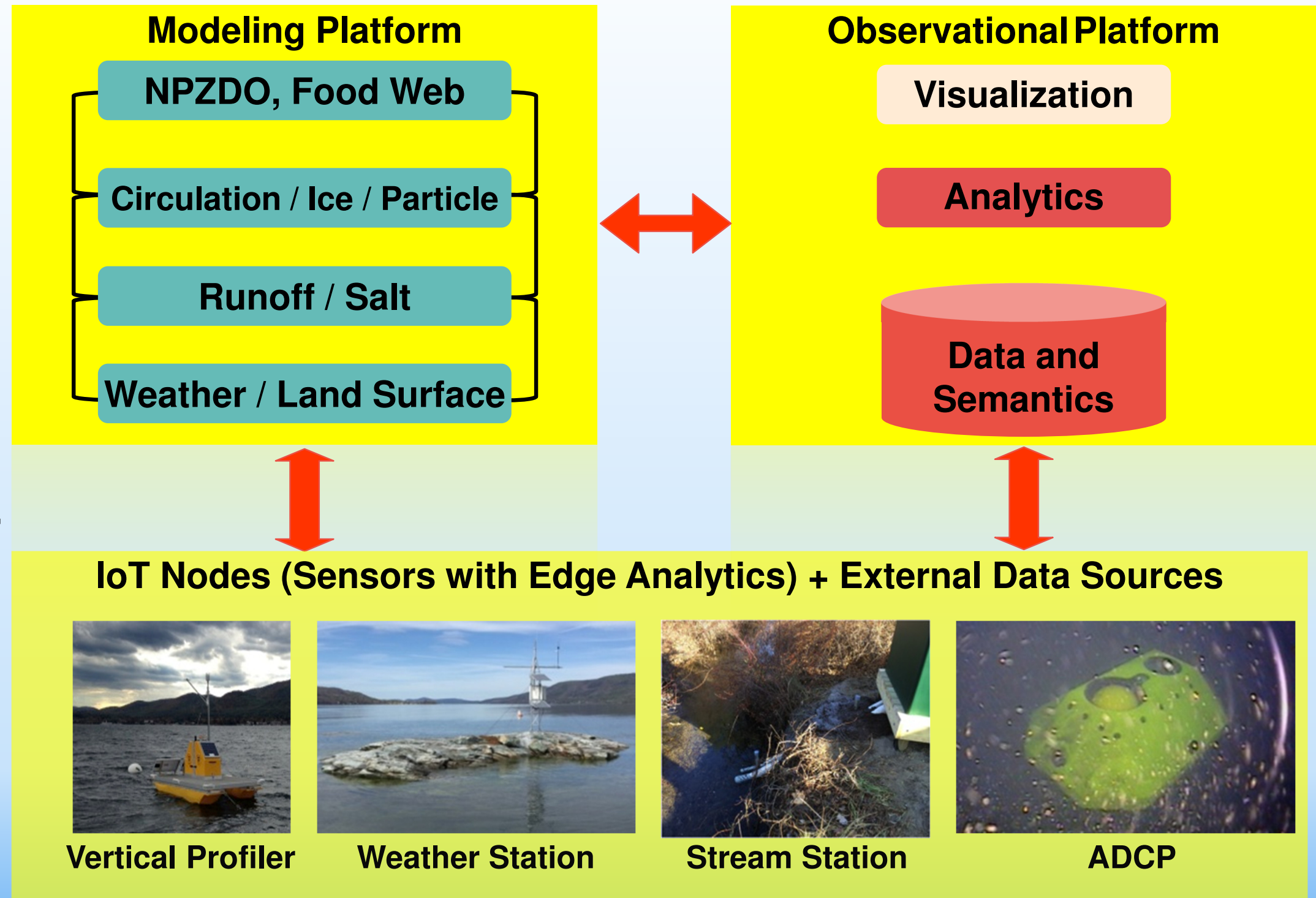
30-year longitudinal study highlighted three environmental stressors

- Road salt (NaCl)
- Nutrient loading via storm water runoff
- Invasive species

- One of the world's most pristine natural ecosystems with \$1B annual tourism industry
- Understand and manage the impact from road salt, storm water runoff and invasive species
- Understand the effects of climate change and the risk for harmful algal blooms
- Make the lake smarter with a coupled observational and modeling platform

<https://jeffersonproject.rpi.edu>

Real-time, event-driven,
distributed, intelligent...



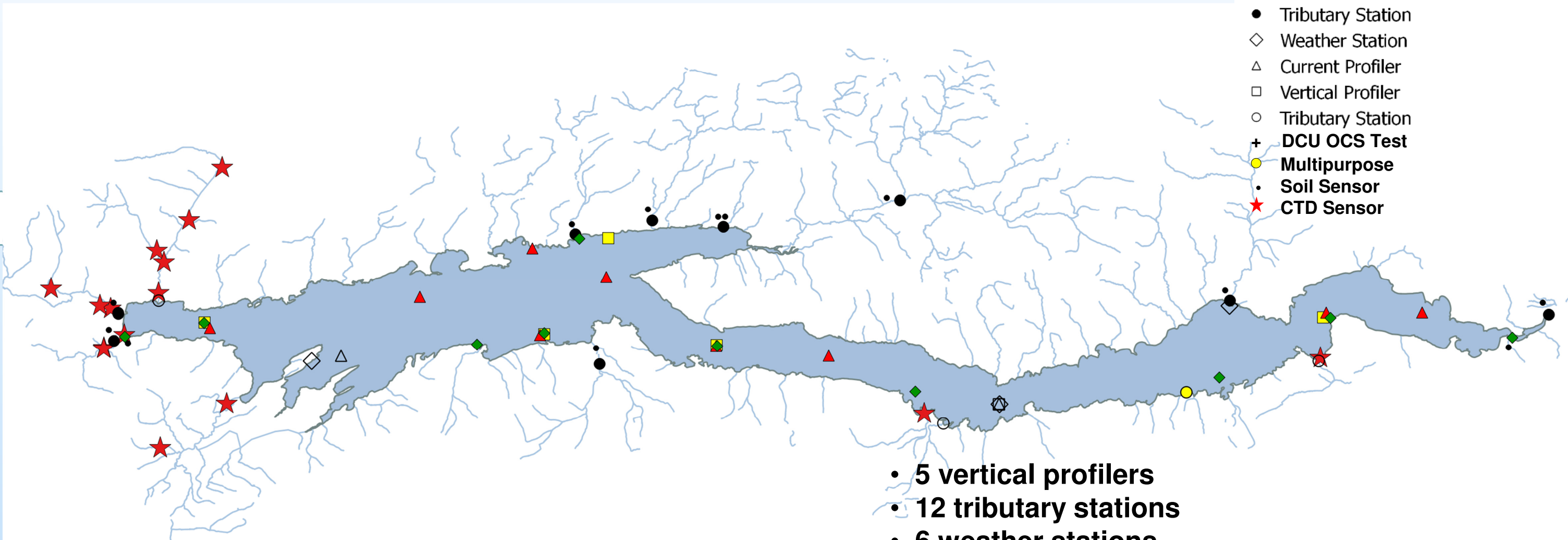
Current Observing System at Lake George



Legend

ModifiedSensorLocations

- ◆ Weather Station
- ▲ Current Profiler
- Vertical Profiler
- Tributary Station
- ◇ Weather Station
- △ Current Profiler
- Vertical Profiler
- Tributary Station
- + DCU OCS Test
- Multipurpose
- Soil Sensor
- ★ CTD Sensor



~50 sensor sites and >500 sensors...

- 5 vertical profilers
- 12 tributary stations
- 6 weather stations
- 1.35B point observations
- >3M CTD observations
- >75GB from ADCPs



Vertical Profilers

- Water quality through entire column (EXO2)
- ADCP (Acoustic Doppler Current Profiler)
- Meteorological sensors



Tributary Stations

- Water quality in streams (EXO2)
- Stream flow
- Soil moisture and conductivity
- Accessory sensors and technology



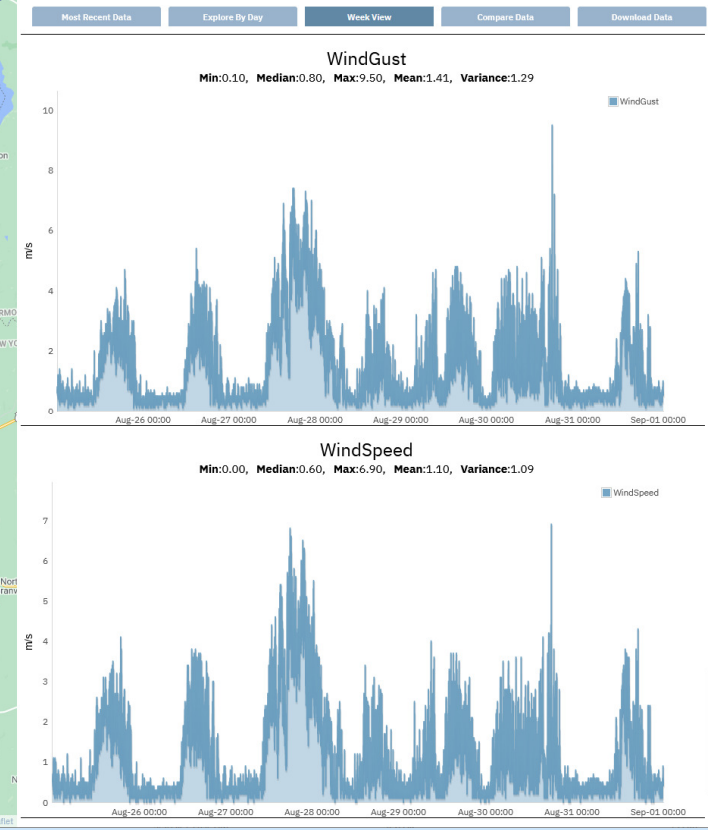
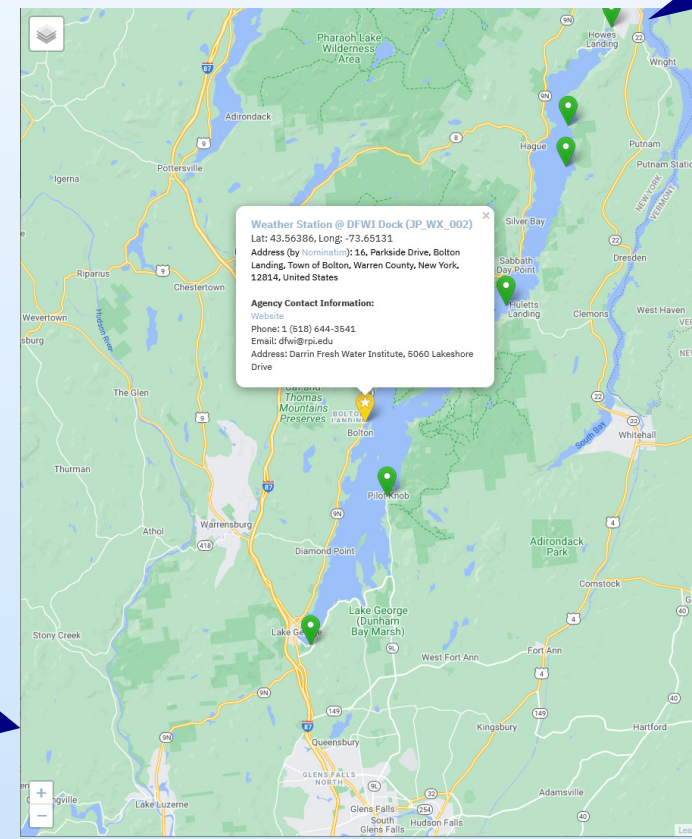
Weather Stations

- Wind speed and direction
- Air temperature and humidity
- Precipitation
- Solar radiation

Observing System Platform

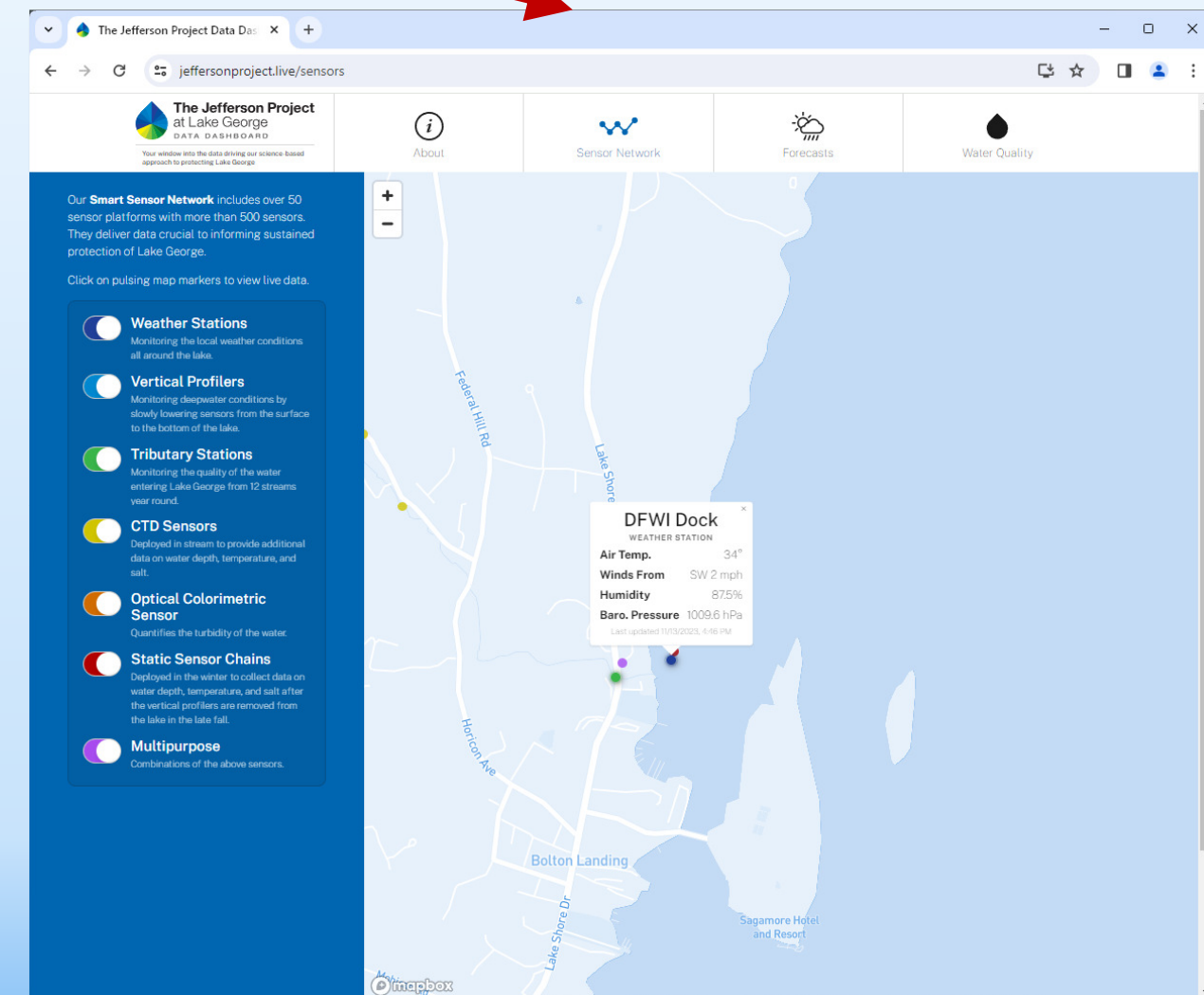
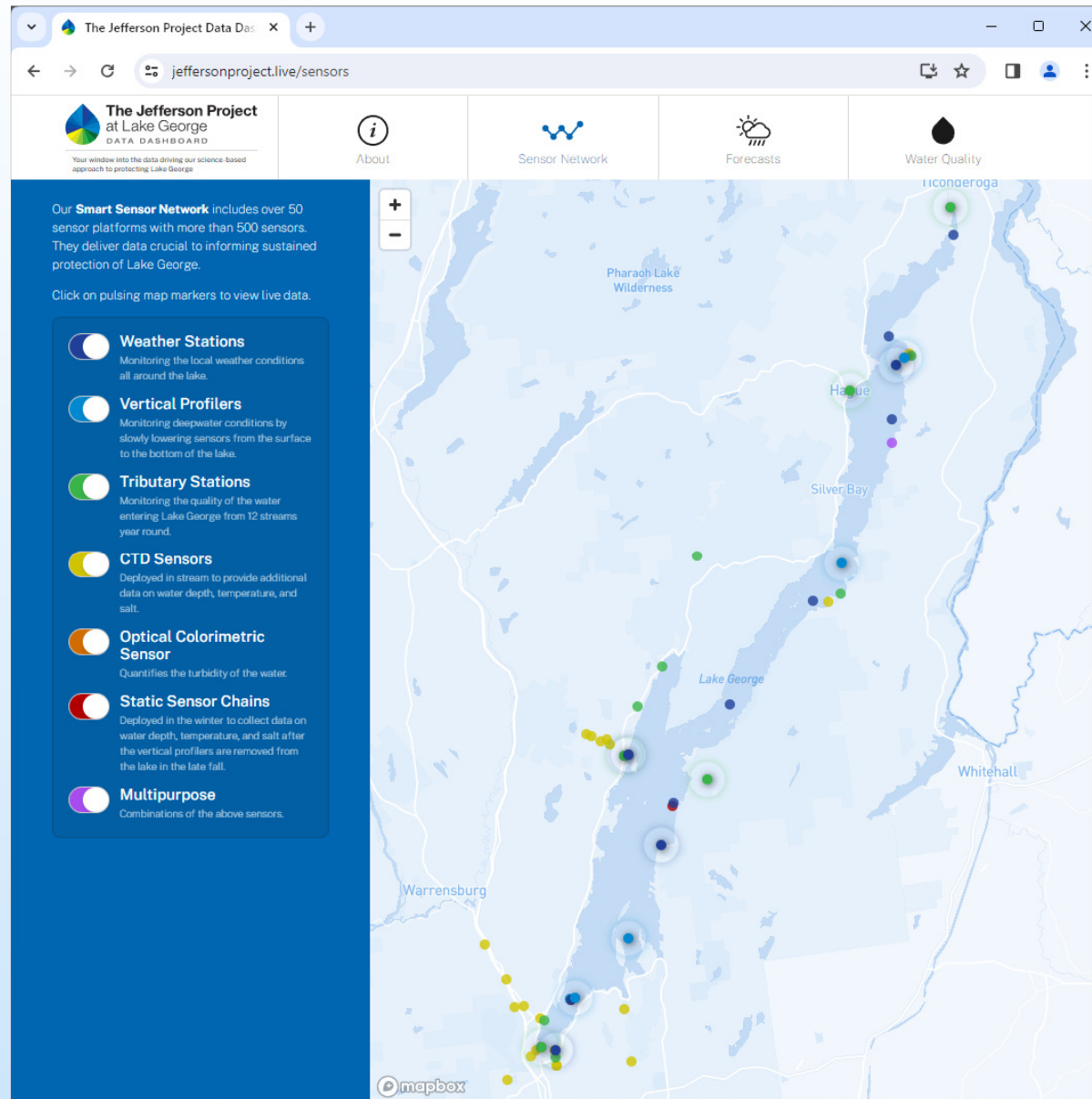
DiamondPoint IoT Edge Node

AssemblyPoint Data and Analytics Platform



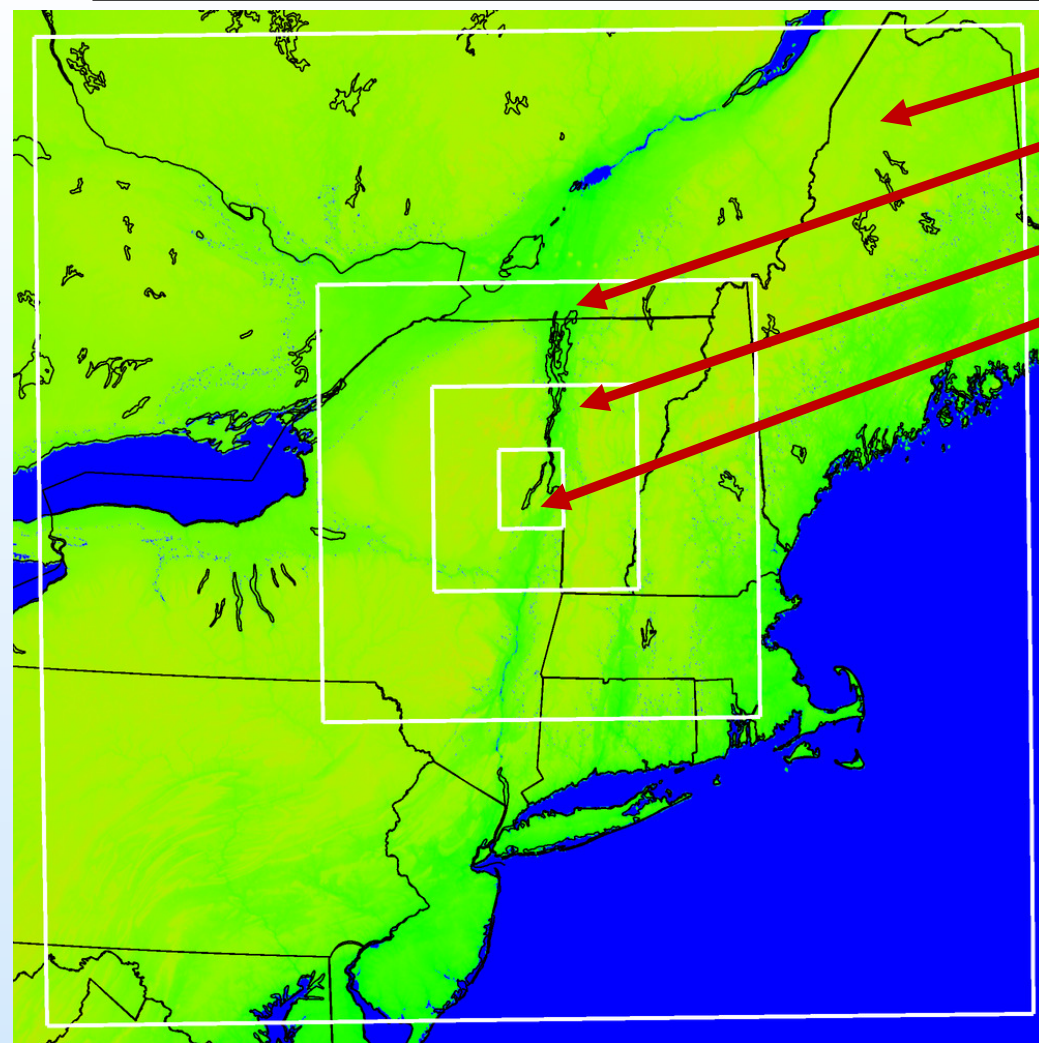
Open Access to Sensor Data for Lake George

Custom data dashboard provided by the Lake George Association



Live access to data via IBM AssemblyPoint
<https://jeffersonproject.live/sensors>

Operational (Production) Coupled Modelling for Lake George IBM



Weather (Deep Thunder™)

[based upon WRF-ARW]

