Macrosystems EDDIE:

Understanding Uncertainty in Ecological Forecasts

Instructor's Manual

Module Description

Ecological forecasting is an emerging approach which provides an estimate of the future state of an ecological system with uncertainty, allowing society to prepare for changes in important ecosystem services. In this module, students will explore the sources of uncertainty in ecological forecasts and how uncertainty affects decision making. This module will introduce students to the basic components of an ecological forecast; how a forecasting model is constructed; what the main contributors to forecast uncertainty are; and why uncertainty is an integral part of an ecological forecast.

Pedagogical Connections

Phase	Functions	Examples from this module
Engagement	Introduce topic, gauge students' preconceptions, call up students' schemata	Short introductory lecture on forecast uncertainty with embedded questions; Pre-class readings and questions
Exploration	Engage students in inquiry, scientific discourse, evidence-based reasoning	Analyzing lake ecosystem data and relationships among variables; Generating forecasts of future lake water temperature while accounting for different sources of uncertainty
Explanation	Engage students in scientific discourse, evidence-based reasoning	Describing how different sources of uncertainty are quantified; Comparing how a single source of uncertainty varies among different forecast models
Expansion	Broaden students' schemata to account for more observations	Comparing how multiple sources of uncertainty compare among different forecast models; Analyzing how representation of forecast uncertainty affects decision-making
Evaluation	Evaluate students' understanding, using formative and summative assessments	In-class discussion on forecast uncertainty and how ecological forecasting can be used to improve ecosystem understanding and natural resource management; Short answer questions embedded throughout module

This module was initially developed by: Moore, T.N., M.E. Lofton, C.C. Carey, and R.Q. Thomas. 12 December 2023. Macrosystems EDDIE: Understanding Uncertainty in Ecological Forecasts. Macrosystems EDDIE Module 6, Version 1. http://module6.macrosystemseddie.org. Module development was supported by NSF grants DEB-1926050 and DBI-1933016. This document last modified: 14 December 202314 December 2023 by M.E. Lofton.

Learning Objectives

By the end of this module, students will be able to:

- Define ecological forecast uncertainty
- Explore the contributions of different sources of uncertainty (e.g., model parameters, model driver data) to total forecast uncertainty
- Understand how multiple sources of uncertainty are quantified
- Identify ways in which uncertainty can be reduced within an ecological forecast
- Describe how forecast horizon affects forecast uncertainty
- Explain the importance of specifying uncertainty in ecological forecasts for forecast users and decision support

How to Use this Module

This entire module can be completed in one 2 to 3-hour lab period, two 75-minute lecture periods, or three 1-hour lecture periods for introductory undergraduate students in Ecology, Environmental Science, Ecological Modelling, and Quantitative Ecology classes. This module can be coupled with other Macrosystems EDDIE ecological forecasting modules: Module 5 "Introduction to Ecological Forecasting"; Module 7 "Using Data to Improve Ecological Forecasts"; or Module 8 "Using Ecological Forecasts to Guide Decision-Making". We found that teaching this module in one longer lab section with short breaks was more conducive for introductory students than multiple 1-hour lecture periods. Lesson structure, depending on the time available for your class:

- Three classes (50-60 minutes)
  - Class 1 – Introductory lecture (30 min.) and completion of Activity A, Objectives 1 and 2 (20 min.)
  - Class 2 – Completion of Activity A, Objectives 3 and 4 (30 min.), and Activity B, Objectives 5 and 6 (20 min.)
  - Class 3 – Completion of Activity B, Objectives 7 and 8 (20 min.), and Activity C (20 min.); wrap-up (10 min.)
- Two classes (75-90 minutes)
  - Class 1 – Introductory lecture and completion of Activity A
  - Class 2 – Activity B & C followed by 10-15 minute presentation discussion of each groups' results
- One class (3 hours)
  - Introductory lecture – 30 mins, Activity A – 45 mins, break – 5mins, Activity B – 45 mins, group presentation/discussion – 15/20mins, Activity C – 30 mins

Quick overview of the activities in this module:

- Activity A: Students build different models to simulate water temperature for their chosen NEON site and generate forecasts without uncertainty.
- Activity B: Students generate multiple forecasts of water temperature with different sources of uncertainty and examine how uncertainty differs among models.

- Activity C: Students generate forecasts that include all sources of uncertainty and partition the contribution of different sources of uncertainty for their forecasts with different models. Students also complete a management scenario case study exploring how the representation of forecast uncertainty affects decision-making.

Module Workflow (for either in-person or virtual instruction)

1. Instructor chooses method for accessing the Shiny app:
    1. In any internet browser, go to: **https://macrosystemseddie.shinyapps.io/module6/**
        1. This option works well if there are not too many simultaneous users (<20)
        2. The app generally does not take a long time to load but requires consistent internet access
        3. It is important to remind students that they need to save their work as they go, because this webpage will time-out after 15 idle minutes. It is frustrating for students to lose their progress, so a good rule of thumb is to get them to save their progress after completing each objective
    2. The most stable option for large classes is downloading the app and running locally, see instructions at: **https://github.com/MacrosystemsEDDIE/module6**
        1. Once the app is downloaded and installed (which requires an internet connection), the app can be run offline locally on students' computers
        2. This step requires R and RStudio to be downloaded on a student's computer, which may be challenging if a student does not have much R experience (but this could be done prior to instruction by an instructor on a shared computer lab)
        3. If you are teaching the module to a large class and/or have unstable internet, this is the best option

Regardless of which option you pick, all module activities are the same!

2. Give students their handout ahead of time to read over prior to class or ask students to download the handout from the module Shiny app page when they arrive to class. The module is set up for students to complete discussion questions in the student handout (a Microsoft Word document) as they navigate through the R Shiny app activities. As they navigate through the app, students will be prompted to answer questions in their handout, as well as download plots that they generate within the app and copy-paste them into their handout. The handout can be submitted to the instructor at the end of the module for potential grading.

3. Instructor gives a PowerPoint presentation that introduces ecological forecasting, sources of forecast uncertainty, and the different models they will be using (~30 mins).
4. After the presentation, the students divide into pairs. Each pair selects their own NEON site and visualizes their site's data (Activity A Objectives 1 and 2). The two students within a pair each build their own different models for predicting water temperature (Activity A Objective 3), and generate forecasts with no uncertainty using each of their models (Activity A Objective 4). For virtual instruction, we recommend putting two pairs together (n=4 students) in separate breakout rooms during this activity so the two pairs can compare results.
5. The instructor can ask students to wait until all students are finished Activity A and then they will all begin Activity B together. For virtual instruction, this would entail having the students come back to the main room for a short check-in.
6. In Activity B, the students work in their pairs to generate forecasts which include different sources of uncertainty (Activity B Objectives 5-8). Students may compare their forecasts with their partners and work together to answer questions embedded throughout this activity about why the different models are affected differently by the different sources of uncertainty.
7. In Activity C, student pairs generate forecasts including all sources of uncertainty and compare how the contributions of different sources of uncertainty to total forecast uncertainty varies among models (Activity C Objective 9). They then complete a management scenario individually and discuss with their partner how the uncertainty visualizations provided in the scenario affected their management decisions (Activity C Objective 10).

Important Note to Instructors:

The R Shiny app used in this module is regularly updated, so these module instructions will periodically change to account for changes in the code. If you have any questions or have other feedback about this module, please contact the module developers (see “We’d love your feedback” below).

We highly recommend that instructors familiarize themselves with the Shiny app prior to the lesson. This will enable you to be more prepared to answer questions related to certain areas of the app’s functionalities.

Things to do prior to starting the instructor’s presentation

- Have the students read through the student handout, especially the ‘Why macrosystems ecology and ecological forecasting?’ and ‘Today’s focal question’ sections.
- Optionally, have students complete the pre-class activity, in which they read a case study about river water level predictions which were provided prior to a severe flooding event on the Red River, USA in 1997, and answer questions about how the representation of uncertainty associated with future predictions affects decision-making.

Guide to Introductory PowerPoint Presentation

Note: the numbers below match the PowerPoint slide numbers. The text for each slide is also in the "Notes" of the PowerPoint, so can be viewed when projecting in Presenter View.

