The Prediction Lab

Predicting Harmful
Algal Blooms

The Problem

Toxins produced by Harmful Algal Blooms (HABs) are a health hazard to humans and animals, and incur a large financial burden to water utilities and other water quality stakeholders

Algae like cyanobacteria can create toxins

T
P
L

The Problem

Toxins produced by Harmful Algal Blooms (HABs) are a health hazard to humans and animals, and incur a large financial burden to water utilities and other water quality stakeholders

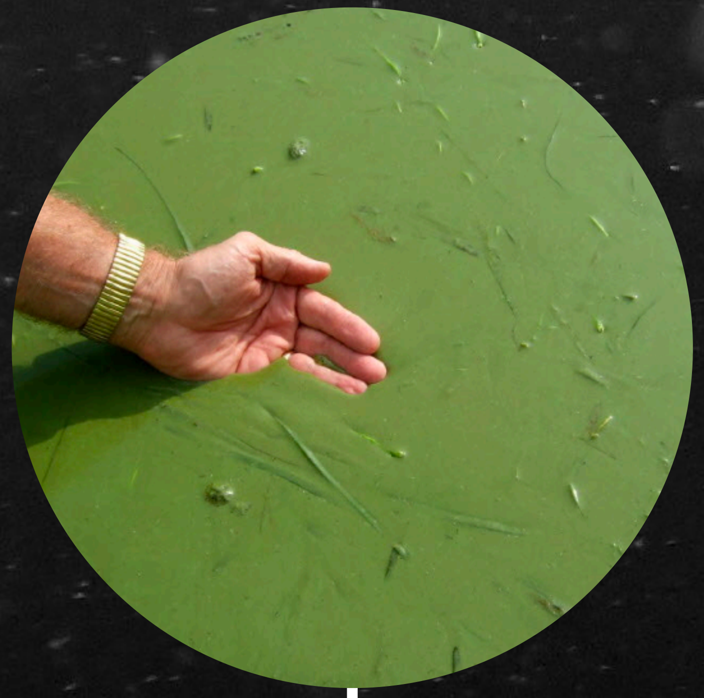
Algae like cyanobacteria can create toxins

Just like the weather, value is in knowing IN ADVANCE that a bloom is coming

T
P
L

The Problem

Toxins produced by Harmful Algal Blooms (HABs) are a health hazard to humans and animals, and incur a large financial burden to water utilities and other water quality stakeholders



Algae like cyanobacteria can create toxins

Just like the weather, value is in knowing IN ADVANCE that a bloom is coming

So we collect data

Lots of data: water samples, CYAN, qPCR…etc

TPL

The Problem

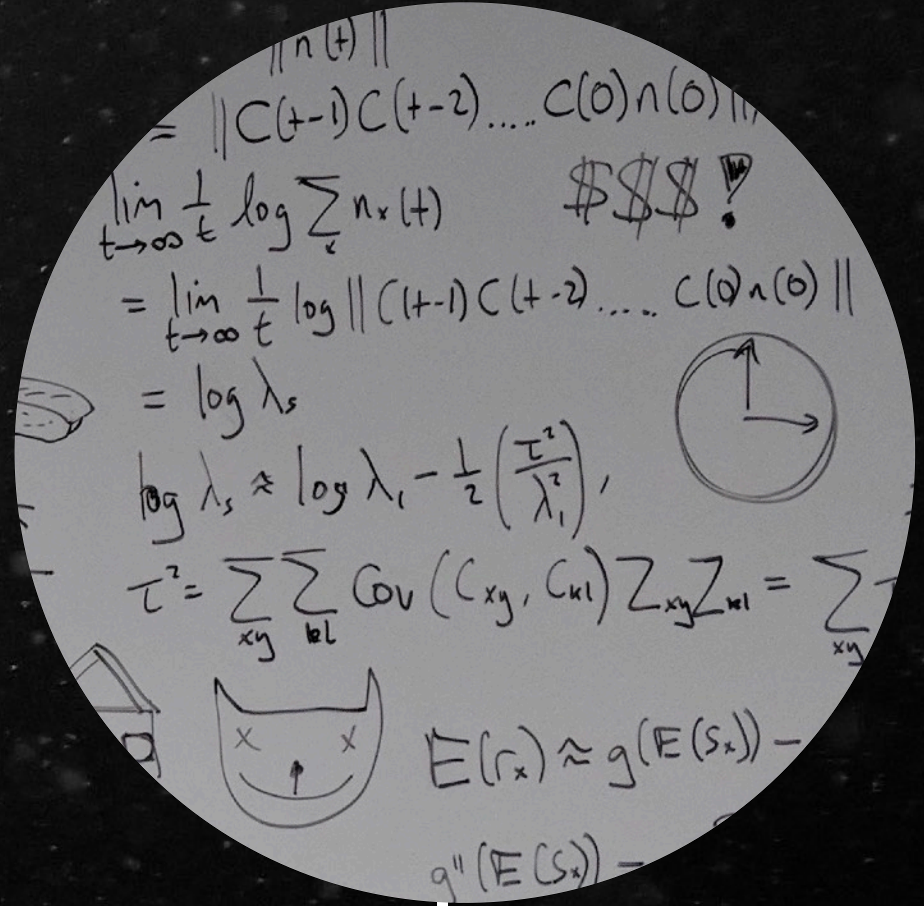
Toxins produced by Harmful Algal Blooms (HABs) are a health hazard to humans and animals, and incur a large financial burden to water utilities and other water quality stakeholders

Algae like cyanobacteria can create toxins

Just like the weather, value is in knowing IN ADVANCE that a bloom is coming

So we collect data

Lots of data: water samples, CYAN, qPCR…etc



Then we use math and algorithms, but…

What's The Real Problem?

We collect a lot of data, we develop a lot of models… but how do we make the data we have actionable?

T
P
L

What's The Real Problem?

We collect a lot of data, we develop a lot of models… but how do we make the data we have actionable?

Given identification of the drivers of HABs, can we reduce their frequency?

T
P
L

What's The Real Problem?

We collect a lot of data, we develop a lot of models… but how do we make the data we have actionable?

Given identification of the drivers of HABs, can we reduce their frequency?

If a HAB is going to happen, how best can we prepare?

T P L

What's The Real Problem?

We collect a lot of data, we develop a lot of models… but how do we make the data we have actionable?

Given identification of the drivers of HABs, can we reduce their frequency?

If a HAB is going to happen, how best can we prepare?

Can we stop HABs once they've started?

LG SONIC
Leading in ultrasonic algae control

T
P
L

What's The Real Problem?

We collect a lot of data, we develop a lot of models… but how do we make the data we have actionable?

Given identification of the drivers of HABs, can we reduce their frequency?

If a HAB is going to happen, how best can we prepare?

Can we stop HABs once they've started?

T
P
L

Challenges

Our algorithms focus on forecasting HAB occurrence: These are the major challenges we've faced:

T
P
L

Challenges

Our algorithms focus on forecasting HAB occurrence: These are the major challenges we've faced:

Ecosystem Uniqueness (no general solution possible)

TPL

Challenges

Our algorithms focus on forecasting HAB occurrence: These are the major challenges we've faced:

Ecosystem Uniqueness (no general solution possible)

Differential Sampling (data rich vs data poor)

Challenges

Our algorithms focus on forecasting HAB occurrence: These are the major challenges we've faced:

Ecosystem Uniqueness (no general solution possible)

Differential Sampling (data rich vs data poor)

Ecosystem Connectivity (open vs closed systems)

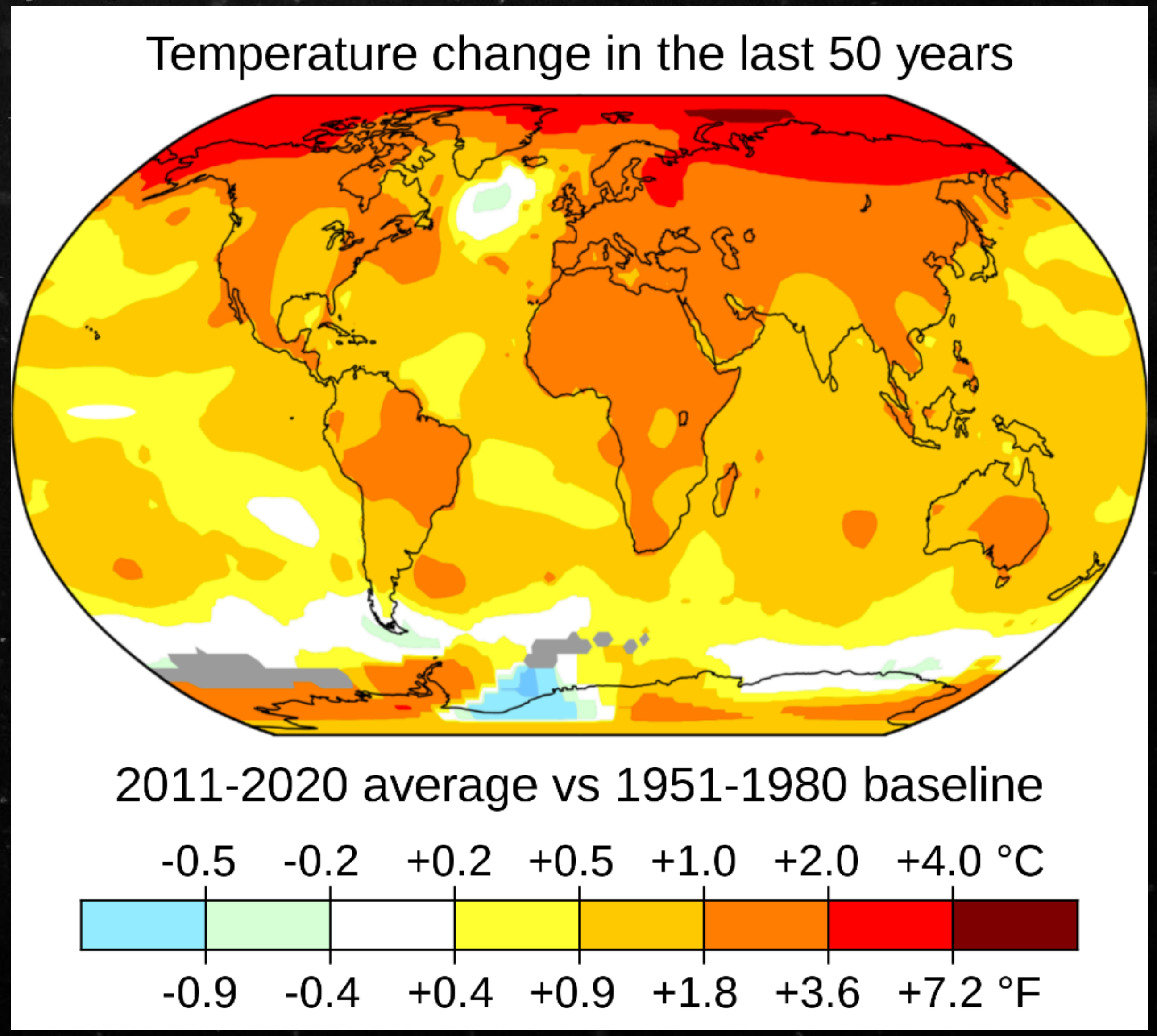
Challenges

Our algorithms focus on forecasting HAB occurrence: These are the major challenges we've faced:

Ecosystem Uniqueness (no general solution possible)

Differential Sampling (data rich vs data poor)

Ecosystem Connectivity (open vs closed systems)



Timescales (weeks, months, years, decades)

