Can you predict the future? Introducing the NEON Ecological Forecasting Challenge

Ecological Forecasting Initiative Meeting 2024

Monday 10th June 2024

Workshop leads: Freya Olsson, Quinn Thomas, Cayelan Carey, Mary Lofton
Virginia Tech, USA

Workshop overview

Objectives:

1. Introduce NEON and the Forecasting Challenge

2. Provide tools and resources to contribute to the Challenge

3. Introduce the Challenge cyberinfrastructure

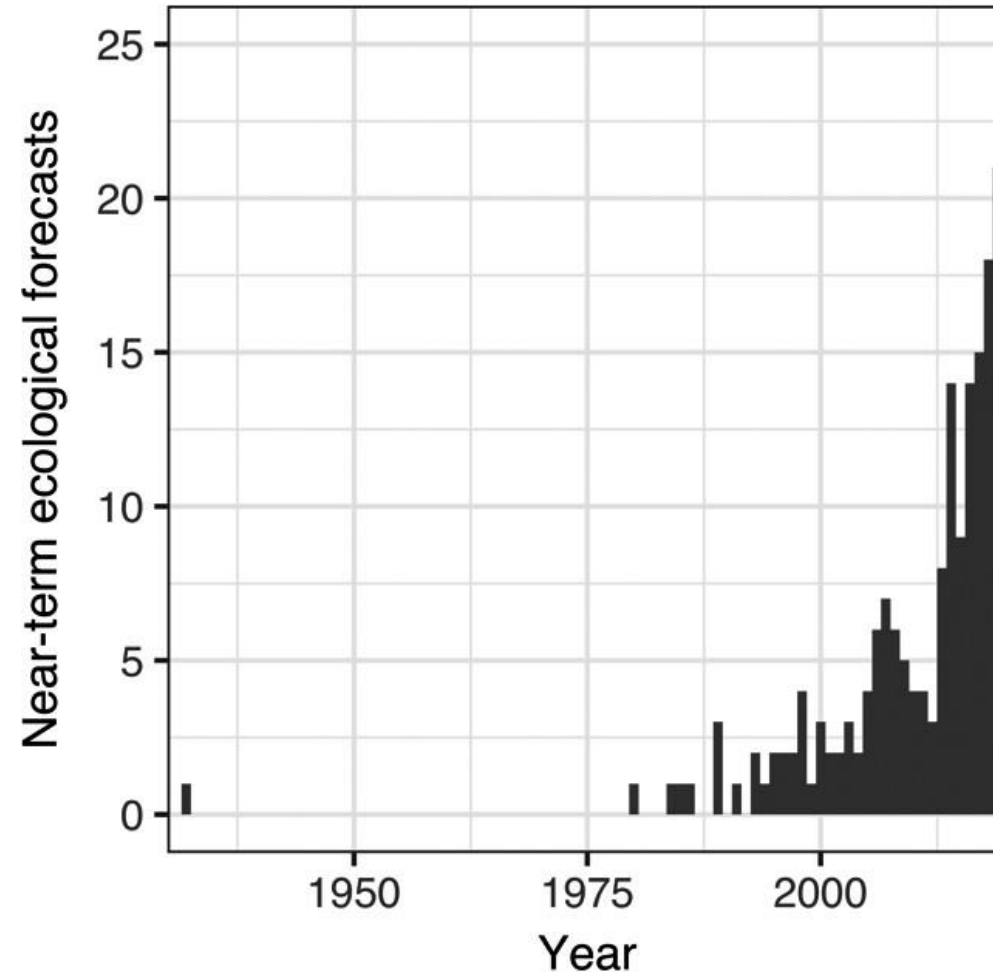4. Explore the existing forecast catalogue

Workshop Overview

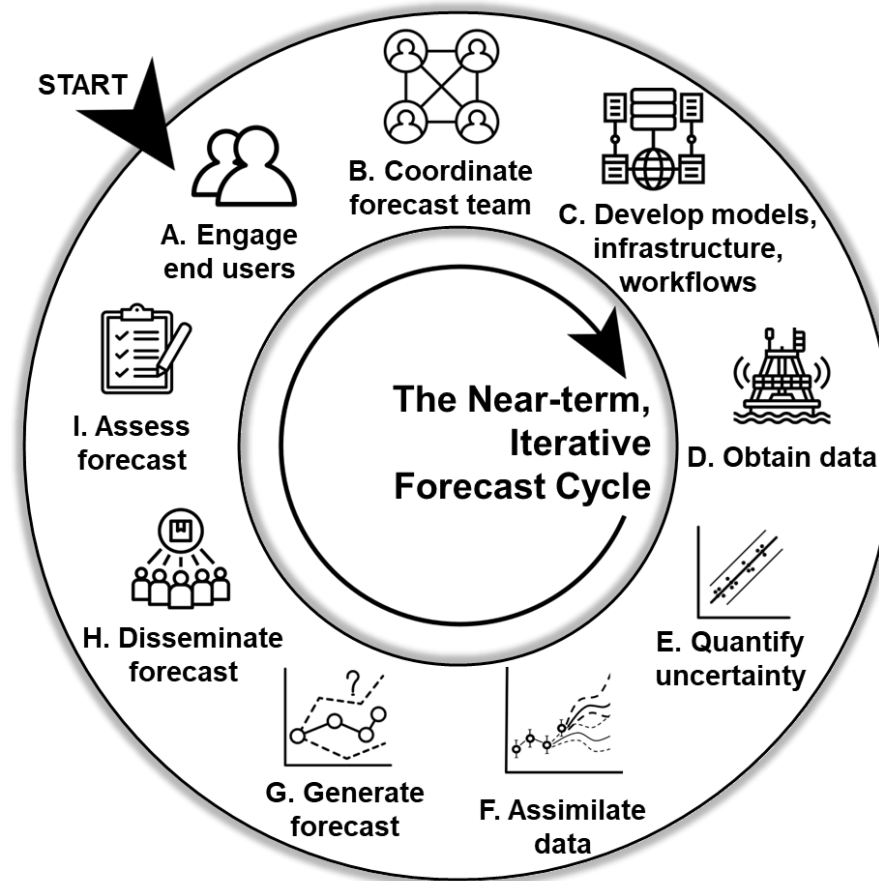|	
9:30-10.00	Introductory presentation
10:00-11.00	Challenge **contributor** tutorial
11.00-11.15	Break
11:15-11:45	Challenge **organizer** walk-through
11.45-12:15	**Challenge synthesis** example
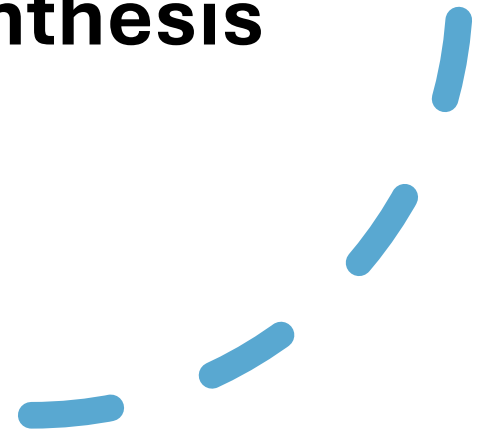
Ecological forecasting

