Macrosystems EDDIE:
Using Data to Improve Ecological Forecasts

Lofton, M.E., Moore, T. N., Thomas, R. Q., Carey, C.C. 07 March 2024.
Macrosystems EDDIE: Using Data to Improve Ecological Forecasts.
Macrosystems EDDIE Module 7, Version 1.
https://serc.carleton.edu/eddie/macrosystems/module7
Module development supported by NSF DEB-1926050; NSF DBI-1933016

Macrosystems EDDIE
Environmental Data-Driven Inquiry & Exploration in Macrosystems Ecology

NSF | neon Operated by Battelle

SERC the Science Education Resource Center at Carleton College

Overview of today

- Introduce the concepts of ecological forecasting, forecast accuracy and uncertainty, and data assimilation
- **Activity A:** Access and explore data from an ecological site of your choice in the National Ecological Observatory Network, then fit a model and generate a forecast of lake chlorophyll-a
- **Activity B:** Explore how updating model predictions with data affects forecast accuracy, including the effects of data observation uncertainty and temporal frequency
- **Activity C:** Explore the effect of assimilating sensor data at different frequencies on management decision-making.

Ecosystems are changing worldwide…

- In response to changes in climate and land use, ecological functioning in many aquatic and terrestrial systems is changing
- Lakes and reservoirs provide many ecological services; thus, understanding how these ecosystems will change in the short-term will help us better manage these vital resources
- Ecological forecasting is a potentially powerful tool to help lake and reservoir managers preemptively prevent or mitigate water quality concerns

Image: Wikimedia commons

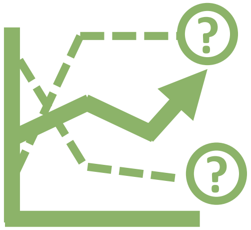
Before we start:

What is an Ecological Forecast?

A prediction of future environmental conditions with uncertainty

- Events have not yet occurred
- Gives a probability or a likelihood of the event to occur (uncertainty)
- Actionable

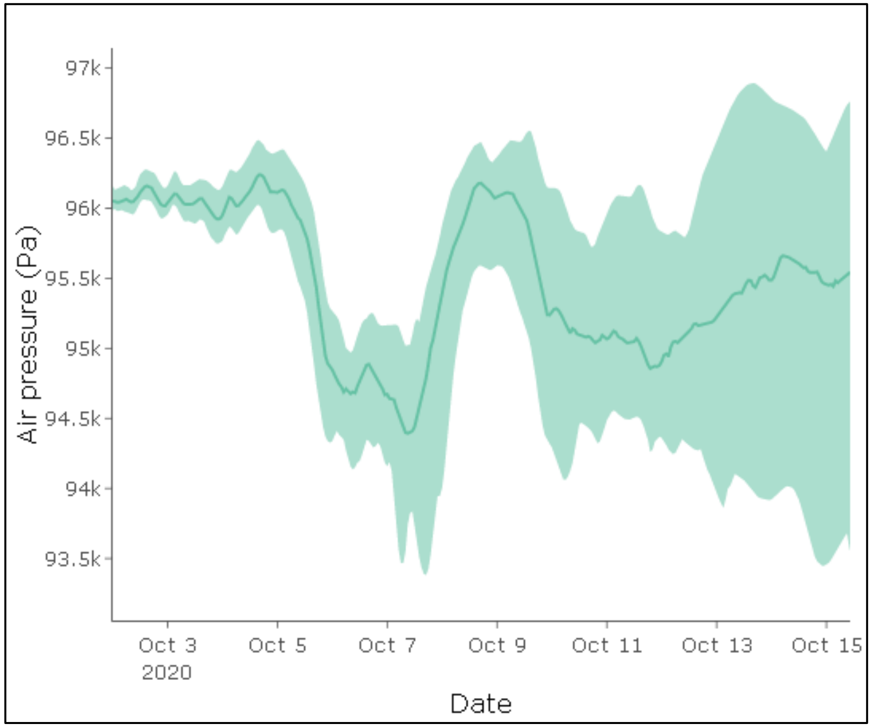
What is **uncertainty?**

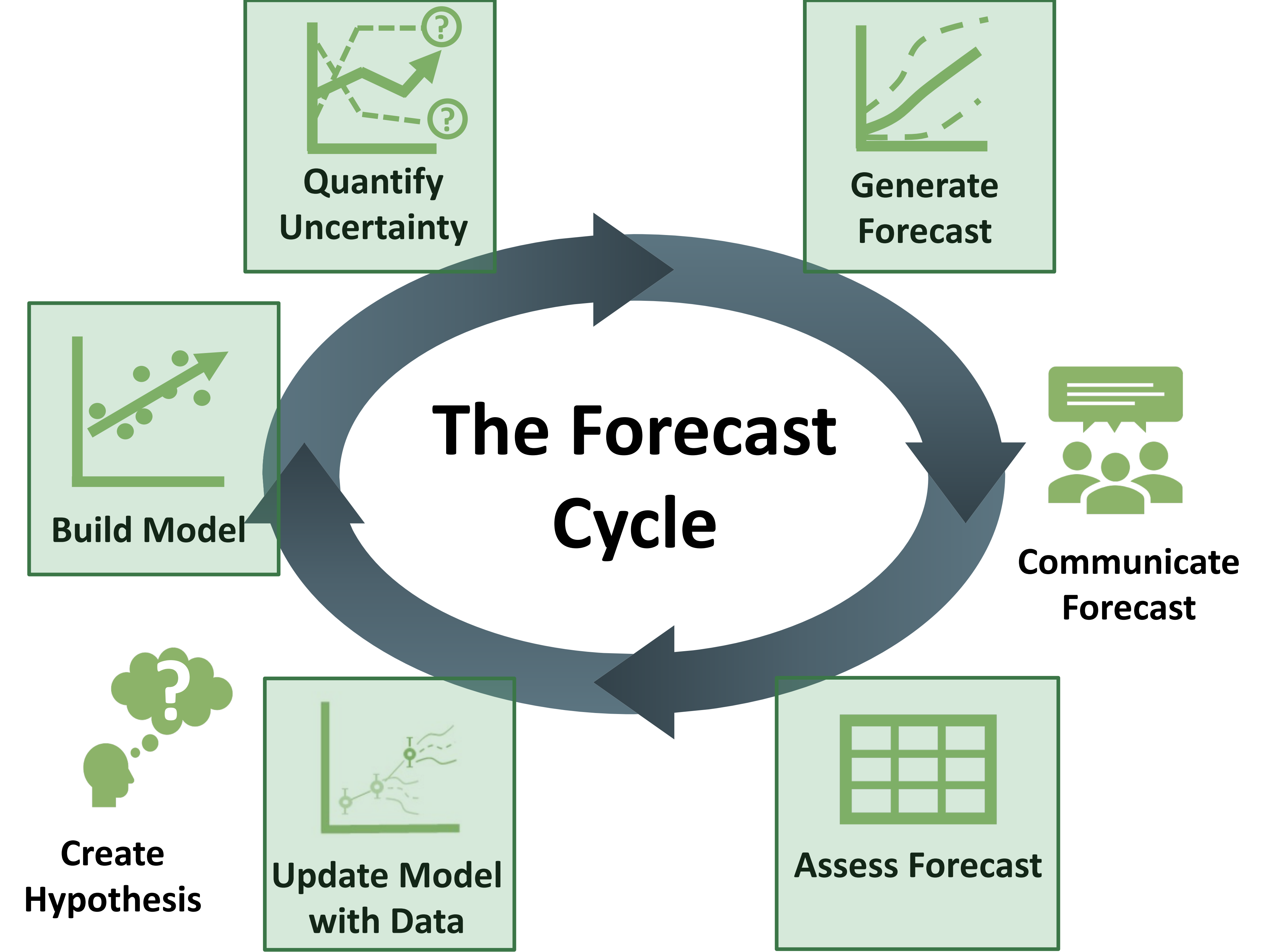


Quantify Uncertainty

- **<u>Forecast uncertainty:</u>** the set of possible alternate future conditions predicted by a model and the chance of observing each outcome.
- We generate multiple different predictions of the future because the future is inherently unknown.
- Uncertainty generally increases as the **forecast horizon** increases.
  - **Forecast horizon:** the length of time into the future for which a forecast is generated