T
P
L

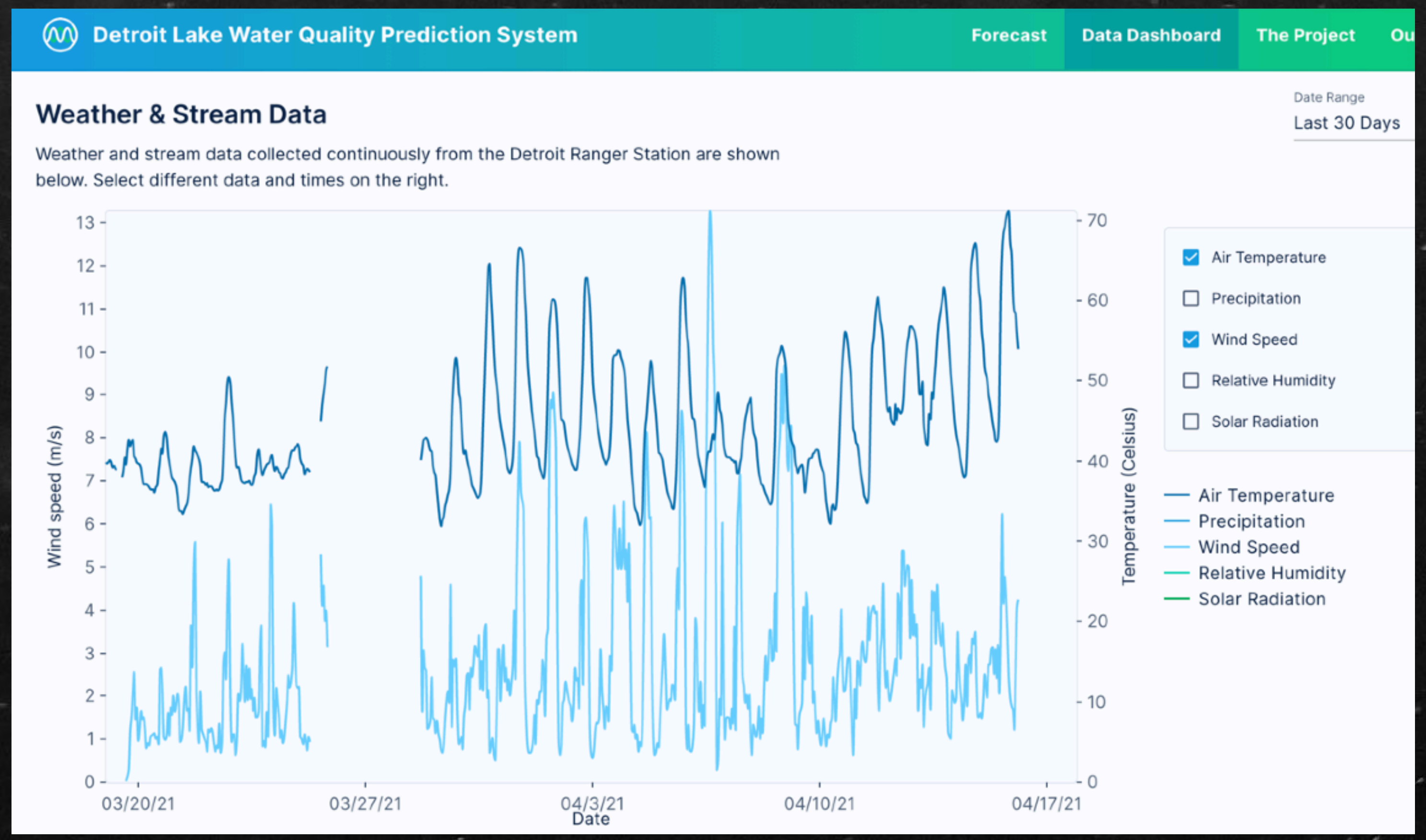
Solutions

We have been developing cyberinfrastructure and a suite of data modeling tools that help address these HAB challenges:

T
P
L

Solutions

We have been developing cyberinfrastructure and a suite of data modeling tools that help address these HAB challenges:

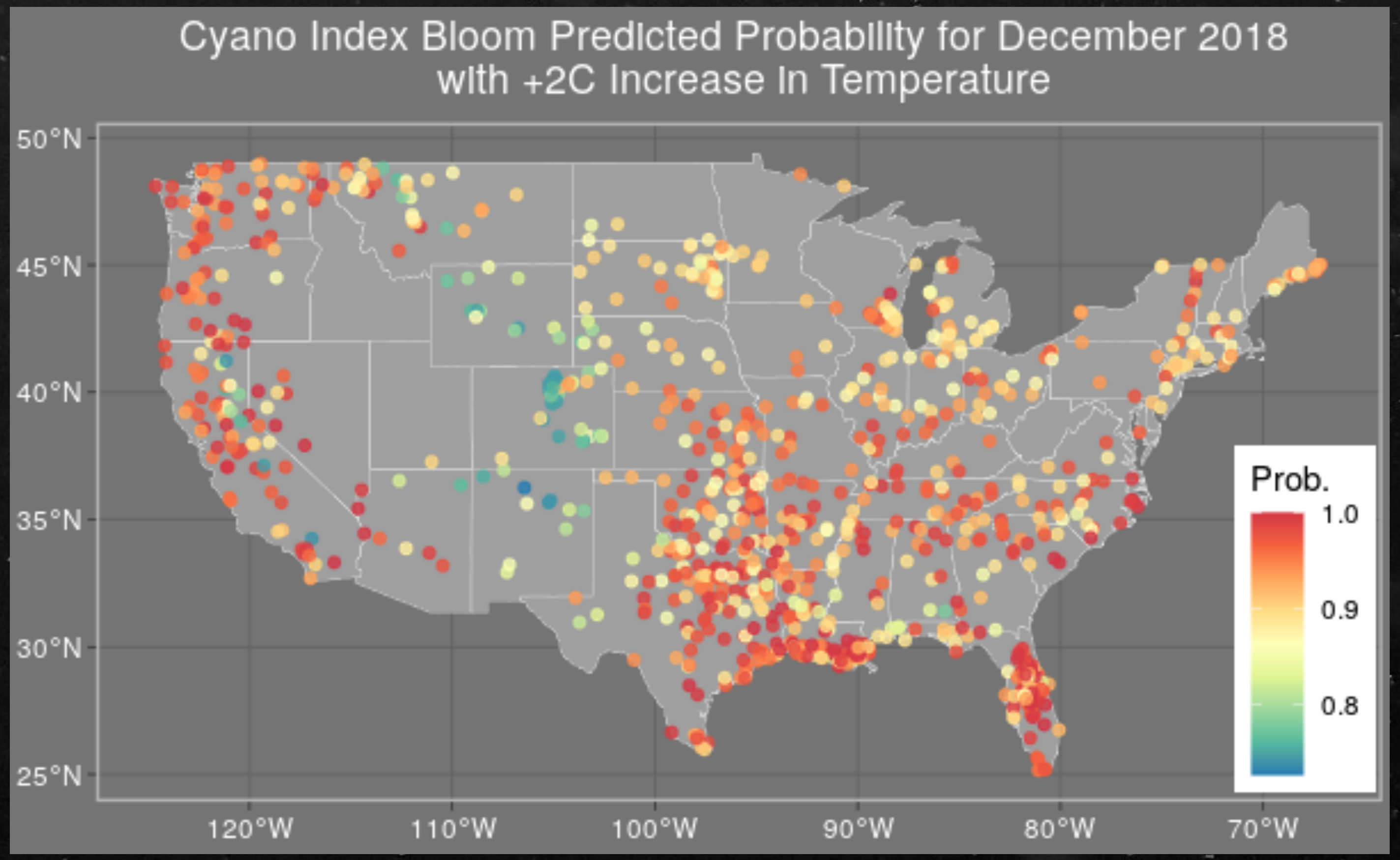


Operational
(Detroit Lake)

Solutions

We have been developing cyberinfrastructure and a suite of data modeling tools that help address these HAB challenges:

Operational (Detroit Lake)



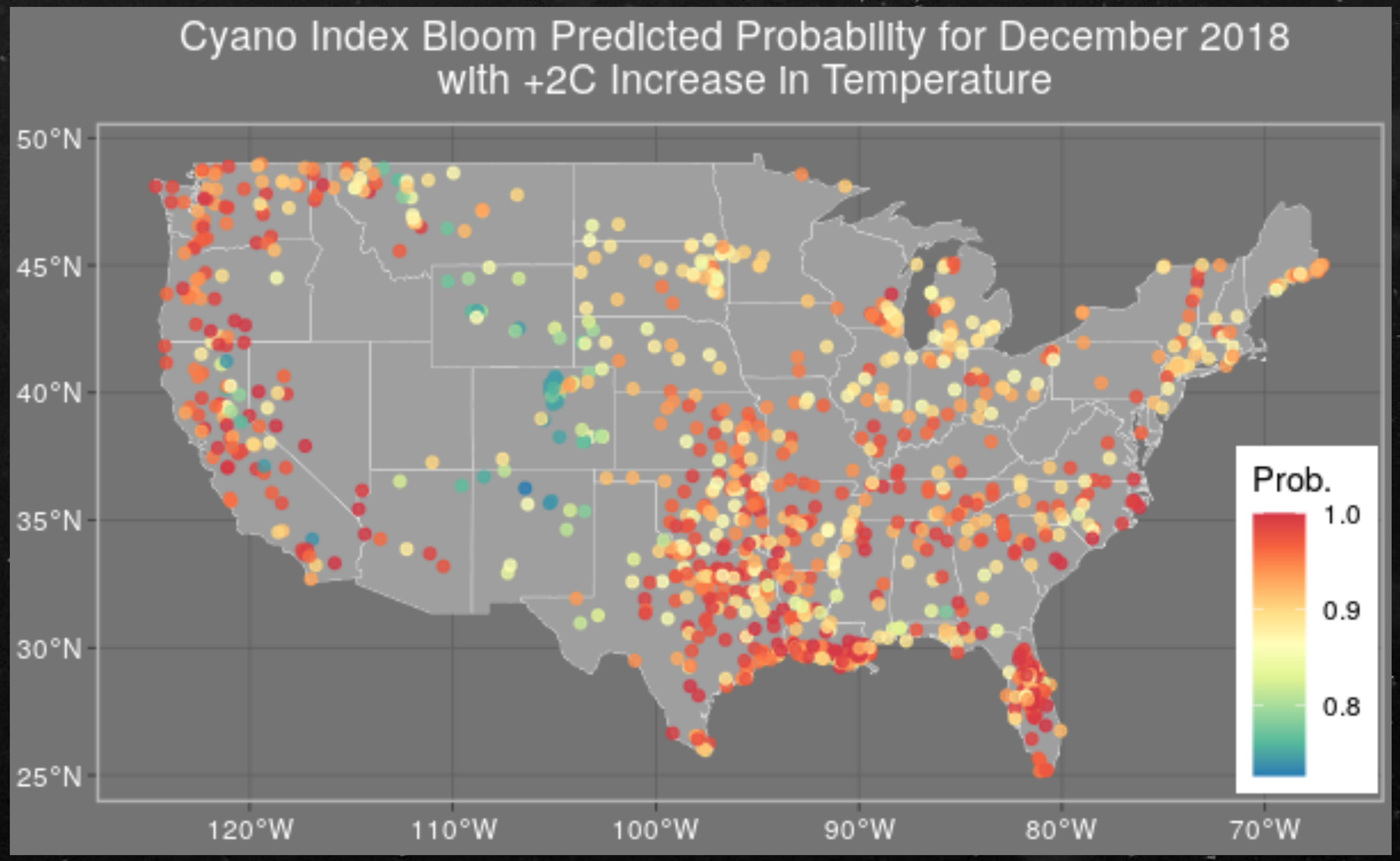
Seasonal and Climate sensitivity (CYAN)

TPL

Solutions

We have been developing cyberinfrastructure and a suite of data modeling tools that help address these HAB challenges:

Operational (Detroit Lake)



Seasonal and Climate sensitivity (CYAN)