<https://www.mmm.ucar.edu/models/wrf>

Weather resolutions
9km
3km

1km
0.333km

Hydrology resolution

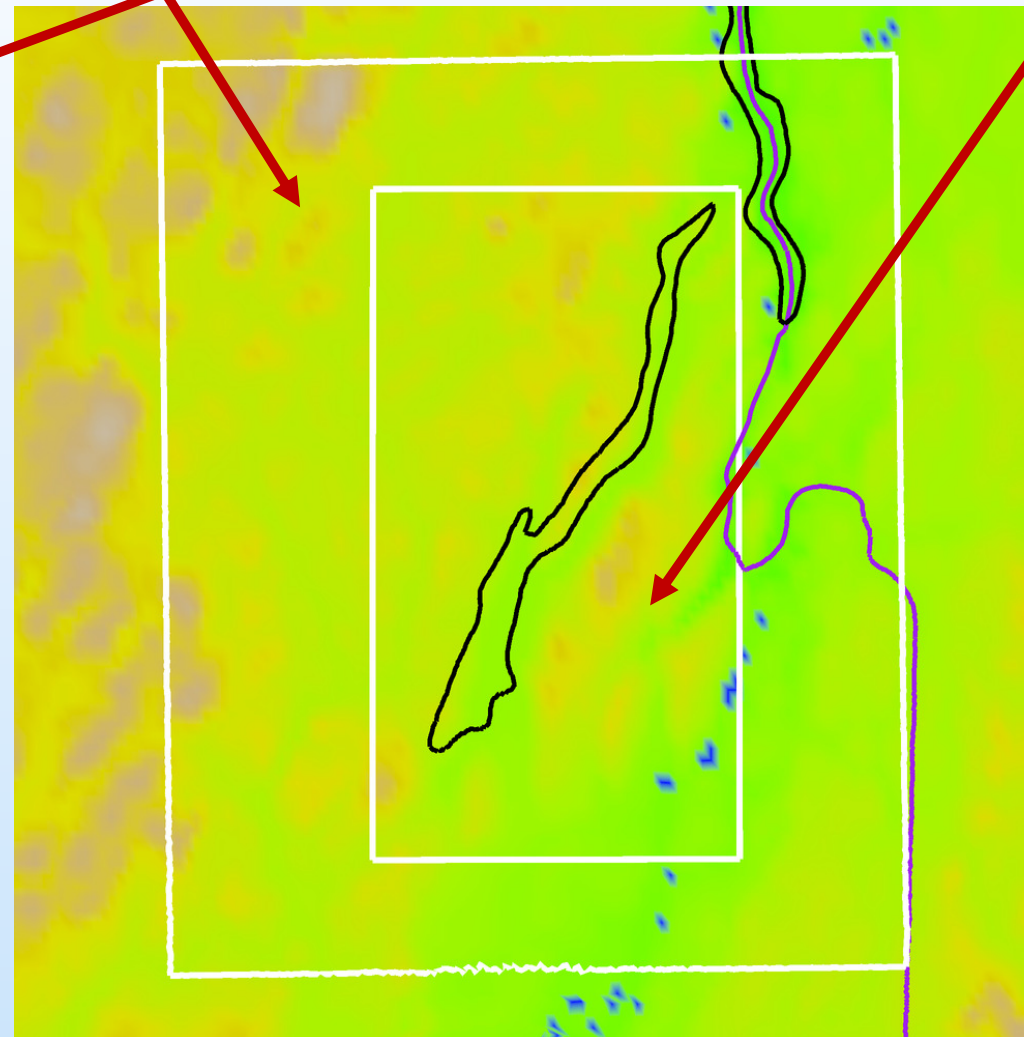
41.7m

**Hydrodynamics
(Deep Circulation)**

[based upon SUNTANS]

<https://sourceforge.net>

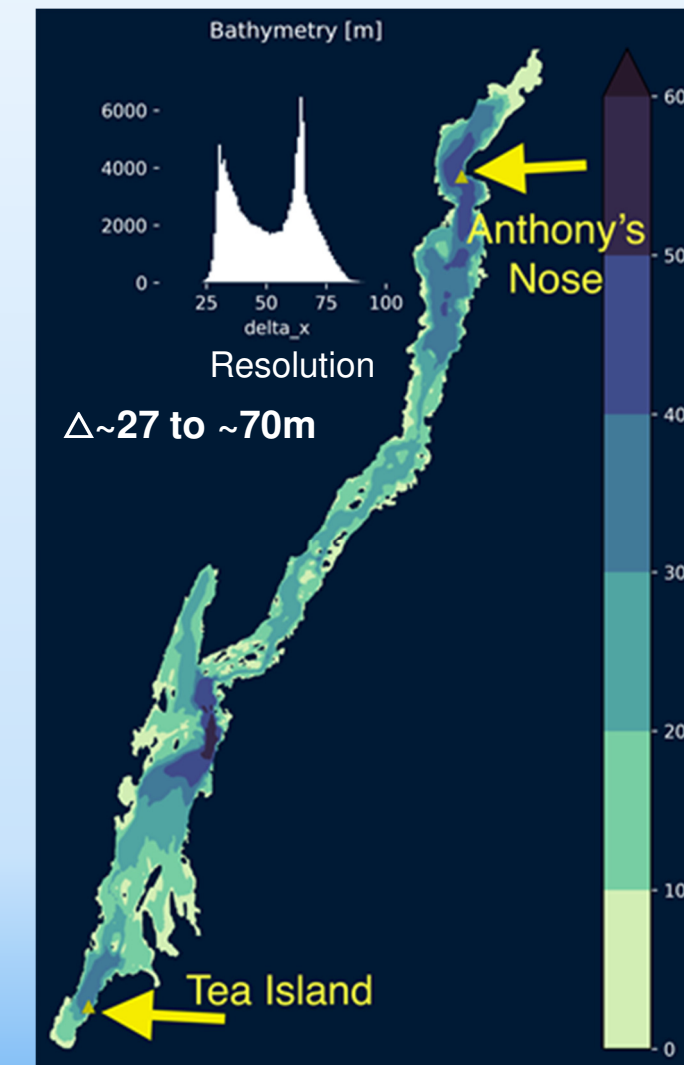
[/p/suntans/wiki/Home/](https://sourceforge.net/p/suntans/wiki/Home/)



**Weather (Deep Thunder™) and
Hydrology (Deep Runoff)**

[based upon WRF-Hydro]

https://ral.ucar.edu/projects/wrf_hydro/model-code

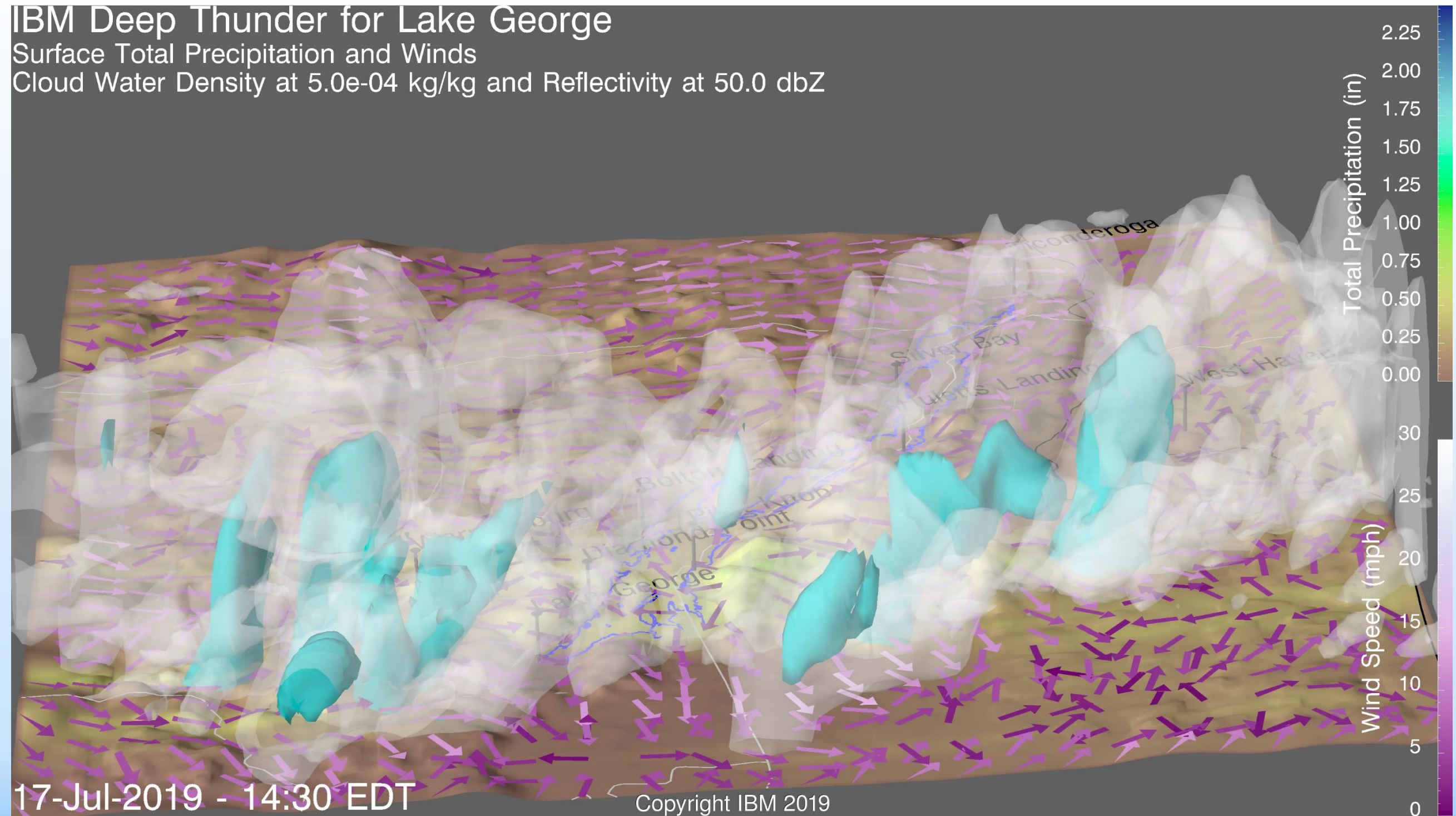


Example Targeted Weather Forecast for Lake George, NY

IBM Deep Thunder for Lake George

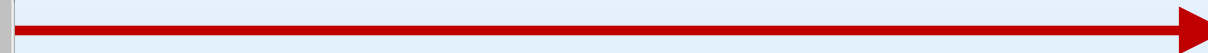
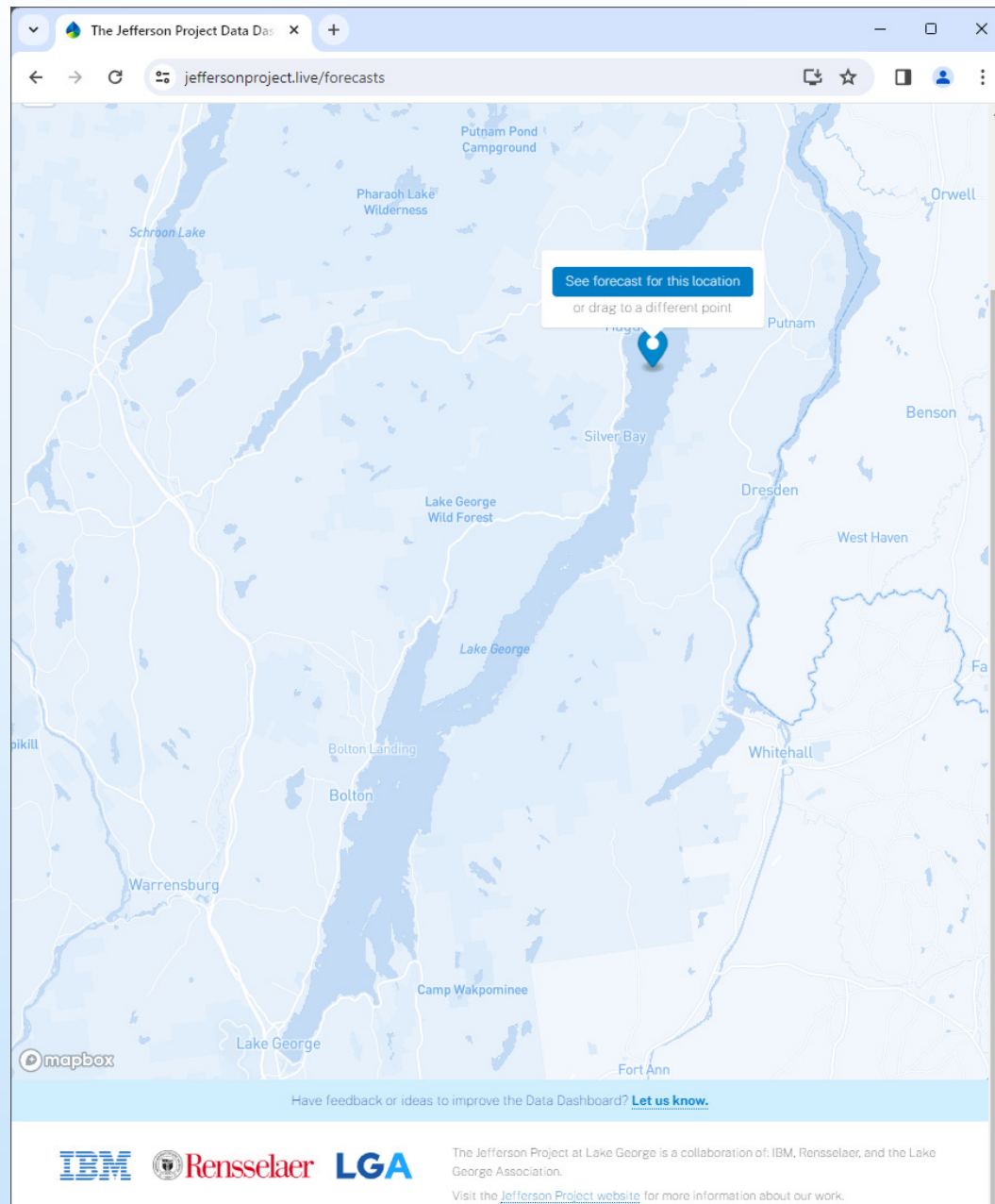
Surface Total Precipitation and Winds

Cloud Water Density at $5.0e-04$ kg/kg and Reflectivity at 50.0 dbZ

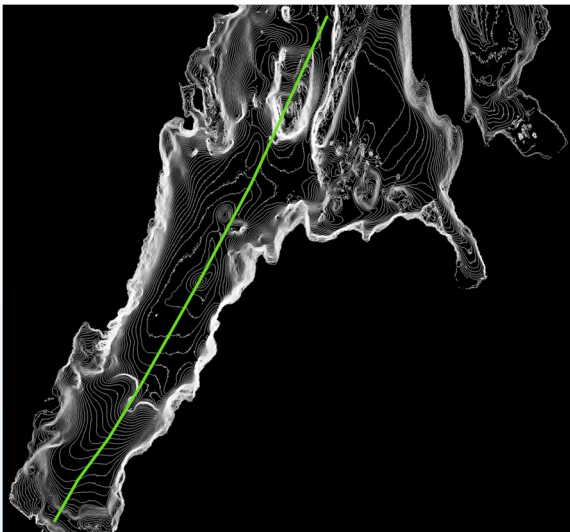


Predicted Clouds, Thunderstorm Cells, Wind and Precipitation – 333m horizontal resolution

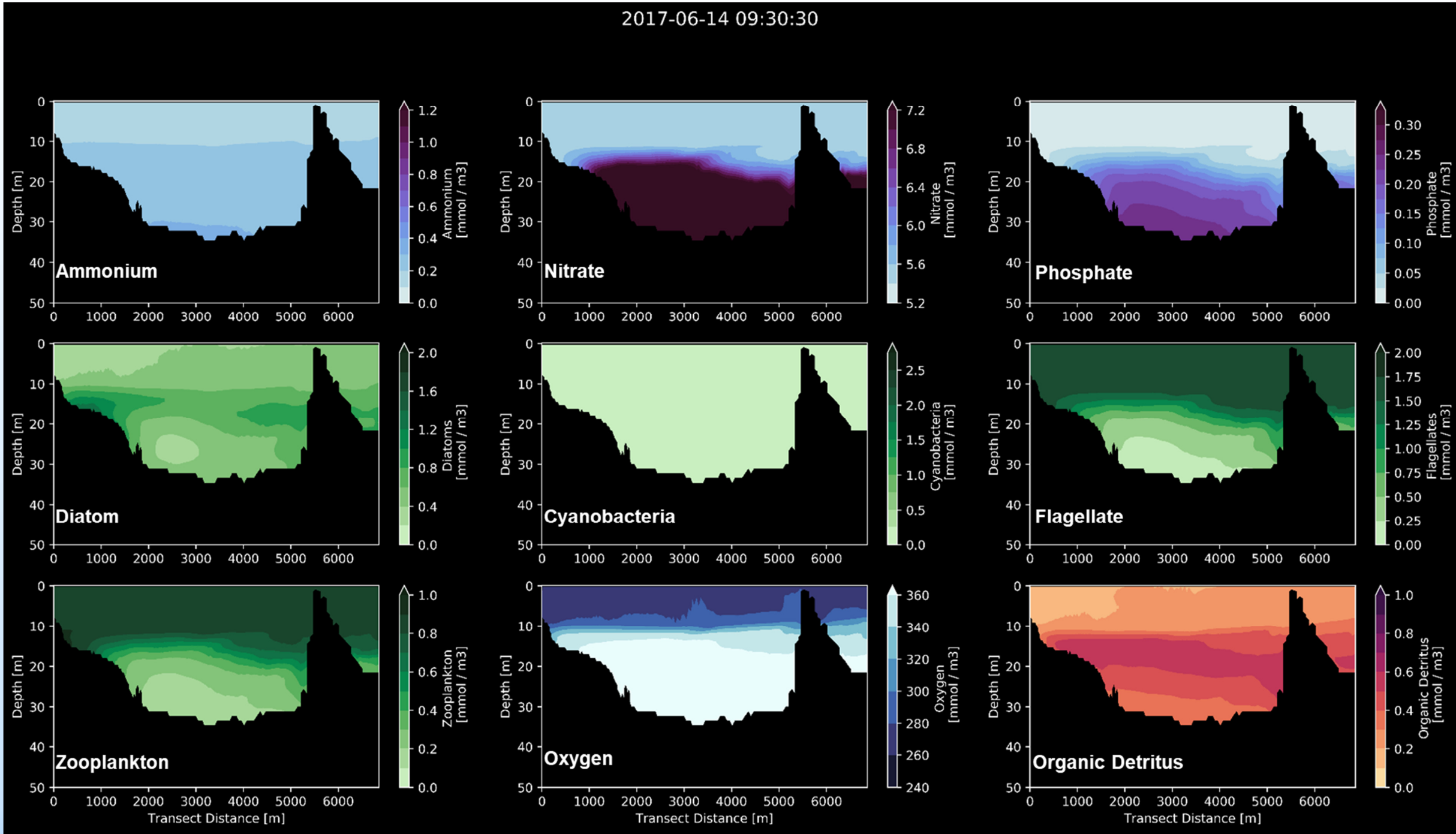
Open Access to Weather Forecast Data for Lake George



- Custom data dashboard provided by the Lake George Association
- Live and open access to data
<https://jeffersonproject.live/forecasts>



**Example
Ecosystem
Simulation Result
across a Transect
in the South
Basin in
Lake George**



Multiple Visualizations Using the Unity Gaming Engine



000000 00 00 20 4000000
Number of tiles 127
200.11 Mb memory in use

Layers
(Esc will access menu)

☒ Weather

☒ Lake Surface Tracks

☒ Cross Sections

☒ LIDAR

Terrain Resolution

50 cm (default)

Jump To

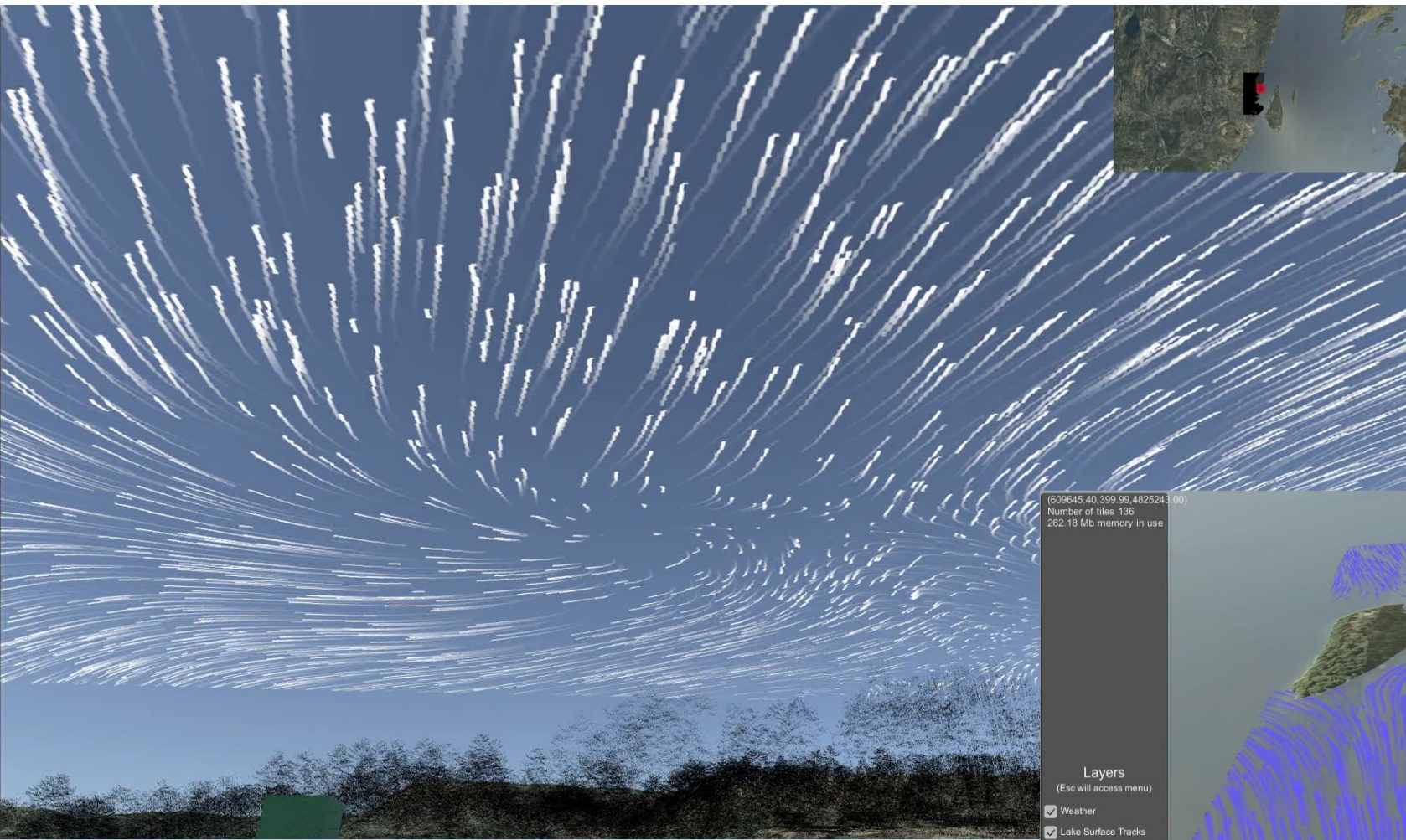
DFWI

Change Height

Lake Surface (default)

Resume

Quit



(609645.40, 399.99, 4825243.00)
Number of tiles 136
262.18 Mb memory in use

Layers
(Esc will access menu)

☒ Weather

☒ Lake Surface Tracks

☒ Cross Sections

☒ LIDAR

Terrain Resolution

50 cm (default)