Plot showing 16-day-ahead forecast of air pressure with shaded regions showing 95% confidence interval and the solid line represents the median

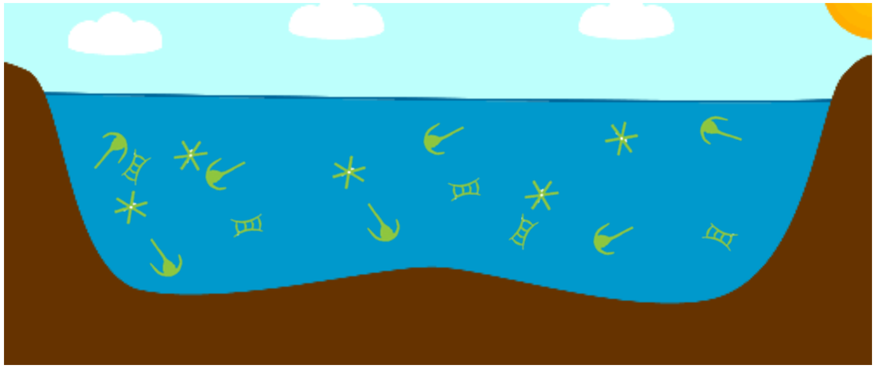



The Forecast Cycle
Quantify Uncertainty
Generate Forecast
Communicate Forecast
Assess Forecast
Update Model with Data
Create Hypothesis
Build Model

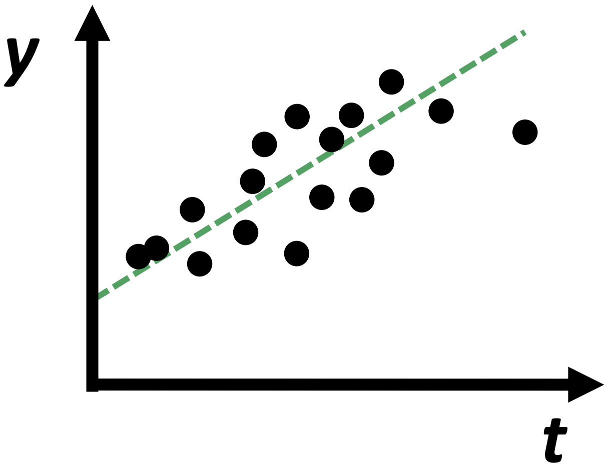
Ecological models

A model is a simplified representation of a real phenomenon, with the goal of understanding and **predicting** that phenomenon

Predicting chlorophyll-a concentrations in a lake



$$y = mx + b$$

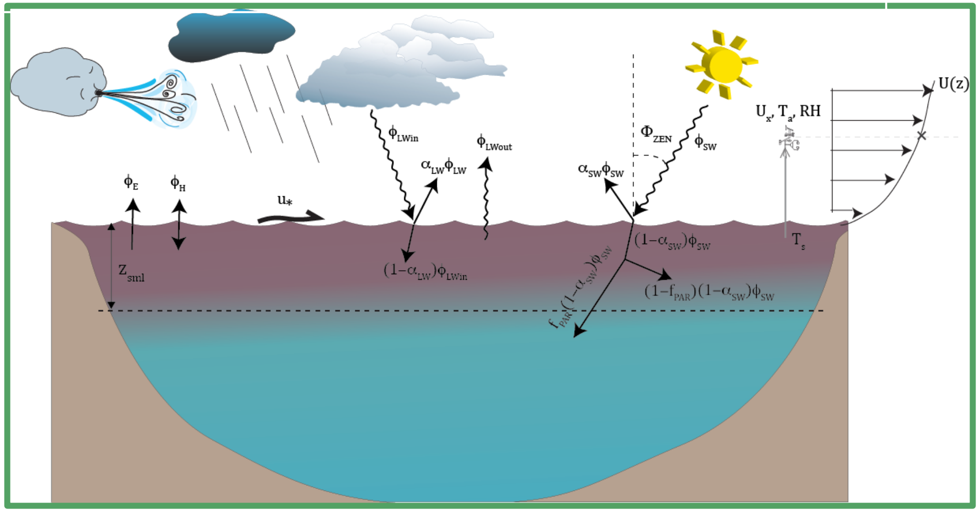


Models as imperfect representations of complex ecosystems

- Models for ecological forecasting can be thought of as imperfect representations of the ecosystems they simulate

Lake

Lake Model (imperfect representation of real lake)



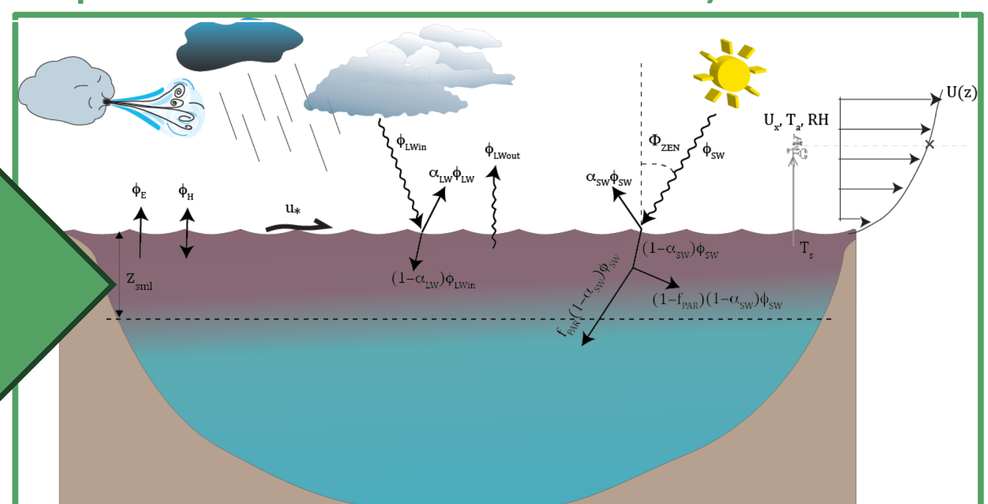
Model diagram courtesy of the General Lake Model (GLM) developers
https://aed.see.uwa.edu.au/research/models/glm/#

Models as imperfect representations of complex ecosystems

- Models for ecological forecasting can be thought of as imperfect representations of the ecosystems they simulate
- How do we make sure the model lake resembles the actual ecosystem? (e.g., the algal bloom in the lake appears in the model)

Lake

Lake Model (imperfect representation of real lake)



Model diagram courtesy of the General Lake Model (GLM) developers
https://aed.see.uwa.edu.au/research/models/glm/#

Important note: data are also imperfect!