- An emerging and growing field

Lewis, A. S. L et al. (2022). Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecol. Appl.*, *32*(2), e02500.

Forecasting Challenges

- Forecasting is challenging!


- A Challenge to catalyze progress

Lofton, M. E., Howard, D. W., Thomas, R. Q., & Carey, C. C. (2023). Progress and opportunities in advancing near-term forecasting of freshwater quality. *Glob. Change Biol.*, *29*(7), 1691–1714

Why a forecasting challenge? – the power of many forecasts!

1. A community of forecasting

   - Standards

   - Development of tools and infrastructure

   - A forecasting **platform**

2. Answer questions of predictability
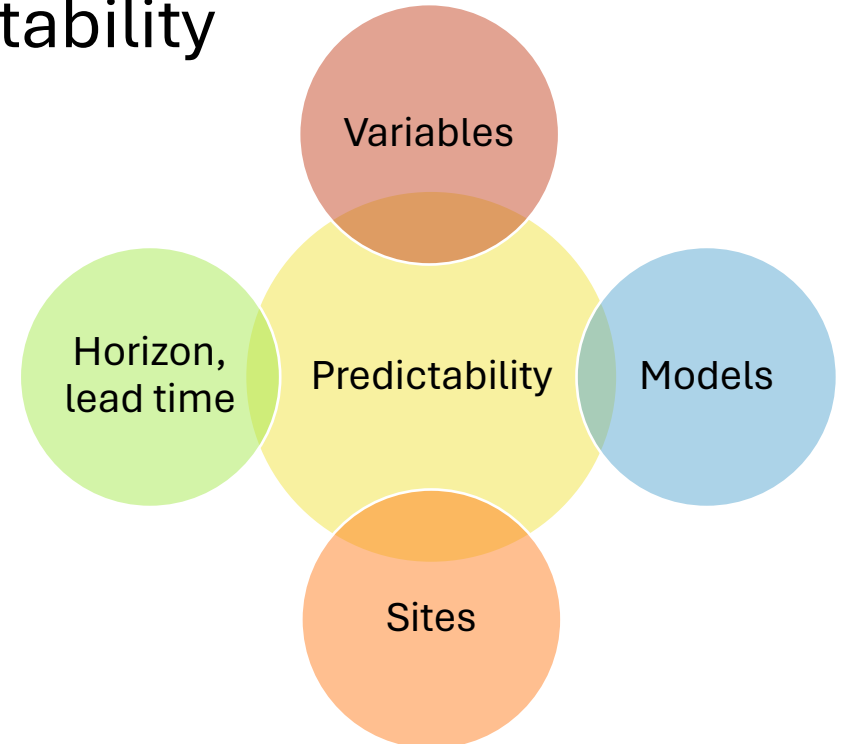
Established the **EFI-NEON Forecasting Challenge** (2021)!

Ecological Forecasting Initiative Research Co-ordination Network

- EFI RCN Goals
  - lower barriers
  - community building
  - infrastructure
  - platform development

NSF-sponsored
Ecological Forecasting Initiative Research Coordination Network
5-year project

Create a community of practice that builds capacity for ecological forecasting by leveraging NEON data products.

https://ecoforecast.org/rcn/

Funded by the National Science Foundation (DEB-1926388)

What is **neon**?