Jump To

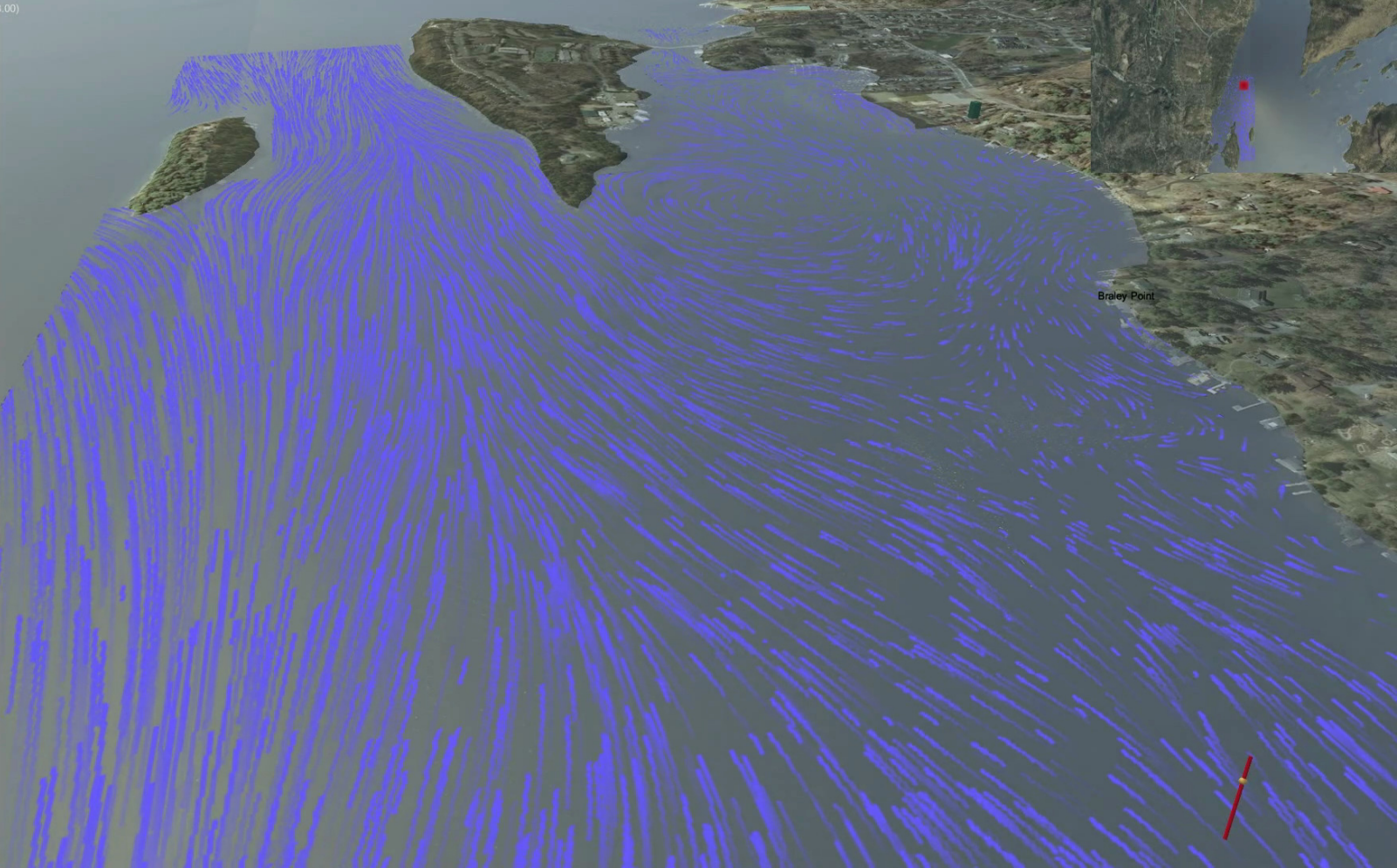
DFWI

Change Height

Lake Surface (default)

Resume

Quit



Adaptive Modeling and Visualization of the Impact of Climate Change and Urbanization

- **Creation of new urban integrated field laboratories that target biogeochemical, energy and water flows between urban areas and surrounding ecosystems**
 - **Four, five-year awards, totaling \$94M, involving over 25 institutions**
- **Exploratory, multi-disciplinary research to understand “climate and environmental predictability” where most people live**
- **Intersection of physics, infrastructure and society towards informing “sustainable, resilient, and equitable solutions”**
- **Enable research that combines novel observations with high-resolution urban modeling**
- **Quantify the benefits of equitable solutions that are applied to heterogeneous urban regions in addressing the impacts of climate change**
- **Provide opportunities to inspire, train and support leading scientists who have an appreciation for the global climate and energy challenges of the 21st century**

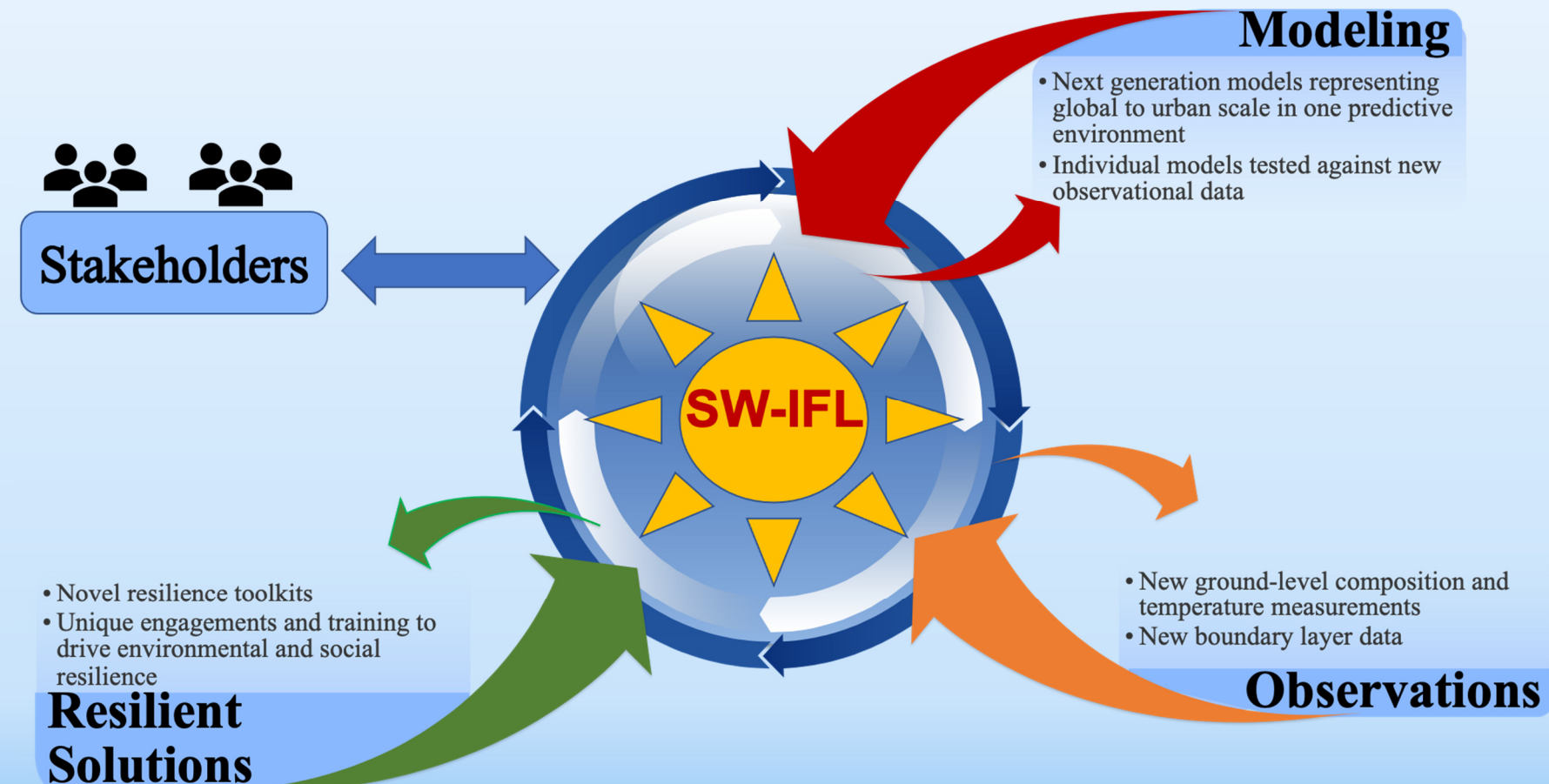
<https://ess.science.energy.gov/urban-ifls/>

Southwest Urban Corridor Integrated Field Laboratory (SW-IFL)

Adaptive modeling & visualization of the impact of climate change & urbanization

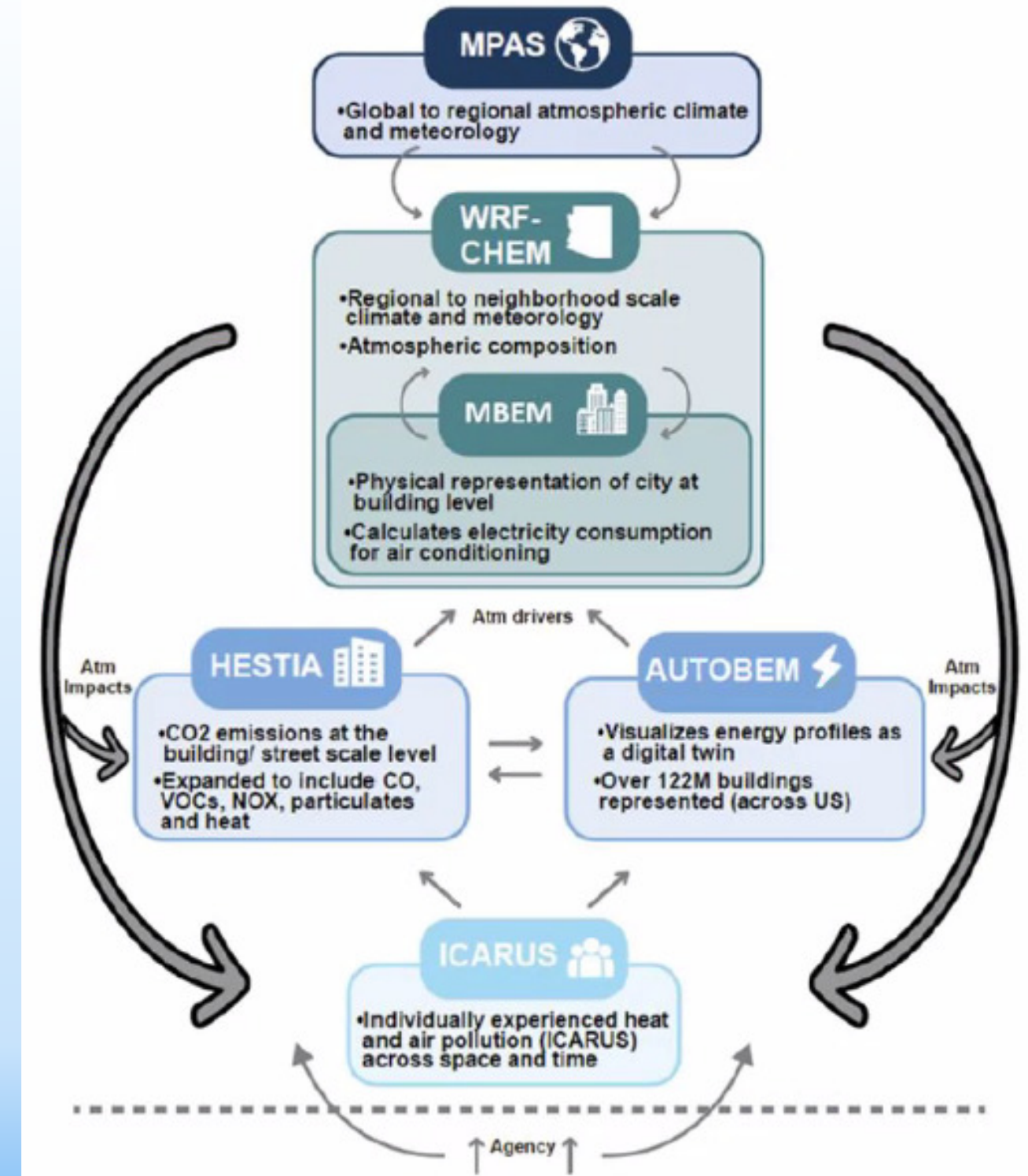
Novel observational and modeling capabilities for growing cities in the arid southwest urban corridor that promote equitable policy interventions to enhance resilience to the interconnected challenges of extreme heat, local-to-global emissions of harmful pollutants, and water use in the built environment

- Integrate observations, modeling and engagement
- Provide new knowledge address extreme heat and related challenges
- Empower government and public to respond
- Inform new policies and solutions that effective, equitable and generalizable



<https://sw-ifl.asu.edu/>

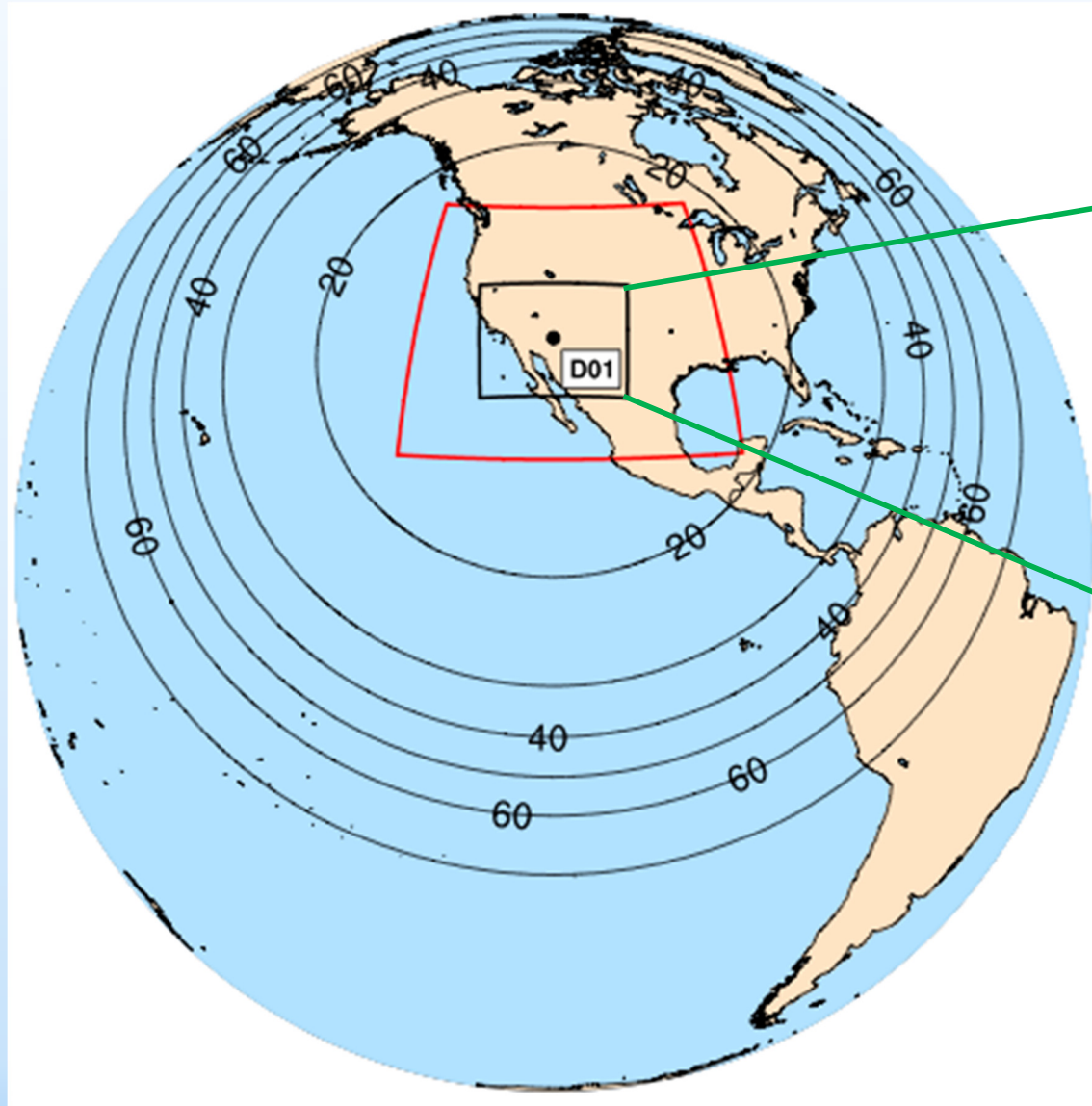
- Deploy models from global to human scales
- Integrate models of exposure, energy use, emissions, atmospheric dynamics and chemistry
- Couple with observations to improve and validate
- Assess disproportional impacts of heat and water stress on disadvantaged communities
- Develop visualizations to communicate, analyze and inform
 - Explain model uncertainty
 - Enable more effective decision support with stakeholders



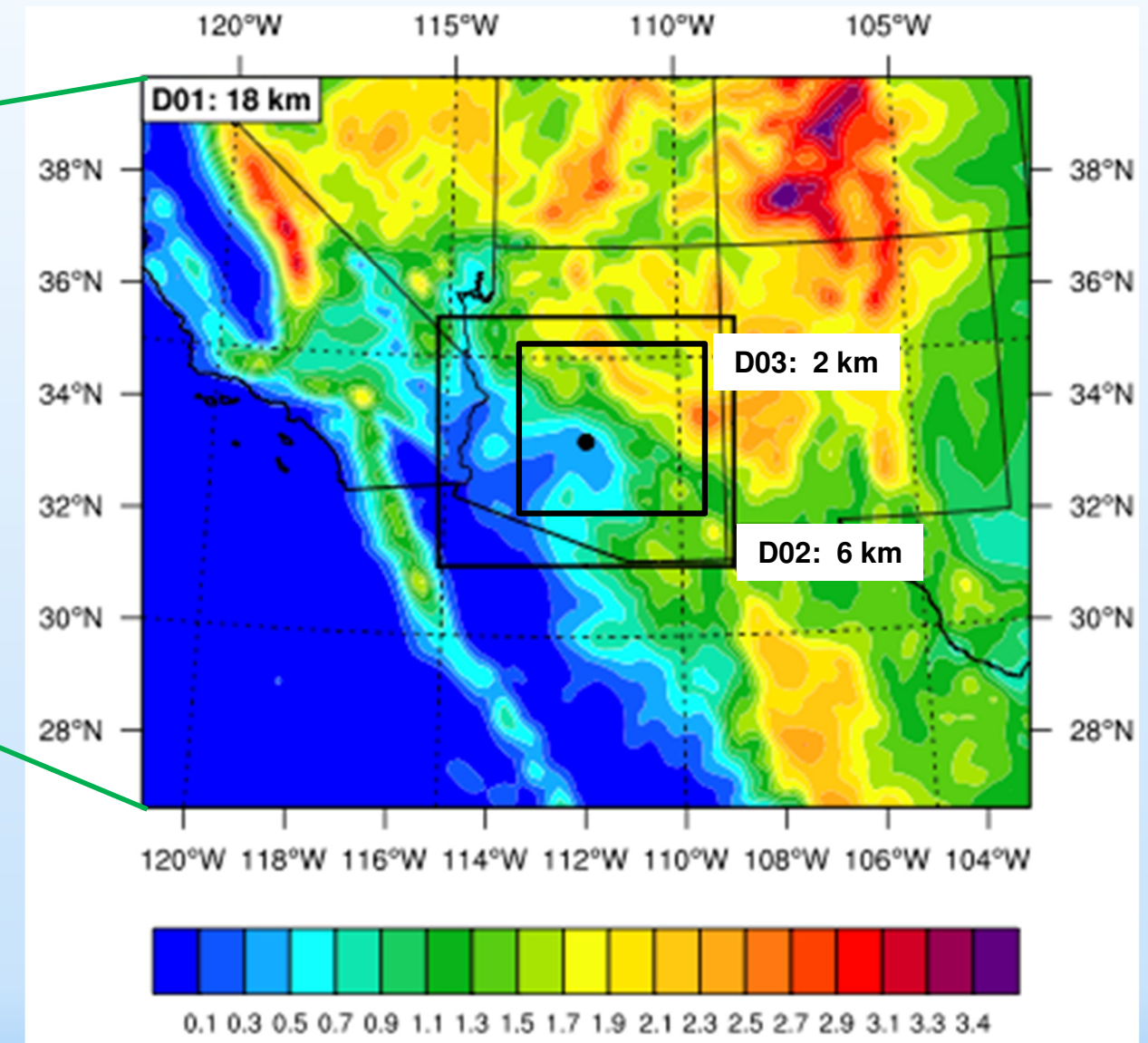
Example: Coupling Global to Regional Scales

Adaptive modelling for global downscaling by coupling MPAS and WRF

MPAS mesh regions resolution (92 to 25km)
and WRF domain



WRF domain (18, 6 and 2km) focused on
Phoenix, AZ



This approach to downscaling reduces computational and storage requirements while increasing the fidelity of the results

Targeted Weather Forecasting for Extreme Events

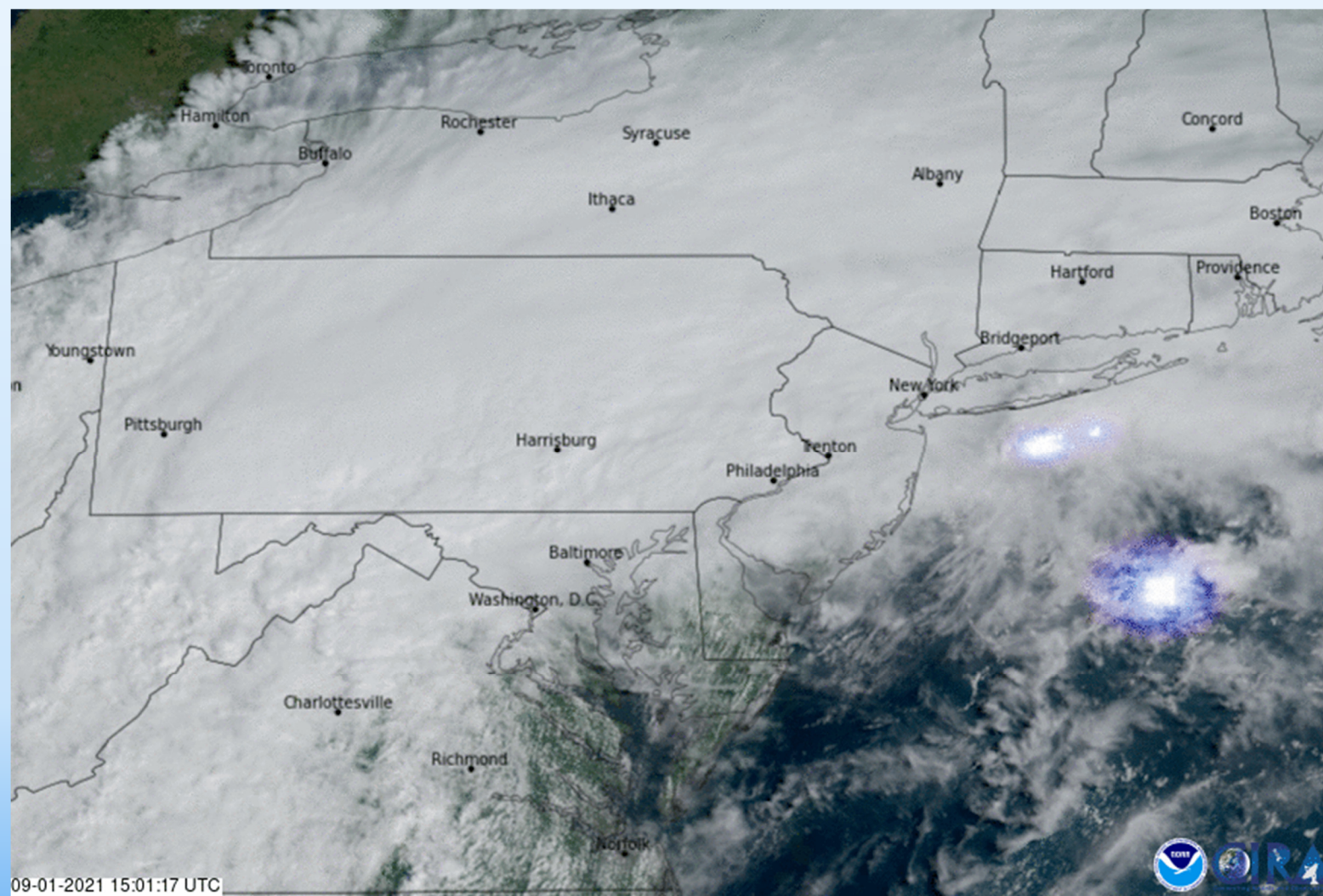
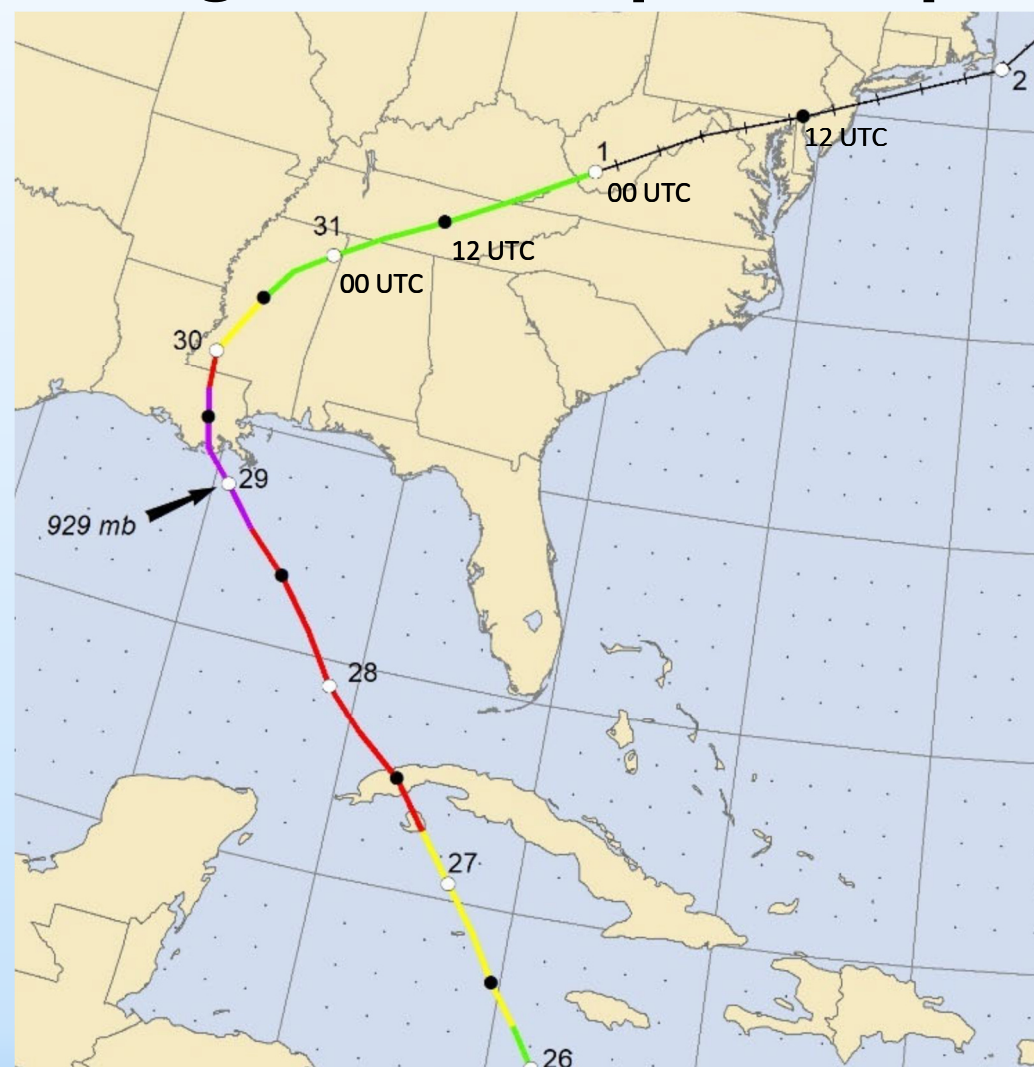
Predictability of the Precipitation from the Remnants of Hurricane Ida in New York City