- It can be tempting to view data as “the truth”
- But observational data are imperfect too:
  - Human error
  - Sensor malfunction
  - Limits to measurement accuracy
- We must account for uncertainty in data
- Ultimately, model predictions and data BOTH have value in producing accurate forecasts

A fouled chlorophyll-a sensor.
Photo credit: Adrienne Breef-Pilz

Our focal question for today:

How can we use data to improve ecological forecasts?

Updating forecasts with data

- **<u>Data assimilation:</u>** the process of updating models with data
- In ecological forecasting, data assimilation is the process of updating ecological forecasting models with new environmental data as they become available

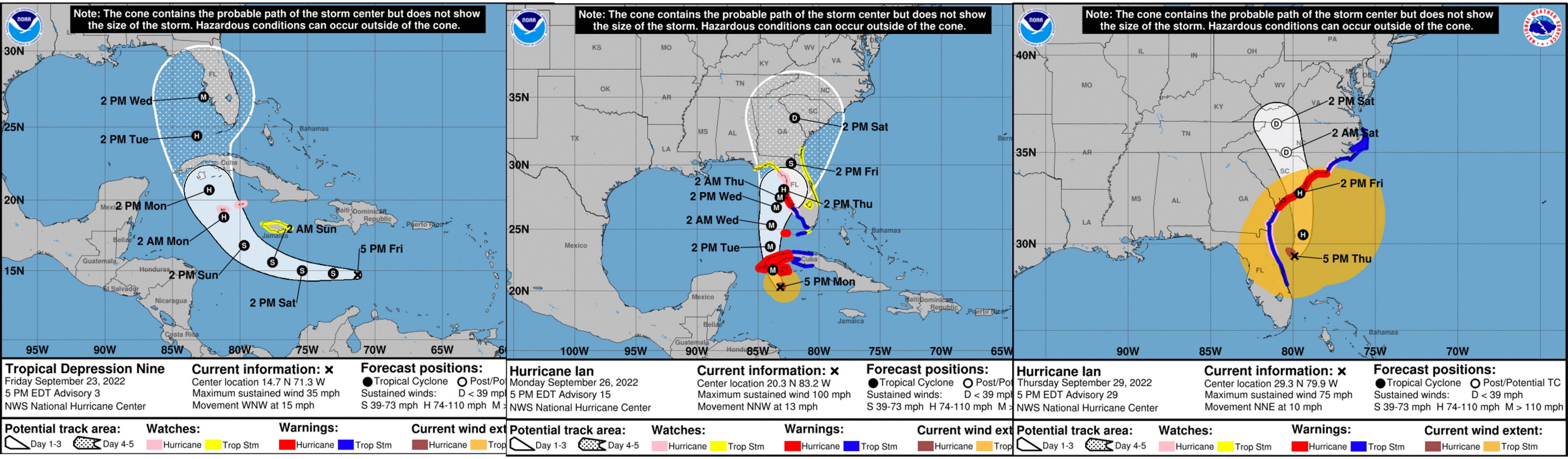


Image credit: US NOAA

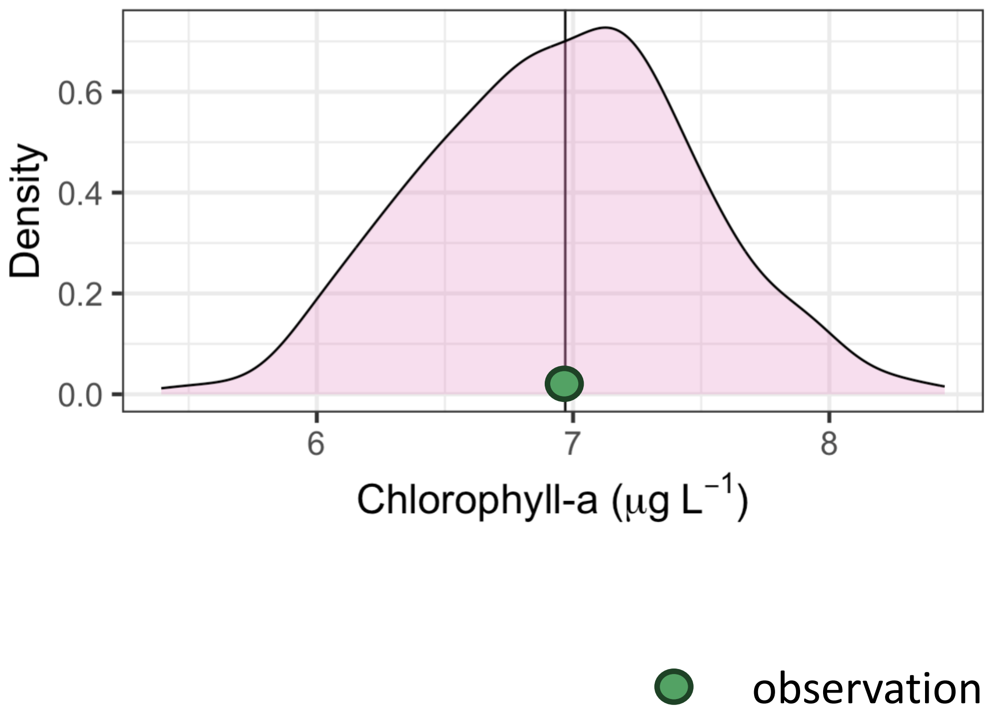
Steps of forecasting with data assimilation

Step 1: Generate a forecast with uncertainty → Step 2: Update the forecast using data → Step 3: Generate a new forecast

Using Distributions to Represent Uncertainty in Forecasts

Step 1: Generate a forecast with uncertainty

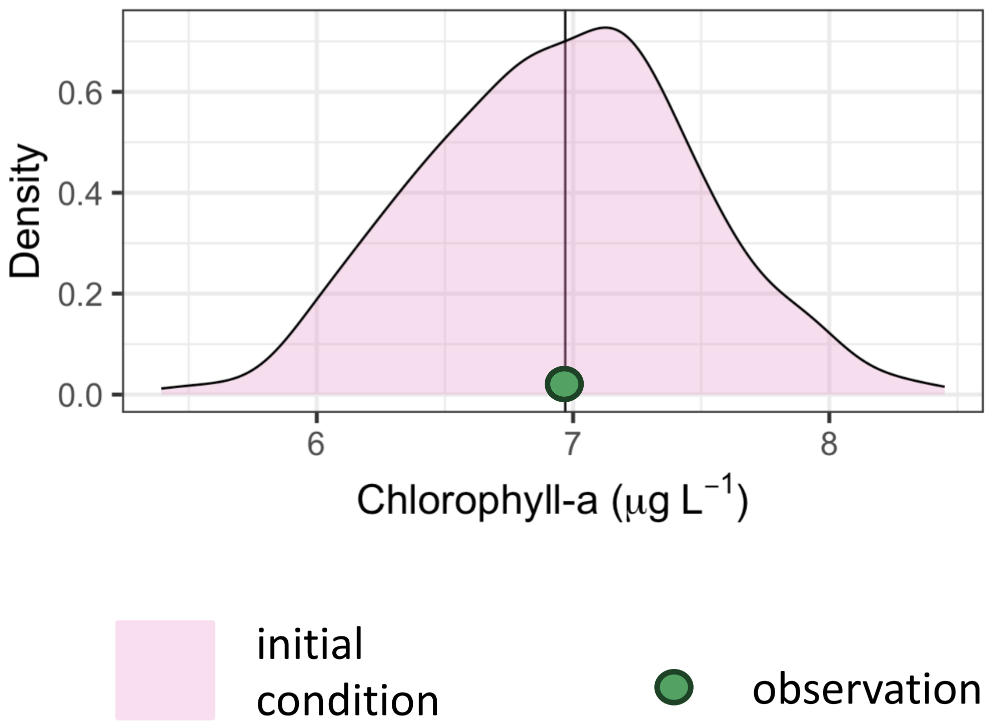
- **How do we represent uncertainty in a forecast?**
- Uncertainty can be represented as a **distribution** of all possible values of a variable
- Here, uncertainty is represented as a normal distribution around the mean (observed value) of 6.95 ug/L.