1. Welcome the students to class.
    a. It is important at this point to emphasize that there will be lots of new material covered during this module, and that going slowly and asking for help is very much encouraged!
2. Quick road map of what will be covered in this lesson
    a. Overview slide for the day (will require instructor edits if adapting for different class lengths!)
    b. Three classes (50-60 minutes)
        i. Class 1 – Introductory lecture (30 min.) and completion of Activity A, Objectives 1 and 2 (20 min.)
        ii. Class 2 – Completion of Activity A, Objectives 3 and 4 (30 min.), and Activity B, Objectives 5 and 6 (20 min.)
        iii. Class 3 – Completion of Activity B, Objectives 7 and 8 (20 min.), and Activity C (20 min.); wrap-up (10 min.)
    c. Two classes (75-90 minutes)
        i. Class 1 – Introductory lecture and completion of Activity A
        ii. Class 2 – Activity B & C followed by 10-15 minute presentation discussion of each groups' results
    d. One class (3 hours)
        i. Introductory lecture – 30 mins, Activity A – 45 mins, break – 5mins, Activity B – 45 mins, group presentation/discussion – 15/20mins, Activity C – 30 mins
3. Big picture, we frame ecological forecasts in the context of changing climate and land use. Our focus today is on aquatic ecosystems: management of these resources could be improved by having advance knowledge of how they could potentially change in the near-term (e.g., 1 to 30 days in the future).
    a. Why do we want to generate ecological forecasts? Answer: Because there is lots of variability in how climate change is occurring globally and lakes provide critical ecosystem services for humans, so ecological forecasts are critical to help management of these resources.
4. Ask the class "What is a Forecast?" – if teaching virtually, prompt them to either type answers into the chat or raise their hand to ask the question. Key aspects of a forecasts are listed in the slide.
    a. Actionable implies that the information is given in a time frame that allows for a response. For example, you could forecast air temperature five minutes into the future with relatively low uncertainty, but this is not much use compared to a forecast for air temperature three days into the future.
    b. **Highlight the inclusion of uncertainty, which is critical to give decision-makers a clear picture of the probability of different forecasted outcomes, and the focus of the module today.**
5. **<u>Forecast uncertainty:</u>** the range of possible alternate future conditions predicted by a model.

    a. We generate multiple different predictions of the future because the future is inherently unknown.
    b. Uncertainty generally increases with time into the future.
    c. Here is a plot showing 16-day forecast of air pressure with shaded regions showing 95% confidence interval and the solid line represents the median. The confidence interval represents the uncertainty in the forecast.
6. This leads to our focal question for this module: where does forecast uncertainty come from and how can it be quantified and reduced?
    a. "Quantified" means that we are calculating a numerical value for our uncertainty
7. Briefly (!!) introduce each point in the forecast cycle. Highlight that it is "iterative" which means that it is a repetitive process, hence why it is described as a cycle
    a. Create Hypothesis – use the example of how primary productivity is affected by water temperature and underwater light.
    b. Build Model – use data to build a mathematical model to describe the observations of water temperature. Driver data are variables (such as future water temperature and underwater light) which can be used to drive the model.
    c. Quantify uncertainty – this is a key step. There are several different types of uncertainty sources that could be included: e.g., driver, process, parameter, initial conditions.
    d. Generate Forecast – using the model built for generating a forecast of primary productivity.
    e. Communicate Forecast – potential forecast users could include a water resource manager for a drinking water reservoir. There are different ways to communicate a forecast, but this is a critical step as a forecast is not effective for helping management if it is not communicated effectively to end users.
    f. Assess forecast – As time goes on, new data are collected and then this can be compared to previous forecasts to assess how accurate the forecast was.
    g. Update model – if the model is not accurately predicting observations, the model can be updated by changing the model parameters, adding a new driver variable; additionally, a model can be updated via data assimilation, which is using the most recent observations to update the model's initial, or starting, conditions to ensure the most accurate possible forecast
    h. In this module, we will be focusing on building a model, quantifying the uncertainty associated with our forecast, and generating a forecast.
8. We will use ecological models to generate our forecasts
    a. An **ecological model** is a simplified representation of nature, with the goal of understanding and predicting environmental dynamics
    b. A simple version of a model is a linear model: "y = mx + b", and today we will be using a model to predict water temperatures in a lake, where today's surface water temperature is the "y" in the equation and the "x" will be other drivers such as yesterday's surface water temperature and air temperature.
9. Ecological models are inherently uncertain as our data are imperfect and we cannot perfectly represent the real world. Uncertainty in ecological models can arise from a number of sources.

   a. For example, **process uncertainty** is uncertainty caused by our inability to model all processes as observed in the real world. Here, we are modeling the effect of air temperature on water temperature, but we may be missing other important processes that affect water temperature, such as inflows to the lake, and that leads to process uncertainty.
   b. In a linear regression, process uncertainty can be thought of as uncertainty in the structure of the model equation (y = mx + b) – maybe there is another equation that would better represent water temperature processes.
10. Second, **parameter uncertainty** refers to the uncertainty in the model parameter values, which can be due to uncertainties in the data or the calibration process used. For example, if we were to refit this linear model with an additional year of data or using data from a different type of temperature sensor, our parameter values might change slightly, leading to parameter uncertainty.
    a. In a linear regression, parameter uncertainty refers to uncertainty in the values of the slope (m) and intercept (b) parameters.
11. Third, **initial conditions uncertainty** refers to uncertainty arising because the current conditions in an ecosystem are not precisely known. For example, suppose we have a model that uses today's water temperature to predict tomorrow's water temperature. We would consider today's water temperature to be the **initial condition.** But we may not have a water temperature observation every day, and so we do not know what the current water temperature conditions are in the lake, leading to initial conditions uncertainty. Additionally, even if we do have a measurement every day, there is observation uncertainty associated with that measurement.
    a. In a linear regression, initial conditions uncertainty refers to uncertainty in our y_t value, which in this case is today's water temperature.
12. Finally, **driver data uncertainty** comes from inaccuracies in the forecasted variables used to drive the model. For example, we all know that air temperature forecasts are not perfect. If we are using an air temperature forecast as input to our linear regression to forecast water temperature, imperfections in the air temperature forecast will translate to imperfections in the water temperature forecast, leading to driver data uncertainty.
    a. In a linear regression, driver data uncertainty refers to uncertainty in x, which in this case is air temperature.
13. We can represent uncertainty using **distributions*,*** which describe all the possible values a variable of interest might have and how likely those values are. Here are two examples of distributions of possible water temperature with low uncertainty (A) and high uncertainty (B).
14. Furthermore, we can use distributions to generate probabilistic, as opposed to deterministic, forecasts.
    a. **Deterministic forecasts** do not include uncertainty and predict the future as a single value (represented by a single line in the figure).

    b. **Probabilistic forecasts** do include uncertainty and predict the future as a range of possible outcomes, where the range of possible outcomes corresponds to a distribution, such that some outcomes are more likely than others.