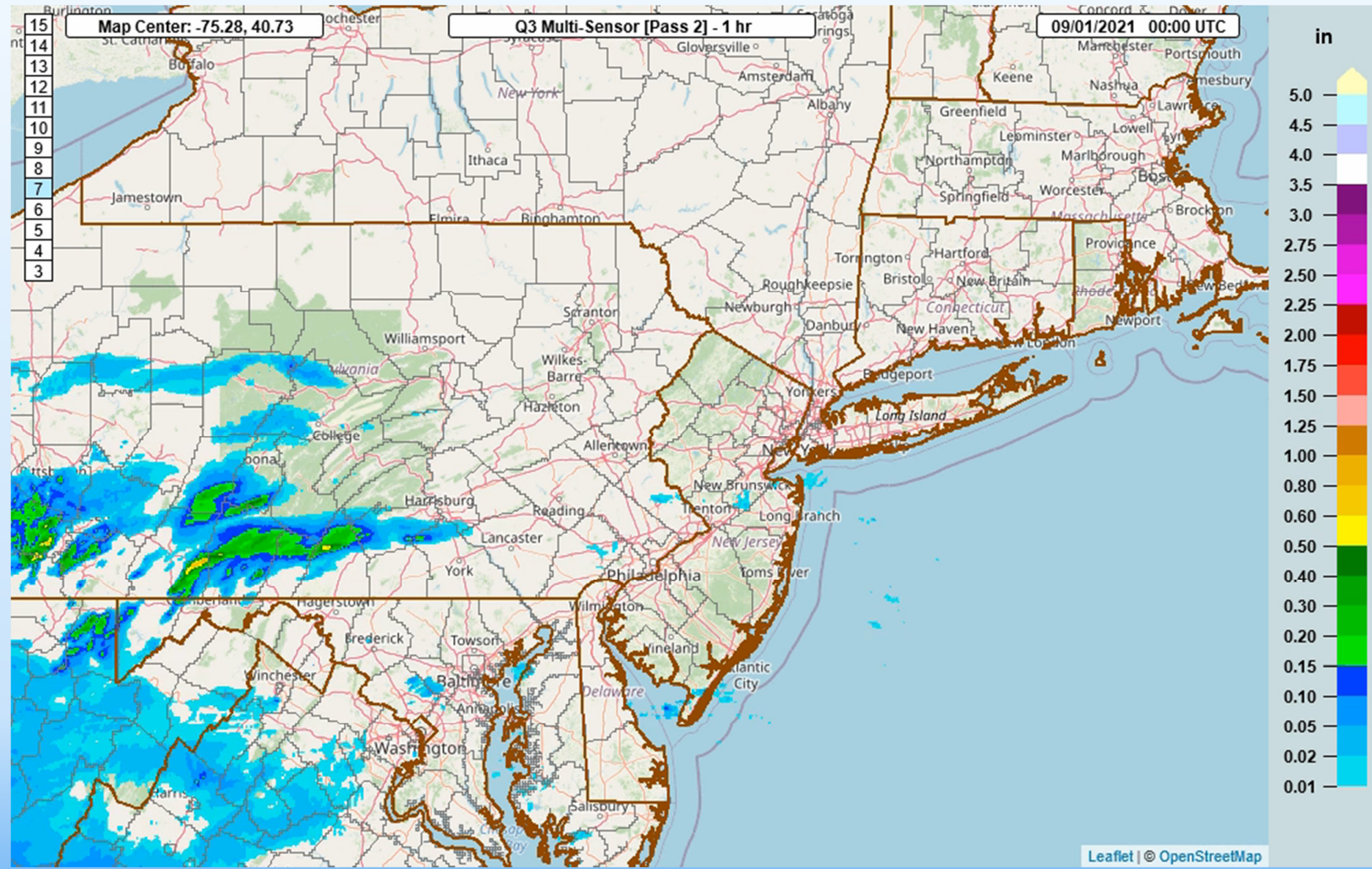
- **Hurricane “Ida” and its remnants -- case study**
- **NWP configuration in the New York City metropolitan area**
- **Experimental results**

Post-Tropical Depression Ida – NYC / 1 September 2021

- Formed as a tropical depression southwest of Kingston, Jamaica on 26 August
- Landfall as a Category 4 hurricane in Louisiana on 29 August
- Degraded to a post-tropical depression by 1 September in West Virginia
- Interacted with a frontal zone, which then impacted the greater NYC metropolitan area



- **Record-breaking rainfall -- over 7" in only a few hours**
 - One-minute rainfall rates as high as 0.08 to 0.12 inches/minute
 - 9.64" at Staten Island
 - 7.19" at Central Park and 3.15" hourly accumulation



Impacts of Post-Tropical Depression Ida on NYC

- Unprecedented flooding: subway platforms and streets looking like rivers
- Over a dozen fatalities from drowning
- For the first time, the National Weather Service issued a flash flood emergency for New York City, “*This is a life-threatening situation.*” “*Seek higher ground now!*”

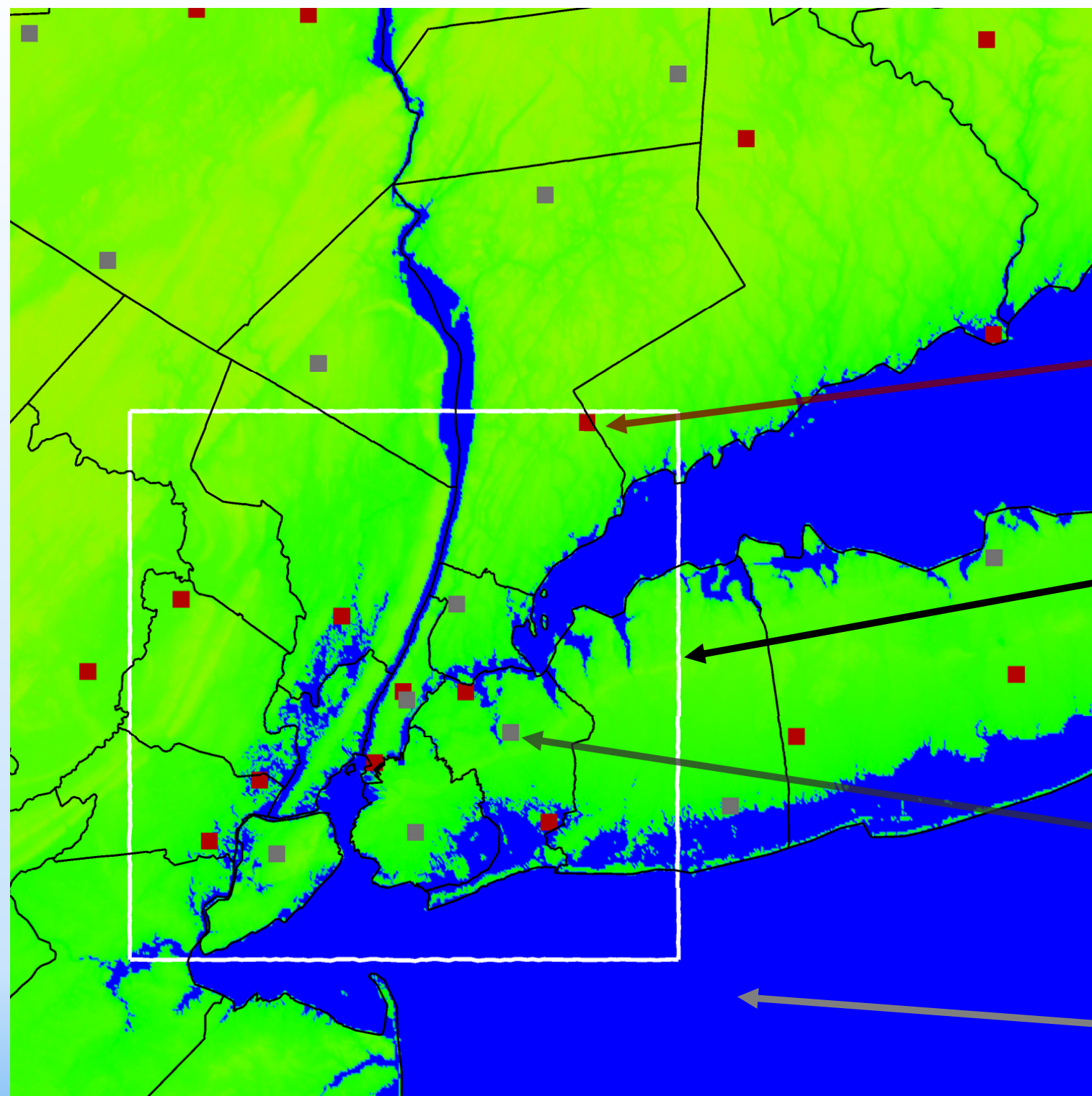


Despite such clear communications many of the impacts may have been avoided with longer prediction lead times.

We examine the potential of this idea with targeted 72-hour models initiated up to a couple of days before the event.

This includes data denial experiments with and without data from the New York State Mesonet (NYSM).

Targeted Weather Model for New York City



**Automated Surface
Observing Systems
(ASOS) site**

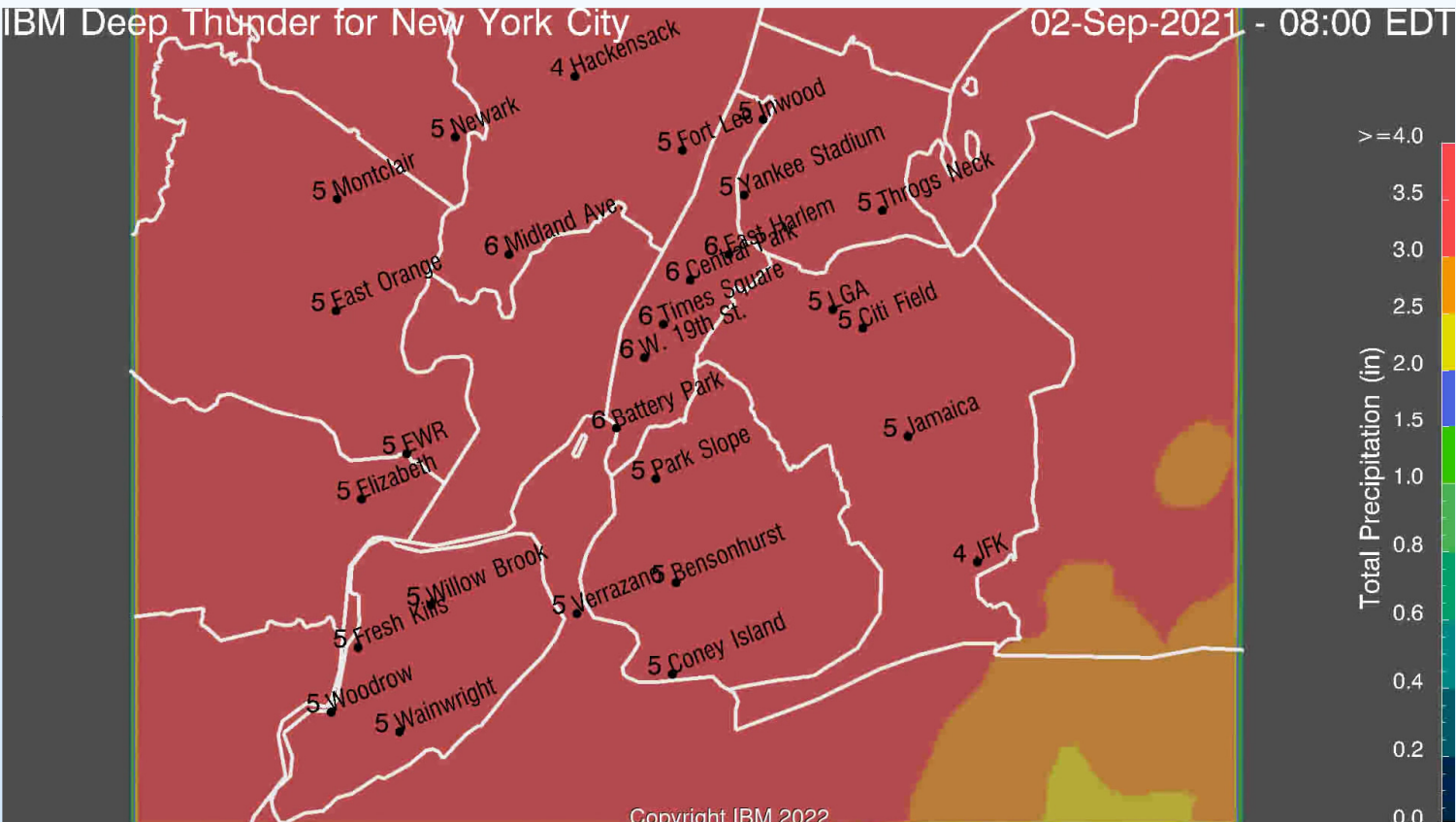
Boundary for 0.667km nest

**New York State
Mesonet (NYSM) site**

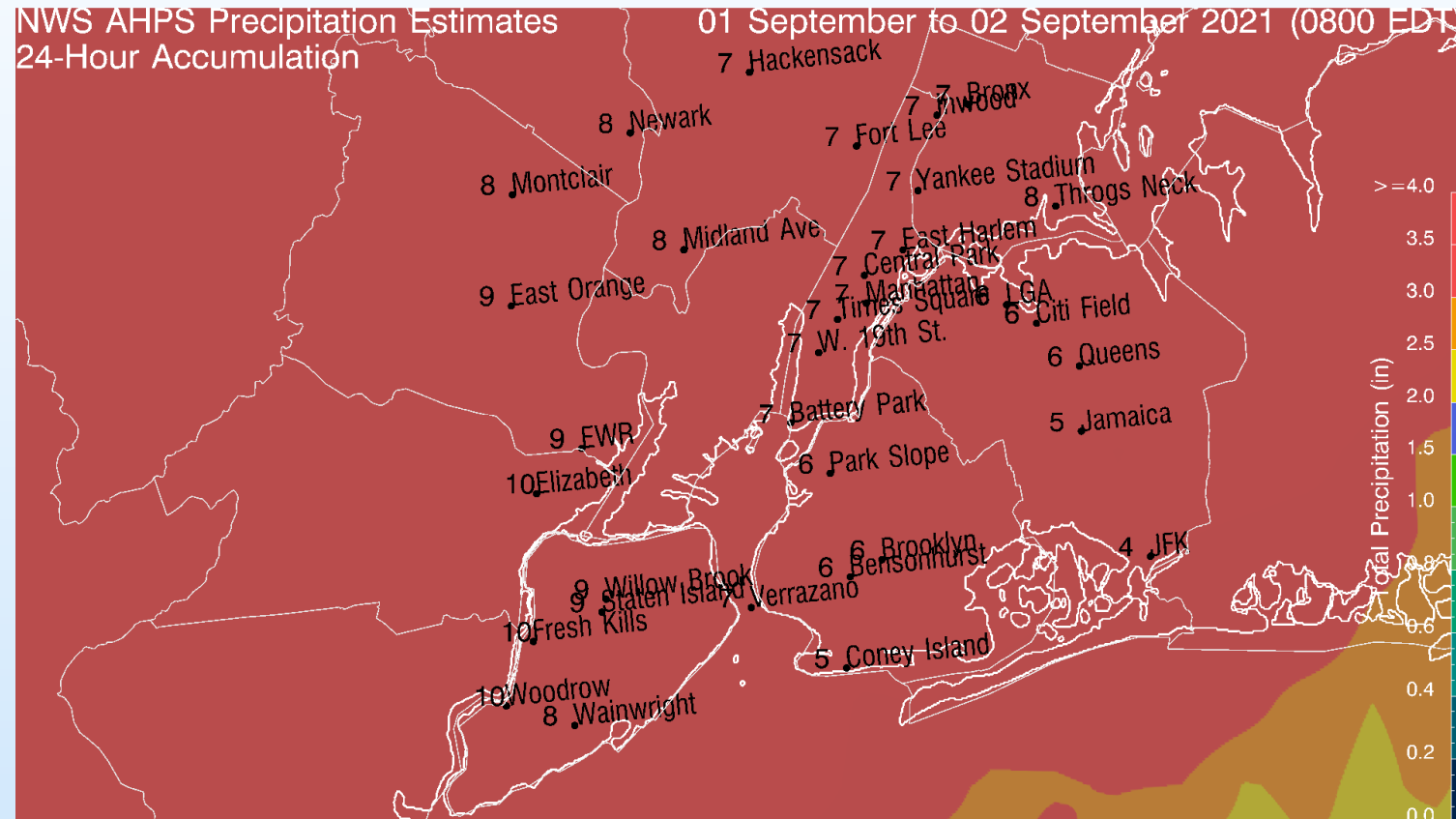
Portion of 2km nest

Accumulated Precipitation: Model vs. Observed

0800 EDT, 1 September to 0800 EDT, 2 September 2021



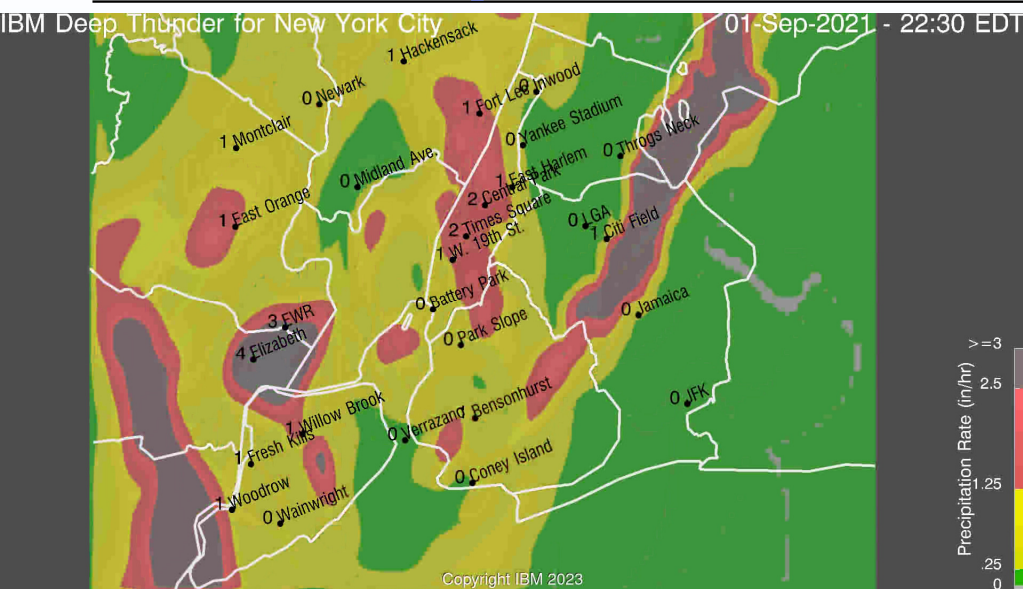
Model



Observed

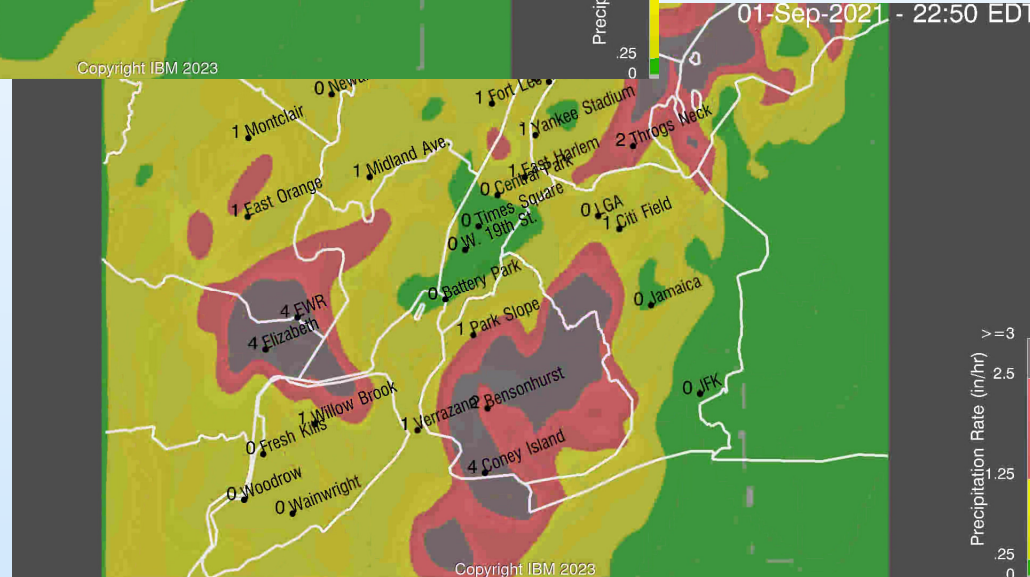
This looks pretty good in comparison to coarse-resolution data from NOAA but given the local impacts, the details need to be examined.