The National Ecological Observatory Network (NEON) is a
continental-scale observation facility

- collect **long-term, open-access ecological data**

- **47 terrestrial and 34 aquatic sites**
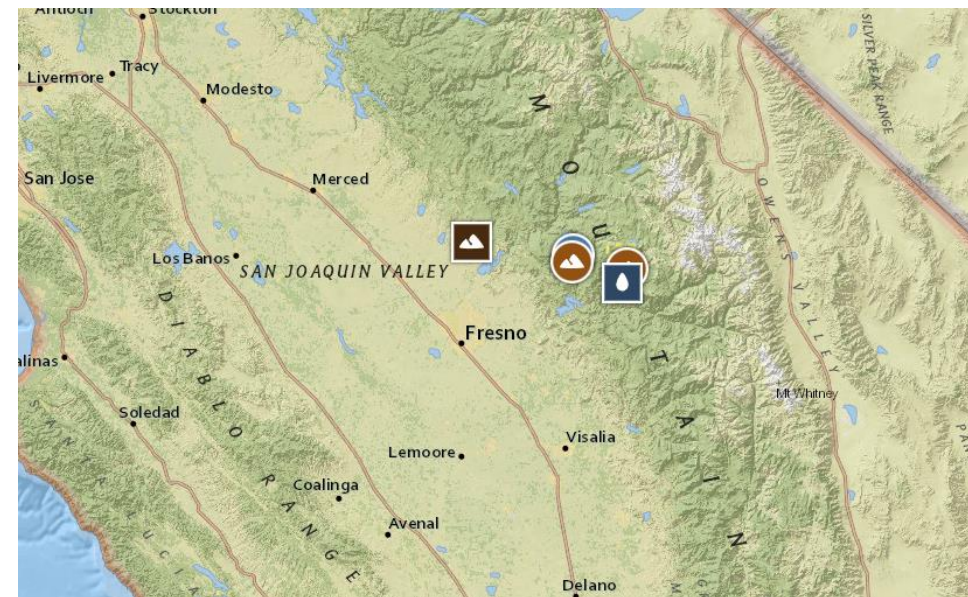
- **194 data products**

NEON sites across the US

NEON Data Collection

- Standardized methodologies

- Co-located sites

- Data available via S3 cloud storage and API

What is neon?

The National Ecological Observatory Network (NEON) is a **continental-scale observation facility**

- collect **long-term, open-access ecological data**

- **47 terrestrial and 34 aquatic sites**

- **194 data products**

2.1. NEON Mission

NEON is a National Science Foundation-sponsored facility for research and education on long-term, large-scale ecological change. NEON's goals are derived from the Integrated Science and Education Plan.
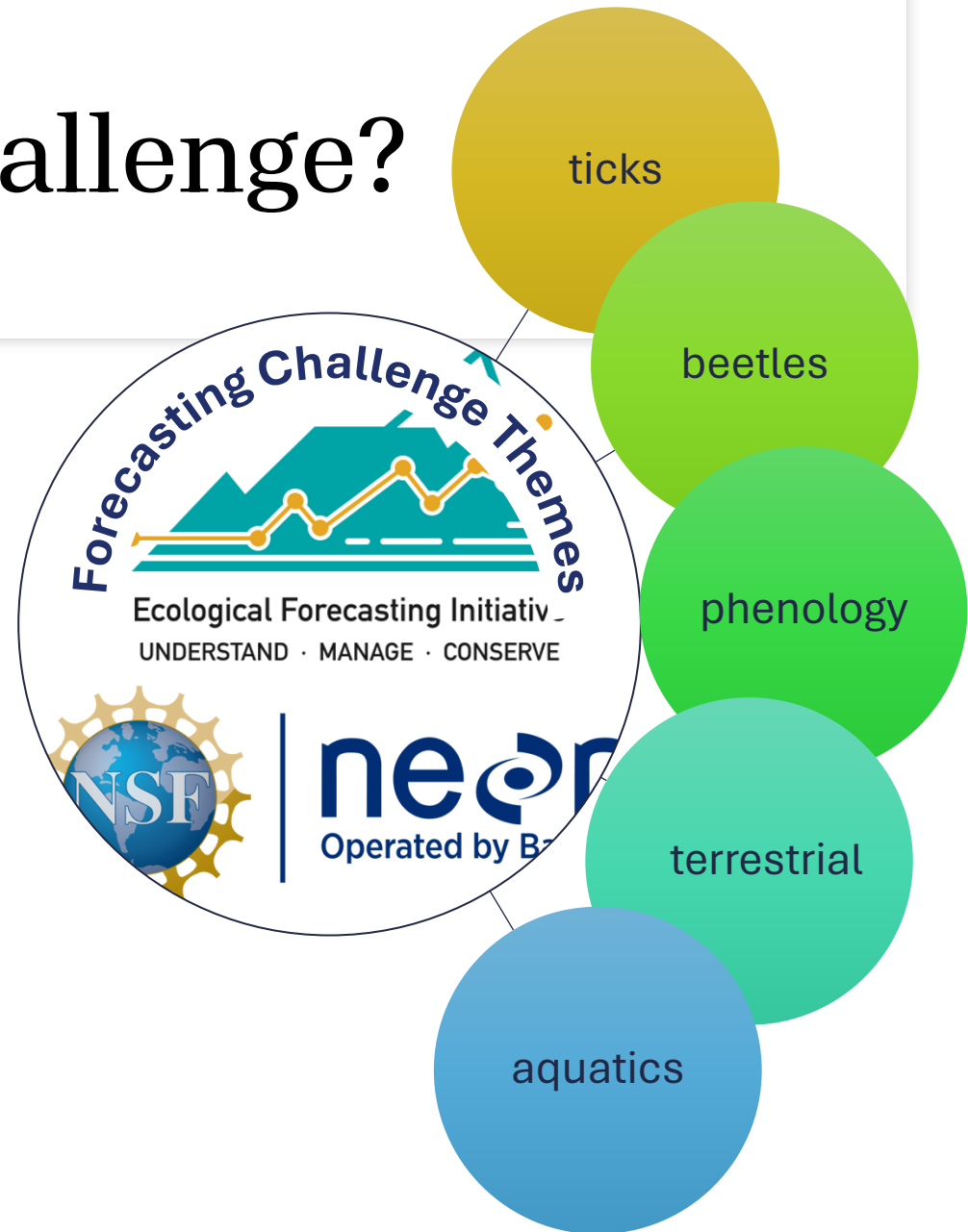
The goals of NEON are to:

- Enable understanding and forecasting of the impacts of climate change, land use change, and invasive species on aspects of continental-scale ecology such as biodiversity, biogeochemistry, infectious diseases, and ecohydrology

- Enable society and the scientific community to use ecological information and forecasts to understand and effectively address critical ecological questions and issues

- Provide physical and information infrastructure to support research, education, and land management.