Using Distributions to Represent Uncertainty in Forecasts

Step 1: Generate a forecast with uncertainty

- Forecasts are generated using the most recent observational data to inform the **initial condition (starting condition)** for the forecast

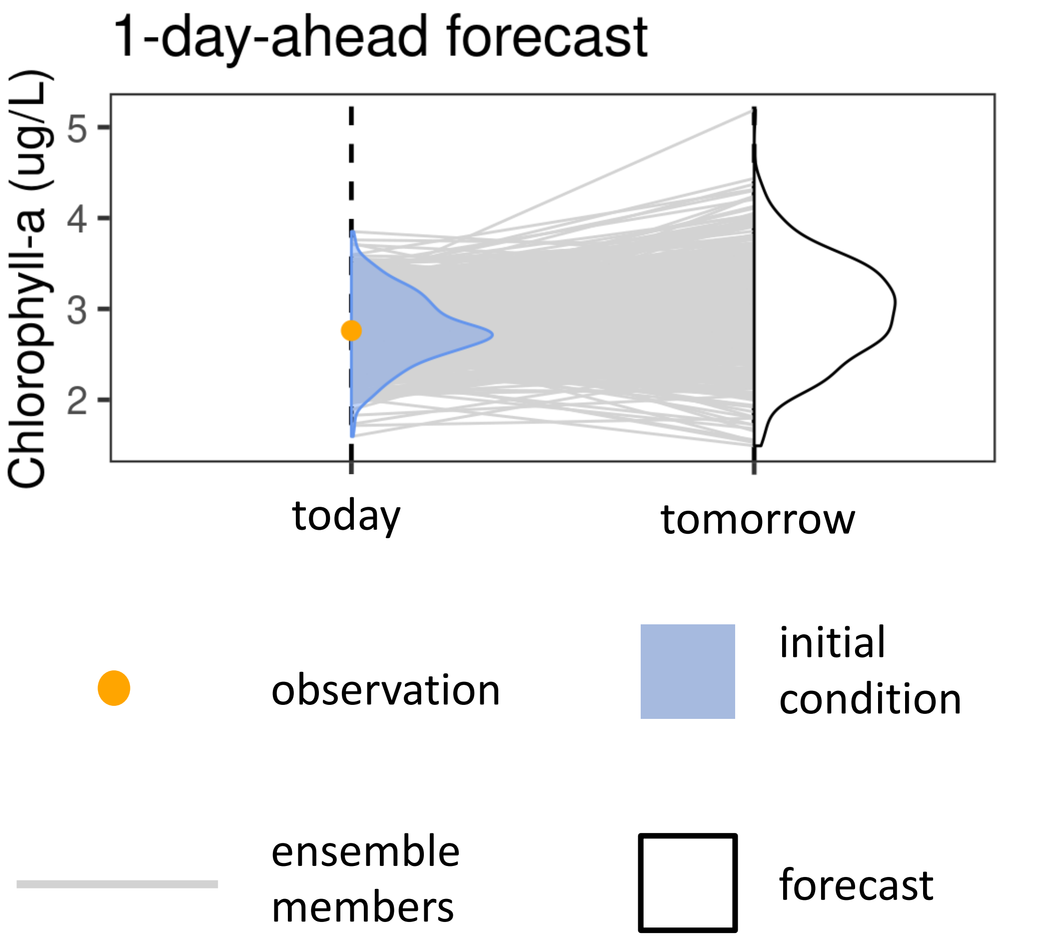


Generate a forecast

Step 1: Generate a forecast with uncertainty

Then, instead of running just a single forecast, we take many draws from our distribution and run a model many times using slightly different conditions.

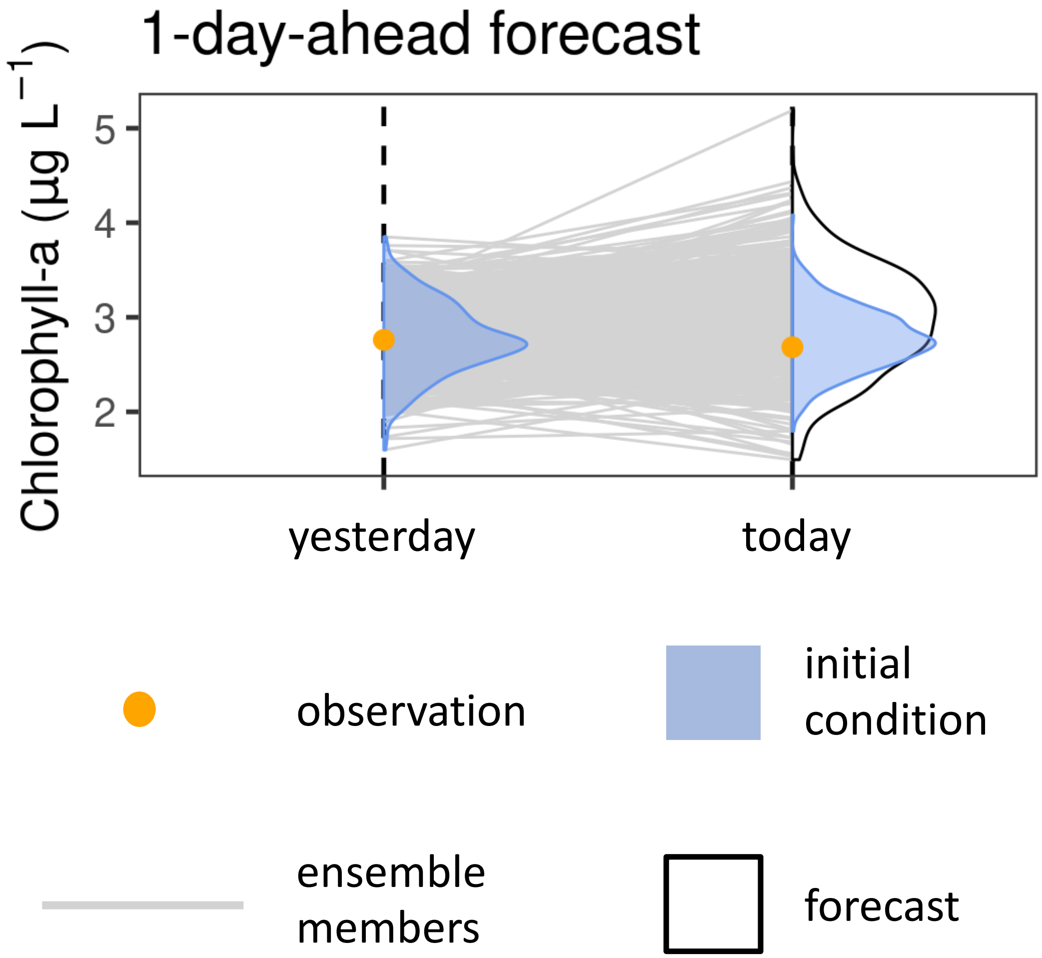
- The complete set of forecasts is referred to as the **ensemble.**
- Individual forecasts within it are **ensemble members.**
- Commonly, each member of the ensemble is **equally likely** to occur.
- **This allows us to generate a forecast with uncertainty!**