Precipitation Rate: Model vs. Observed: 1 September 2021

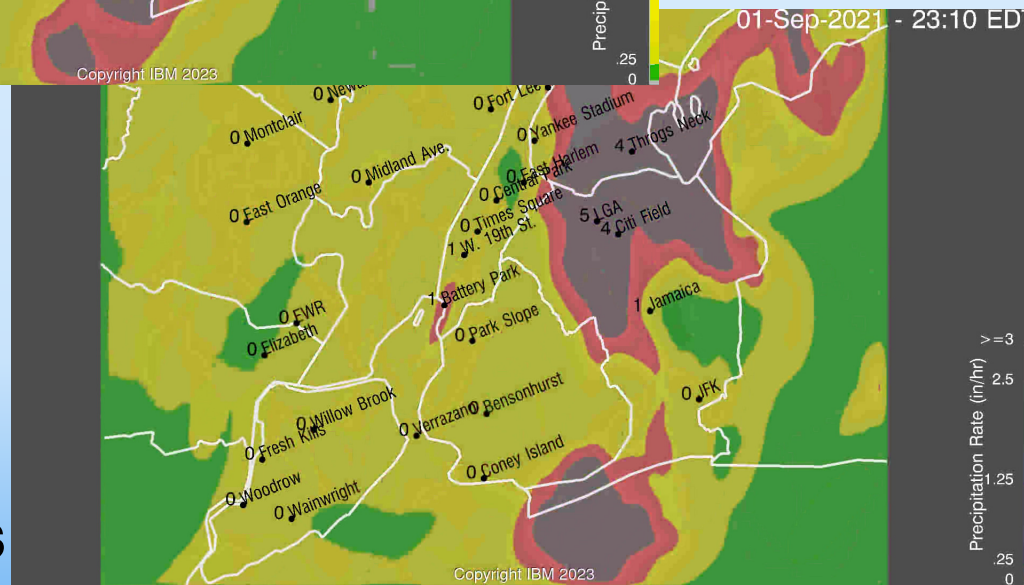


When modelled
instantaneous
rates were
highest

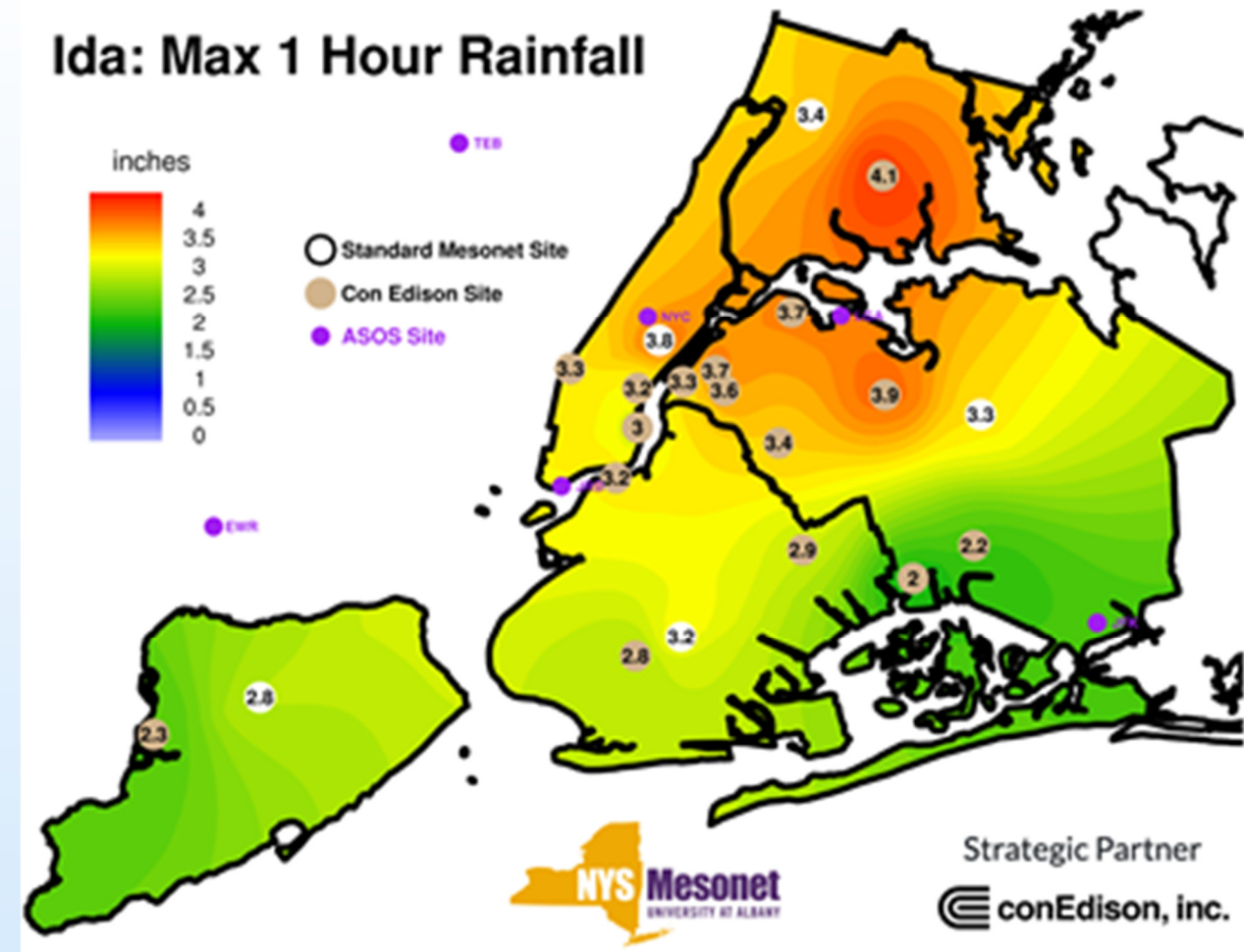
Manhattan



Brooklyn



Bronx and Queens

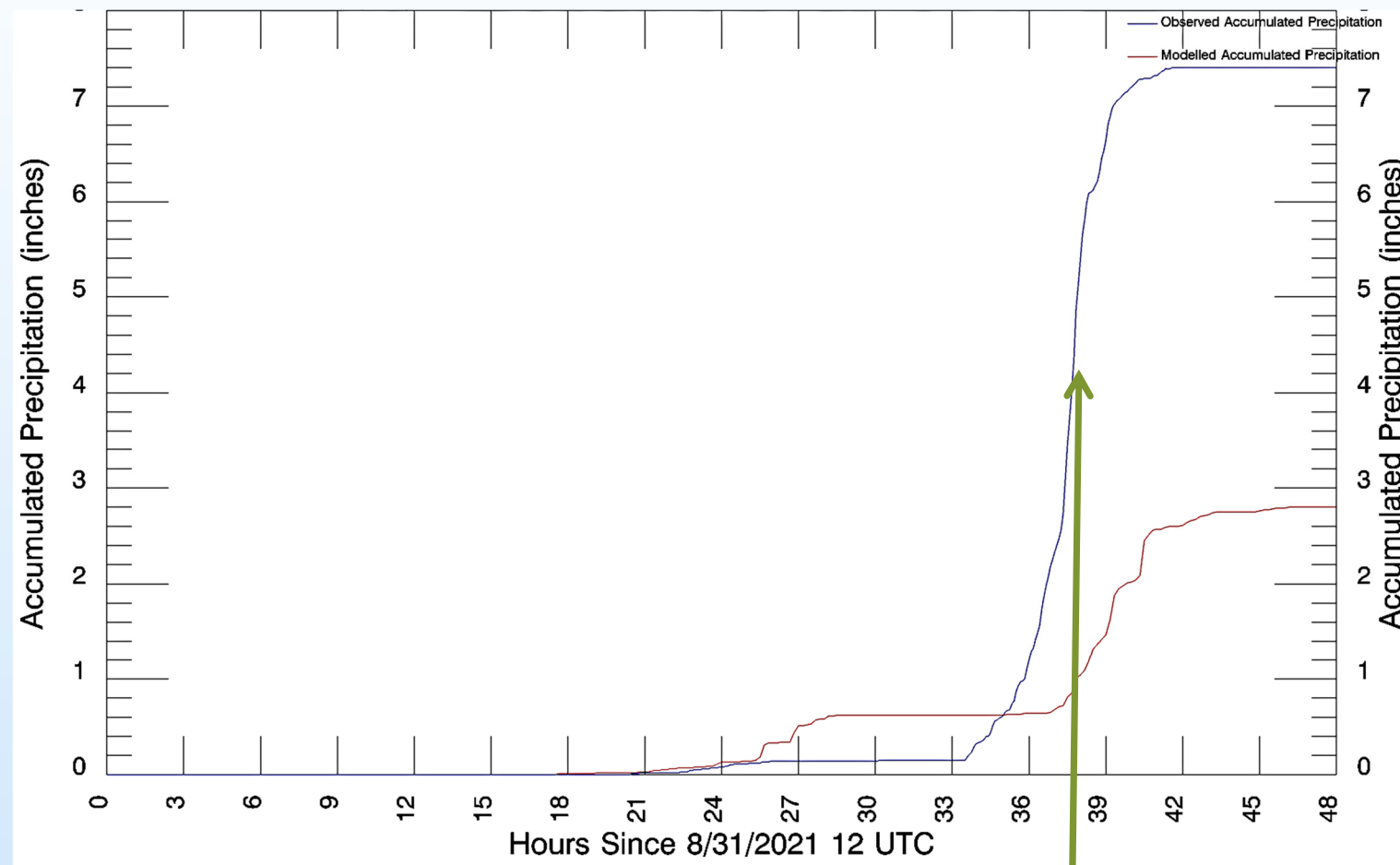


Maximum Observed
Hourly Accumulation

In comparison to the more
detailed NYSM data, higher
precipitation rates in the outer
boroughs are not captured.

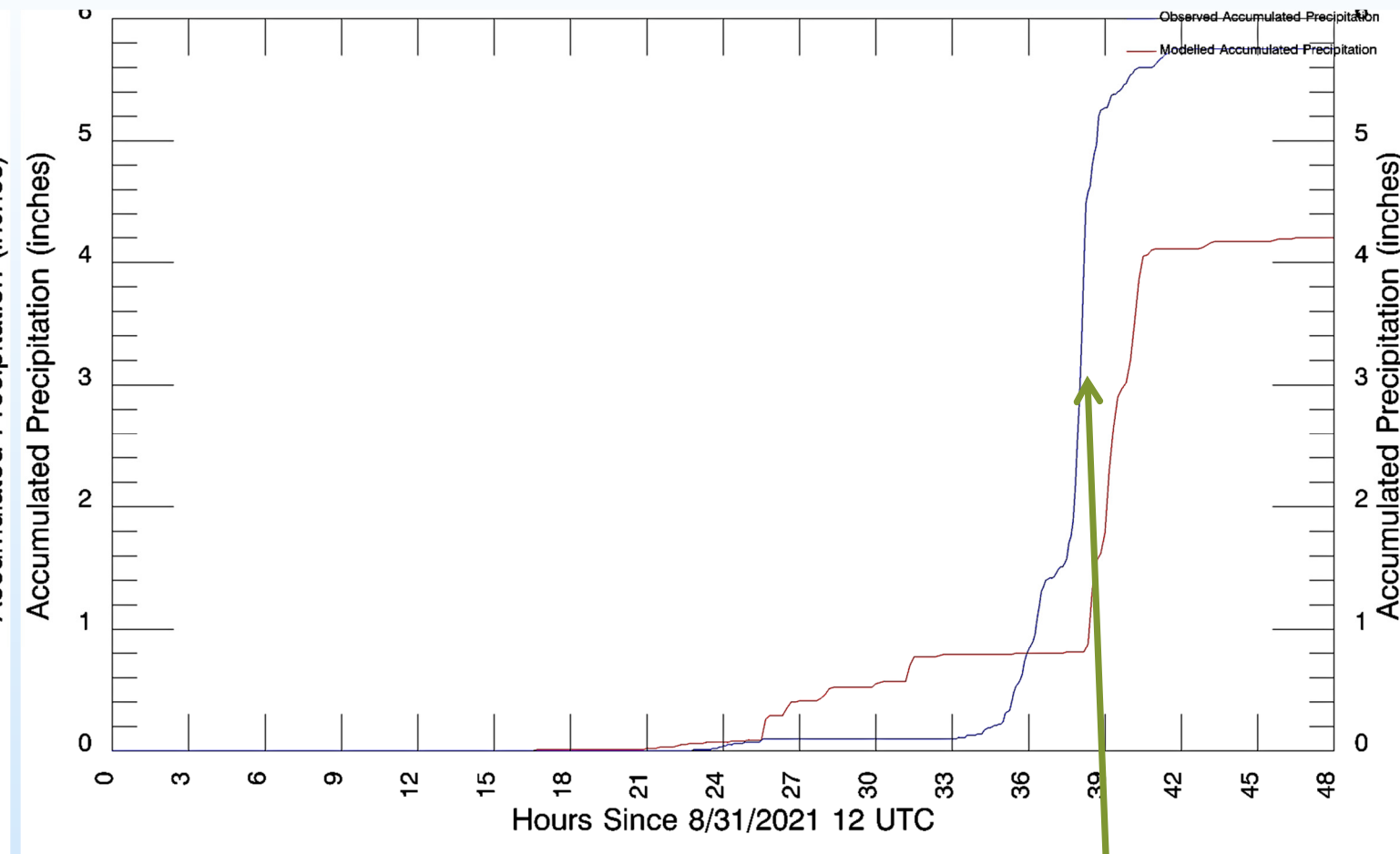
NYSM Site-Specific Model vs. Observed Precipitation (Model Initialized at 12 UTC on 31 August 2021)

— Observed
— Modelled



Bronx

3.43" reported maximum hourly accumulation at 2210 EDT on 1 September



Queens

3.22" reported maximum hourly accumulation at 2250 EDT on 1 September

Model does not capture the intensity and the start time is ~1.5 hours late while the end time is roughly correct.

- **Results from the example simulations illustrate that 12 to 36 hours lead time before the highest impact rainfall is feasible**
- **Geographic distribution of precipitation (intensity and accumulation) was reasonable compared to observations at a coarse level**
- **Errors in model timing for the most intense period were partially corrected with later background data and assimilated observations and larger modelling domain**
- **Precipitation accumulations were lower than observed, but rates were reasonably consistent with observations, just shorter**

Data from the New York State Mesonet were critical to run these experiments and make these assessments

Open Source Software and Data

- Weather and climate models
 - Weather Research and Forecasting Community Model (WRF)
<https://www.mmm.ucar.edu/models/wrf>
 - Model for Predictions Across Scales – Atmospheres Community Model (MPAS-A)
<https://mpas-dev.github.io/>
- Hydrology model: WRF Hydrological Community Model (WRF-Hydro)
https://ral.ucar.edu/projects/wrf_hydro/model-code
- Limnology model: Stanford Unstructured Nonhydrostatic Terrain-following Adaptive Navier-Stokes Simulator (SUNTANS)
<https://sourceforge.net/p/suntans/wiki/Home/>
- Data model: Climate and Forecast (CF)-compliant network Common Data Form (netCDF)
<https://www.unidata.ucar.edu/netcdf/>, <https://www.ogc.org/standard/netcdf/>
- Data model for point observations: Observations Data Model (ODM)
<https://github.com/ODM2/ODM2>
- Data access and visualization: netCDF Web Map Service (ncWMS)
ncwms.sourceforge.net
- Data visualization and analysis: QGIS, OpenDX, Matplotlib, Leaflet, pandas, R, ...
- Observing system data flow: Message Queuing Telemetry Transport (MQTT)
<https://mqtt.org/>
- Observing system alerting and analytics
 - Apache Kafka: <https://kafka.apache.org/>
 - Apache Edgent: <https://incubator.apache.org/projects/edgent.html>

■ NOAA

- **Global Forecast System (GFS)**
<https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>
- **RAPid Refresh (RAP)**
<https://rapidrefresh.noaa.gov/>
- **Meteorological Assimilation Data Ingest System (MADIS)**
<https://madis.ncep.noaa.gov/index.shtml>
- **Vegetation (Visible Infrared Imaging Radiometer Suite Green Vegetation Fraction, VIIRS GVF)**
<https://www.ospo.noaa.gov/Products/land/gvf/>

■ NASA

- **Terrain (Shuttle Radar Topography Mission, SRTM)**
<https://www2.jpl.nasa.gov/srtm/>
- **Land Use and Land Cover (Moderate Resolution Imaging Spectroradiometer, MODIS)**
<https://modis.gsfc.nasa.gov/data/dataproduct/mod12.php/>
- **Sea Surface Temperature (Short-term Prediction Research and Transition Center, SPoRT SSTs)**
<https://weather.ndc.nasa.gov/sport/sst/>
- **Soil conditions (Land Information System, LIS)**
<https://lis.gsfc.nasa.gov/data>

■ ECMWF

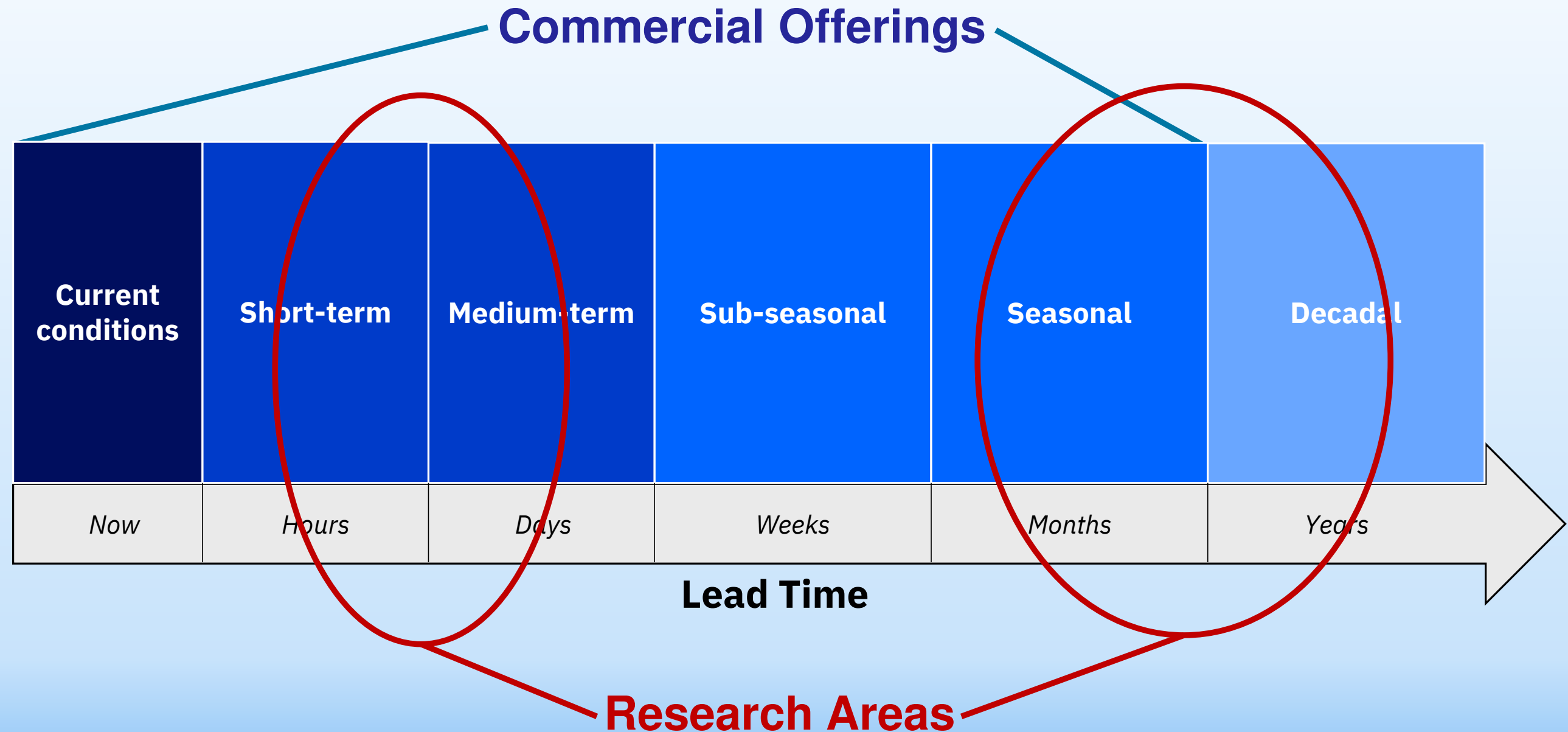
- **Reanalysis (ERA5)**
<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>

Questions ?

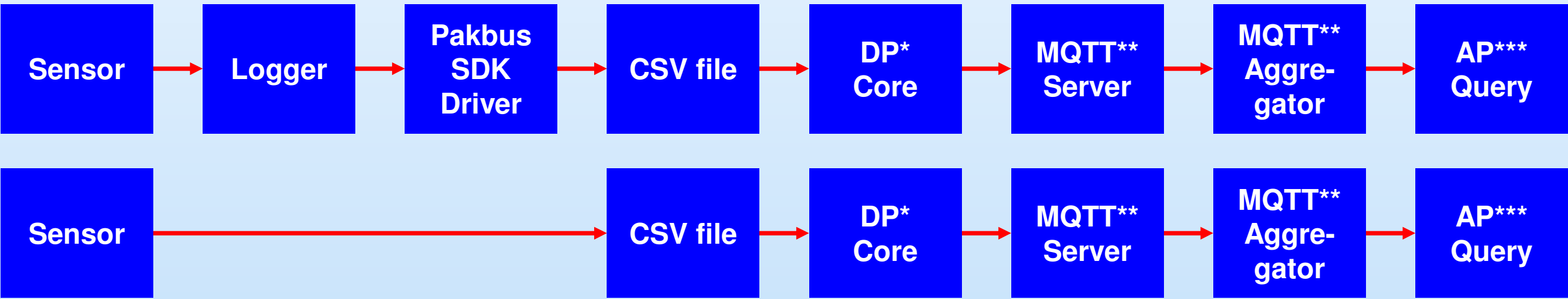
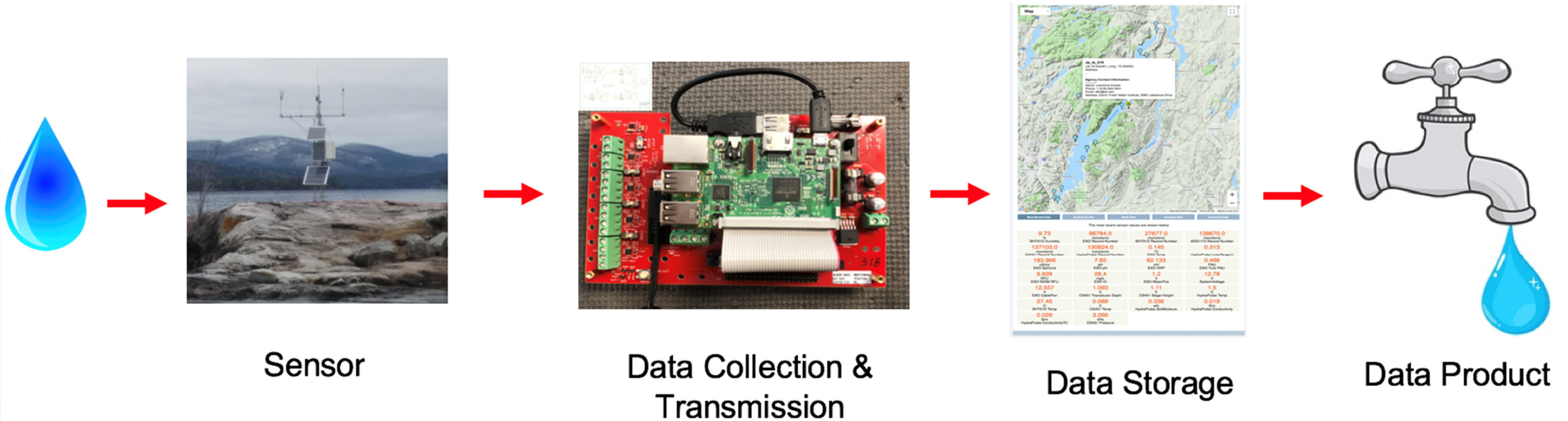
Backup Slides



We generate predictions for (almost) every lead time and every point in the world...