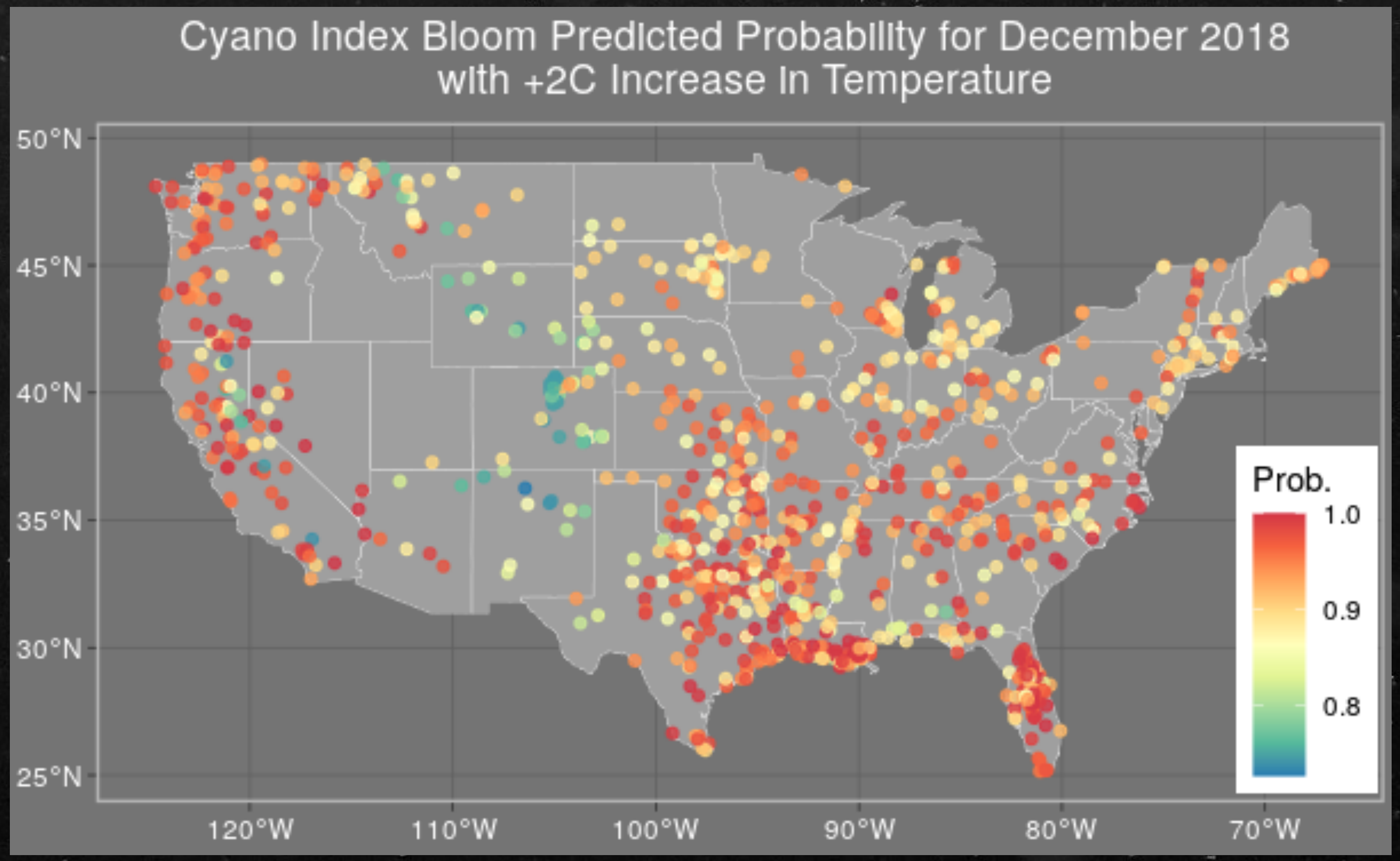
Next Gen Analytics

TPL

Solutions

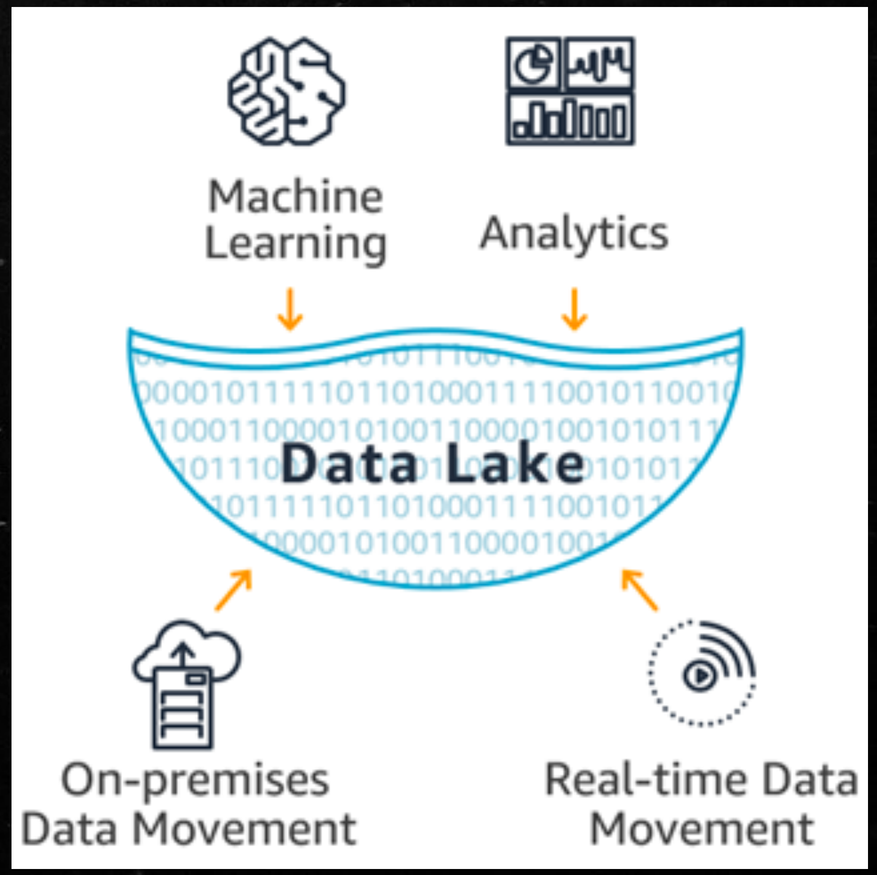
We have been developing cyberinfrastructure and a suite of data modeling tools that help address these HAB challenges:

Operational (Detroit Lake)



Seasonal and Climate sensitivity (CYAN)

Next Gen Analytics



Cloud Infrastructure

Operational Forecasts

Detroit Lake: Bayesian Model Averaging framework applied to the in situ data collected from the lake to provide daily 1-week and 2-week forecasts of cyanobacteria and toxin concentrations

T
P
L

Operational Forecasts

Detroit Lake: Bayesian Model Averaging framework applied to the in situ data collected from the lake to provide daily 1-week and 2-week forecasts of cyanobacteria and toxin concentrations


Ecological Applications, 19(7), 2009, pp. 1805–1814
© 2009 by the Ecological Society of America

Bayesian model averaging for harmful algal bloom prediction

Grant Hamilton,[1,4] Ross McVinish,[2] and Kerrie Mengersen[3]

[1]*School of Natural Resource Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001 Australia*
[2]*Mathematics Department, The University of Queensland, Brisbane, Queensland 4072 Australia*
[3]*School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001 Australia*

We expanded the Bayesian Model Averaging approach to include neural nets in addition to linear and non-linear functions

Based on empirical samples

T P L

Operational Forecasts

Detroit Lake: Bayesian Model Averaging framework applied to the in situ data collected from the lake to provide daily 1-week and 2-week forecasts of cyanobacteria and toxin concentrations


Ecological Applications, 19(7), 2009, pp. 1805–1814
© 2009 by the Ecological Society of America

Bayesian model averaging for harmful algal bloom prediction

Grant Hamilton,[1,4] Ross McVinish,[2] and Kerrie Mengersen[3]

[1] *School of Natural Resource Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001 Australia*
[2] *Mathematics Department, The University of Queensland, Brisbane, Queensland 4072 Australia*
[3] *School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001 Australia*

We expanded the Bayesian Model Averaging approach to include neural nets in addition to linear and non-linear functions

Operational Dashboard (Detroit Lake)

Based on empirical samples

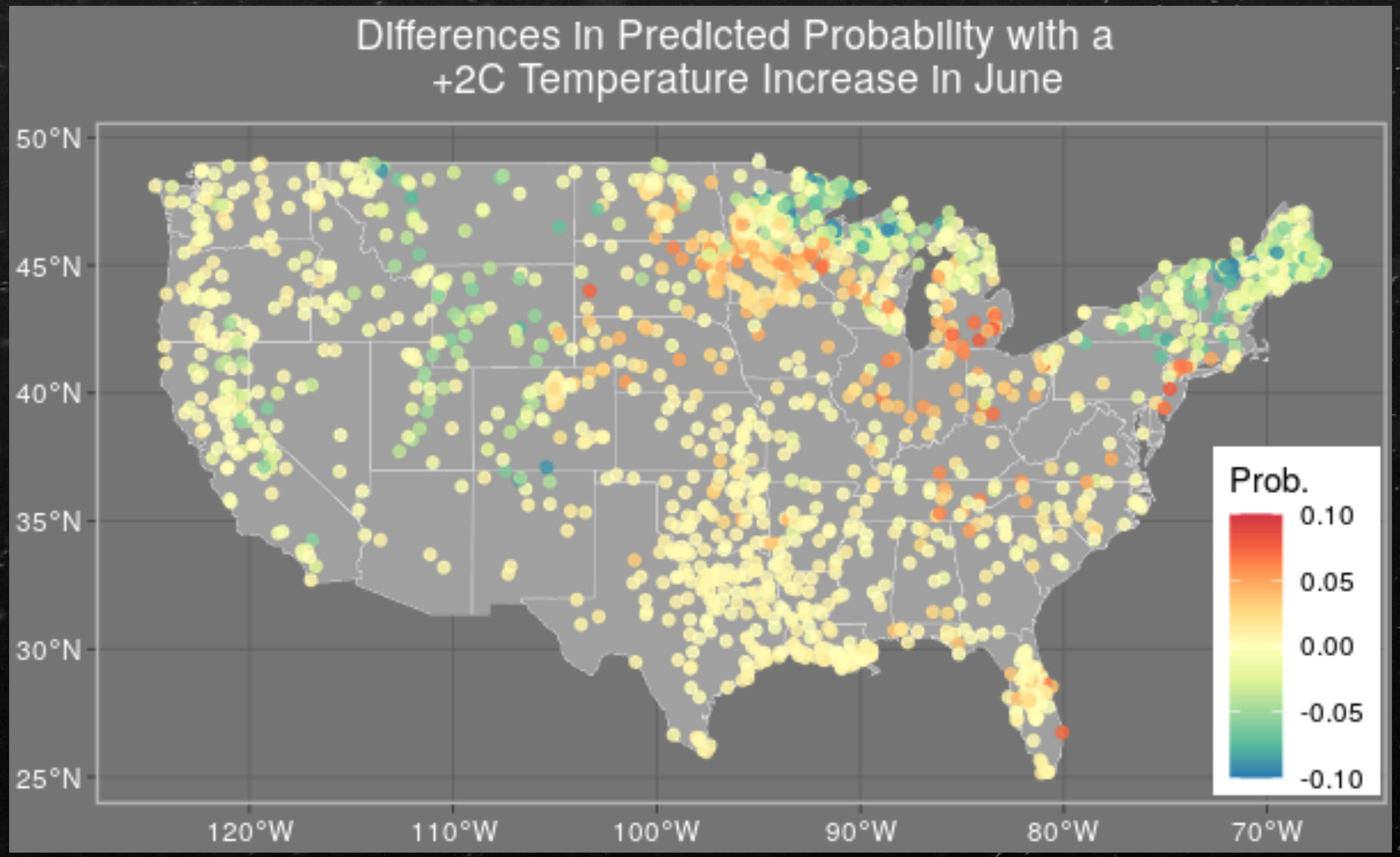
TPL

Long-term Forecasts

In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).

Long-term Forecasts

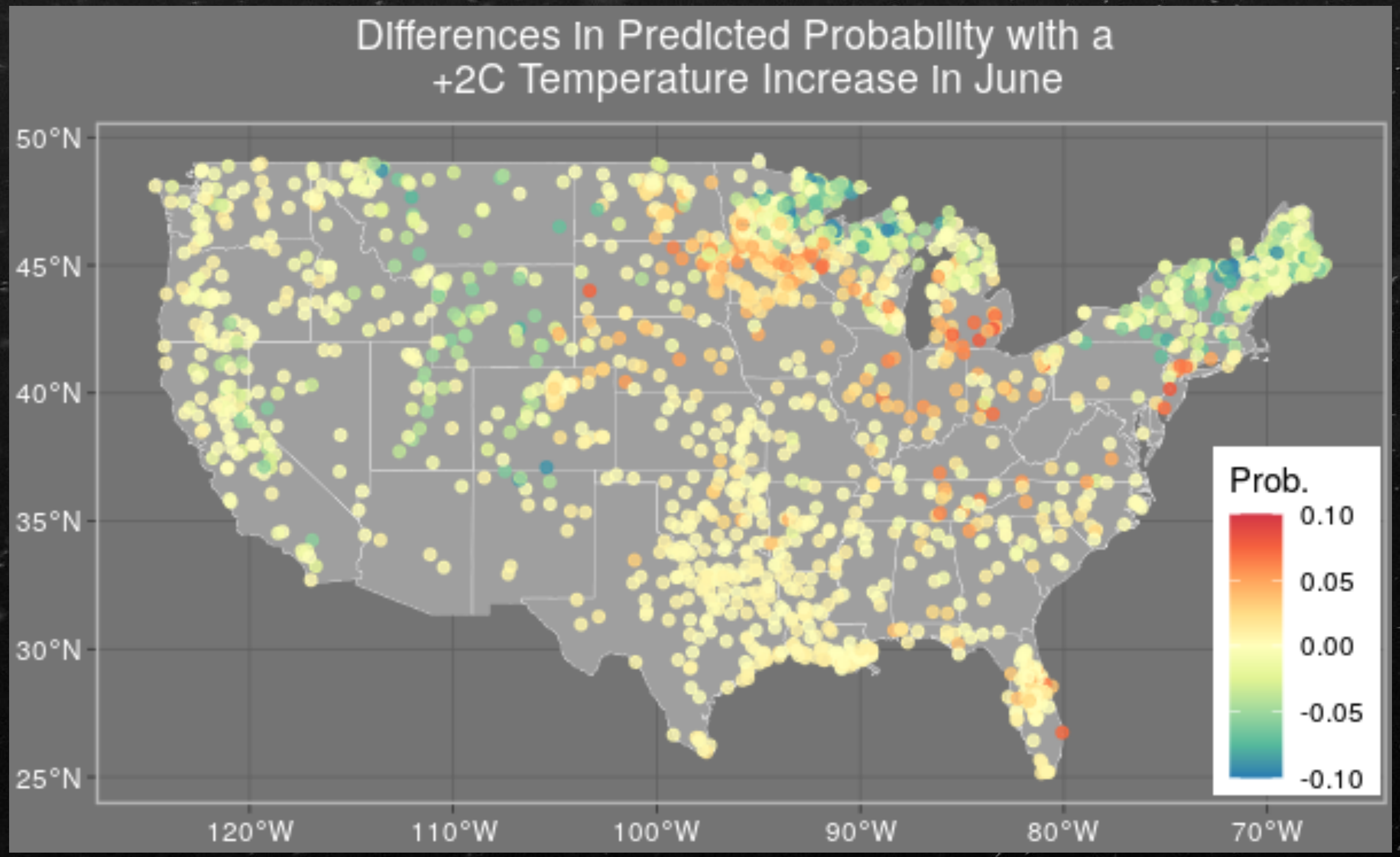
In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).