Updating initial conditions

Step 2: Update the forecast using data

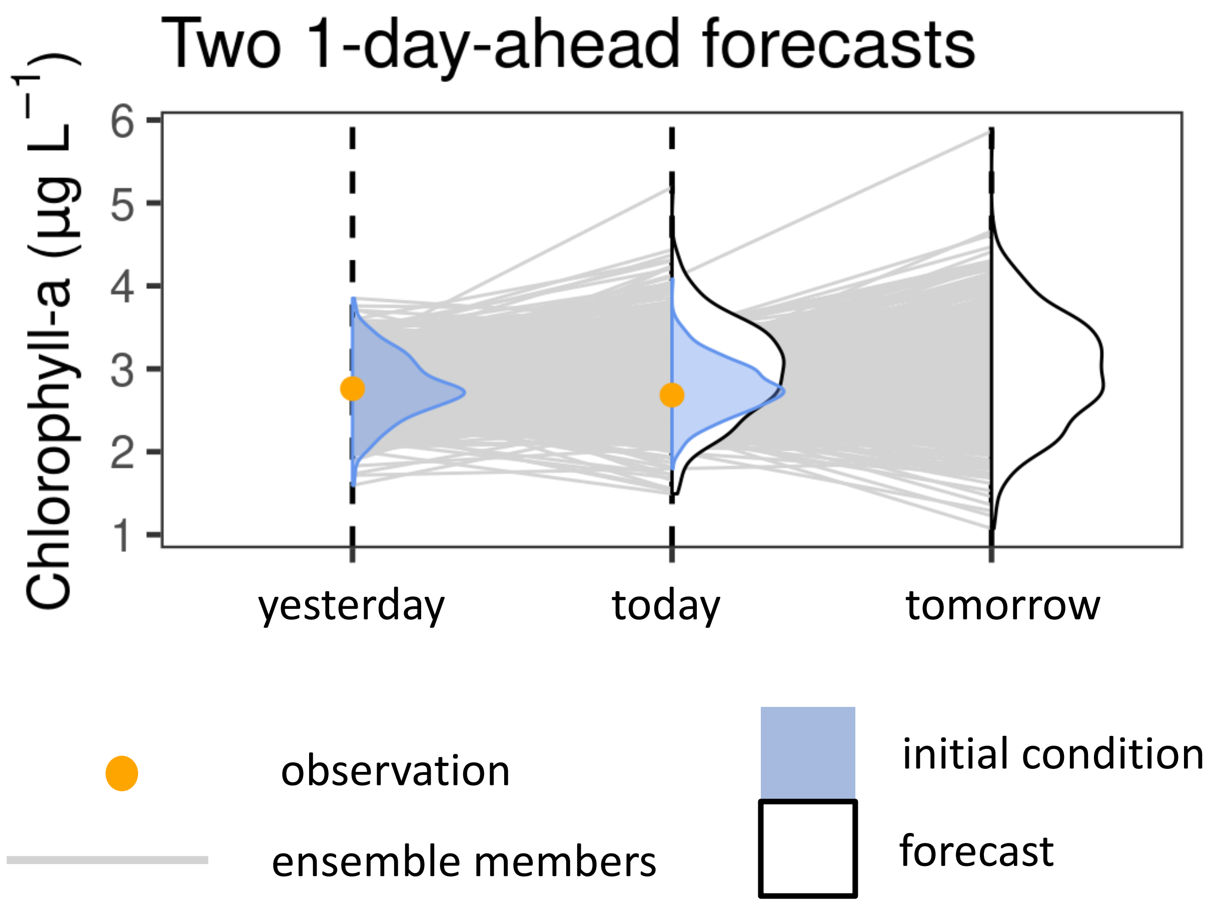
- New data are used to update the **initial condition**
- This updating can be done using a technique called an **ensemble Kalman filter**, which accounts for uncertainty in both model predictions and new observations



Generate a new forecast

Step 3: Generate a new forecast

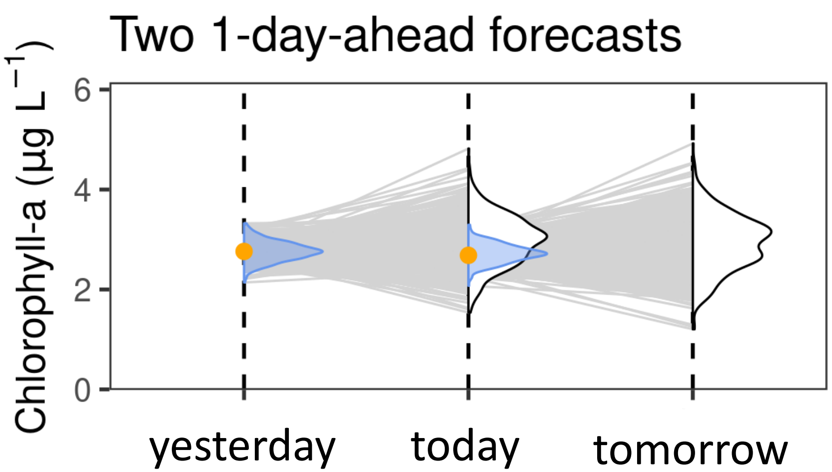
- A new forecast is generated using the updated initial condition



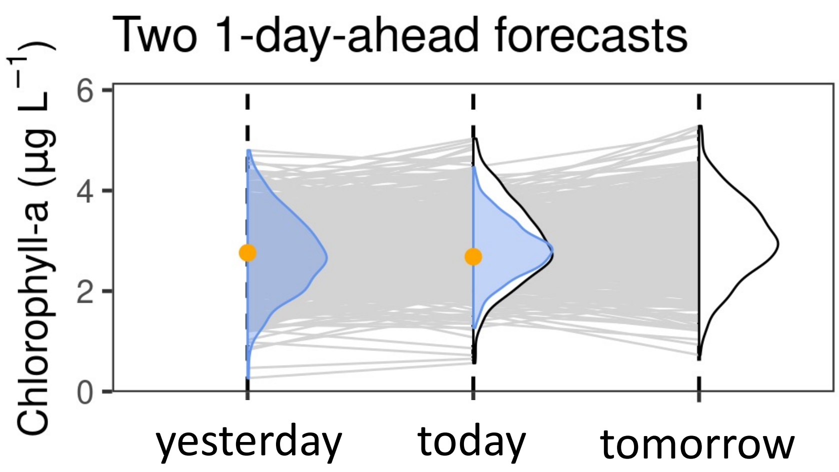
Uncertainty in models and observations

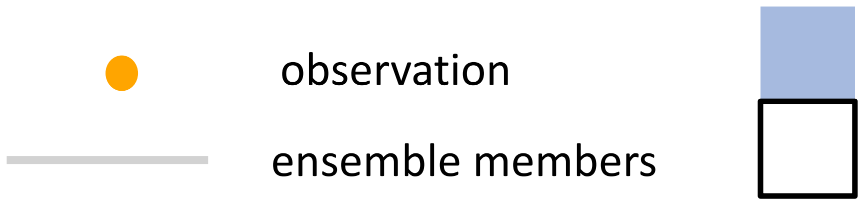
- The amount of uncertainty in model predictions and observations affects how much we adjust our forecasts based on observations