From: https://www.neonscience.org/sites/default/files/NEON_Strategy_2011u2_0.pdf

What is the EFI-NEON Challenge?

*"A platform for the community to make predictions of conditions at NEON sites **before the data are collected**"*

- All 81 sites
- 5 themes

Forecasting Challenge Themes

Ecological Forecasting Initiative
UNDERSTAND · MANAGE · CONSERVE

NSF | neon
Operated by B...

ticks

beetles

phenology

terrestrial

aquatics

What is the EFI-NEON Challenge?

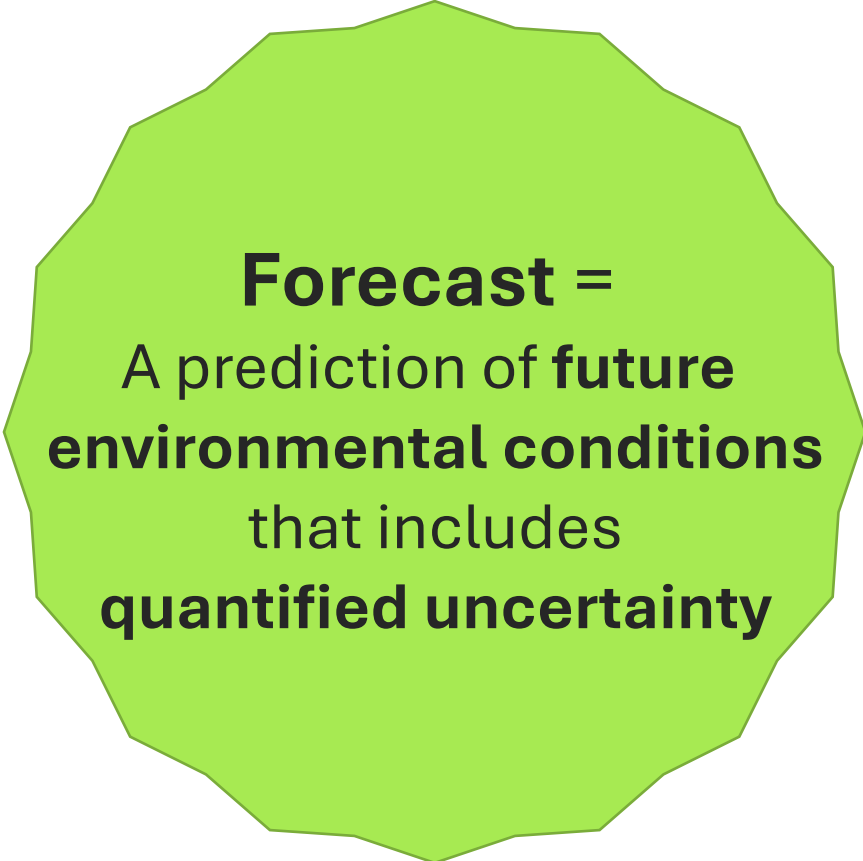*"A platform for the community to make predictions of conditions at NEON sites **before the data are collected"***