Observing System Platform Data Flow

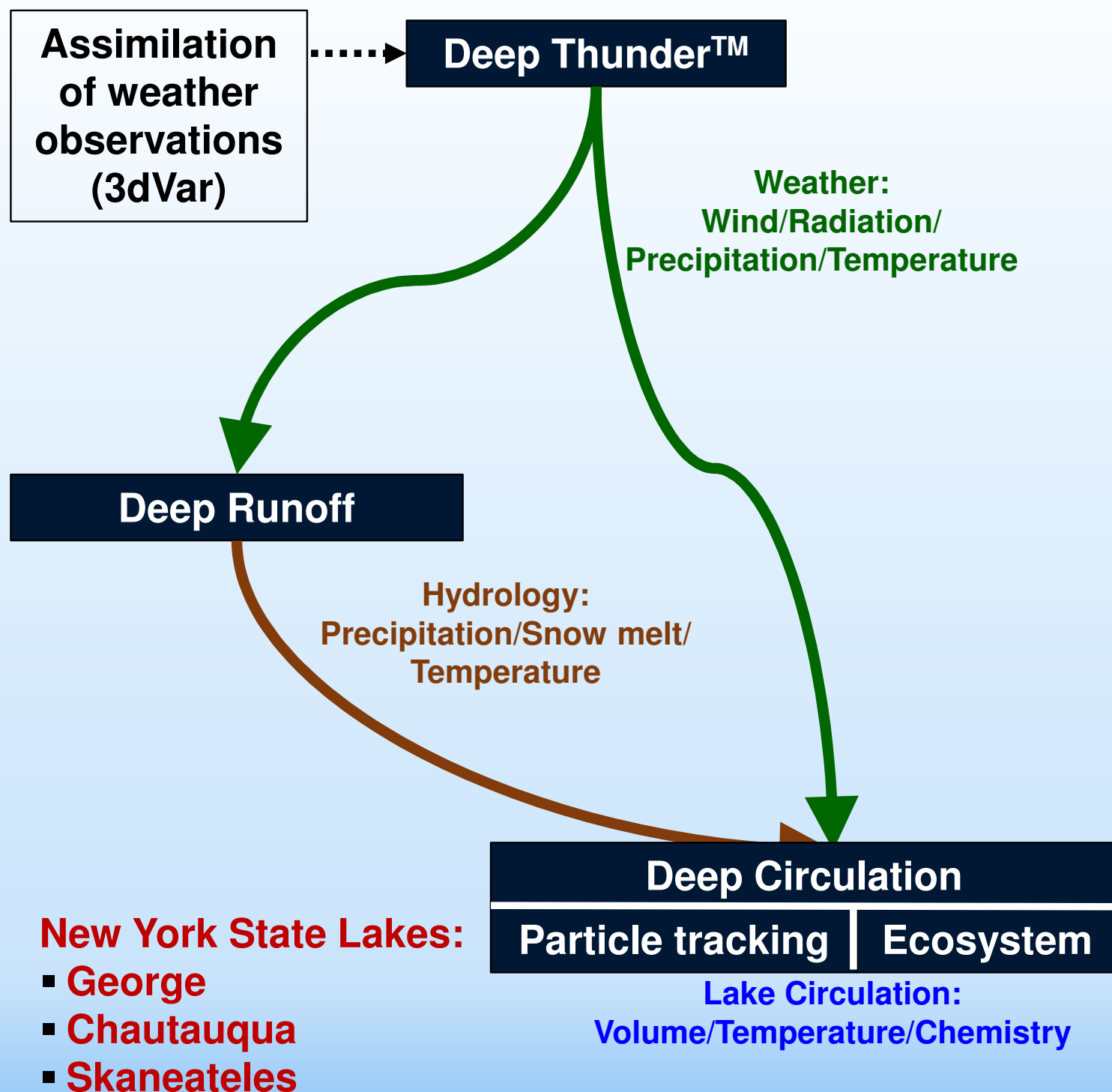


*DP = DiamondPoint

**MQTT = Message Queuing Telemetry Transport

***AP = AssemblyPoint

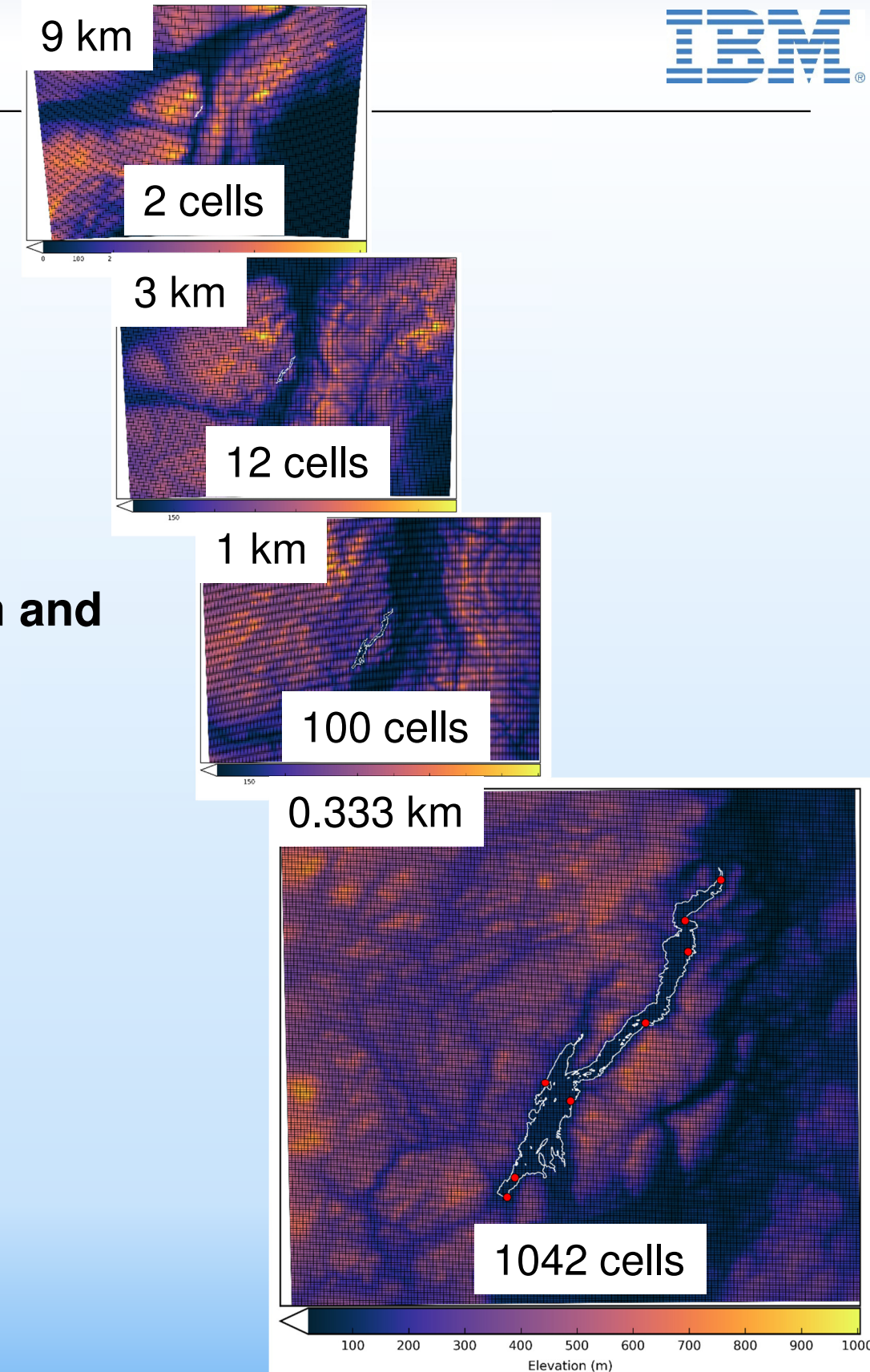
Operational (Production) Coupled Modelling (> 500TB of Data)



- Weather model enables atmospheric forcing for other models, predictions for operations and field work, and public outreach
 - *Deep Thunder:* $\Delta 333\text{m}$ [based upon WRF-ARW]
- Hydrology model to understand terrestrial runoff (precipitation and snowmelt, sediments, nutrients, contaminants such as salt) into the lake:
 - *Deep Runoff:* $\Delta 41\text{m}$ [based upon WRF-Hydro, using NOAH-MP]
- Lake circulation model to understand transport (sediments, nutrients, contaminants, hydrodynamics and water quality)
 - *Deep Circulation:* $\Delta \sim 27$ to $\sim 70\text{m}$ [based upon SUNTANS]
- Daily forecasts initiated at 0 UTC with monitoring via Slack (36- and 72-hour)
- Automated validation
- Interactive and production 2d and 3d visualization
- Web-based dissemination via ncWMS

Deep Thunder Weather Model Configuration

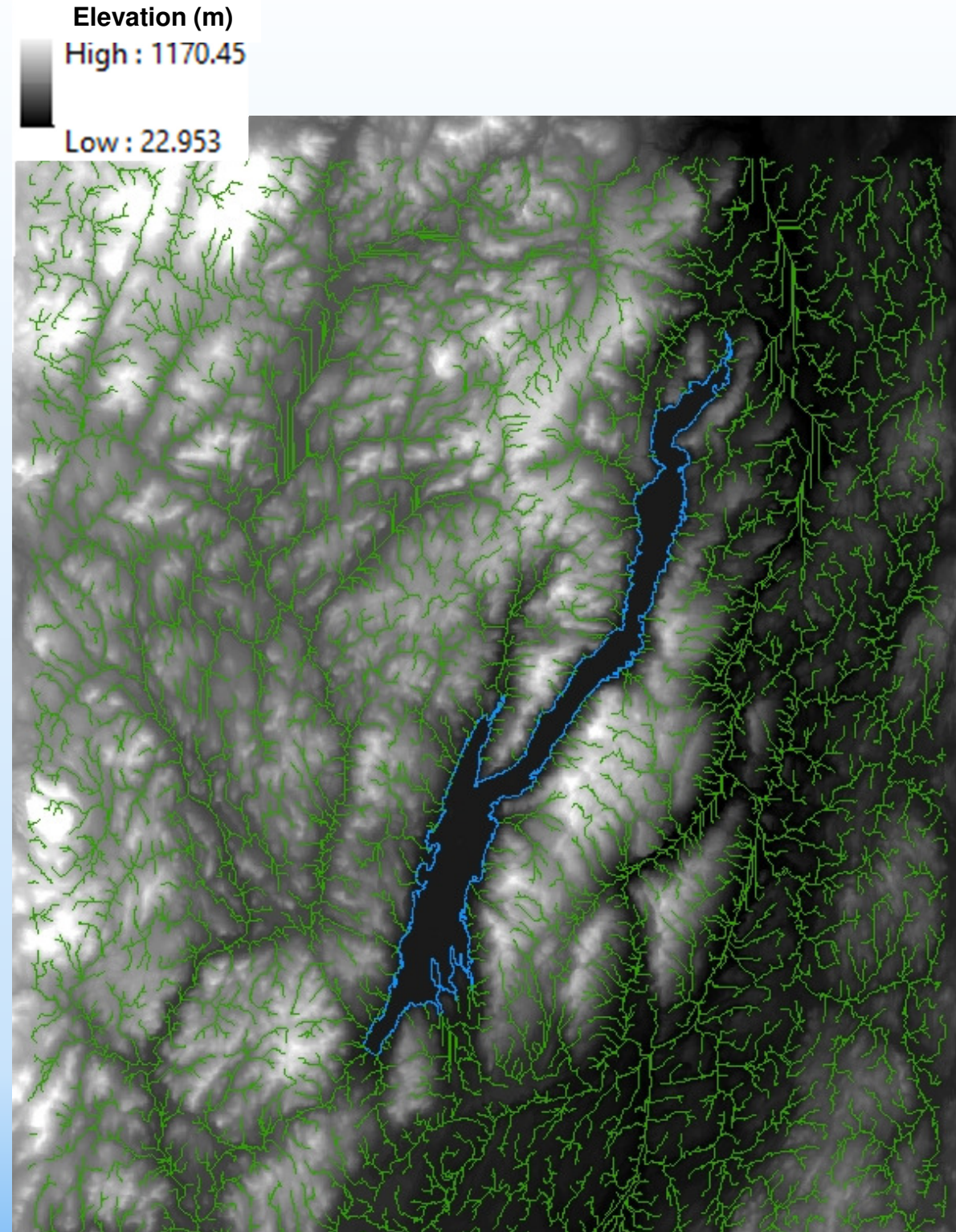
- Based upon WRF-ARW v3.9.1.1
<https://www.mmm.ucar.edu/models/wrf>
- Four one-way nested domains: 9 km, 3 km, 1 km, 0.333 km [173x173/136x136/190x190/178x220]
- 40 vertical levels with 14 typically below 1 km
- Assimilates data from ~ 1500 local observing sites using 3D-Var
- Daily forecasts (initialized at 00 UTC) of 36-hour duration at .333km and 72-hour duration at 1km
- Physics configuration
 - Thompson microphysics
 - MYNN PBL and surface layer schemes
 - NOAH LSM
 - Grell-Devenyi ensemble cumulus physics for outer nest
 - RRTMG long- and short-wave radiation
 - Large Eddy Simulation (LES) for innermost nest (.333km)
- NOAA RAP for background fields and hourly GFS for lateral boundaries
- NASA data sets for terrain, land use, topography, SSTs, etc.



- Based upon WRF-Hydro 5.1

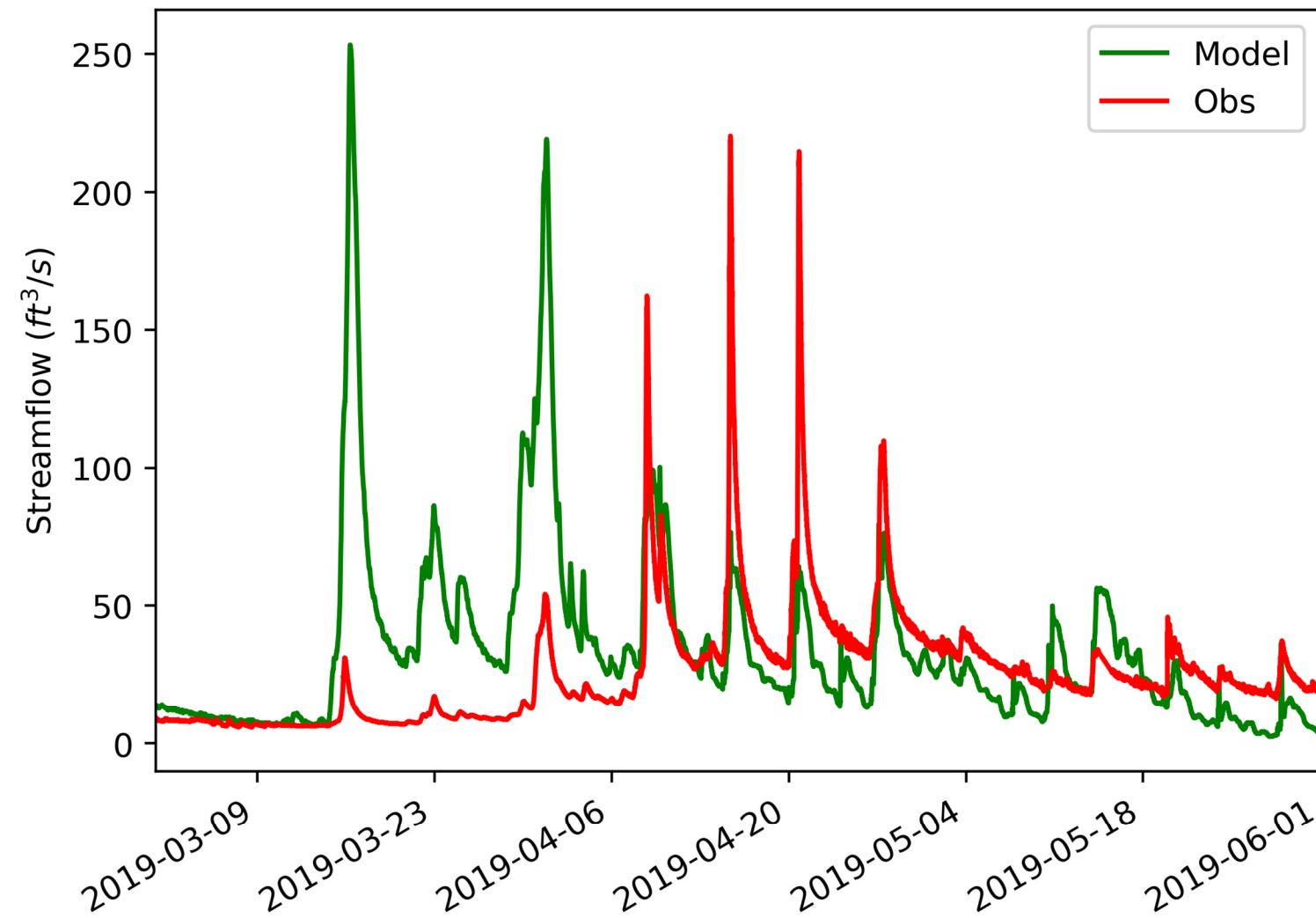
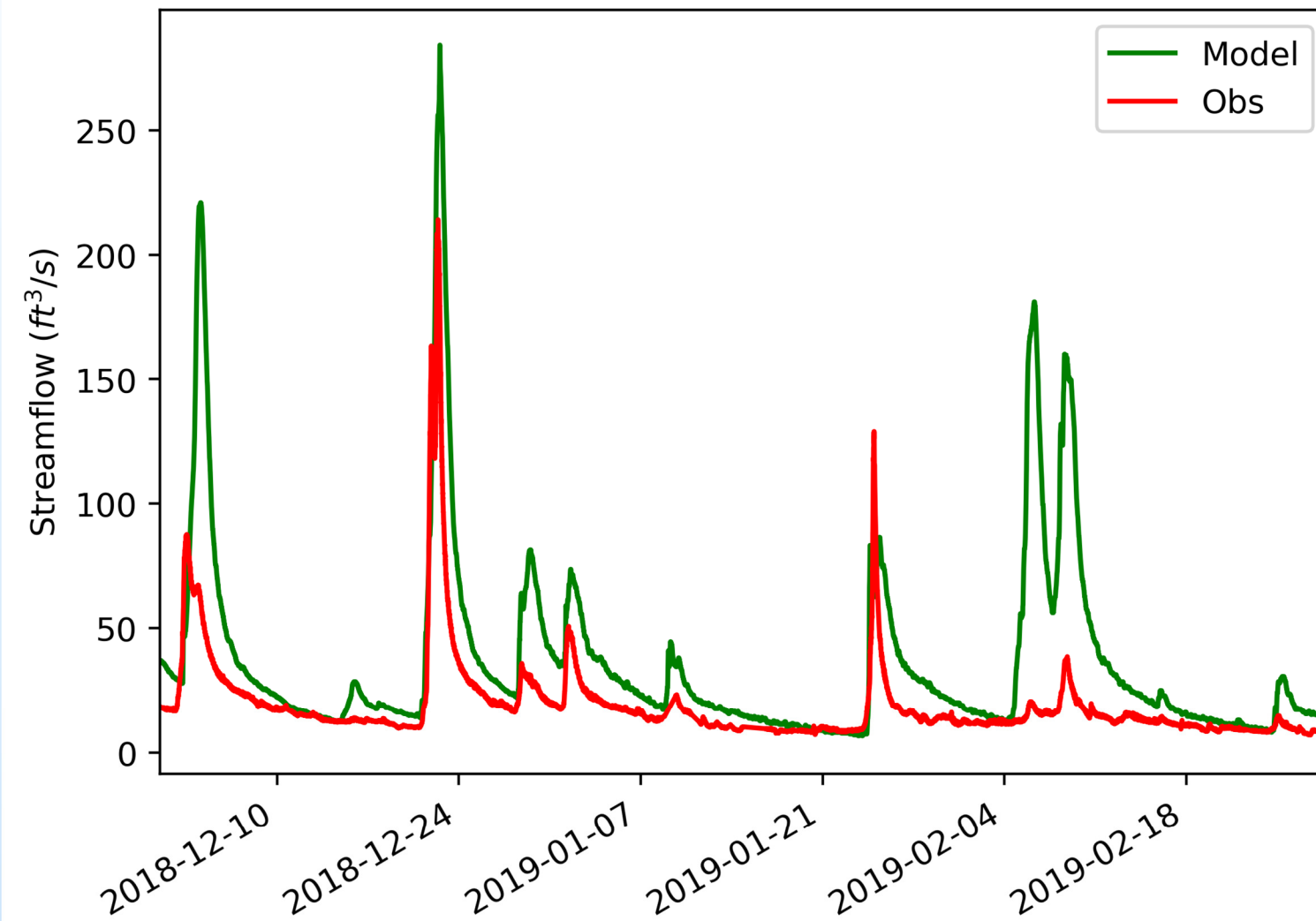
https://ral.ucar.edu/projects/wrf_hydro/model-code

- Gridded stream routing
- Horizontal resolution of 41.7m [1416 x 1752] (driven by 8:1 ratio from the 333m weather model nest)
- Land surface interactions simulated by NOAH-MP
- Simulating flow through 1563 individual stream segments (reaches) in the Lake George watershed
- 182 streams that enter Lake George and are coupled to the circulation model
- Extensions for salt runoff



**Watershed topography
(grayscale) with
overlays of stream
network (green and lake
boundary (light blue))**

Example Predicted Streamflow at Hydrograph Location: West Brook (Lake George)



Green = Model
Red = Observations

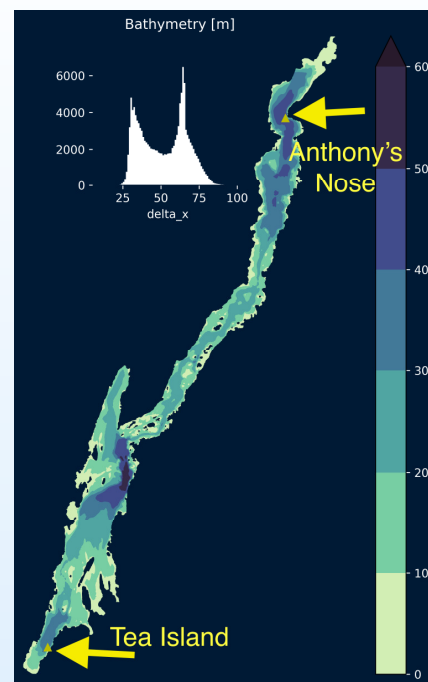
Deep Circulation Limnology Model Configuration

- Based upon the Stanford Unstructured Nonhydrostatic Terrain-following Adaptive Navier-Stokes Simulator (SUNTANS)

<https://sourceforge.net/p/suntans/wiki/Home/>

- Horizontal resolution (~27 to ~70m, average ~47m)
- 57 vertical layers, ranging from 0.5m to 1.5m
- COARE V3.0 heat flux equations
- Mellor-Yamada 2.5 turbulence closure scheme
- Forced by 333m weather model nest