Low observation uncertainty; Forecast adjusted *a lot* by point outside prediction interval



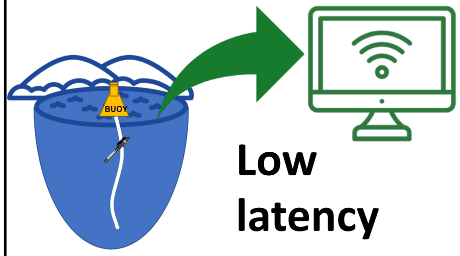
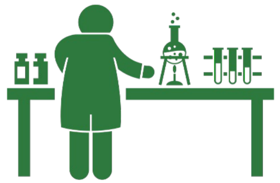
High observation uncertainty; Forecast adjusted a little by point outside prediction interval





Data assimilation frequency

- We can only update our models with data when data are available
- Data availability can also be affected both by **data frequency** and **data latency**
  - **Data frequency:** how often data are collected (i.e., temporal frequency)
  - **Data latency:** the time between when an observation is collected and when it is ready to be used in a forecast model

Field equipment	Data frequency and latency
 Sensor	 High frequency  Low latency
 Sample bottle and cooler	 Low frequency  High latency

Today…

We are going to generate forecasts of **phytoplankton biomass** in lakes using an **ecological model,** and then use real data from the **National Ecological Observatory Network** to update the model through data assimilation.

Phytoplankton are microscopic primary producers that live in the water. They are the dominant primary producers in many lakes and reservoirs.

What are the drivers of phytoplankton biomass in a lake?

- Light
- Water temperature
- Available nutrients
- Zooplankton

Image: Wikimedia commons

Today…

We are going to generate forecasts of **phytoplankton biomass in lakes** using an **ecological model,** and then use real data from the **National Ecological Observatory Network** to update the model through data assimilation.

Why forecast **phytoplankton biomass in lakes?**

Excess phytoplankton biomass can lead to harmful blooms, which compromise water quality through:

- Production of toxins,
- Production of taste and odor compounds,
- Creation of anoxic zones, leading to fish kills

Image: Wikimedia commons

Today…

We are going to generate forecasts of **phytoplankton biomass in lakes** using an **ecological model,** and then use real data from the **National Ecological Observatory Network** to update the model through data assimilation.

Chlorophyll-a data are a useful proxy for phytoplankton biomass in lakes.

- Chlorophyll-a is a molecule that phytoplankton use for photosynthesis.
- Measuring chlorophyll-a concentrations in water can give a bulk estimate of phytoplankton concentrations.
- Chlorophyll-a data can be collected at a high frequency by automated sensors.

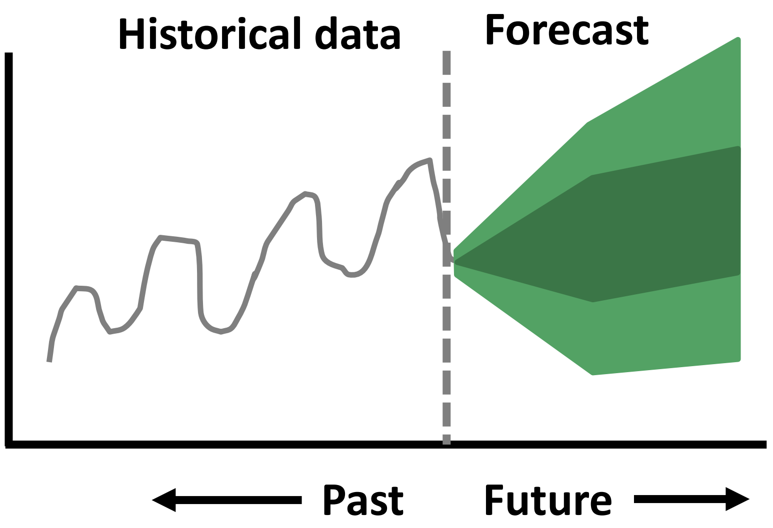
Image: Wikimedia commons

Today…

We are going to generate forecasts of **phytoplankton biomass** in lakes using an **ecological model,** and then use real data from the **National Ecological Observatory Network** to update the model through data assimilation.

We will fit an **autoregressive model** to chlorophyll-a data from a lake site of your choice.

An **autoregressive model** uses past values of a variable to predict future values.



Today…

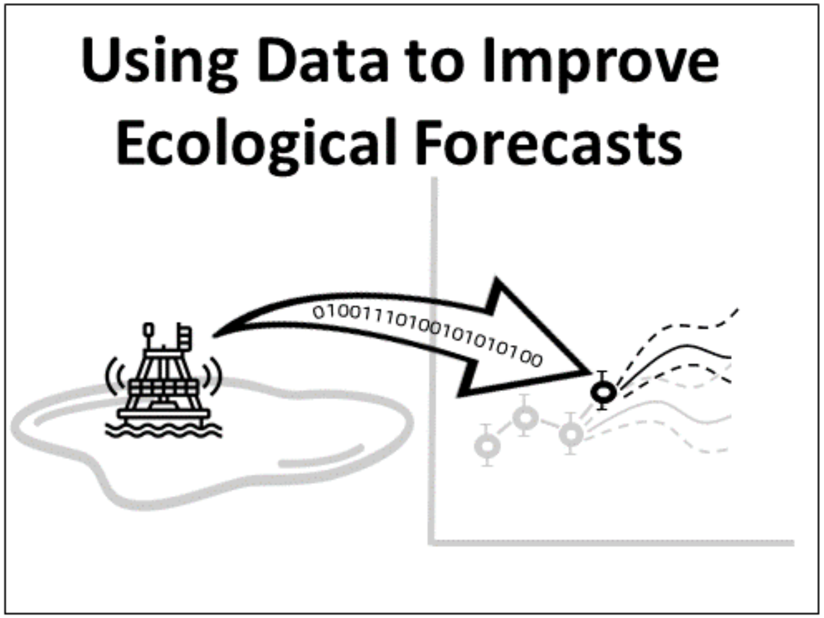
We are going to generate forecasts of **phytoplankton biomass** in lakes using an **ecological model,** and then use real data from the **National Ecological Observatory Network (NEON)** to update the model through data assimilation.