- All 81 sites
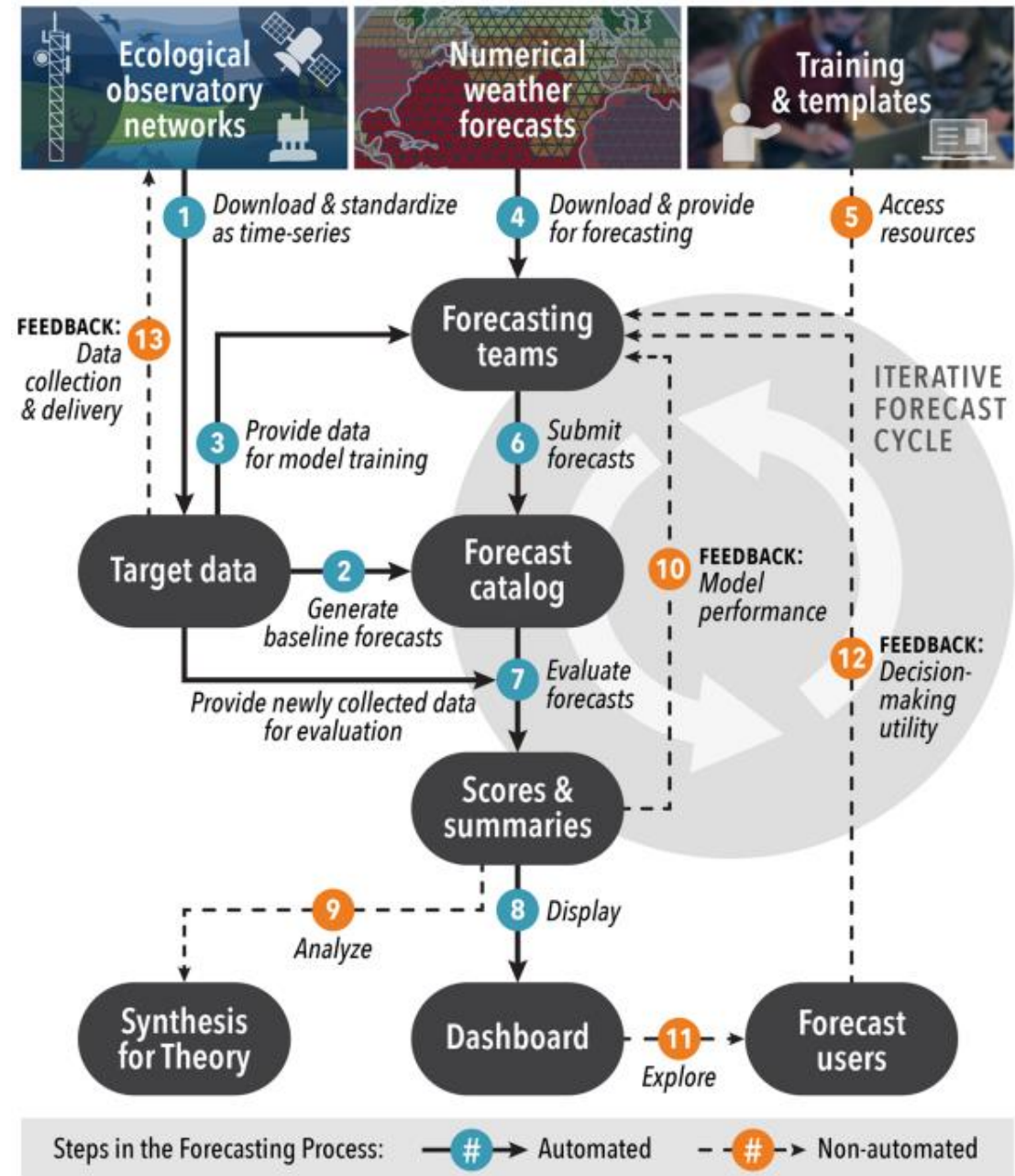- 5 themes
- > 30,000 forecast submitted!

Forecast =
A prediction of **future environmental conditions** that includes **quantified uncertainty**

Thomas, et al. (2023). The NEON Ecological Forecasting Challenge. *Front. Ecol. Environ.*, *21*(3), 112–113. https://doi.org/10.1002/fee.2616

Challenge overview

Thomas, et al. (2023). The NEON Ecological Forecasting Challenge. *Front. Ecol. Environ.*, *21*(3), 112–113. https://doi.org/10.1002/fee.2616

Workshop overview

Objectives:

1. Introduce NEON and the Forecasting Challenge

2. **Provide tools and resources to contribute to the Challenge**

3. Introduce the Challenge cyberinfrastructure

4. Explore the existing forecast catalogue and scores

Forecast contributor:

- **Aquatics theme – Can we predict how water temperature will change over the next month?**
  - Water temperature in lakes
  - NEONs water temperature data product (DP1.20264.001)
  - 30 day forecast horizon
  - Data latency of 2-3 days
- **Simple baseline** model to build on

Water temperature = key variable in driving many biogeochemical cycles and habitat available for thermal-sensitive species

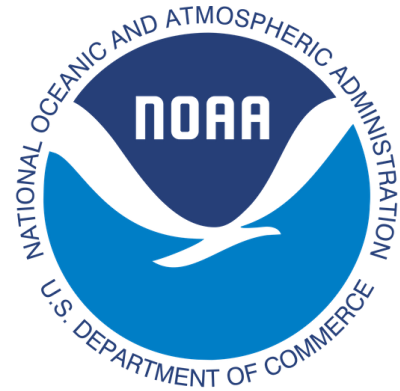NEON Buoy at Crampton Lake (Land O'Lakes, WI)

Diversity of NEON aquatic sites

Intro to key terms

- Targets
- NOAA weather
- Uncertainty
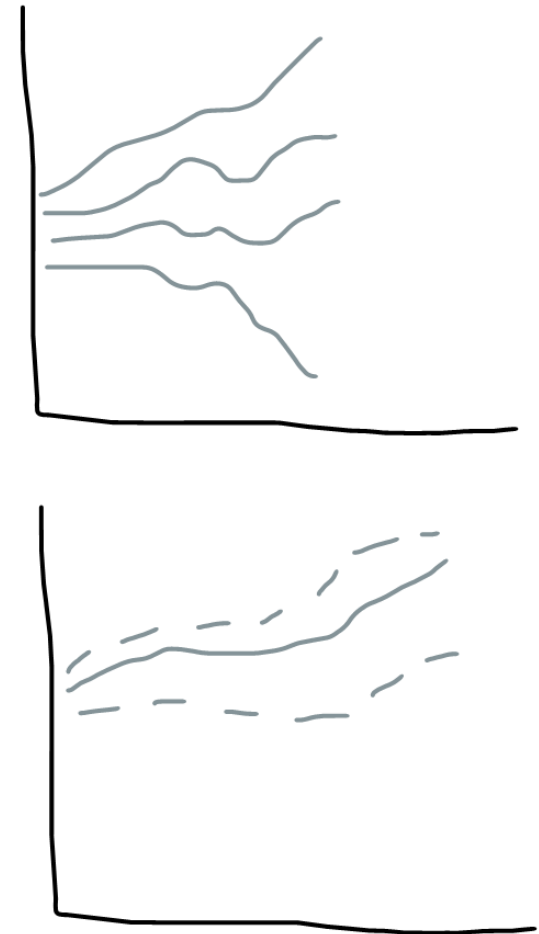- Forecast standards

Weather forecast covariates

- National Oceanic and Atmospheric Administration

- 3 NOAA forecast data products available via the vera4castHelpers R package:

  - Stage_1: **raw forecasts** from NOAA.
    30 member ensemble forecast.