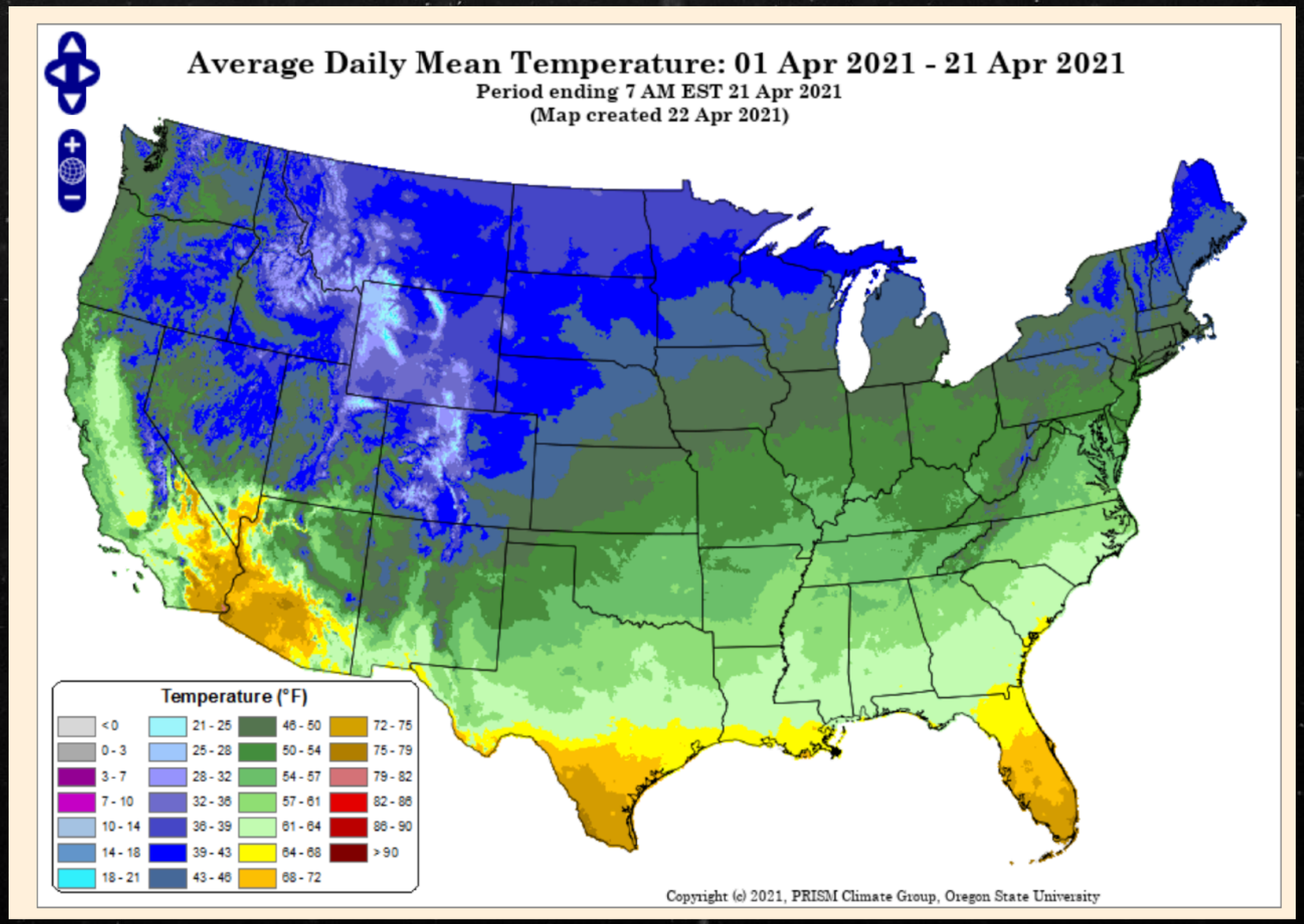
CYAN national data

Long-term Forecasts

In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).



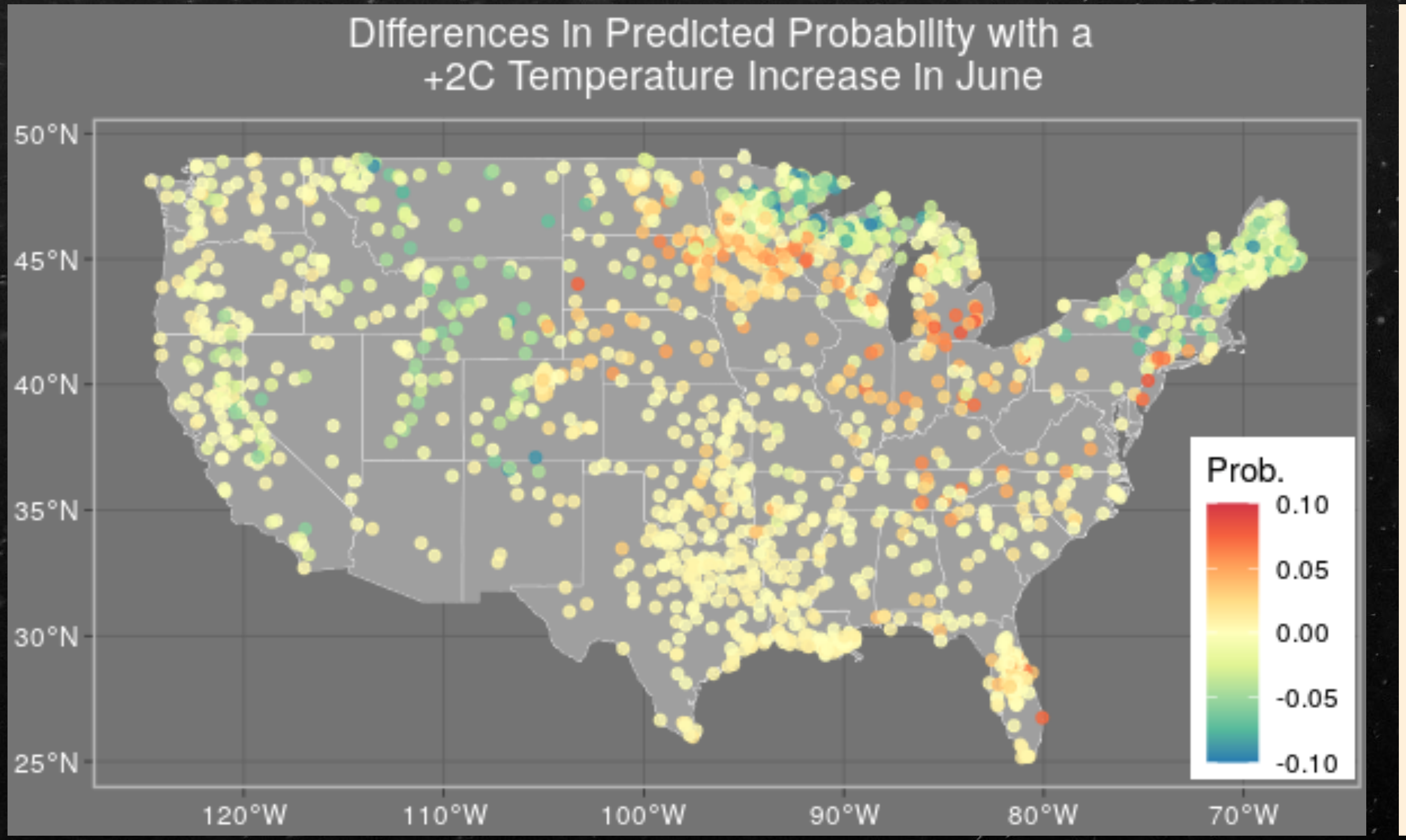
CYAN national data



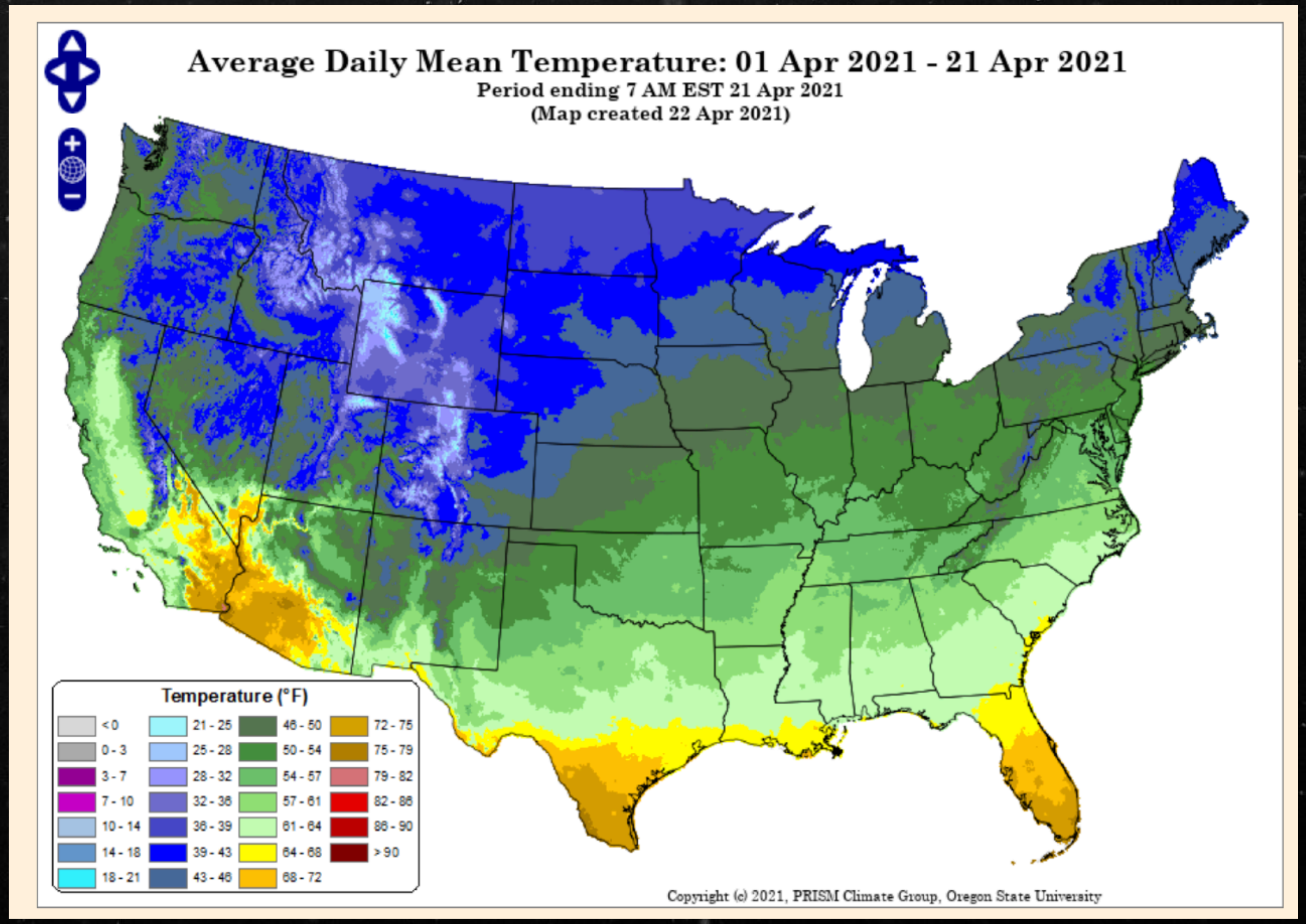
PRISM (OSU) downscaled weather/climate data