Example Predicted Current Velocity and Temperature at Two Different Depths

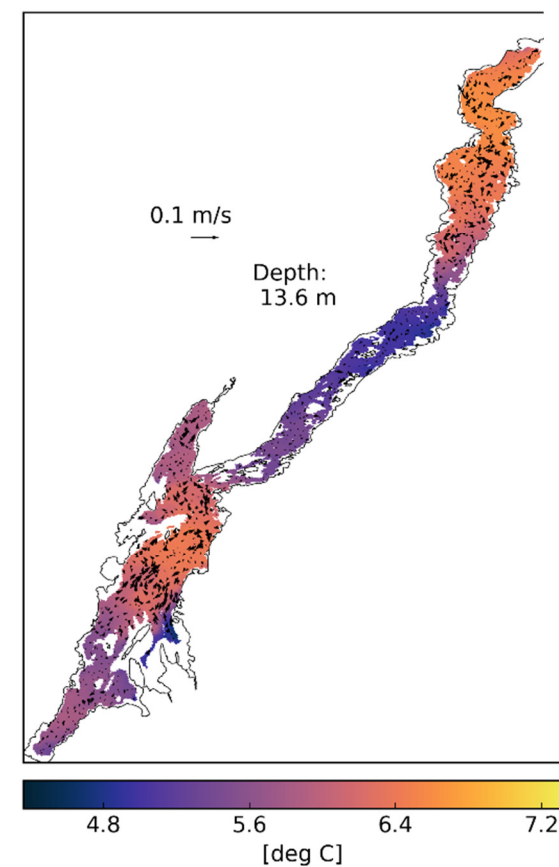
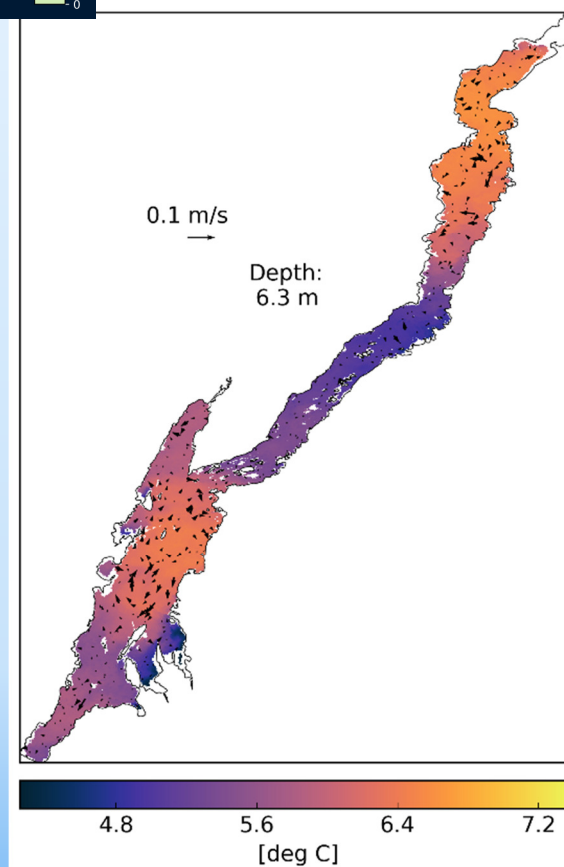
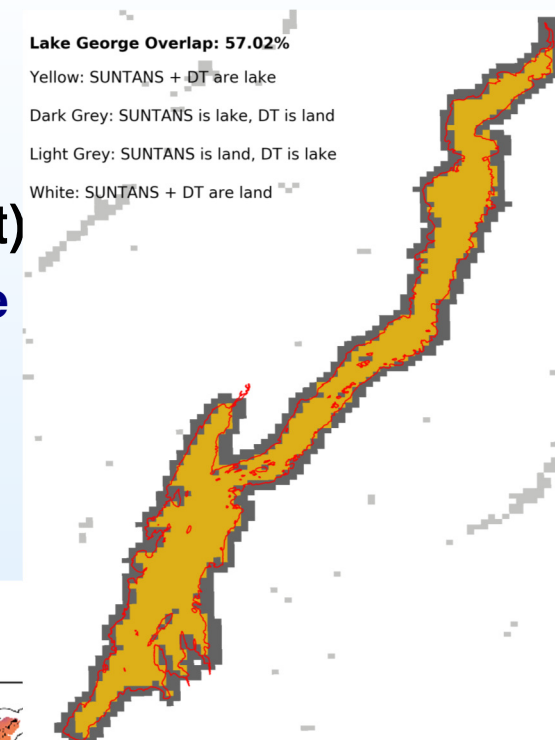


- Model mesh bathymetry (left)

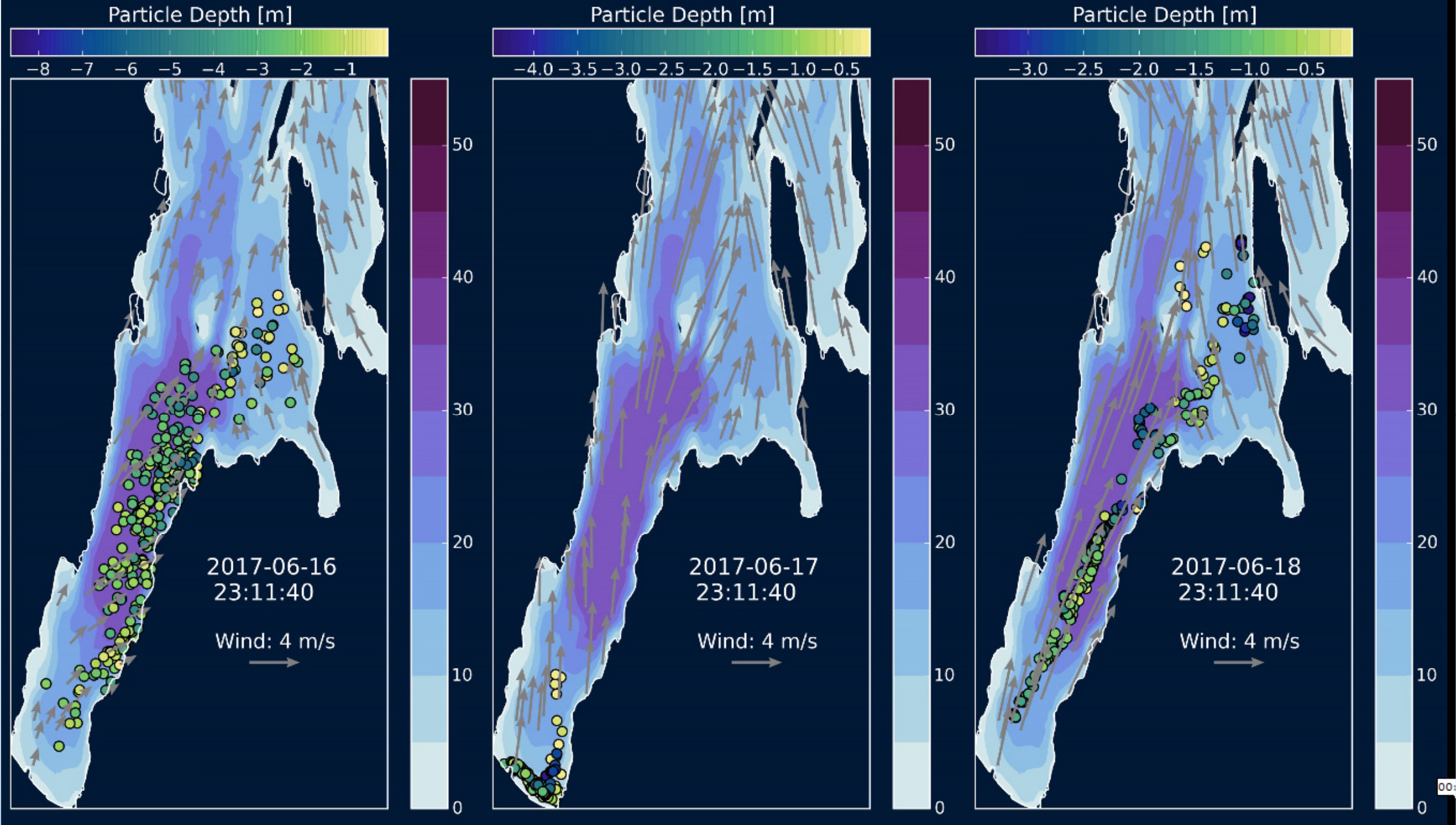
- Weather and circulation land masks (right)

- Yellow (both circulation and weather are lake – 57%)
- Dark Grey (circulation lake, weather land)
- Light Grey (circulation land, weather lake)
- White (weather – land)

2018-12-02 01:20:00



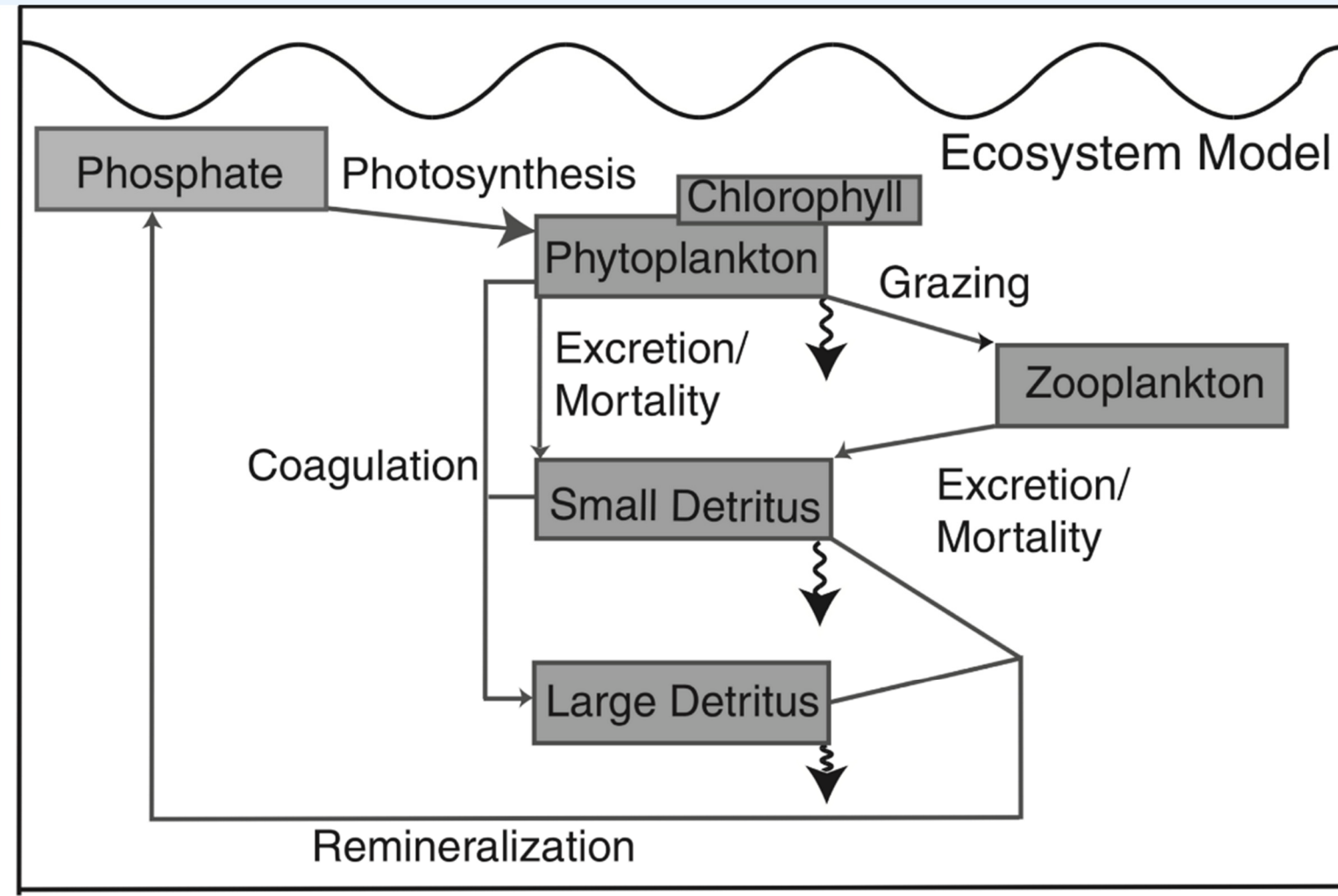
Particle Tracking to Determine How Pollutants Move (Lake George)



Pollutant Releases in the Southern Basin for Three Consecutive Days

Ecosystem Model

- With simple starting conditions
- Using meteorological and hydrodynamic model output
- Nutrients, Phytoplankton ($\propto$ chlorophyll a), Zooplankton, Detritus, O_2 (Dissolved)
- Sediments, Ammonium, Flagellates, Cyanobacteria
- High concentration of phytoplankton sinks over time



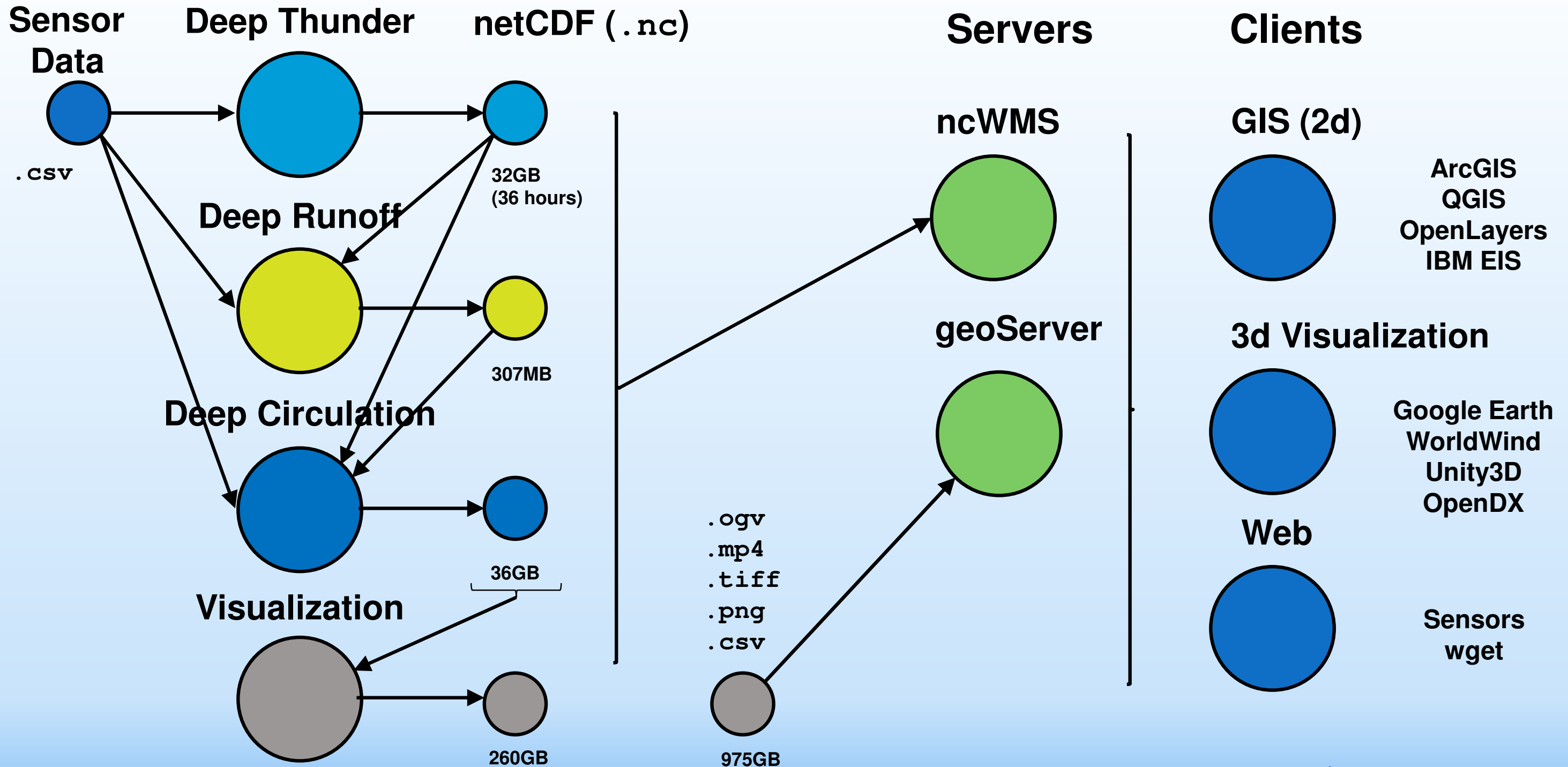
Coagulation with other elements

$$\frac{\partial P}{\partial t} = v \cdot P - m_p \cdot \text{graz} \cdot Z - \text{coag}(D_s + P) - s_p \frac{\partial P}{\partial z}$$

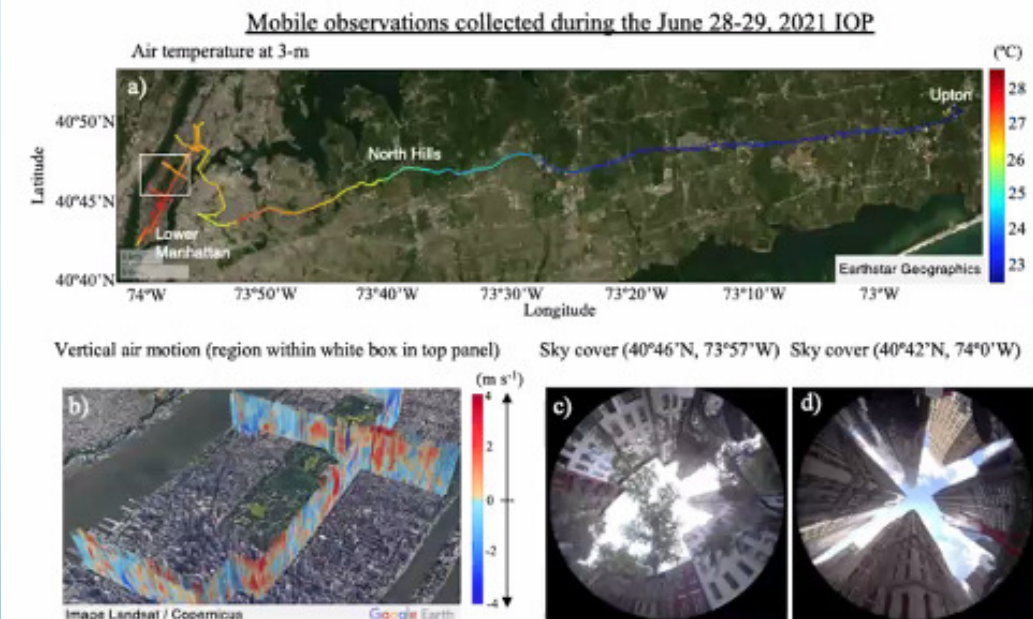
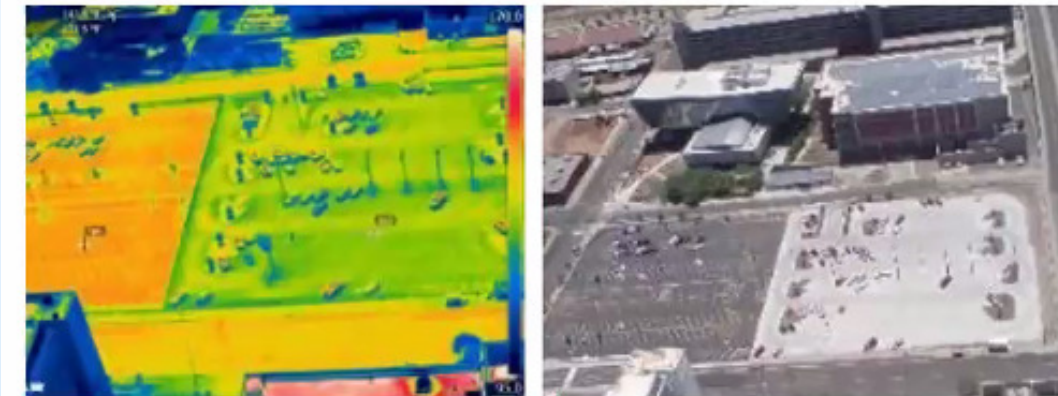
Growth of phytoplankton
Phytoplankton fed upon
Sinking of phytoplankton

- CF-compliant netCDF (generalized data model as abstraction for self-describing multidimensional arrays)
(<https://www.unidata.ucar.edu/netcdf/>)
- Software interface and applications available on many platforms plus many APIs
- Interfaces to many relevant data sets, data models and physical models
- Metadata conventions for some complex topologies and coordinate systems
- Familiar in relevant communities (e.g., atmospheric, hydrologic, oceanographic)
- http access via ncWMS (ncwms.sourceforge.net)

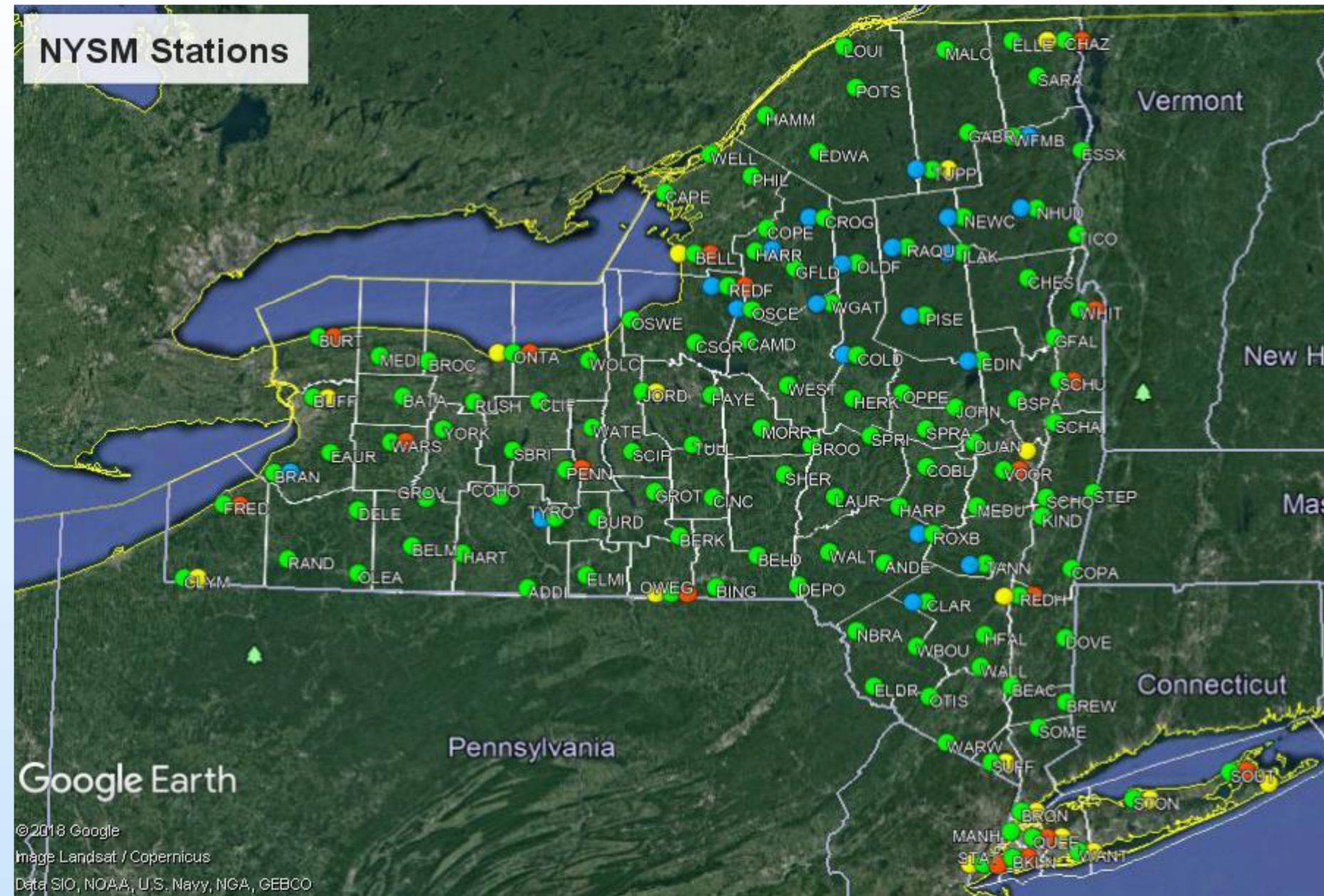
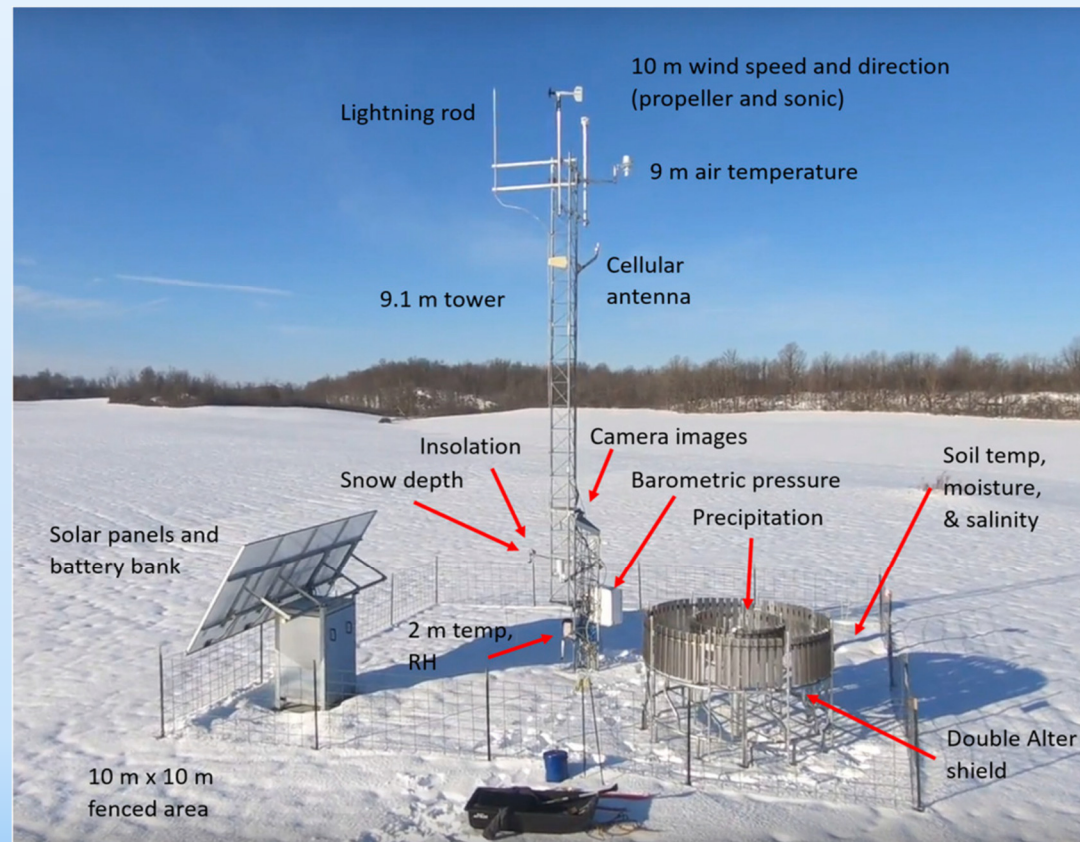
Model Data to Client Access (Lake George)