NEON is a continental-scale observatory designed to collect long-term open access ecological data to better understand how U.S. terrestrial and aquatic ecosystems are changing

Image: Map of NEON sites, neonscience.org

Learning objectives of today's module:

1. Define data assimilation
2. Generate an ecological forecast for chlorophyll-a in a lake
3. Describe how to assess ecological forecast accuracy
4. Describe how data assimilation affects forecast accuracy and uncertainty
5. Explain how updating models with data collected at different time scales (e.g., daily, weekly) and with different levels of associated uncertainty affects ecological forecasts



Activity A: Select a site, fit a model, and forecast

With a partner (work in pairs):

Objective 1: Select and learn about a NEON lake site

Objective 2: Explore chlorophyll-a data collected at your lake site

Objective 3: Fit an autoregressive forecast model to chlorophyll-a data

Objective 4: Generate a one-day-ahead forecast of chlorophyll-a with uncertainty

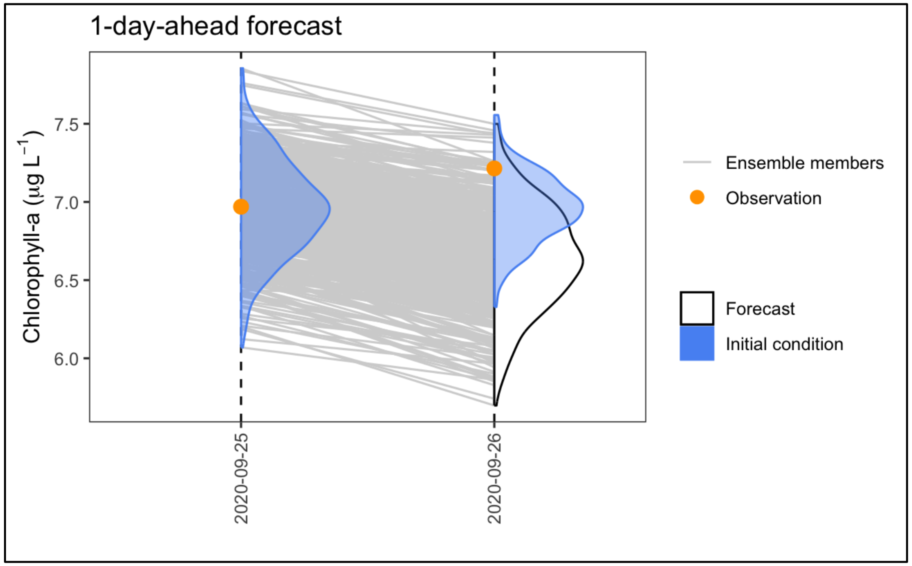
Activity B: Examine how data assimilation affects forecasts

With a partner (work in pairs):

Objective 5: Compare one-day-ahead forecasts generated with and without data assimilation.

Objective 6: Compare forecasts updated using data with low vs. high observation uncertainty.

Objective 7: Compare a series of forecasts updated with data at different frequencies (e.g., weekly, daily).

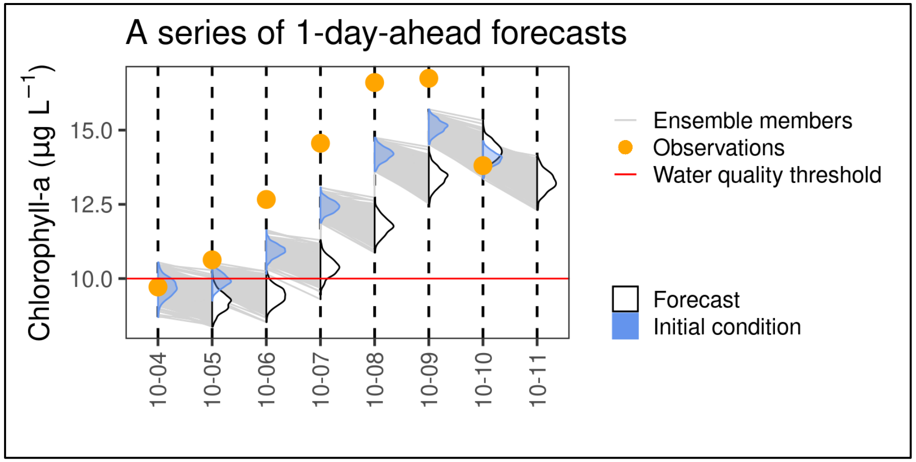


Activity C: Make a management decision using forecasts with data assimilation

With a partner (work in pairs):

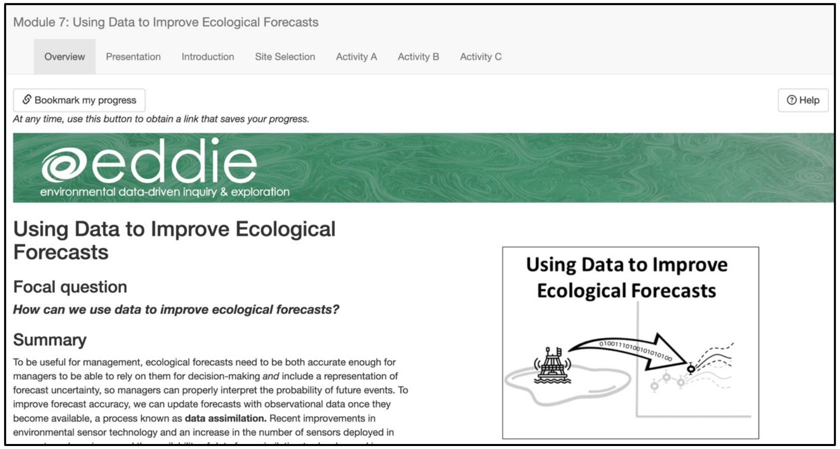
Objective 8: Make management decisions using forecasts generated with different frequencies of data assimilation.

You may have the opportunity to present your choice and the resulting forecasts to the class.



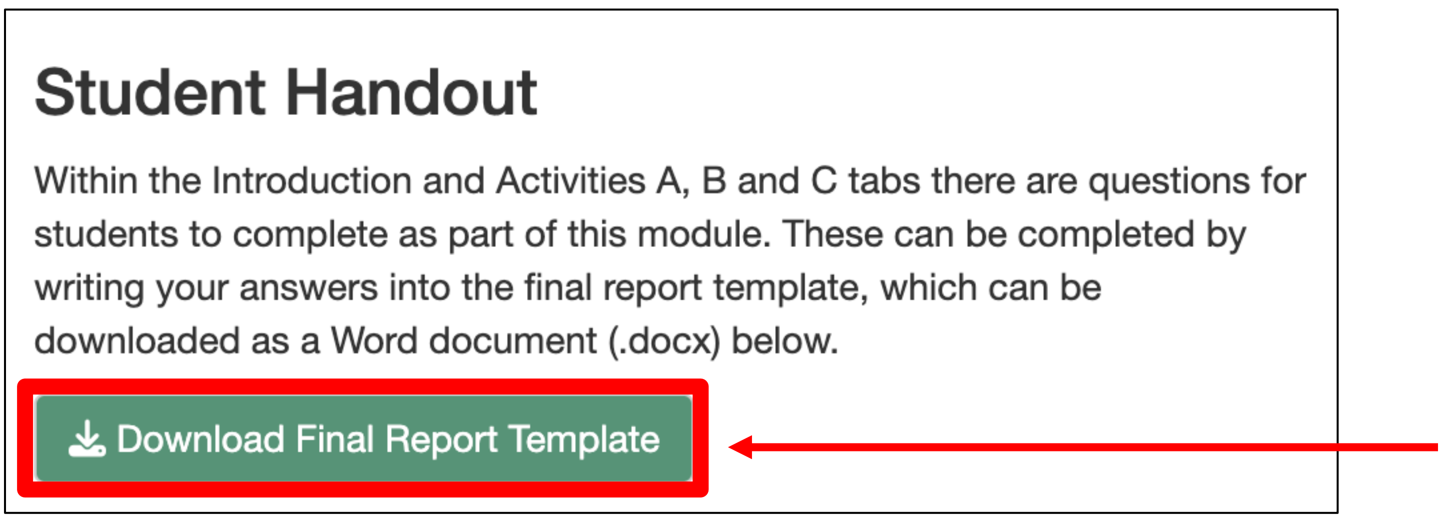
Shiny App

- The module can be accessed at: https://macrosystemseddie.shinyapps.io/module7/
- This is an interactive webpage built using R
- It has interactive plots and options embedded which allow you to build your own personal model, visualize and explore the data, and answer questions