Long-term Forecasts

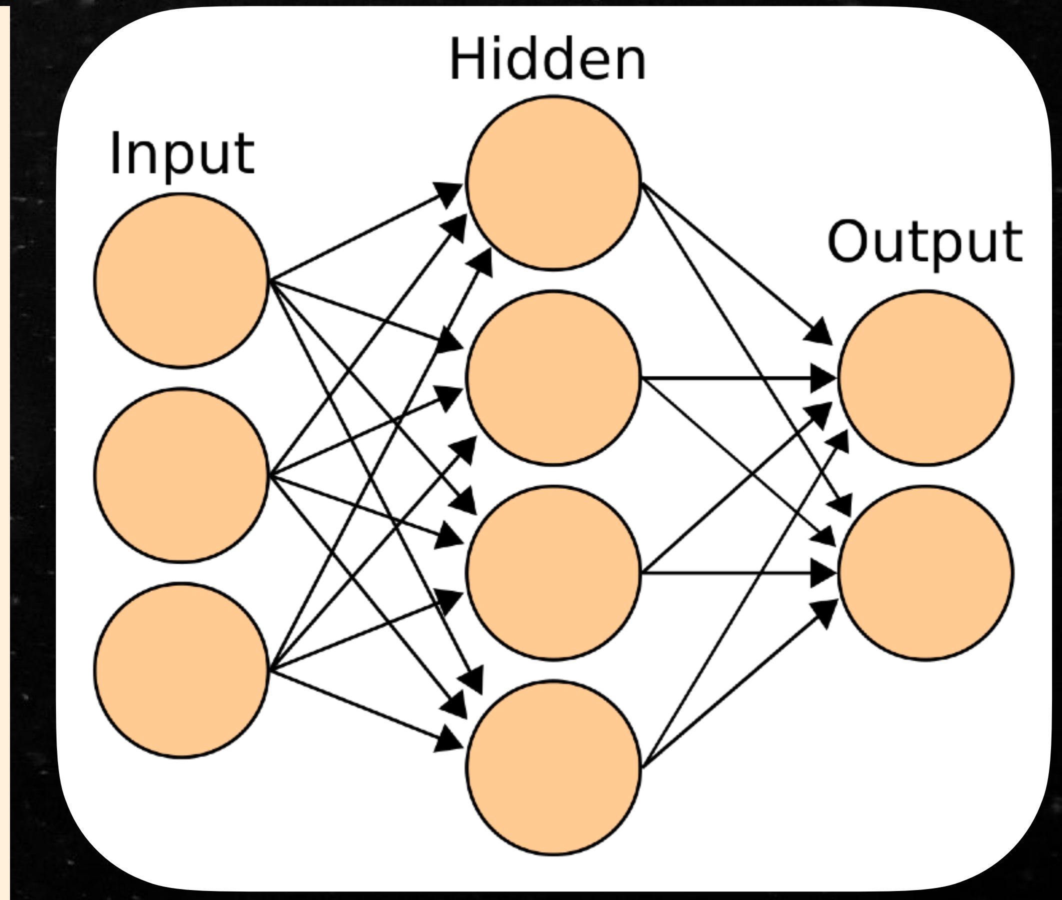
In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).



CYAN national data



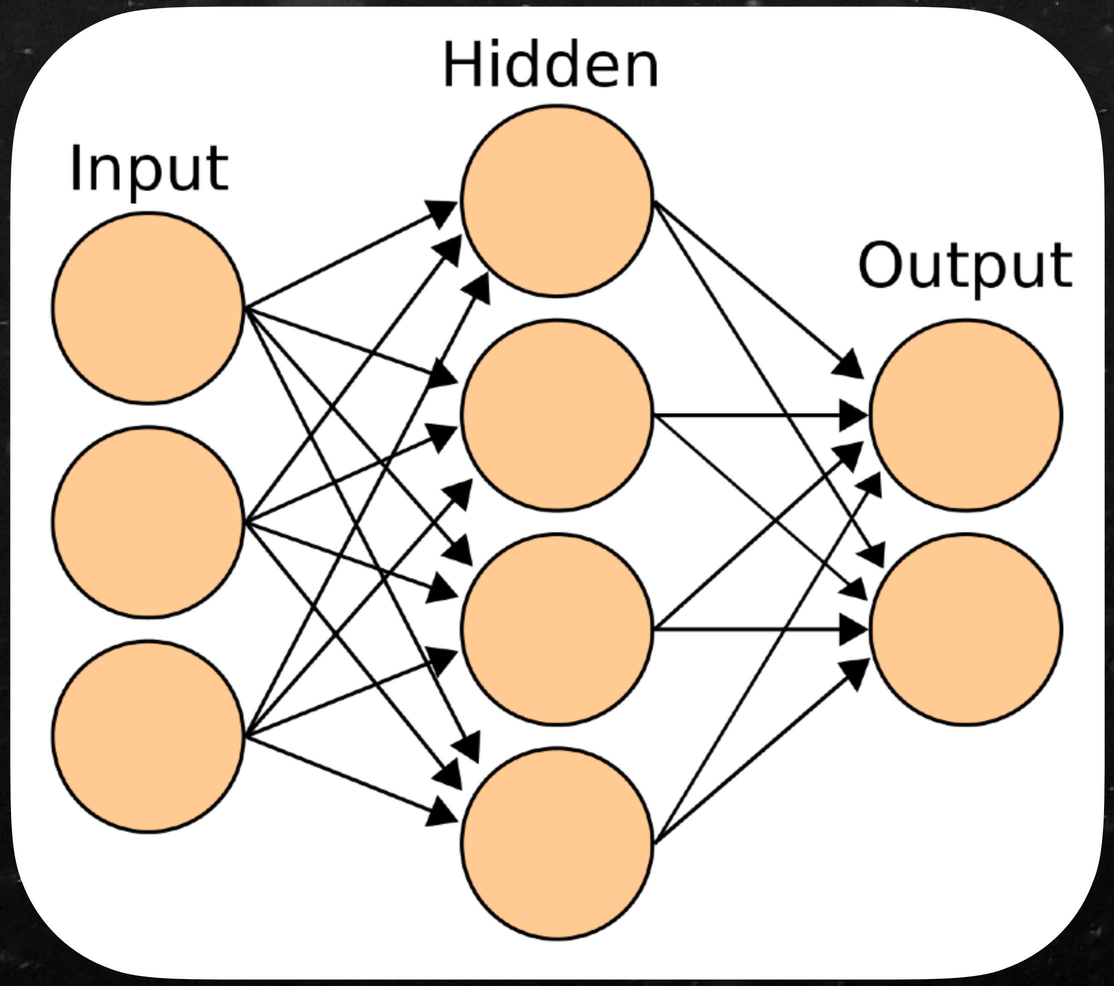
PRISM (OSU) downscaled weather/climate data



Machine learning modeling

Long-term Forecasts

In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).



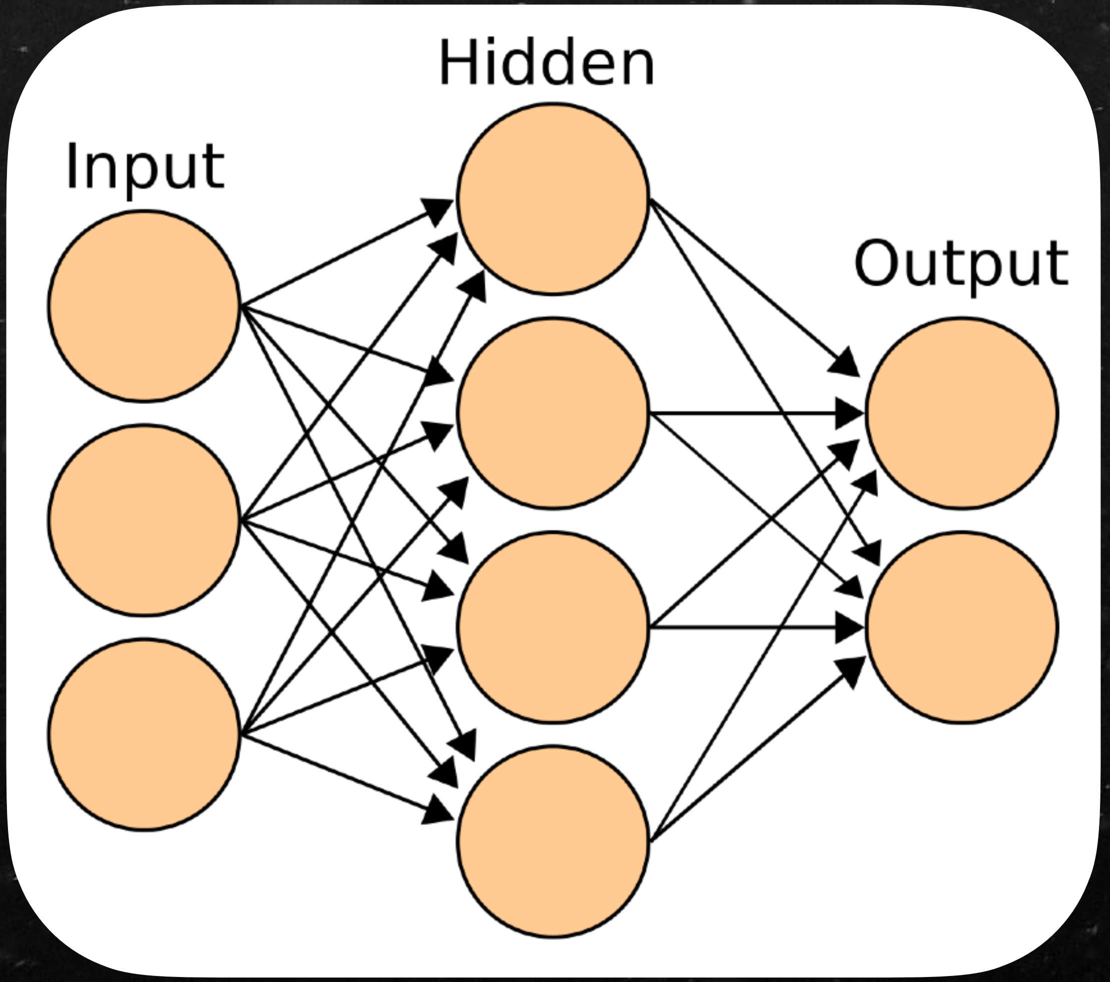
Machine learning modeling

TPL

Long-term Forecasts

In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).

Long-range weather forecasts

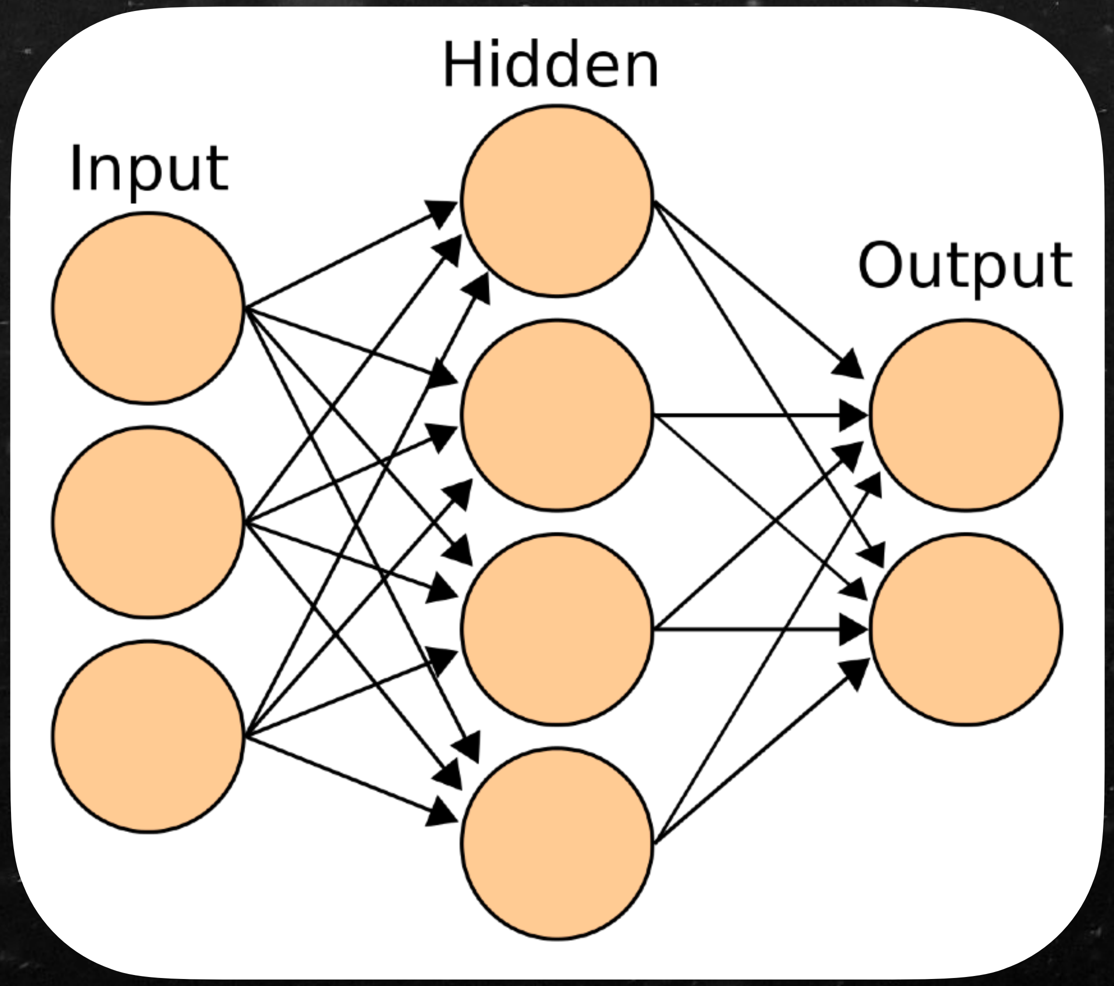


Machine learning modeling

T P L

Long-term Forecasts

In addition to operational forecasts (1-2 week timescales), we have been developing seasonal (6 month) and decadal forecasts (e.g. 2030, 2050, 2100).