15. Ensemble forecast – talk through the animation. Emphasize that each member of the ensemble is a potential future path and has an equally likely chance of occurring.
    a. To generate an **ensemble forecast,** instead of running just a single forecast, a model is run multiple times with slightly different conditions, often based on a distribution of possible values
    b. The complete set of forecasts is referred to as the **ensemble**
    c. Individual forecasts within it are **ensemble members**
    d. Commonly, each member of the ensemble is **equally likely** to occur
    e. **This allows us to generate a probabilistic forecast**
16. Often, it is beneficial to reduce the uncertainty in a forecast to give decision-makers a clearer picture of what the future might hold. There are two steps to managing forecast uncertainty. The first step is to quantify uncertainty, which involves determining the magnitude of uncertainty in a forecast and from which sources it arises.
17. The second step is reducing forecast uncertainty. Methods for reducing forecast uncertainty depend on which source of uncertainty you are trying to reduce.
    a. For example, if you determine that process uncertainty is a large source of uncertainty to your forecast, you might try to improve the forecast model by changing the model structure to better reflect processes that are occurring in the real world. In our case, we might try to change the model structure to account for the effect of inflows on lake water temperature, for example.
18. If you determine that parameter uncertainty is a large source of uncertainty to your forecast, you might try to improve the forecast model by collecting more data on key parameters. For example, if you are trying to forecast abundance of fish in a lake, you might collect additional data on fish growth and mortality rates to better parameterize your model.
19. If you determine that initial conditions uncertainty is a large source of uncertainty to your forecast, you might try to reduce forecast uncertainty by collecting more frequent data so that current conditions in the ecosystem at any given time are more precisely known. For example, if you are trying to forecast water temperature, you might install a sensor that collects water temperature data every hour, reducing initial conditions uncertainty.
20. If you determine that driver data uncertainty is a large source of uncertainty in your forecast, you might try to reduce it by improving the accuracy of your driver data forecast. For example, if you are using an air temperature forecast as input to generate a water temperature forecast, improving the quality of the air temperature forecast will reduce driver data uncertainty in the water temperature forecast.
21. We are going to generate forecasts of **water temperature in lakes** using **ecological models** calibrated to real data from the **National Ecological Observatory Network (NEON)**.

    a. We will **explore** the sources of your forecast's uncertainty, **quantify** its uncertainty, and learn how to **manage** the **uncertainty** of your **forecast**.
22. Water temperature exerts a major influence on biological activity and growth, has an effect on water chemistry, can influence water quantity measurements, and governs the kinds of organisms that live in water bodies.
    a. For example, trout do best in a certain range of water temperature, and above this temperature they are more likely to exhibit stress signals.
    b. Freshwater ecosystems are currently experiencing a multitude of stressors such as land use change and climate change.
    c. Being able to predict how such systems can change in the short-term (up to 7 days into the future) will provide natural resource managers with critical information to take pro-active actions to prevent degradation of water quality.
23. Today, we will be forecasting future water temperature using four simple models. First, we will use a persistence model, which states that water temperature tomorrow (t+1) is equal to water temperature today (t).
24. Next, we will use two linear regression model: one that uses today's water temperature to predict tomorrow's water temperature with some slope (m) and intercept (b), and one that uses a forecast of tomorrow's air temperature to predict tomorrow's water temperature, with some slope (m) and intercept (b).
25. Finally, we will use a multiple linear regression model that uses both today's water temperature and tomorrow's forecasted air temperature to forecast tomorrow's water temperature. In this model, there are three model parameters: two coefficients for the air temperature and water temperature predictors (beta 1 and beta 2) and an intercept (beta 3).
26. We will be forecasting water temperature at National Ecological Observatory Network (NEON) sites
    a. 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. Today we will be forecasting water temperature at NEON lake sites.
27. Learning objectives!
    a. Talk through these with the students one by one: use the embedded animations to sequentially show each of the six bullet points.
28. Introduce Activity A, which has four objectives (have students work in pairs).
    a. Select a lake site
    b. Explore water and air temperature data
    c. Build water temperature forecast models
    d. Generate deterministic forecasts
29. Activity B – continue working in pairs and answer the questions individually. Explore forecast uncertainty and compare how different sources of uncertainty affect the different models. Each objective will explore one source of uncertainty.
    a. Process uncertainty

    b. Parameter uncertainty
    c. Initial conditions uncertainty
    d. Driver uncertainty
30. Activity C – this activity has two objectives
    a. Generate forecasts with total forecast uncertainty and quantify the different contributions of each source of uncertainty to forecasts
    b. Decide from which depth to release water from a reservoir using forecasts which include uncertainty
31. Shiny App:
    a. The module can be accessed at: https://macrosystemseddie.shinyapps.io/module6/
    b. This is an interactive webpage built using R code
    c. It has interactive plots and options embedded which allow you to build your own personal model, visualize and explore the data, and answer questions
32. Generating the student report. At this point, the instructor may choose to navigate to the Shiny app and demonstrate its features while screen sharing the app in a browser. Alternatively, the instructor may cover the material on slides 32-34 in PowerPoint.
33. Saving & resuming progress in the Shiny app.
34. We recommend that you save your progress often!
35. Help students transition over to the Shiny app and get started. Instructors may use the table to record which students are working with which lake sites, if helpful.
36. Thank you for participating! If you would like to learn more, please check out the other Macrosystems EDDIE ecological forecasting modules.

Guide to Shiny App

Overview
This is the landing page of the Shiny app. It gives an overview of the module - there are no questions students need to complete on this tab.

Presentation
Here is a recap of some of the key points from the presentation. There is some text and also the main slides from the presentation.

Introduction
The introduction outlines the workflow for the module and provides instructions about how to save and resume progress. This is also where they will download the module report into which they should type their answers. Students should answer questions 1-2 (hereafter, denoted as Q1-2).

Site Selection

Exploration allows the students to explore data associated with different National Ecological Observatory Network (NEON) lake sites. They will answer questions about their site characteristics and patterns and relationships in the data at their site.

Tips for Site Selection:

- When they are done with Site Selection, they should have chosen a NEON lake site to work with for Activities A and B. If they do not choose a site, the app will display warnings when they proceed to Activity A to go back and choose a site.

Activity A: Build A Model With Uncertainty

Activity A challenges the students to construct different models to predict water temperature and explore model parameters.

Tips for Activity A:

- **Important: Tell the students to read through the detailed text and embedded slides in each section as this will explain what is happening within each objective and help answer questions.**
- When they are done with Objective 1, they should have chosen a NEON lake site to work with for the rest of the module. If they do not choose a site, the app will display warnings when they proceed to subsequent objectives to go back and choose a site.
- In Objective 2, students will be asked for the first time to download a plot from the Shiny app and copy-paste it into their final report. You may need to help students navigate to their Downloads folder and retrieve the plot file, which is a .png file that will have the corresponding question and the date of download in the filename (e.g. Q8a-plot-2023-09-21.png). Once they open the file, they can just Ctrl+C and then Ctrl+V (or equivalent on Mac) to insert the plot into their Word document.
- Starting in Objective 3, many of the objectives have embedded carousels of scrolling PowerPoint slides which provide important information for the objective (as well as for some of the questions embedded throughout the app!). **Emphasize to students that scrolling through these slides can help them answer questions and understand the module.**
- In Objective 3, students will be asked to fit four different forecast models using data from their chosen lake site. If they do not fit all four models, the app will display warnings when they proceed to subsequent objectives to go back and fit all four models.
- Walk around the pairs/move between breakout rooms and make sure that everyone can follow along the Shiny app successfully.
- **When you close class, remind students to save their progress by clicking the "Bookmark my progress" link at the top left of the app and copy-pasting the link at the top of their student handout. If they do not do regularly save their progress, they may not be able to retrieve the models and plots they have generated during the class period.**

Activity B: Explore Forecast Uncertainty

At this point, students will explore the different sources of forecast uncertainty and see how the sources of uncertainty affect a suite of simple forecast models. Through this they will gain understanding about how process, parameter, initial conditions and driver uncertainty affect forecasts of water temperature individually.

Tips for Activity B:

- **Important: Tell the students to read through the detailed text within each objective. We have embedded lots of directions, hints, and troubleshooting help within the Shiny app text! We encourage instructors to read and work through the Shiny app before teaching the module so that you are familiar with all of the steps of this activity.**
- If you are continuing from a previous lesson, it is good to show the students how to reload the their progress in the app by copy-pasting their link into their web browser.
- For Objectives 5-8, the slide carousels at the top of each objective explain how each form of uncertainty is quantified. Emphasize to students that these slides contain important information to understand each objective.
- Students will be generating forecasts with four different models. Sometimes, a particular source of forecast uncertainty does not apply to every model. For example, parameter uncertainty is uncertainty in model parameters – but a persistence model doesn't have any parameters! So this model does not have any parameter uncertainty. Encourage students to think critically about the patterns of uncertainty they see across models.
- Remind students to download plots and answer questions in their Word documents as they go.
- Walk around the pairs/move between breakout rooms and make sure that everyone can follow along the Shiny app successfully.
- **When you close class, remind students to save their progress by clicking the "Bookmark my progress" link at the top left of the app and copy-pasting the link at the top of their student handout. If they do not do this, they will not be able to retrieve the models and plots they have generated during the class period.**

Activity C: Managing Uncertainty

This is a relatively short activity with two objectives. First, students generate forecasts which include all sources of uncertainty, and they then can quantify the different contributions of each source of uncertainty to forecasts made using different models. Second, students are presented with a management scenario and must make a decision using a forecast with uncertainty. These two objectives can provide fodder for group discussion about how visualization of uncertainty can affect decision making.