  - Stage_2: processed from stage_1.
    Recommended for **future forecasts (hourly inputs)**

  - Stage_3: the **historic data product**.
    A 'stacked' data set taking every 1 day-ahead forecast.
    Useful for **model training/calibration**.

- Eight weather variables
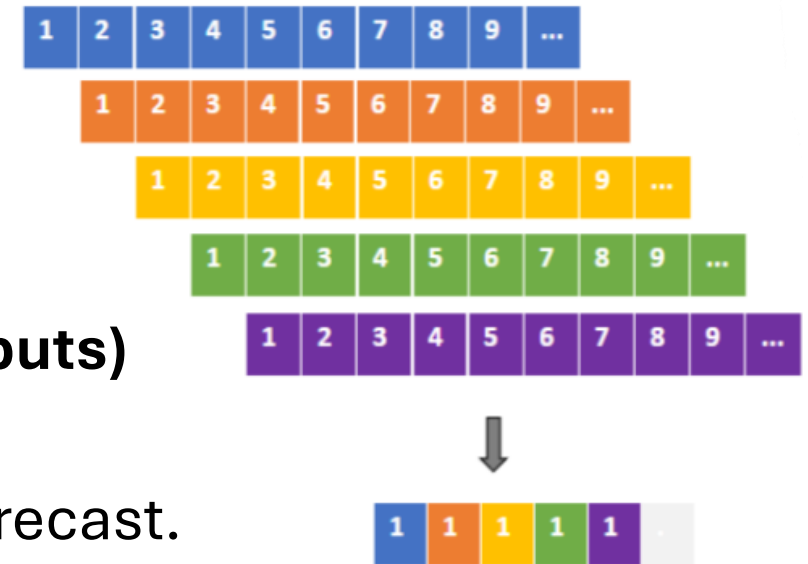
Forecast uncertainty

- Forecasts must include an estimate of uncertainty.

- The uncertainty can be represented using:
  - different model runs (**ensemble members**) or
  - a **distribution** (e.g. a normal distribution with a mean and standard deviation).
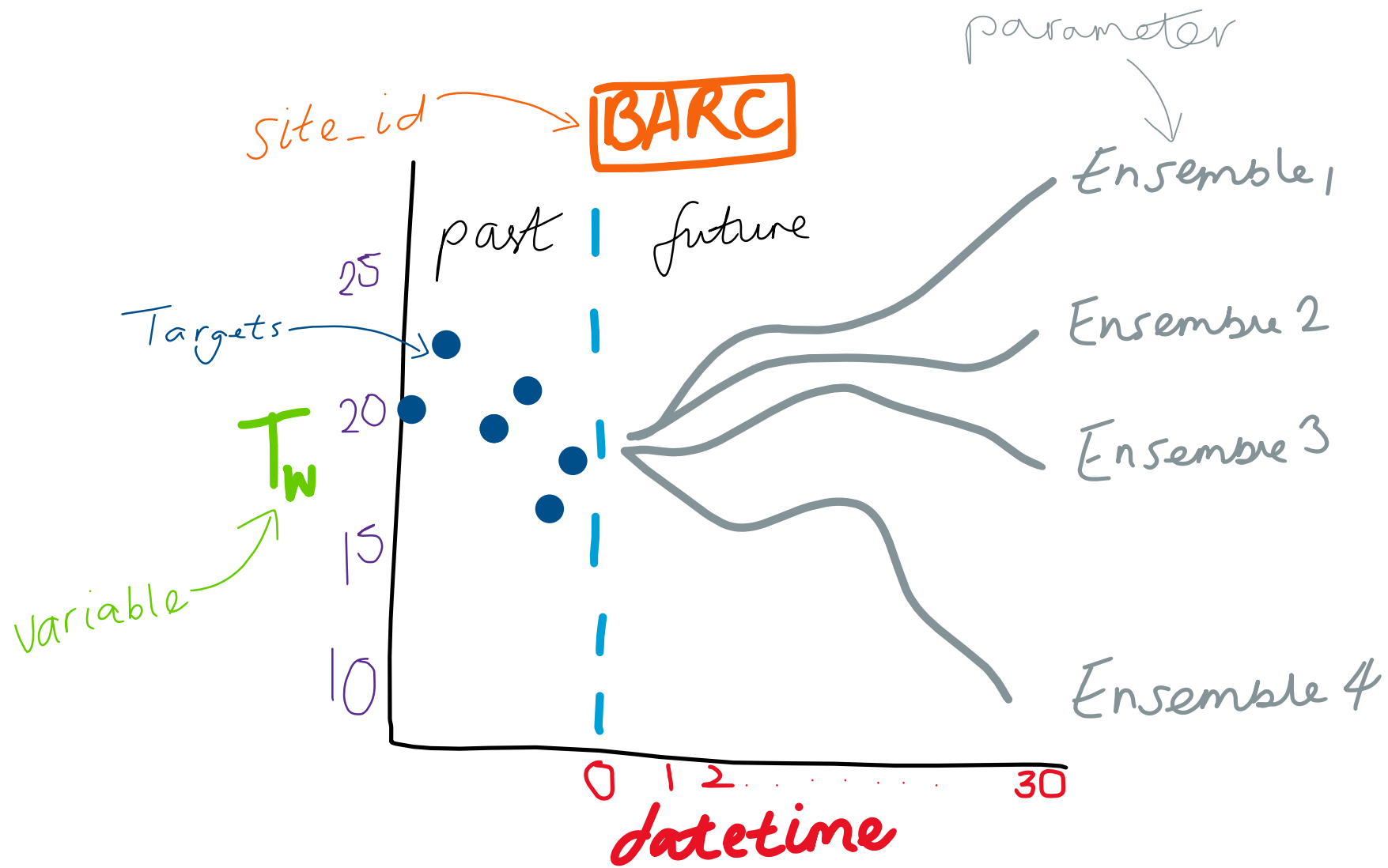
Forecast standards