Machine learning modeling

Long-range weather forecasts

CMIP6 Climate Projections

Next Gen: Hybrid Machine Learning

Purely data driven machine learning (e.g. our Bayesian Model Averaging, or Neural Nets, or Random Forests) do not explicitly respect the laws of physics or biology.

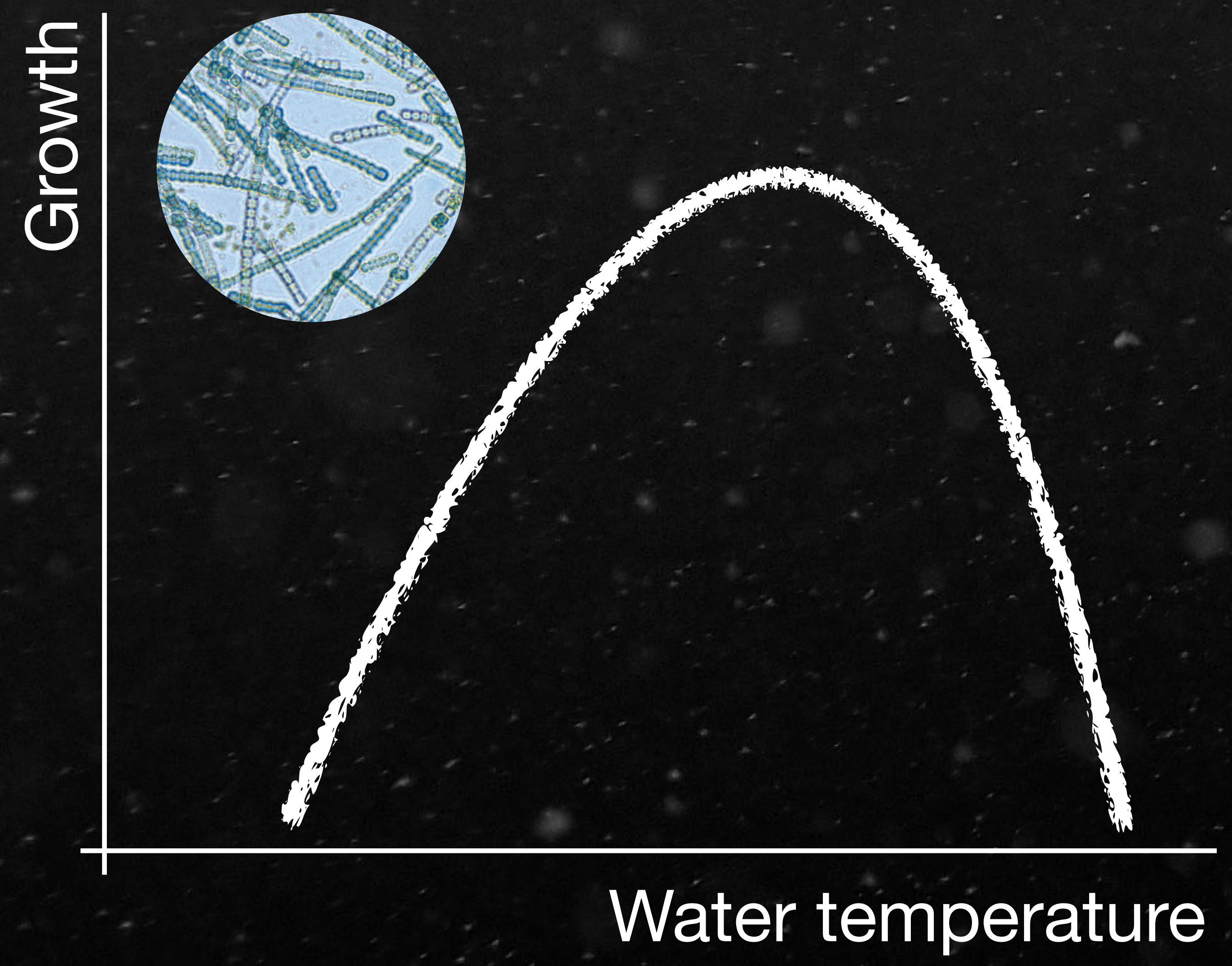
But, hybrid machine learning models do. Case study - this year!

T
P
L

Next Gen: Hybrid Machine Learning

Purely data driven machine learning (e.g. our Bayesian Model Averaging, or Neural Nets, or Random Forests) do not explicitly respect the laws of physics or biology.

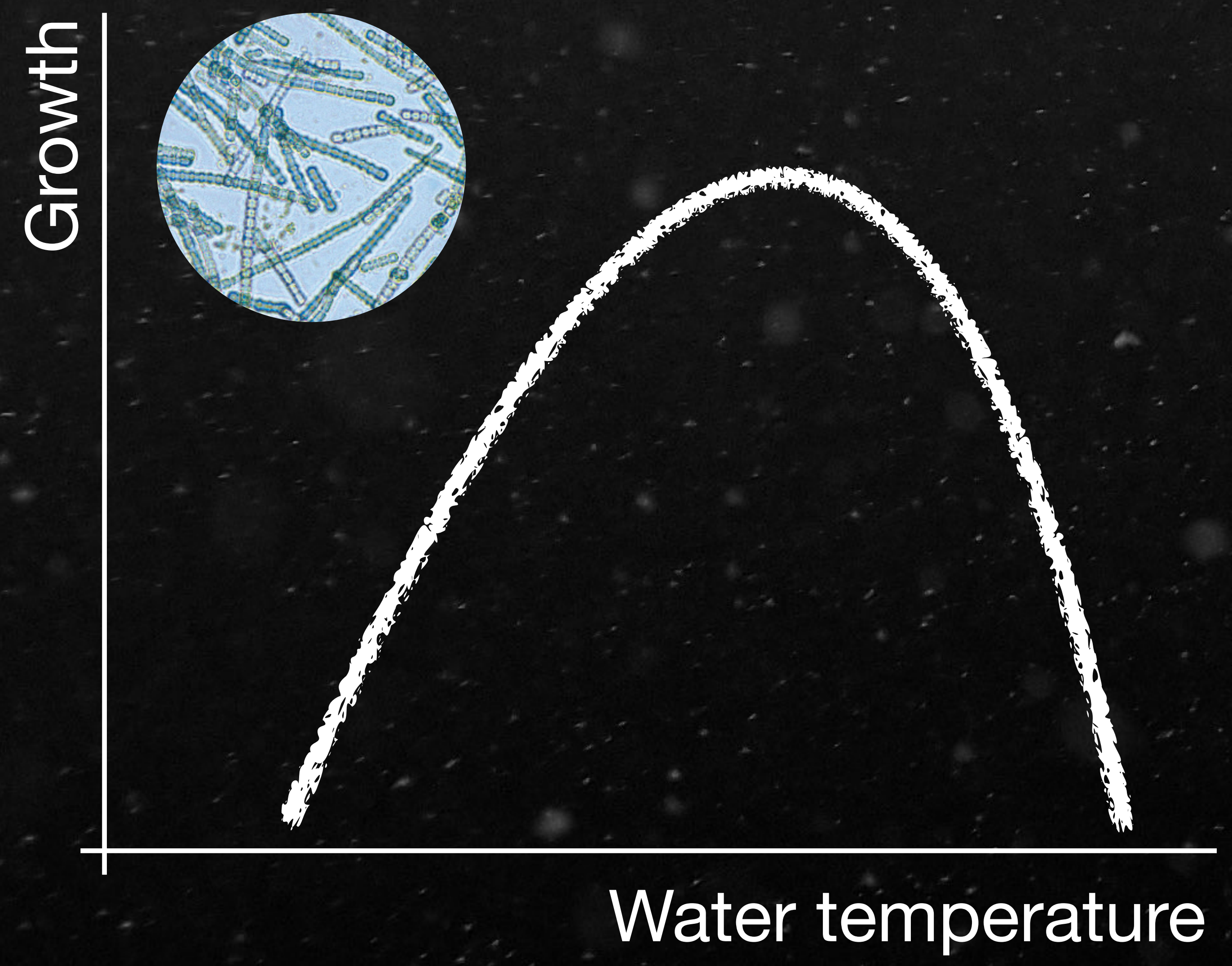
But, hybrid machine learning models do. Case study - this year!



Next Gen: Hybrid Machine Learning

Purely data driven machine learning (e.g. our Bayesian Model Averaging, or Neural Nets, or Random Forests) do not explicitly respect the laws of physics or biology.

But, hybrid machine learning models do. Case study - this year!