Resources and References

Optional pre-class readings and videos:

Articles:

- Dietze, M., & Lynch, H. (2019). Forecasting a bright future for ecology. Frontiers in Ecology and the Environment, 17(1), 3. https://doi.org/10.1002/fee.1994
- Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., Larsen, L. G., Loescher, H. W., Lunch, C. K., Pijanowski, B. C., Randerson, J. T., Read, E. K., Tredennick, A. T., Vargas, R., Weathers, K. C., & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences, 115(7), 1424–1432. https://doi.org/10.1073/pnas.1710231115
- Jackson, L. J., Trebitz, A. S., & Cottingham, K. L. (2000). An Introduction to the Practice of Ecological Modeling. BioScience, 50(8), 694. https://doi.org/10.1641/0006-3568(2000)050[0694:aittpo]2.0.co;2

Videos:

- NEON's Ecological Forecast: The Science of Predicting Ecosystems
- Fundamentals of Ecological Forecasting Series
  - Why Forecast?

Recent publications about EDDIE modules:

- Carey, C. C., R. D. Gougis, J. L. Klug, C. M. O'Reilly, and D. C. Richardson. 2015. A model for using environmental data-driven inquiry and exploration to teach limnology to undergraduates. Limnology and Oceanography Bulletin 24:32–35.
- Carey, C. C., and R. D. Gougis. 2017. Simulation modeling of lakes in undergraduate and graduate classrooms increases comprehension of climate change concepts and experience with computational tools. Journal of Science Education and Technology 26:1-11.
- Klug, J. L., C. C. Carey, D. C. Richardson, and R. Darner Gougis. 2017. Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy. Ecosphere 8:e01733.
- Farrell, K.J., and C.C. Carey. 2018. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. Ecology and Evolution 8:7744-7751. DOI: 10.1002/ece3.4363
- Carey, C. C., Farrell, K. J., Hounshell, A. G., & O'Connell, K. 2020. Macrosystems EDDIE teaching modules significantly increase ecology students' proficiency and confidence working with ecosystem models and use of systems thinking. Ecology and Evolution. DOI: 10.1002/ece3.6757

- Moore TN, Thomas RQ, Woelmer WM, Carey CC. Integrating Ecological Forecasting into Undergraduate Ecology Curricula with an R Shiny Application-Based Teaching Module. *Forecasting*. 2022; 4(3):604-633. https://doi.org/10.3390/forecast4030033

We'd love your feedback!

We frequently update this module to reflect improvements to the code, new teaching materials and relevant readings, and student activities. Your feedback is incredibly valuable to us and will guide future module development within the Macrosystems EDDIE project. Please let us know any suggestions for improvement or other comments about the module at http://module6.macrosystemseddie.org or by sending an email to MacrosystemsEDDIE@gmail.com

Answer Key

The following plots are indicative of what student model output should look like (approximately) if the module is run correctly. We note that answers may vary depending on which lake and model the students run in the module. Answers are given below as bullet points beneath each question.

Pre-class activity: Explore how uncertainty in predictions can affect decision-making

Read the following paper, which you can either access independently online or obtain from your instructor:

Pielke, R. A. (1999). Who decides? Forecasts and responsibilities in the 1997 Red River flood. Applied Behavioral Science Review, 7(2), 83–101.

Refer to the paper you read to answer the questions below.

A. Ahead of the 1997 Red River flood in East Grand Forks, the National Weather Service (NWS) provided two river crest predictions of 47.5 and 49 ft. How were these two different predictions made? (What was the difference between the two predictions?).

   Answer: The two predictions are based on two different scenarios: the first assumes average temperatures for that time of year and no precipitation during the prediction period, while the second assumes average temperatures and additional precipitation. Because the Red River is a snowmelt-driven system, it is possible to make river crest predictions using temperature.

B. Pielke reports that many people misinterpreted the two river crest predictions provided by the NWS. Describe two ways in which the river crest predictions were incorrectly interpreted.

   Answer: Some people interpreted the two predictions as a range; some people interpreted the larger number as a maximum possible crest.

C. The NWS did not quantify or report the uncertainty associated with its river crest predictions for the 1997 Red River flood event. Referring to Fig. 2 in the paper, what do you think would be a reasonable estimate of the uncertainty associated with NWS river crest predictions? Explain your reasoning.

   Answer: Answers will vary; estimates between 0-10 feet (exclusive) would probably be reasonable based on Fig. 2 in the Pielke (1999) paper.

D. Pielke concludes that “Confusion about the uncertainty of the [river crest] predictions led to misplaced responsibility for flood fight decision making.” Explain what is meant by this statement in your own words.

   Answer: Answers will vary: the idea is that because the NWS provided a prediction with two numbers that were not clearly explained and with no quantified uncertainty, local decision-makers did not have enough information to make informed decisions. As a result, responsibility for flood decision making was implicitly transferred from local decision-makers (where it belongs) to the NWS (where it does not belong).

E. Reflecting on what you have read, explain how reporting, or failing to report, the uncertainty associated with future predictions of natural phenomena can affect decision-making.

   Answer: Answers will vary; student should recognize that inclusion of uncertainty is critical to giving decision makers an accurate picture of the likelihood of possible future outcomes so that they can make an informed decision.

Now navigate to the [Shiny interface] to answer the rest of the questions.

The questions you must answer are written both in the Shiny interface as well as in this handout. As you go, you should fill out your answers in this document.

Think about it!

Answer the following questions:

1. What is meant by the term 'uncertainty' in the context of ecological forecasting?

   Answer: The range of possible alternate future conditions predicted by a model (defined on slide 5 of the instructor PowerPoint)

2. How do you think knowing the uncertainty in a forecast helps natural resource managers? For example, if a drinking water manager received a toxic algal bloom forecast with high vs. low uncertainty in a bloom prediction, how might that affect their decision-making?

   Answer: Answers will vary; if uncertainty is high, a manager might not allow the forecast to influence their decision as much as when the forecast uncertainty is low.

Note that student answers will vary depending on which lake they choose; provided below are example answers if students were to choose Prairie Pothole Lake (PRPO).

Activity A - Build Models and Generate Forecasts

Build different models to simulate water temperature for your chosen NEON site.

1) Objective 1: Select and view a NEON site

Select a NEON site from the table, then click on the 'View latest photo' button to load the latest image from that site. Follow the link at the bottom of the 'About Site' section to find out more about the site.

---

3. Fill out information about your selected NEON site:

   a. Name of selected site: Prairie Pothole

   b. Four letter site identifier: PRPO

   c. Latitude: 47.129839

   d. Longitude: -99.253147

   e. Lake area (km2): 1.4 km^2

   f. Elevation (m): 579 m

2) Objective 2: Explore water temperature

Explore the water and air temperature data measured at the selected site. This is data that has been downloaded from the NEON Data Portal. The variables shown have been selected for this module but there are a wide range of variables collected at each NEON site.

---

4. List two potential impacts on lakes and inland water bodies as a result of increasing water temperature.

   Answer: Answers will vary; possible changes include changes in dissolved oxygen, nutrient, metal, and algae concentrations (see text in R Shiny app Activity A, Objective 2)

5. Describe any patterns you see in the surface water temperature data for your lake.

   Answer: Answers will vary; students will likely notice a seasonal pattern in water temperatures with warmer temperatures in summer and cooler in winter

6. Fill out the table in your Word document with the summary statistics of water temperature:

   Table 1. Water Temperature Statistics

Variable	Mean	Minimum	Maximum
Surface water temperature	19.07	-0.1	27.42

   Note that while negative water temperatures are not physically possible, the negative minimum values can be a good teaching moment to illustrate to students that environmental sensors are not perfect!