Help maintain consistency in forecast generation, submissions and scoring

Need to submit a forecast in a standardized format

- file format (csv)
- file name format ([theme]-[reference_datetime]-[team_name].csv)
- specific column names
- column format (datetime/character/integer/etc.)

datetime	reference_datetime	duration	site_id	depth_m	family	parameter	variable	prediction	model_id	project_id
2024-05-02	2024-05-01	P1D	fcre	1.6	ensemble	1	temperature	22.63	example_ID	neon4cast
2024-05-02	2024-05-01	P1D	fcre	1.6	ensemble	2	temperature	26.88	example_ID	neon4cast
2024-05-02	2024-05-01	P1D	fcre	1.6	ensemble	3	temperature	24.57	example_ID	neon4cast
2024-05-02	2023-05-01	P1D	fcre	1.6	ensemble	4	temperature	23.17	example_ID	neon4cast
...	...		...	...	...		...	...	...	...
2024-05-31	2023-05-01	P1D	fcre	1.6	ensemble	1	temperature	19.40	example_ID	neon4cast
2024-05-31	2023-05-01	P1D	fcre	1.6	ensemble	2	temperature	20.31	example_ID	neon4cast
2024-05-31	2023-05-01	P1D	fcre	1.6	ensemble	3	temperature	25.61	example_ID	neon4cast
2024-05-31	2023-05-01	P1D	fcre	1.6	ensemble	4	temperature	22.74	example_ID	neon4cast

Filename = **aquatics-2024-06-10-example_ID.csv**
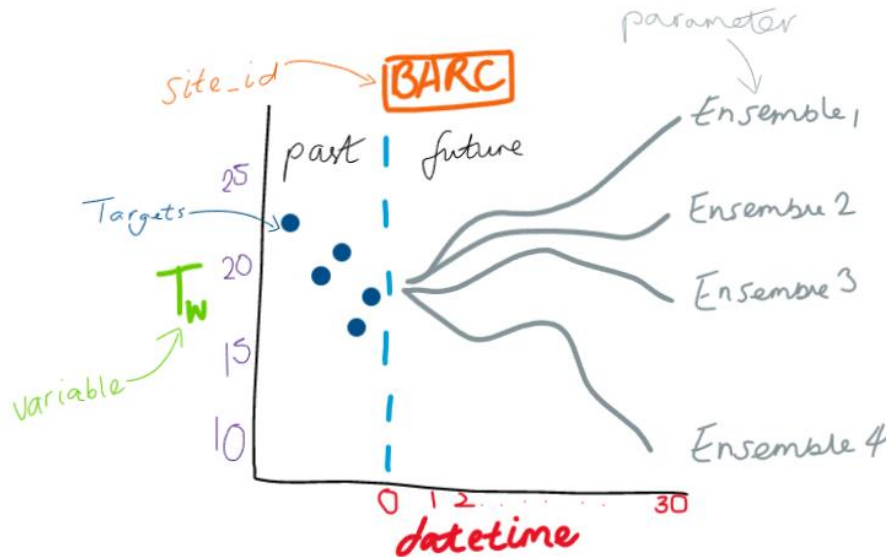
Basic workflow to submit a forecast

1. Read EFI-NEON Challenge documentation (https://projects.ecoforecast.org/neon4cast-dashboard/)

2. Investigate the forecast target variables

3. Build/apply your model!

4. Produce forecast of future conditions – SUBMIT TO THE CHALLENGE!

5. Register, complete model description, and submit your forecasts

6. Wait for the scores to come in and revel in the glory of predicting the future (~5-day before first evaluation)

7. Use new data to update the model

Workflow automation

8. Submit another forecast! And another...!

Key points:

Include uncertainty
- Distribution
- Ensemble
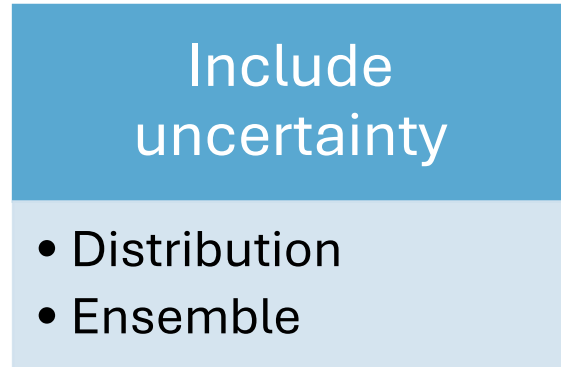
Defined in the family and parameter columns

e.g.
family = normal
parameter = mu + sigma

What period does the forecast cover? (e.g. daily = P1D)

When is the forecast for?

Use multiple ideas of time
- Duration
- Datetime
- Reference_datetime

When was the forecast generated?

