

This document provides a brief introduction to some terminology that might be used in discussion around the EFI-NEON Forecast Challenge.

Firstly, what is the **EFI-NEON Forecast Challenge**?

A challenge to the ecological scientific community to predict conditions at NEON sites before the data are collected. The Challenge revolves around the five theme areas listed below that span aquatic and terrestrial systems, and population, community, and ecosystem processes across a broad range of ecoregions that uses data collected by NEON. From these near-term ecological forecasts we are excited to learn more about the predictability of ecological processes. What modeling frameworks, mechanistic processes, and statistical approaches best capture community, population, and ecosystem dynamics?

Read more about the Challenge here: <https://projects.ecoforecast.org/neon4cast-docs/>

| Term                    | Definition   |
|-------------------------|--|
| Climatology             | A forecast model that utilizes historic data to produce future forecasts. It is also often used as a <i>baseline</i> model which forecast skill can be compared with. An example, used in the aquatics challenge is to use the day of year mean as the forecast mean.  |
| Covariates              | A characteristic or variable that you might use to make a forecast (in addition to the variable you are forecasting). For example, the NOAA weather data could be used as a covariate in a forecast model.   |
| Distributional forecast | A forecast in which the uncertainty is quantified by the probability distribution of the forecast values (mean and standard deviation. At the moment, only normal distribution forecasts are supported. For non-normal distribution, we recommend using an ensemble forecast   |
| EFI                     | Ecological Forecasting Initiative ( <a href="https://ecoforecast.org/">https://ecoforecast.org/</a> ) – “a grassroots consortium aimed at building and supporting an interdisciplinary community of practice around near-term (daily to decadal) ecological forecasts.”  |
| Ensemble forecast       | An ensemble forecast is a set of forecasts that present the range of future conditions, as a way to represent the uncertainty in the forecast. Each ensemble represents an equally likely forecast and may be generate using slightly different initial conditions, model parameters or driving data.  |
| Metadata                | A description of your forecasting approach. This is available as an online form, with multiple choice questions about your modelling method including how you quantified the uncertainty in your forecasts.  |
| NOAA data               | National Oceanic and Atmospheric Administration weather forecast data. This has been compiled by Challenge organizers to provide historic and future weather forecasts for all NEON sites that can be used in forecasting workflows.<br>3 data products can be accessed using the neon4cast R package:<br>1. Stage_1: raw forecasts from NOAA. 30 member ensemble forecast of the next 35 days. The first 10 |

|                         |  |
|-------------------------|--|
|                         | <p>days have 3 hourly forecast and 6 hourly for the remainder of the forecast period</p> <ol style="list-style-type: none"> <li>2. Stage_2: processed from stage_1 to meet international conventions and interpolated to 1 hour forecasts. Recommended for future forecasts</li> <li>3. Stage_3: the historic data product. A 'stacked' data set taking every 1 day ahead forecast. An accurate representation of historic weather conditions</li> </ol>   |
| NEON                    | National Ecological Observatory Network – “a continental-scale observation facility designed to collect long-term open access ecological data to better understand how U.S. ecosystems are changing.” The EFI-NEON Forecast Challenge leverages this open data set to challenge the forecasting community to forecast this data before they are collected  |
| Random walk/persistence | The forecast values are set to the last observation. It is often used as a baseline model which other forecasts skill can be compared with.  |
| Scores                  | The means to assess forecast skill. The challenge uses the continuous rank probability score (crps). This score uses both the accuracy (mean) and precision (sd) of the forecast to calculate the score.   |
| Standards               | To help maintain consistency in forecast generation, submission and scoring forecasts submitted to the Challenge must meet standardized format for file format (csv/NetCDF), file name, column names, and column format. See here for specific details ( <a href="https://projects.ecoforecast.org/neon4cast-docs/Submission-Instructions.html">https://projects.ecoforecast.org/neon4cast-docs/Submission-Instructions.html</a> ).  |
| Targets                 | <p>The data collected by NEON which is to be forecasted. Can be thought of as the “training dataset” that can be used to build and test your forecasting method. The R script Challenge organizers have for generating the targets can be found at: <a href="https://github.com/eco4cast/neon4cast-aquatics">https://github.com/eco4cast/neon4cast-aquatics</a></p> <p>The challenge uses the following NEON data products: - DP1.20264.001 (Temperature at specific depth in surface water for lakes), DP1.20288.001 (Water quality), and DP1.20035.001 (Temperature in surface waters for streams)</p> |
| tsibble                 | The tsibble package extends the tidyverse to temporal data with the objects built on top of tibble objects, based on the tidy principles of data.  |
| Uncertainty             | All predictions contain some element of uncertainty, which comes from a number of factors such as imprecise quantifications of model processes, uncertainty from driving data, and initial conditions. Challenge submission must include an estimate of forecast uncertainty, which can be represented using different model runs (ensemble members) or the statistics of the forecast (mean and standard deviation).  |