- Develop regional data clearinghouse
- Deploy additional field measurement sites, including testbeds
- Conduct intensive observations campaign each summer: mobile observatories, new networks, flux towers



- **126 Standard** ●
- **20 Snow** ●
- **17 Profiler** ●
- **18 Flux** ●



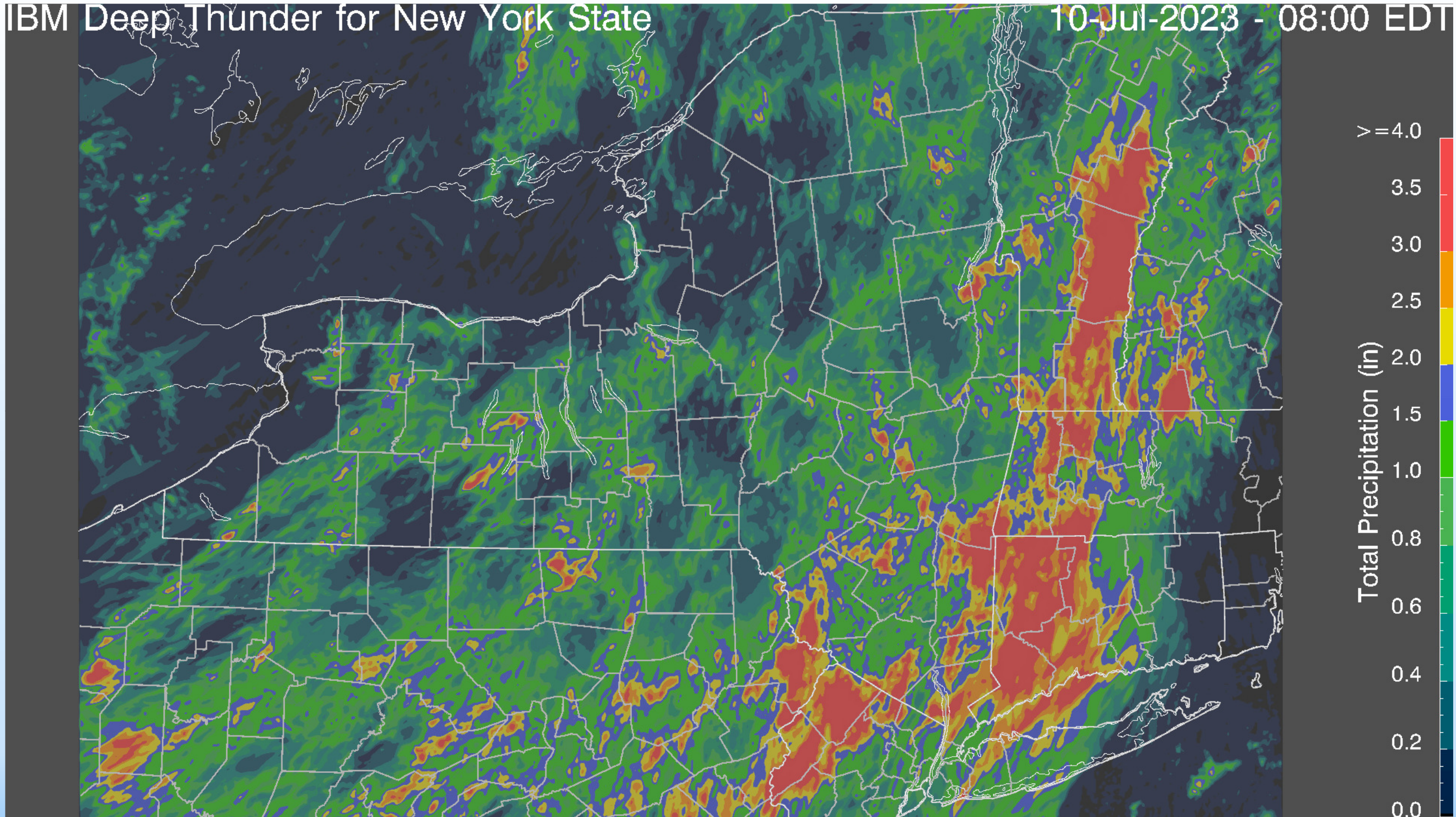
- **Observations available every 5 minutes**
- **≥ 1 standard site per county, average 27km spacing**
- **Data from the lower boundary layer to the subsurface**

<https://www2.nysmesonet.org/>

9-10 July 2023 Flooding Event

IBM Deep Thunder for New York State

10-Jul-2023 - 08:00 EDT



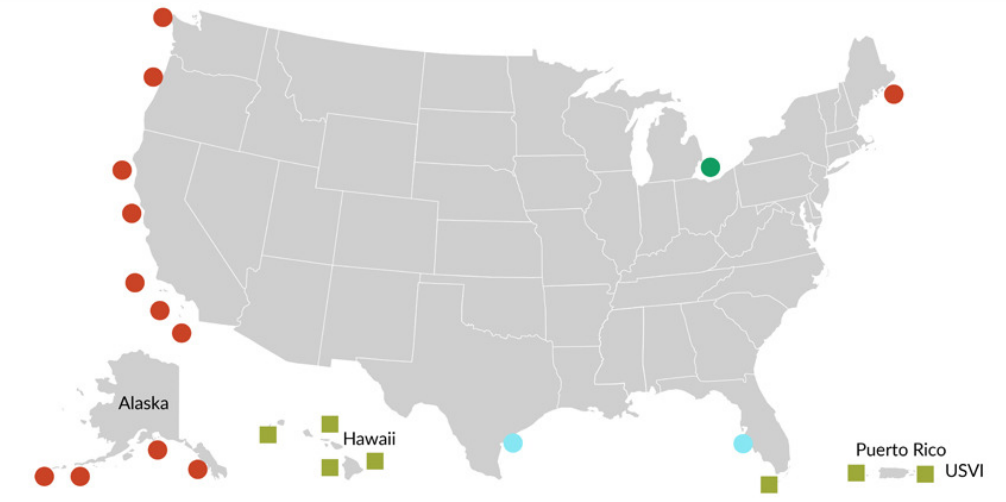
Predicted Precipitation at 1km resolution

Seasonal Forecasting of Harmful Algal Blooms

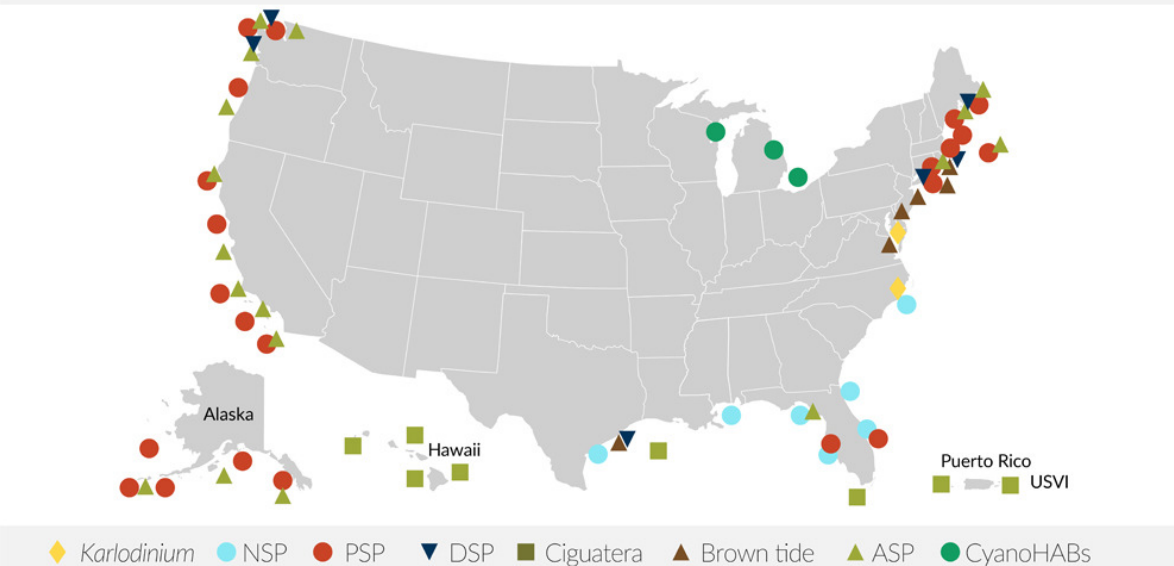
- Phytoplankton are mostly microscopic algae that are a foundation of the world's aquatic food chains
- A few hundred species cause harmful impacts to wildlife, humans, etc., and under certain conditions, can rapidly multiply and accumulate to create visible “blooms”
- Reports of harmful algal blooms world have increased in the last few decades
- The increases are related to changes in land use, agricultural and water treatment practices, coastal development, climate change and other factors, as well as improved detection

Incidents of HABs in US Coastal Waters

Pre-1972



Present

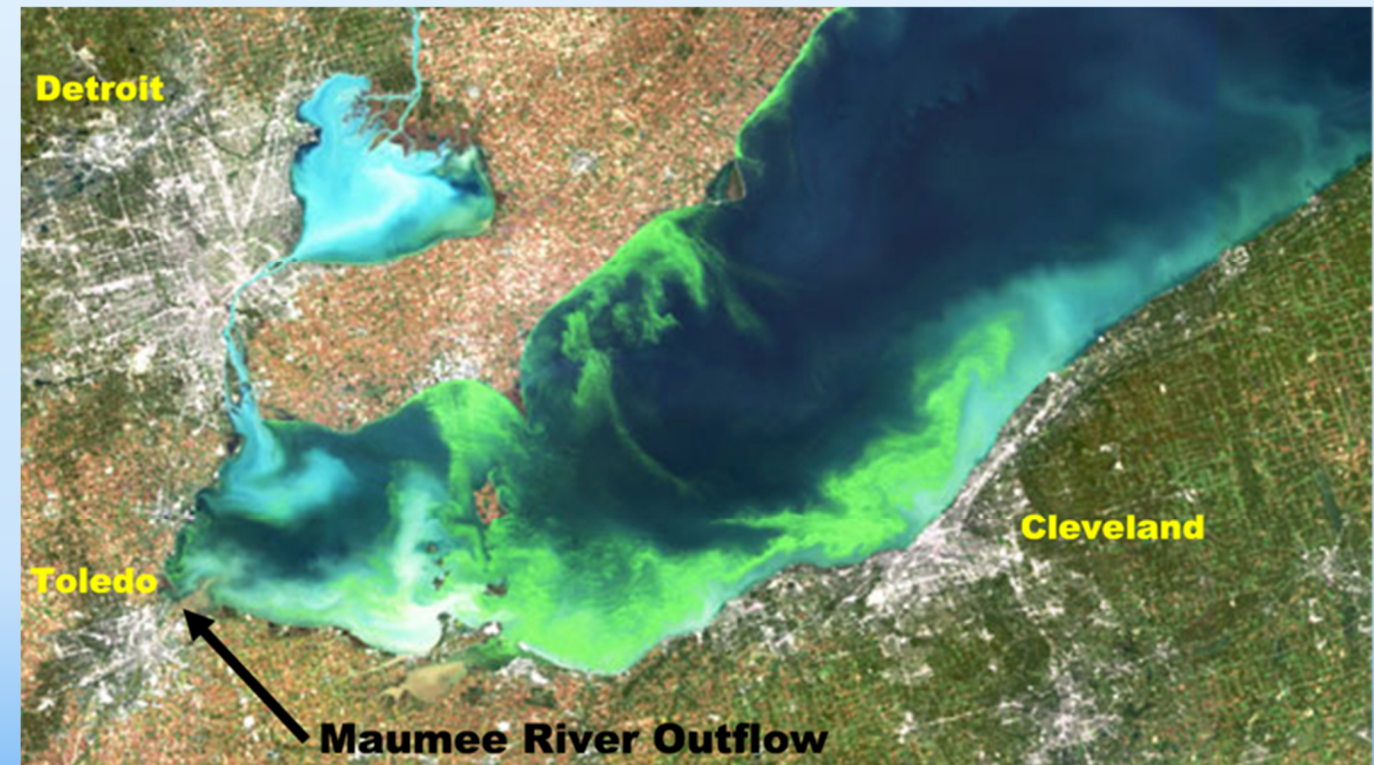


Source: U.S. National Office for Harmful Algal Blooms,
Woods Hole Oceanographic Institution

- As one notable example, consider Lake Erie, the smallest and shallowest of Laurentian Great Lakes of North America
- The impact of such blooms has an annual average cost of O(\$100M) to shoreline communities in Ohio, Michigan, and Ontario (especially for recreation and tourism)



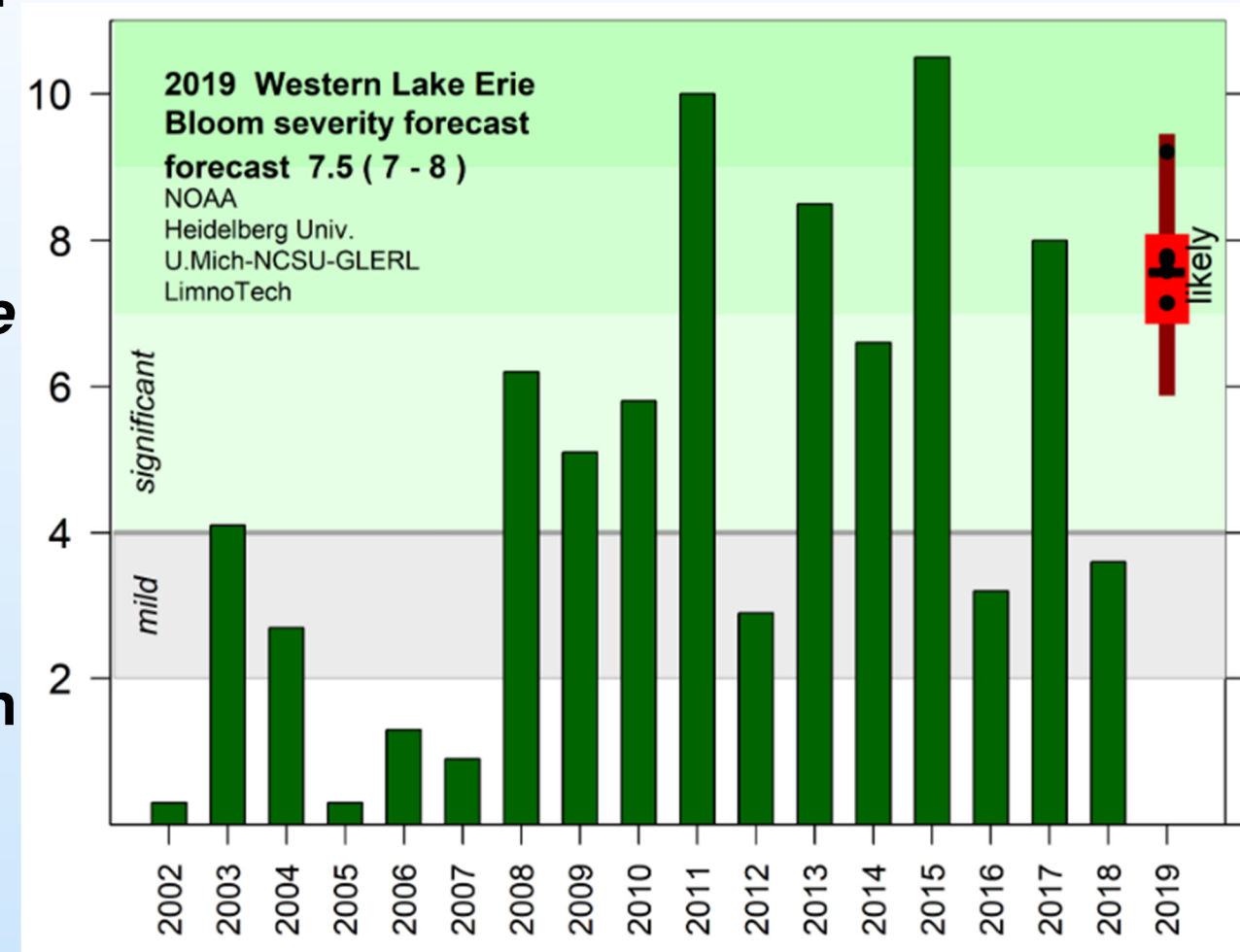
A toxic Microcystis bloom washes up on the shore of Maumee Bay in western Lake Erie on August 29, 2011. (Photograph available courtesy of S. Bihn, Western Lake Erie Waterkeeper.)



Optical three-band visible image from the Envisat satellite (MERIS_FP) at 300m resolution showing one of the largest blooms from October 2011 extended from Toledo to Cleveland and beyond, and along the Ontario shore. (Image available courtesy of the European Space Agency.)

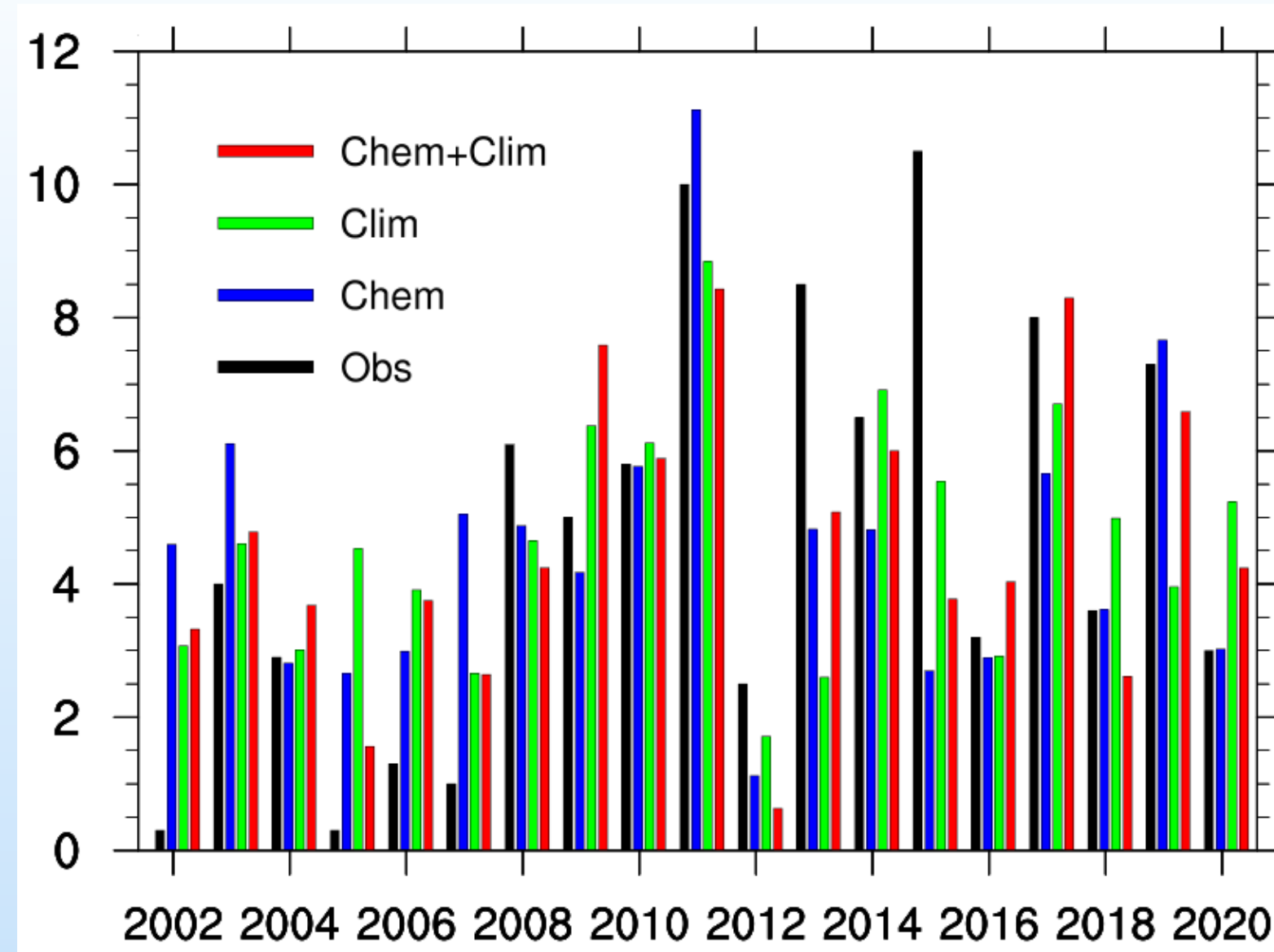
HABs Severity Index for Lake Erie

- NOAA and its research partners forecast HABs severity index for western Lake Erie and work with academic partners to refine and improve the models used to develop the forecasts
- Projections for the season comes in early July every year
- Significant HAB for 2019
- NOAA's July forecast said "*Bloom is expected to measure 7.5 on the severity index but could range between 6 and 9.*"
- NOAA's prediction for 2019 was very close to the actual value of 7.3
- The 2020 bloom had a final severity index (SI) of 3.0, much less than in 2019. The forecasted bloom severity was between 4 and 5, a small overestimate of the measured SI of 3.0
- An index above 5 indicates blooms having greater impact
- The severity index is based on bloom's biomass – the amount of algae – over a sustained period



HABs severity index (SI)

- HABs severity index (SI) from observations (black bar)
- SI prediction using genetic algorithm that is based on nutrient loading only (blue bar)
- SI prediction using genetic algorithm that is based on climate indices (orange bar)
- SI prediction that is based on nutrient loading and climate indices (red bar)



<https://research.ibm.com/blog/toxic-algae-lake-erie>
<https://www.nature.com/articles/s43247-022-00510-w>

■ NOAA

- Climate indices

<https://www.cpc.ncep.noaa.gov/>

<https://psl.noaa.gov/>

■ NCWQR

- Nutrient loading

<https://ncwqr.org/monitoring/>

■ Met Office Hadley Centre

- Sea Surface Temperature (SST)

<https://www.metoffice.gov.uk/hadobs/hadisst/>