7. Fill out the table in your Word document with the summary statistics of air temperature:

   Table 2. Air Temperature Statistics

Variable	Mean	Minimum	Maximum
Air temperature	15.46	-8.45	25.91

8. Assess a possible relationship between air and water temperature data at your lake.

   a. Click 'Download plot' to save the plot of air and water temperature at your lake to your computer; then, copy-paste it into your report.

      Please copy-paste your Q-8a-plot.png image here.

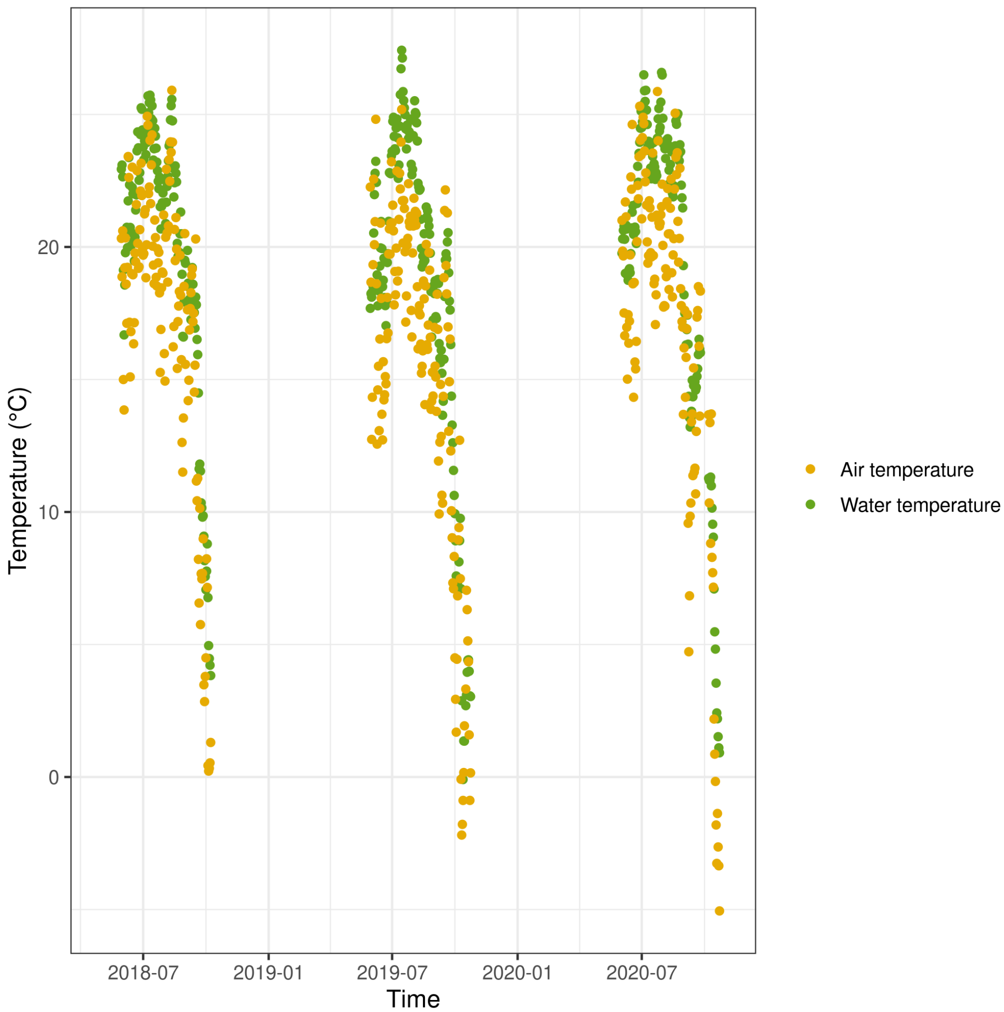


Figure 1. Time series of air and water temperature for your selected NEON lake.

b. Do you think there is a relationship between air temperature and water temperature at this lake? Explain your reasoning.

Answer: Yes, water and air temperature both appear to increase and decrease at the same time.

3) Objective 3: Build models

Use observed water temperature and air temperature data to build linear regression models to predict water temperature.

9. Describe, in your own words, the difference between a linear regression model and a multiple linear regression model.

   Answer: A linear regression model uses a single independent variable (x variable) to predict a dependent variable (y variable), with two parameters, which are the slope (m) and intercept (b). A multiple linear regression model uses multiple independent variables to predict the dependent variable, and each independent variable has its own coefficient parameter (β_1, β_2, etc.), in addition to the model intercept parameter (β_0).

10. Based on your exploration of water and air temperature data at your lake site in the previous objective, do you think it is reasonable to use air temperature as an independent variable in a model to predict water temperature? Why or why not?

    Answer: Yes, because air temperature likely affects water temperature since the two variables track together.

11. Assess the persistence model predictions compared to observations.

a. Click 'Download plot' to save the plot of persistence model predictions and water temperature observations at your lake to your computer; then, copy-paste it into your report.

Please copy-paste your Q-11a-plot.png image here.

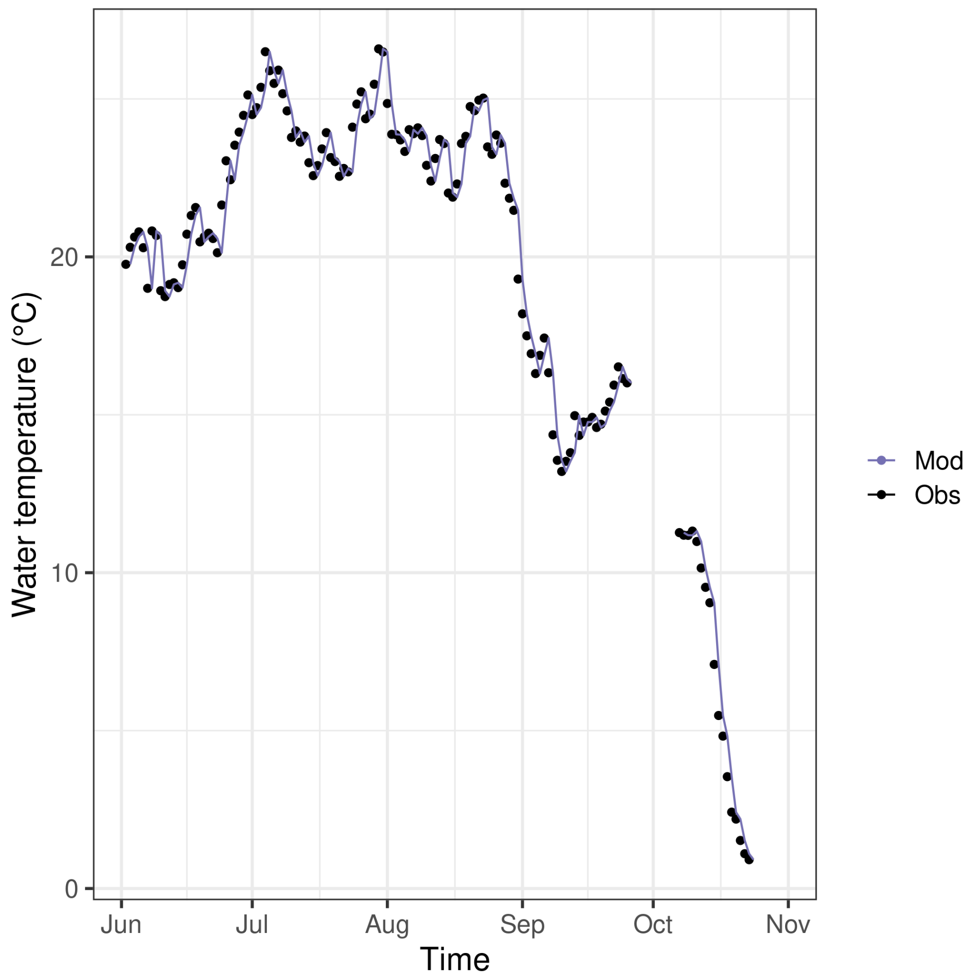


Figure 2. Time series of persistence model predictions and water temperature observations for your selected NEON lake.

b. Based on the plot of observations and model predictions, do you think the persistence model does a good job of predicting model temperature? Why or why not?