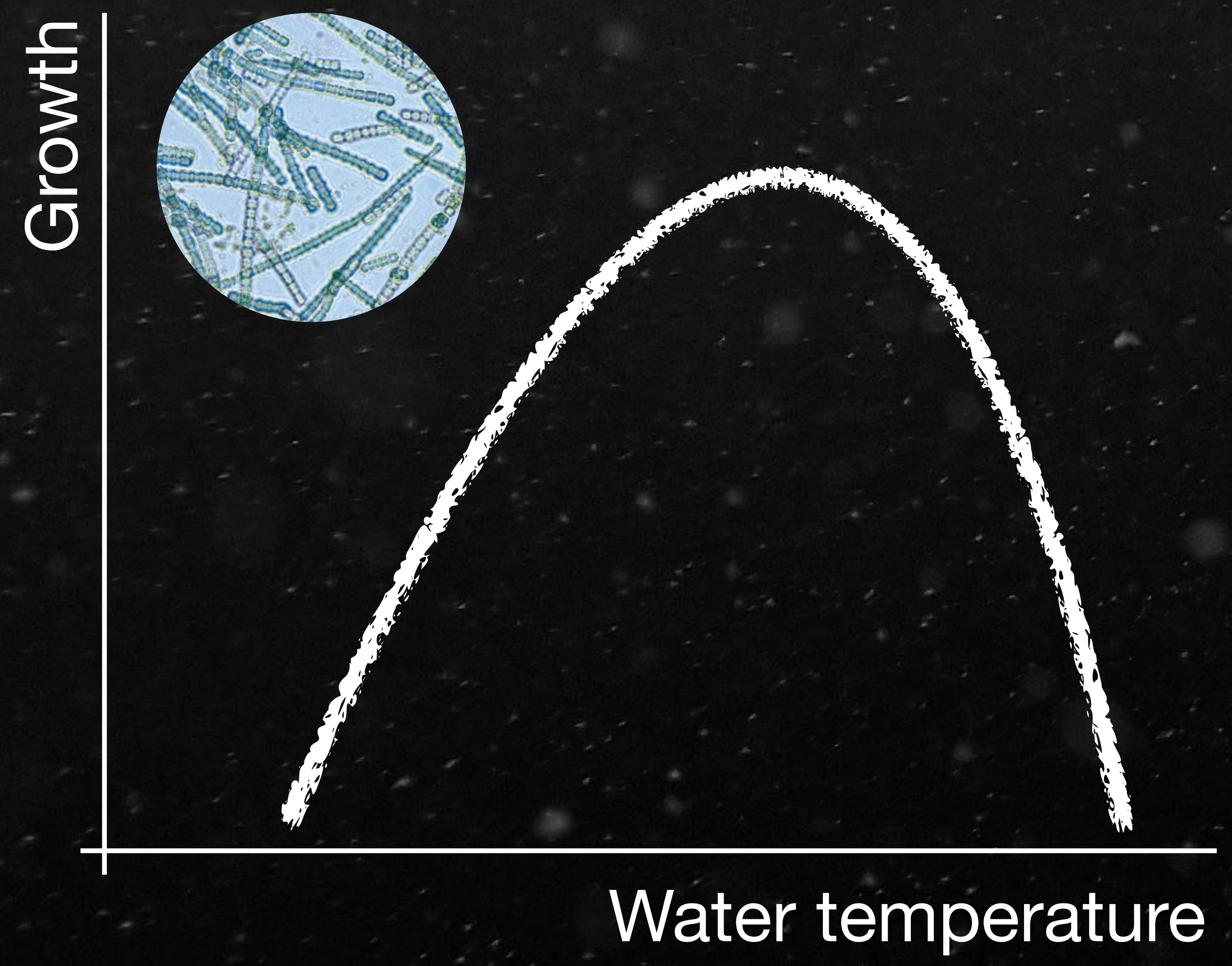


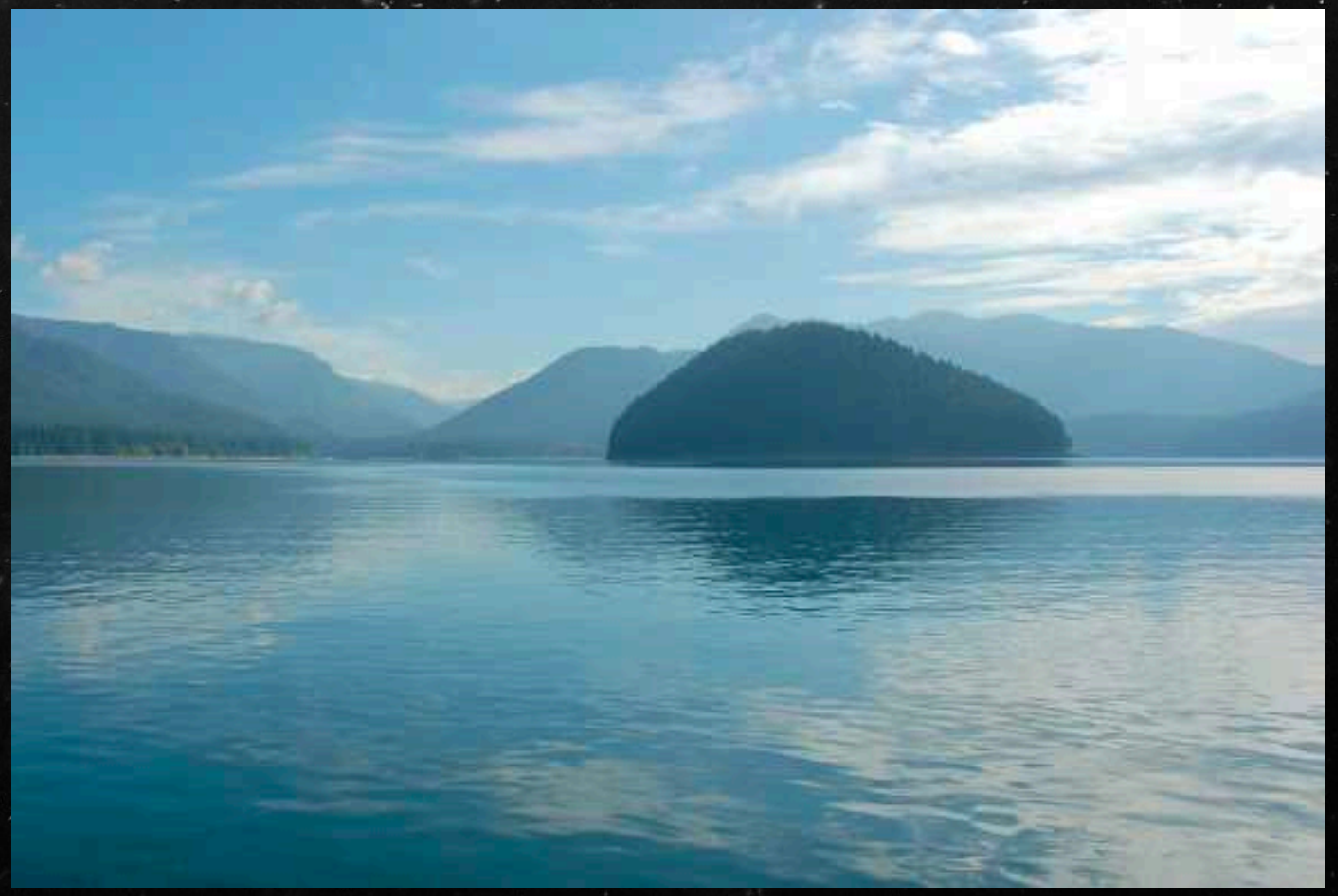
TPL

Next Gen: Hybrid Machine Learning

Purely data driven machine learning (e.g. our Bayesian Model Averaging, or Neural Nets, or Random Forests) do not explicitly respect the laws of physics or biology.

But, hybrid machine learning models do. Case study - this year!





T
P
L

Next Gen: Interpretable AI

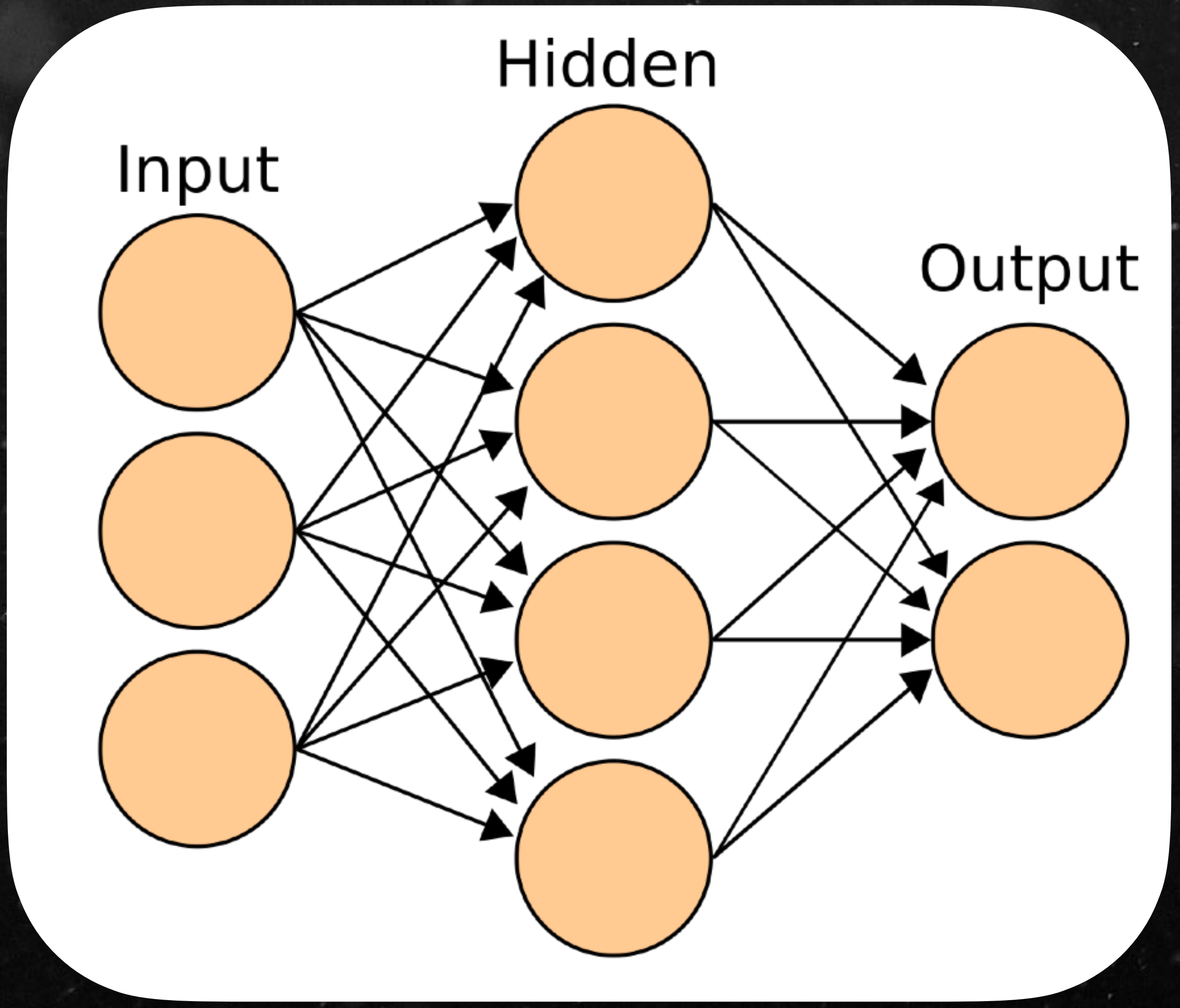
Machine learning (e.g. neural nets) offer accurate predictions, but you don't know what's going on under the hood. Interpretable AI is a class of machine learning that people can understand

The Learning Boundary

Next Gen: Interpretable AI

Machine learning (e.g. neural nets) offer accurate predictions, but you don't know what's going on under the hood. Interpretable AI is a class of machine learning that people can understand