Downloading the Report

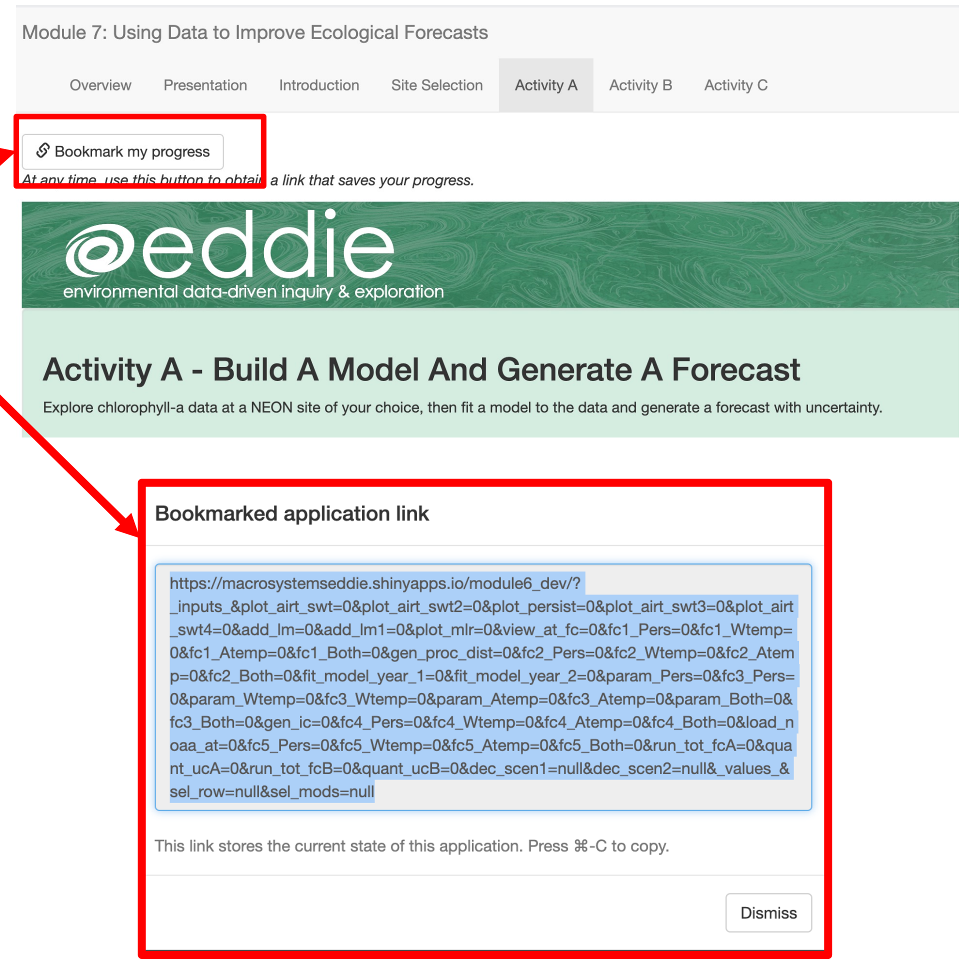
1. Navigate to the “Introduction” tab
2. Click on the “Download Final Report Template” button to download a Word document into which you can type your answers.



Saving & Resuming Progress

Saving Progress

1. Scroll to top of the page.
2. Click on the “Bookmark my progress” button. A pop-up window with a *very long link* will appear.
3. Copy-paste the link and store it at the top of your final report.



Resuming progress

1. Open your browser.
2. Copy-paste the link into your browser.
3. As you navigate through the tabs in the module, your progress will reappear.

We recommend that you save your progress often!

- Because the Shiny app can time out after inactivity (15 minutes) or disconnect if an internet connection is interrupted, we don't want you to lose your work.

- Save your progress as you go, as well as every time you close your computer or close the Shiny app in your internet browser.

- After you save the link somewhere safe, you should be able to resume your progress where you left off!

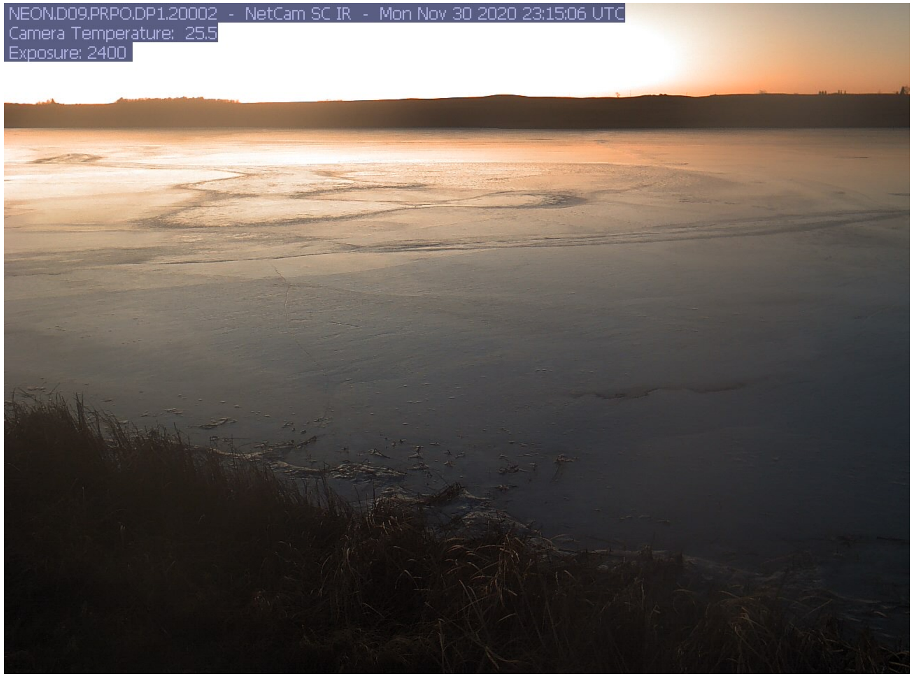
Let's Go!

- For the activity we will work in pairs.
- Each pair selects the same NEON site and works through Activities A, B, and C.
- It is possible that more than one pair will be assigned to the same lake.

Lake name	Students
Crampton Lake	
Suggs Lake	
Barco Lake	
Prairie Pothole	
Little Rock Lake	
Prairie Lake @ Dakota Coteau	

https://macrosystemseddie.shinyapps.io/module7/

Thank you for participating!



Check out other ecological forecasting modules:

- **Module 5:** Introduction to Ecological Forecasting
- **Module 6:** Understanding Uncertainty in Ecological Forecasts
- **Module 8:** Using Ecological Forecasts to Guide Decision-Making

Find out more at: macrosystemsEDDIE.org

Macrosystems EDDIE
Environmental Data-Driven Inquiry & Exploration in Macrosystems Ecology

neon
Operated by Battelle

SERC the Science Education Resource Center at Carleton College