Answer: Students are likely to say yes here, because the predictions seem to track the observations; however, if you look closely, the predictions are always slightly offset from the observations because today's value is used as tomorrow's prediction, so actually, the model does ok if temperature does not change much from today to tomorrow, but it does very poorly if there is a big temperature change over the course of a day.

12. Let's pretend a lake manager has asked you to provide a 30-day forecast of water temperature. Do you think a persistence model would be a good choice for generating such a forecast? Why or why not?

    Answer: No, because you would be using today's water temperature to predict water temperature a month into the future, and water temperatures are likely to change a lot over the course of a month. The persistence model is only able to capture one to a few days maximum in the future.

13. Assess the linear relationship between yesterday and today's water temperature.

    a. Click 'Download plot' to download the scatterplot of today's water temperature vs. yesterday's water temperature and copy-paste it into your report.

       Please copy-paste your Q-13a-plot.png image here.

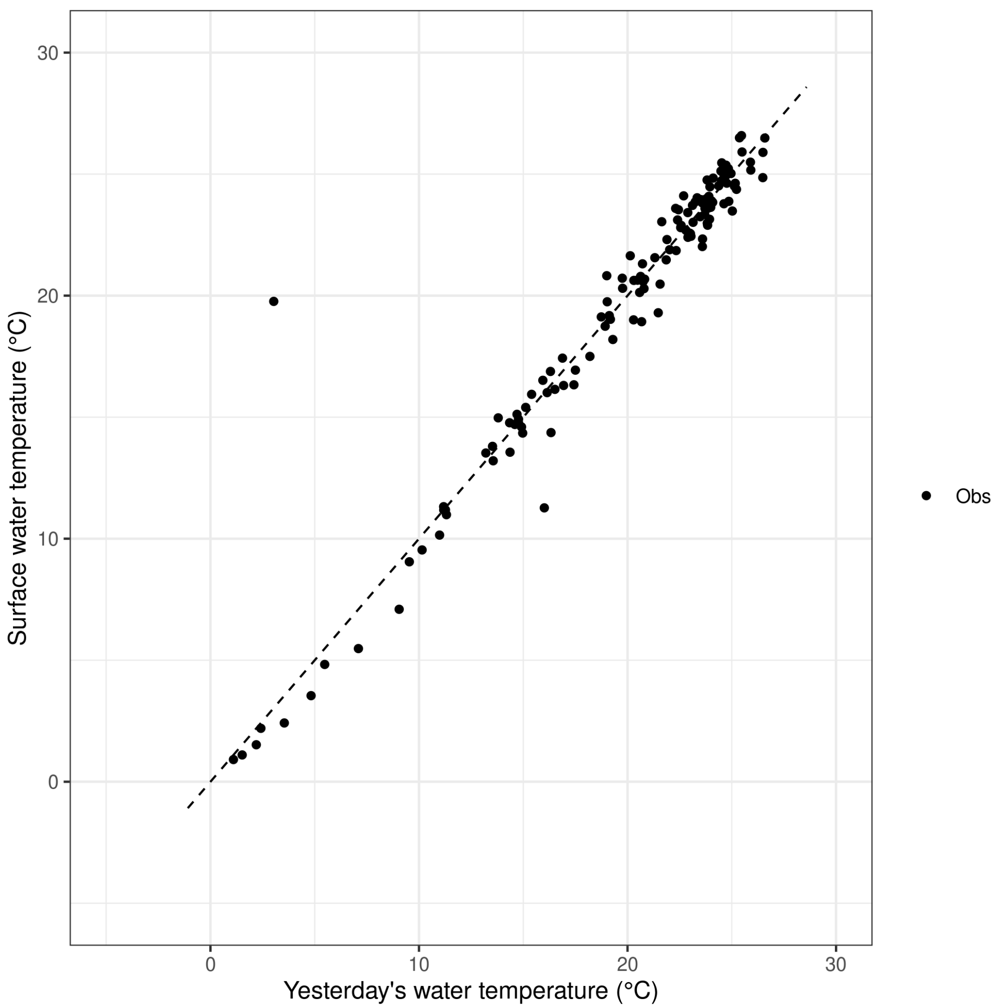


Figure 3. Scatterplot of today's water temperature observations versus yesterday's water temperature observations for your selected NEON lake.

b. Does today's water temperature exhibit a positive or negative linear relationship with yesterday's water temperature? Record your answer in your report.

Answer: Positive

14. Assess the linear relationship between air and water temperature.

a. Click 'Download plot' to download the scatterplot of today's water temperature vs. today's air temperature and copy-paste it into your report.

Please copy-paste your Q-14a-plot.png image here.

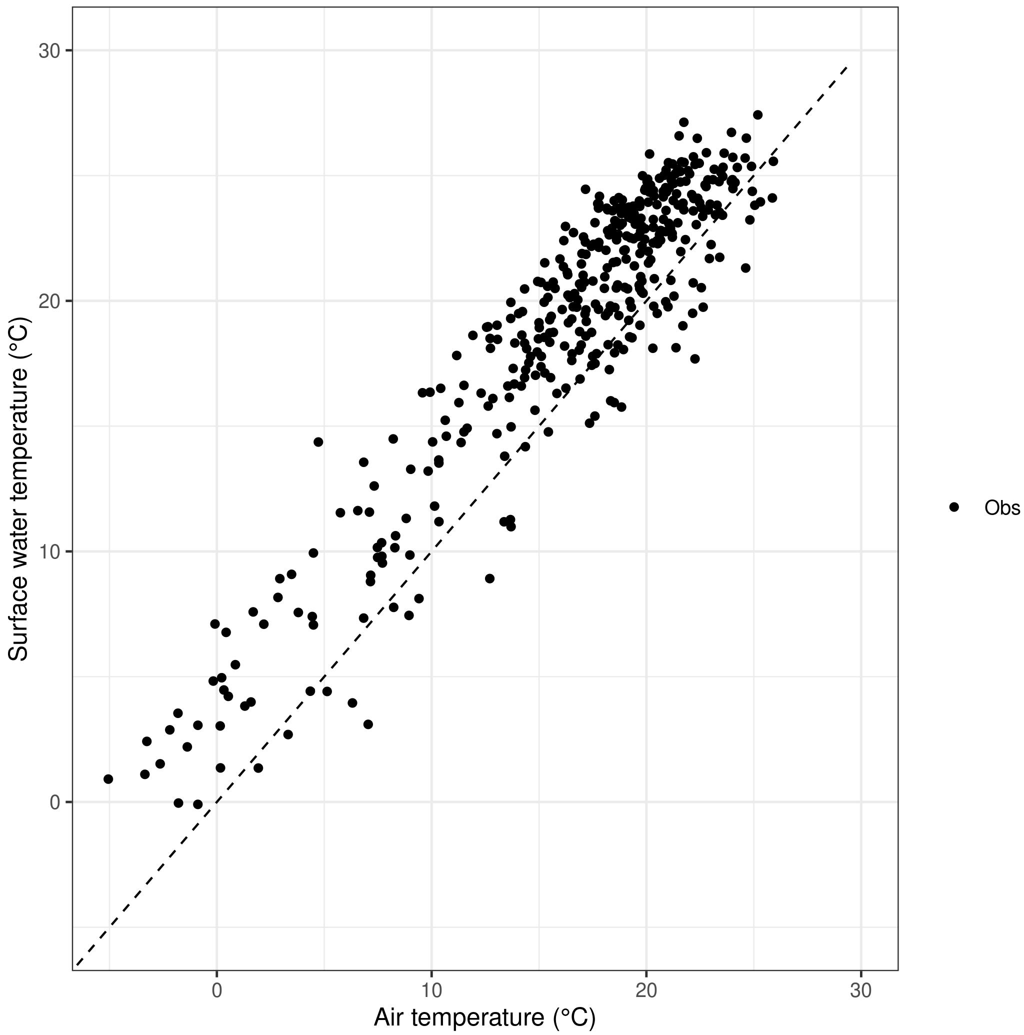
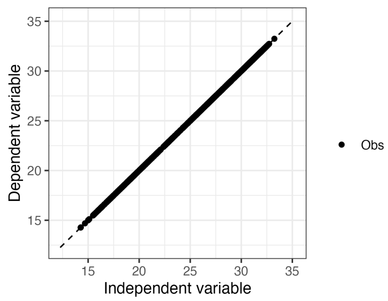


Figure 4. Scatterplot of today's water temperature observations versus today's air temperature observations for your selected NEON lake.

b. Does today's water temperature exhibit a positive or negative linear relationship with today's air temperature? Record your answer in your report.:

Answer: Positive

15. In a scatterplot of a perfect linear relationship, all the points fall exactly on the 1:1 line (see example figure to the right). Knowing this, assess the linear relationships between: 1) today's water temperature and yesterday's water temperature; and 2) today's water temperature and today's air temperature. Which plot do you think exhibits a stronger linear relationship, and why?