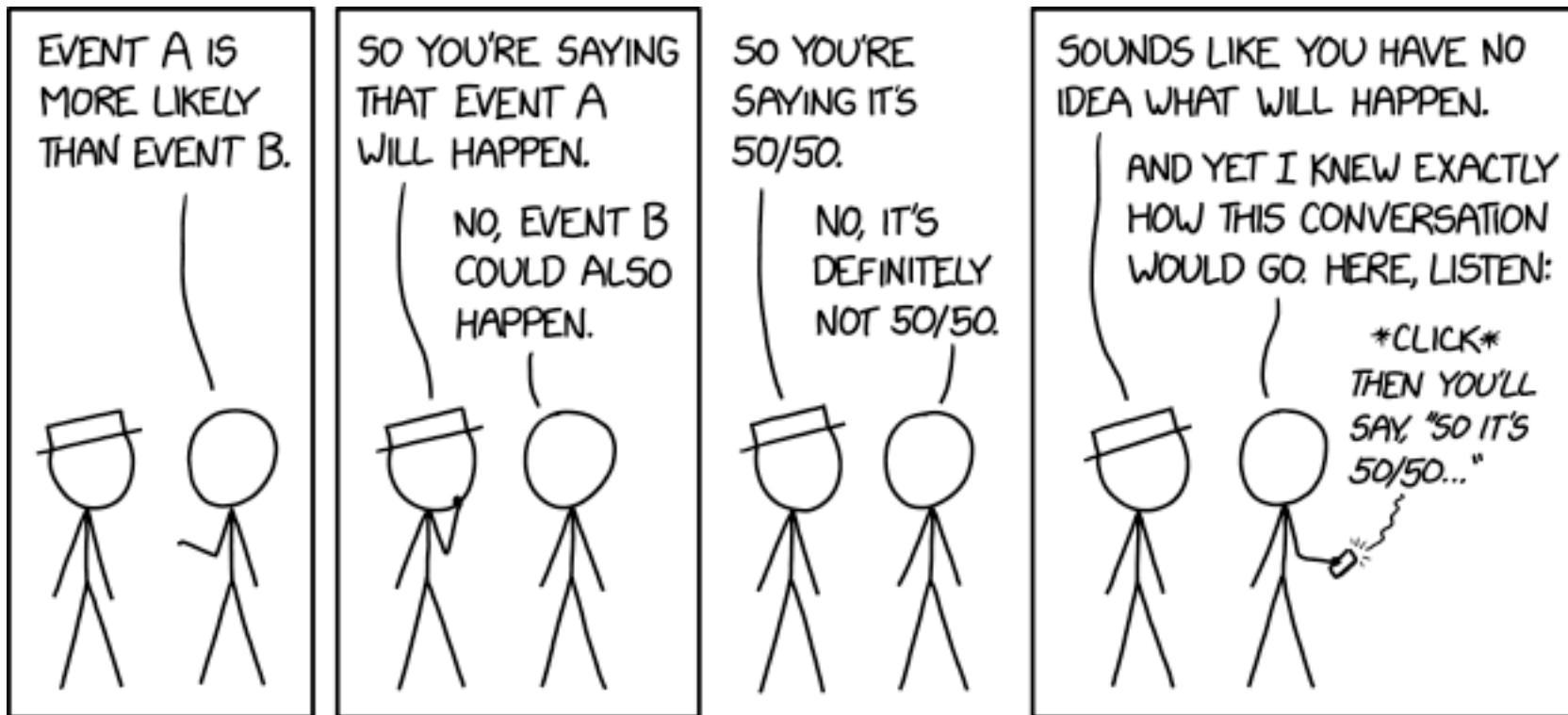## Follow a standard format
- Required columns
- csv file

R functions are available to help you check your forecast format!

Install `neon4cast` from GitHub

More information: https://projects.ecoforecast.org/neon4cast-ci/
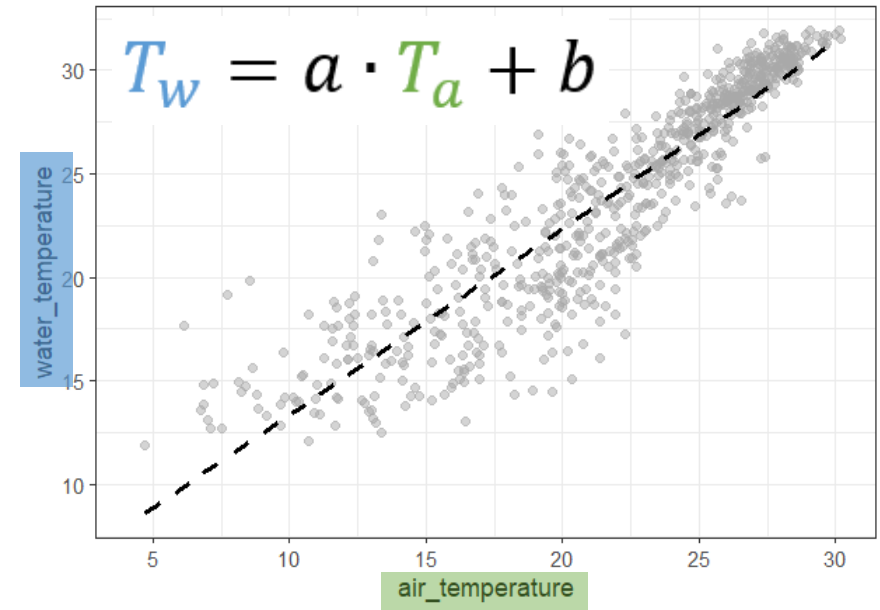
Questions?

https://xkcd.com/2370

Let's forecast!

1. Follow-along R markdown
   - Forecasting water temperature using a Linear model with air temperature

2. Modify the model and submit an automated forecast!

$$T_w = a \cdot T_a + b$$

Installation instructions

https://github.com/eco4cast/NEON-Challenge-EFI2024/