Black box neural network



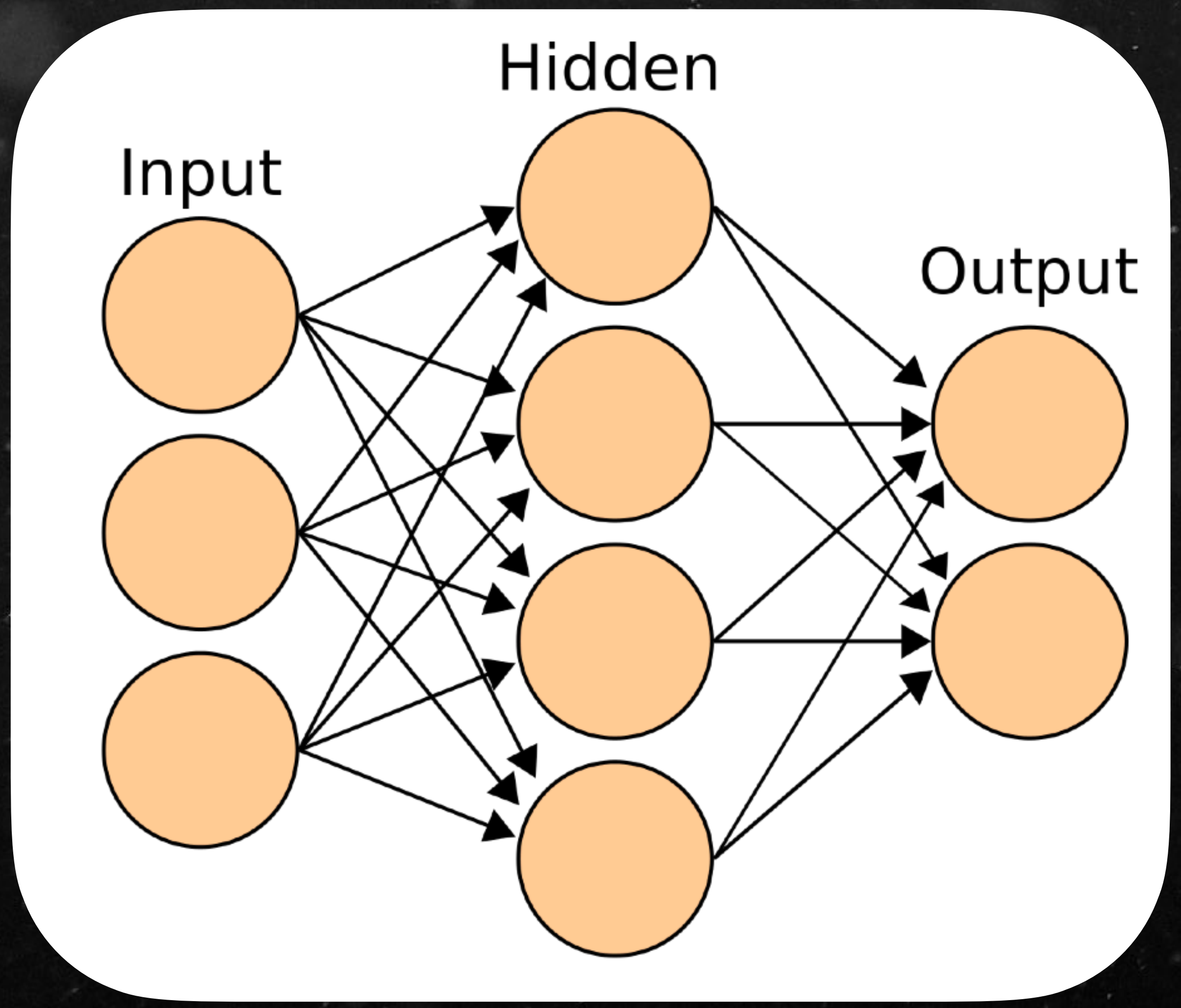
The Learning Boundary

T P L

Next Gen: Interpretable AI

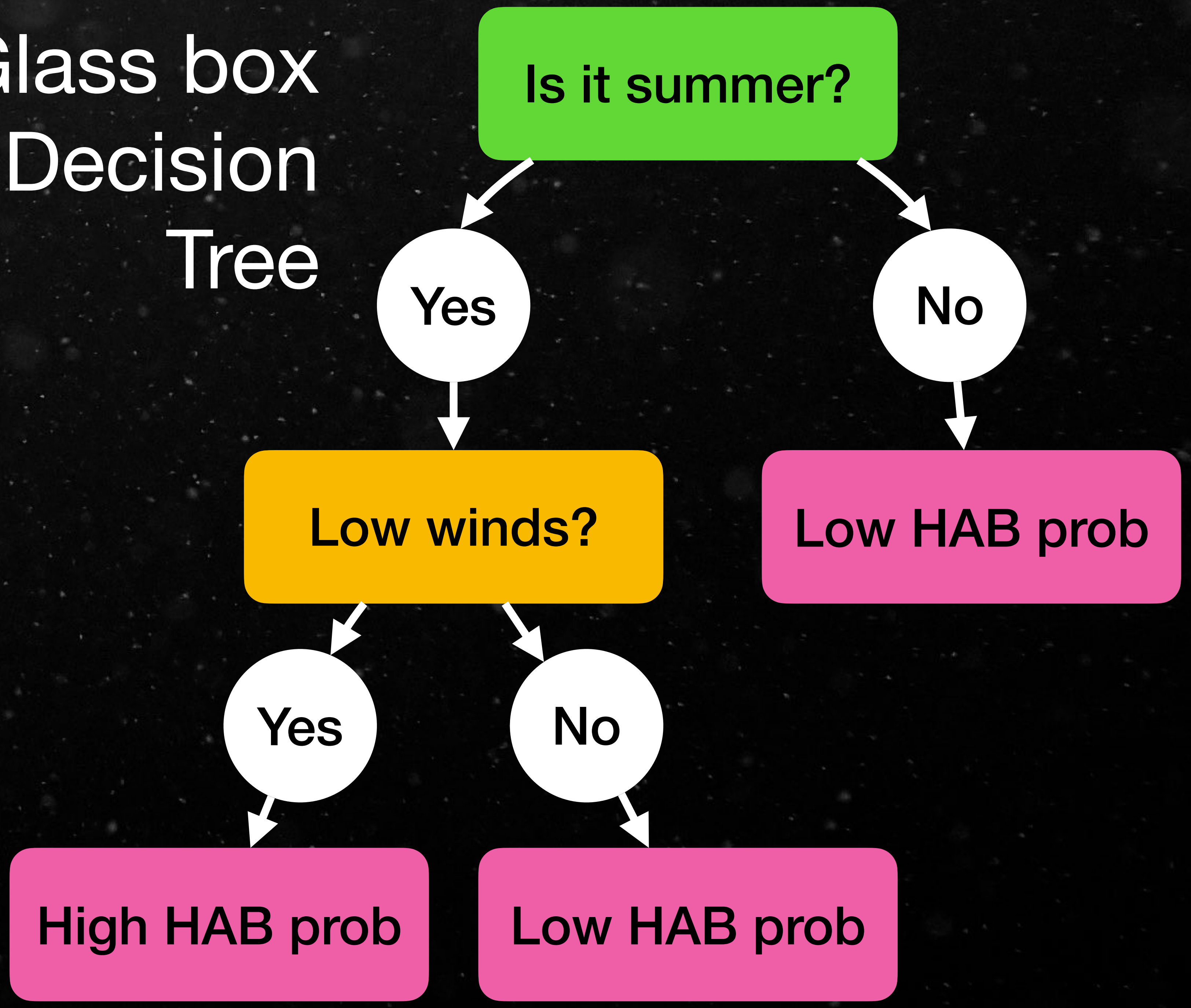
Machine learning (e.g. neural nets) offer accurate predictions, but you don't know what's going on under the hood. Interpretable AI is a class of machine learning that people can understand

Black box neural network



The Learning Boundary

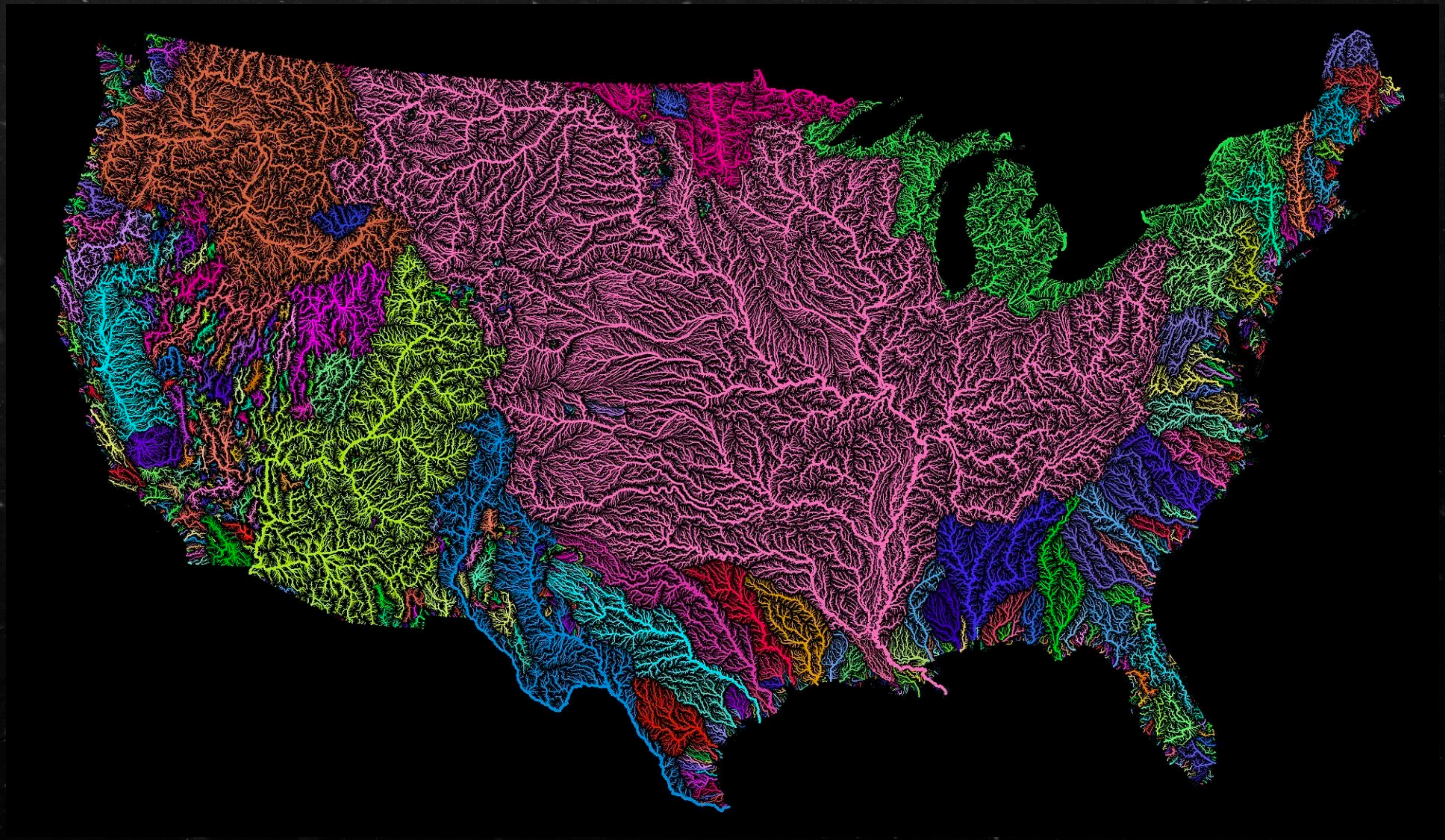
Glass box Decision Tree



T P L

Next Gen: Transfer Learning

Water quality modeling suffers from the "lots of small data" problem (i.e. its not a big data problem).



T
P
L

Next Gen: Transfer Learning

Water quality modeling suffers from the "lots of small data" problem (i.e. its not a big data problem).

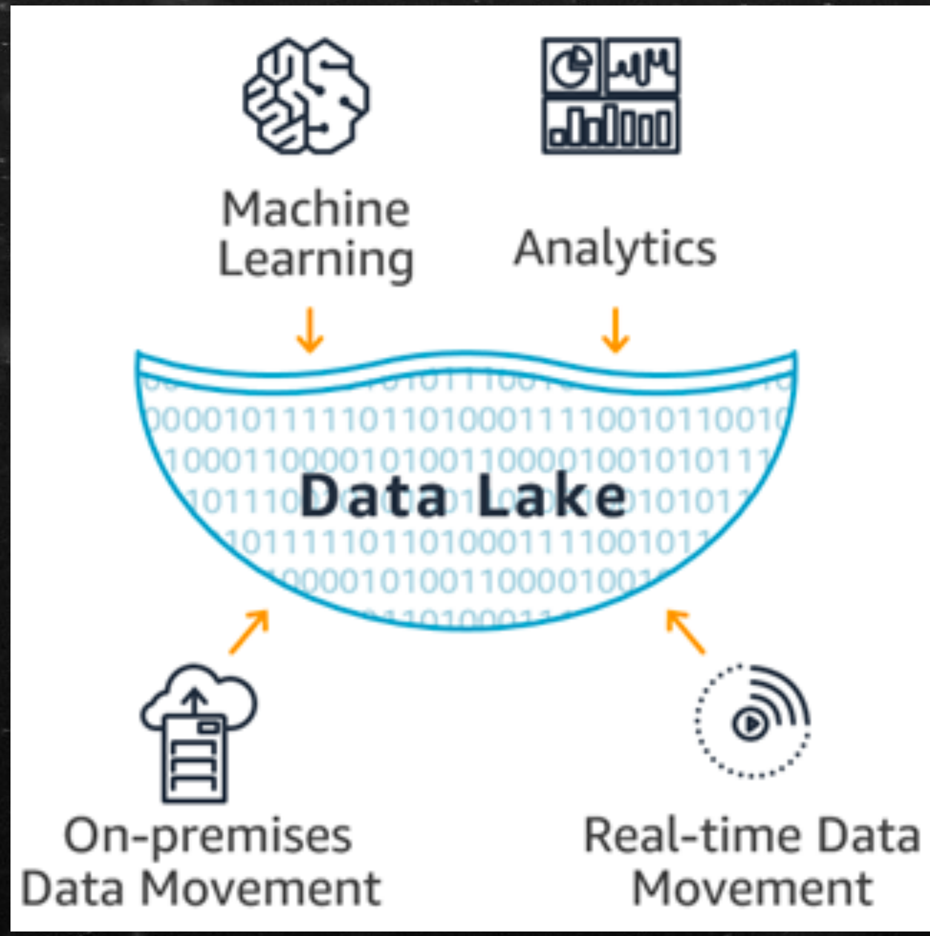
Next Gen: Transfer Learning

Water quality modeling suffers from the "lots of small data" problem (i.e. its not a big data problem).

TPL

Cloud Data Infrastructure

It's not just about modeling though. We're learning how important it is to have a unified, standardized and accessible (via API) database: AWS Data Lake + database…



Unified, standardized Data Lake and database

Collaborators in a 2021 Water Research Foundation project to develop a national water quality database for application to transfer learning

TPL

Conclusions: Measurable Usefulness

In sum, collectively we are doing really well monitoring and modeling our environments. I think the next steps are:

1) synthesize all available data in a standardized database (Cyan, water samples, qPCR, cubesat imagery, USGS… etc etc)

2) developing metrics by which we can measure the impact of our monitoring efforts and modeling.

LAGOS LAke multi-scaled GeOSpatial and temporal database

Publicly accessible lake water quality and ecological context data for the US

University of Michigan

In Summary:

- Multimodal data synthesis and open access
- Models: interpretable AI and transfer learning
- Predictions: at weekly, monthly and decadal timescales
- Measurable impacts

The Prediction Lab

james@thepredictionlab.com

T
P
L