Example of perfect linear relationship

Answer: Answers will vary depending on the lake students have chosen; for PRPO, yesterday's water temperature exhibits a stronger relationship with today's water temperature than today's air temperature does. You can tell because the points in plot Q13a fall closer to the 1:1 line than for plot Q14a.

16. Interpret the parameters of the linear regression models.

    a. Record the values of the model parameters for BOTH linear models in the Q16a table in your report.

    Table 3. Linear Regression Parameters

Model	Slope (m)	Intercept (b)
Water temperature	0.97	0.6
Air temperature	0.9	4.38

    b. A positive linear regression model with a perfect fit has a slope (m = 1) and an intercept (b = 0). Knowing this, use the parameter values of your two linear regression models to assess the fit of: 1) the model using yesterday's water temperature to predict today's water temperature; and 2) the module using today's air temperature and today's water temperature. Which model do you think exhibits a better fit, and why?

    Answer: Answers will vary based on which lake students have chosen; here, the correct answer is the water temperature model, because the slope is closer to 1 and the intercept is closer to 0.

17. Interpret the parameters of the multiple linear regression model.

    a. Record the values of the model parameters for multiple linear regression model in the Q17a table in your report.

       Table 4. Multiple Linear Regression Parameters

Model	β0	β1	β2
Water temp. and air temp.	1.11	0.63	0.36

    b. Because both of our independent variables have the same units (degrees Celsius), we can compare the values of the coefficients to understand which independent variable has a stronger relationship with the dependent variable. Knowing this, is yesterday's water temperature or today's air temperature more strongly associated with today's water temperature?

       Answer: In all cases, for all lakes, yesterday's water temperature is a stronger driver of today's water temperature than today's air temperature; you can tell because the absolute value of the coefficient on water temperature is greater than the coefficient on air temperature.

18. Assess the model fits.

    a. Click ‘Download plot’ to download the timeseries of all model predictions and water temperature observations and copy-paste it into your report.

    Please copy-paste your Q-18a-plot.png image here.

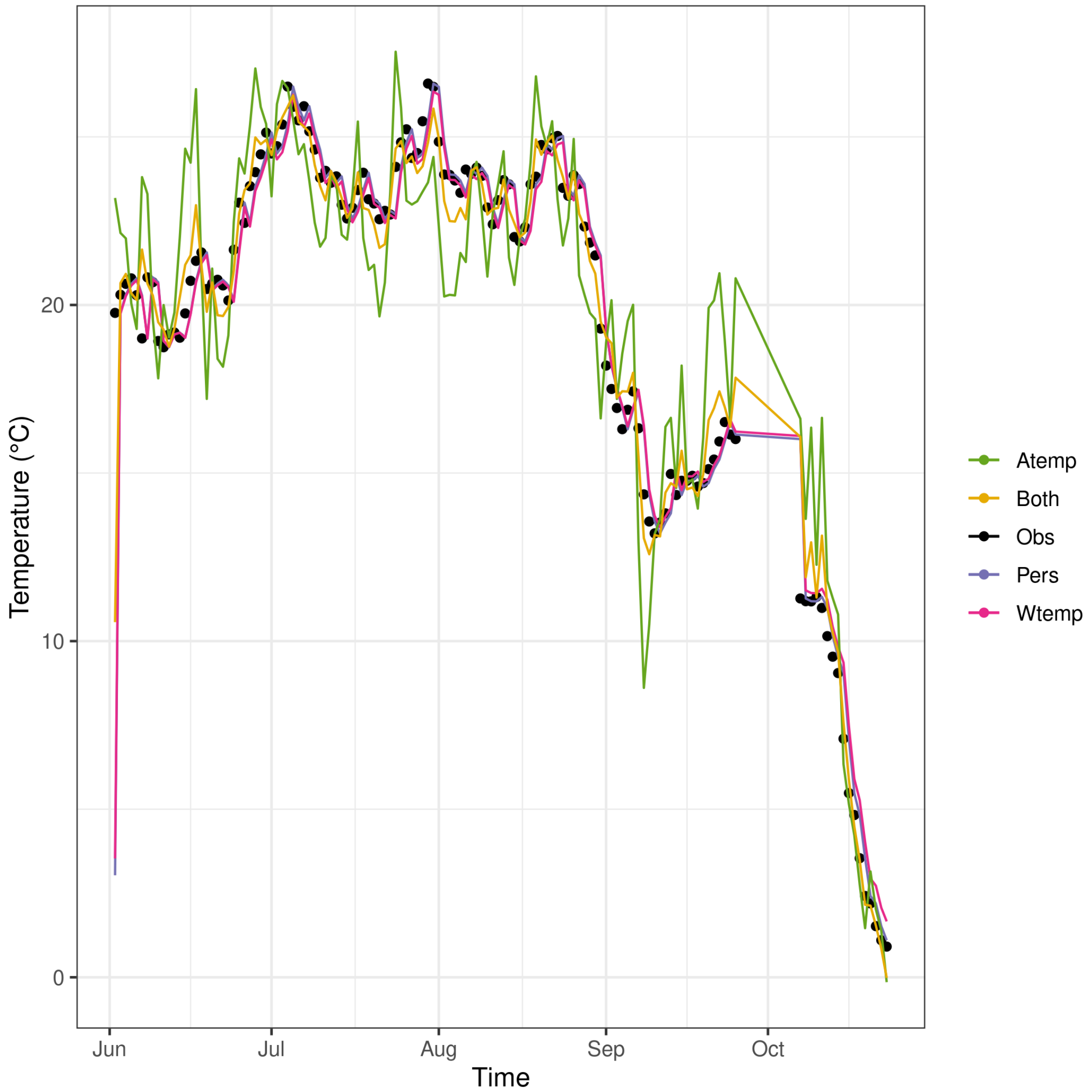


Figure 6. Time series of model predictions and water temperature observations for your selected NEON lake.

    b. Which model do you think will produce the best forecasts of water temperature? Why?

    Answer: Answers will vary, but students should compare model predictions to observations to help justify their responses.

4) Objective 4: Generate Forecasts

Generate multiple forecasts of water temperature.

19. Use your forecasts to estimate future water temperature.

    a. Click 'Download plot' to download the plot of deterministic forecasts and copy-paste it into your report.

    Please copy-paste your Q-19a-plot.png image here.

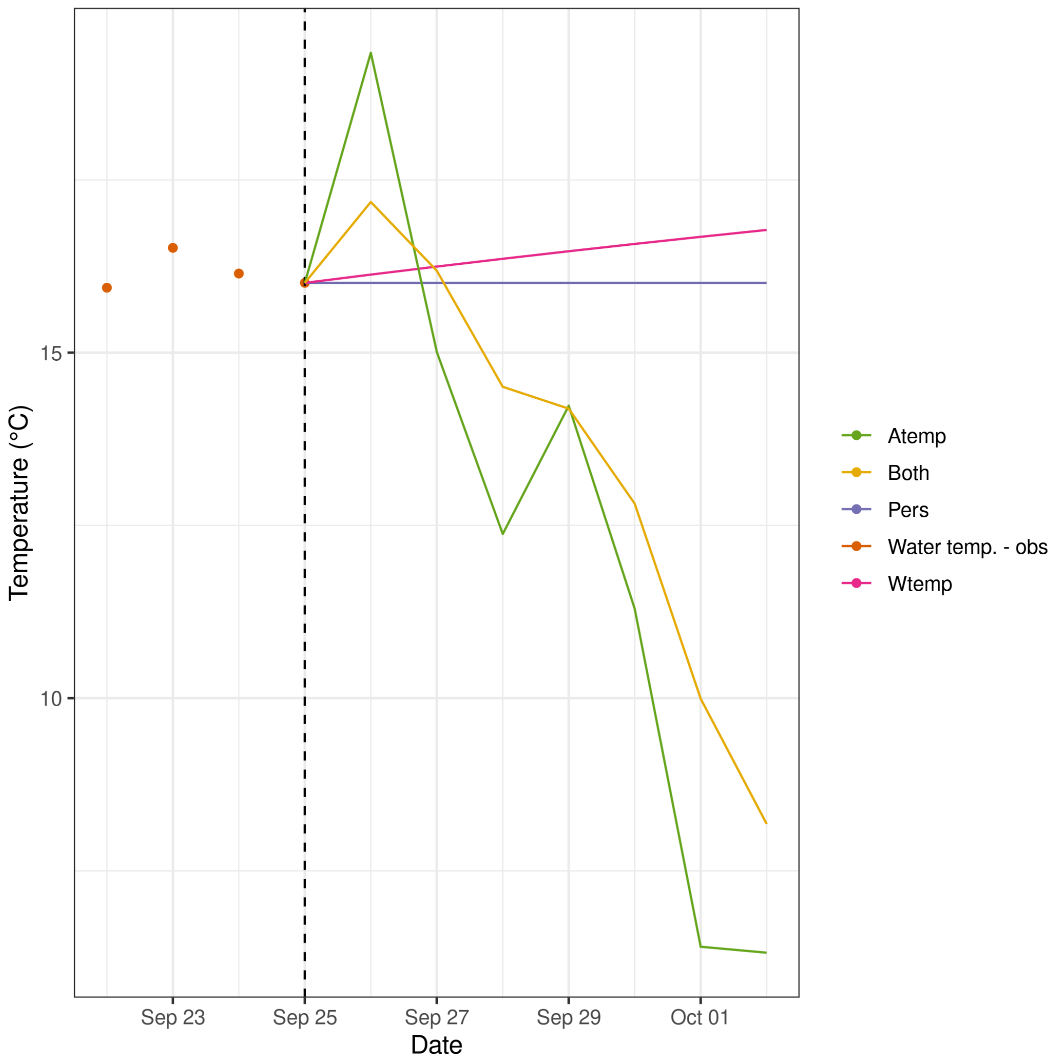


Figure 7. Deterministic water temperature forecasts for your selected NEON lake.

b. Using the forecasts you have generated, what is your best estimate of what the water temperature will be 7 days in the future? Explain your reasoning.:

   Answer: Answers will vary; students should refer to the predictions made by the different forecast models for Oct. 2 to justify their response.

20. Describe, in your own words, the difference between a deterministic and a probabilistic forecast.

    Answer: Deterministic forecasts do not account for uncertainty associated with predictions, while probabilistic forecasts do.

Activity B - Explore Forecast Uncertainty

Generate multiple forecasts of water temperature and examine how different sources of uncertainty affect the different models.

5) Objective 5: Process Uncertainty

Explore how process uncertainty affects the different models when forecasting water temperature.

Note that for Q21-23, students should use the slide carousel at the top of Activity B, Objective 5 to help them answer the questions.

21. What are model residuals?

    Answer: The differences between values predicted by a model and observations.

22. Briefly explain, in your own words, how the residuals of a model can be used to generate a process uncertainty distribution.

    Answer: You can find the standard deviation of the model residuals and use that as the standard deviation of a process uncertainty distribution (with a mean of 0).

23. Define 'ensemble forecast' in your own words.

    Answer: An ensemble forecast is generated by running a model many times with slightly different conditions to produce many possible predicted outcomes; as a result, ensemble forecasts account for uncertainty in model predictions.

24. Interpret the process uncertainty distribution plot.

a. Click 'Download plot' to download the process uncertainty distribution figure and copy-paste it into your report.:

Please copy-paste your Q-24a-plot.png image here.

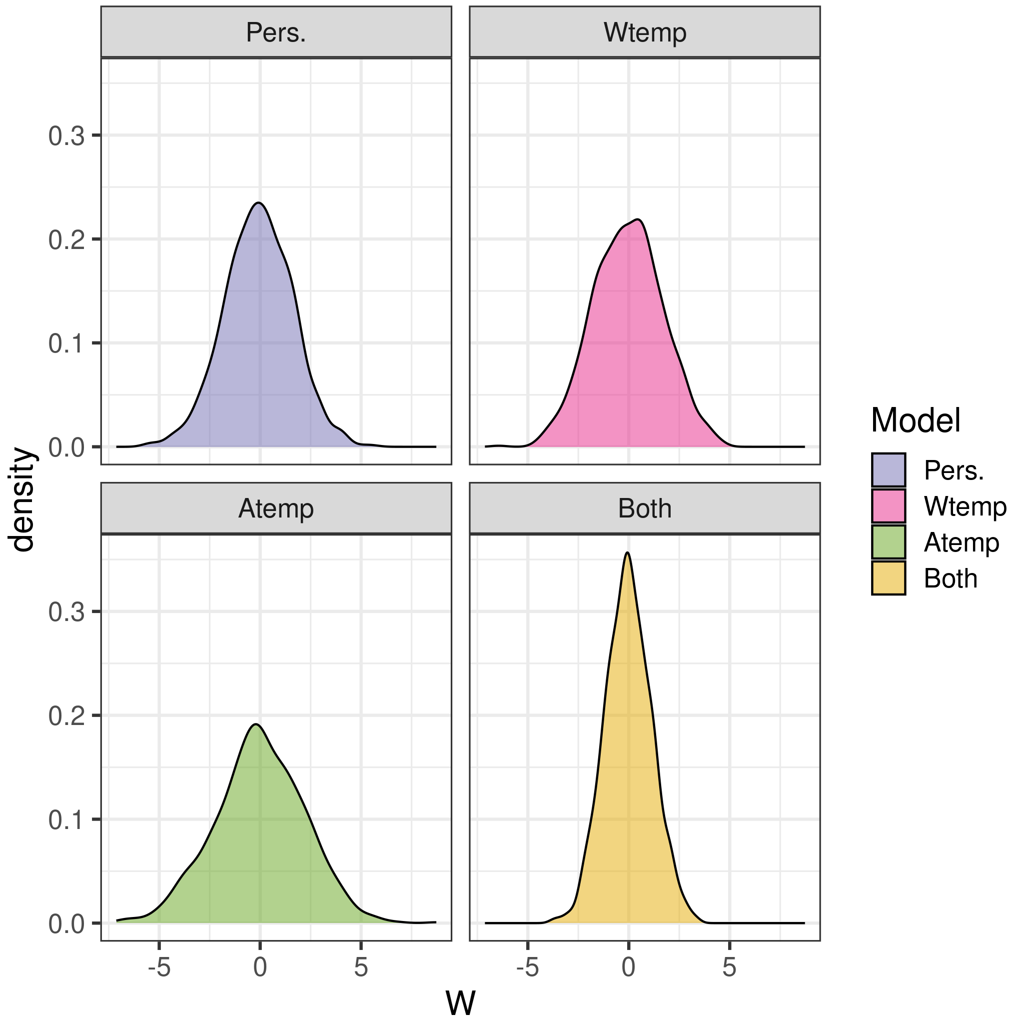


Figure 8. Distributions of process uncertainty for all four models.

b. Examine the plot of process uncertainty distributions on the right. Which model do you think has the smallest amount of process uncertainty? Explain your reasoning.:

Answer: In this figure, the "Both" model has the smallest amount of process uncertainty, because its distribution is the narrowest (smallest standard deviation of residuals).

25. Describe how forecast uncertainty changes into the future.

a. Click 'Download plot' to download the plot of forecasts with process uncertainty and copy-paste it into your report.

Please copy-paste your Q-25a-plot.png image here.

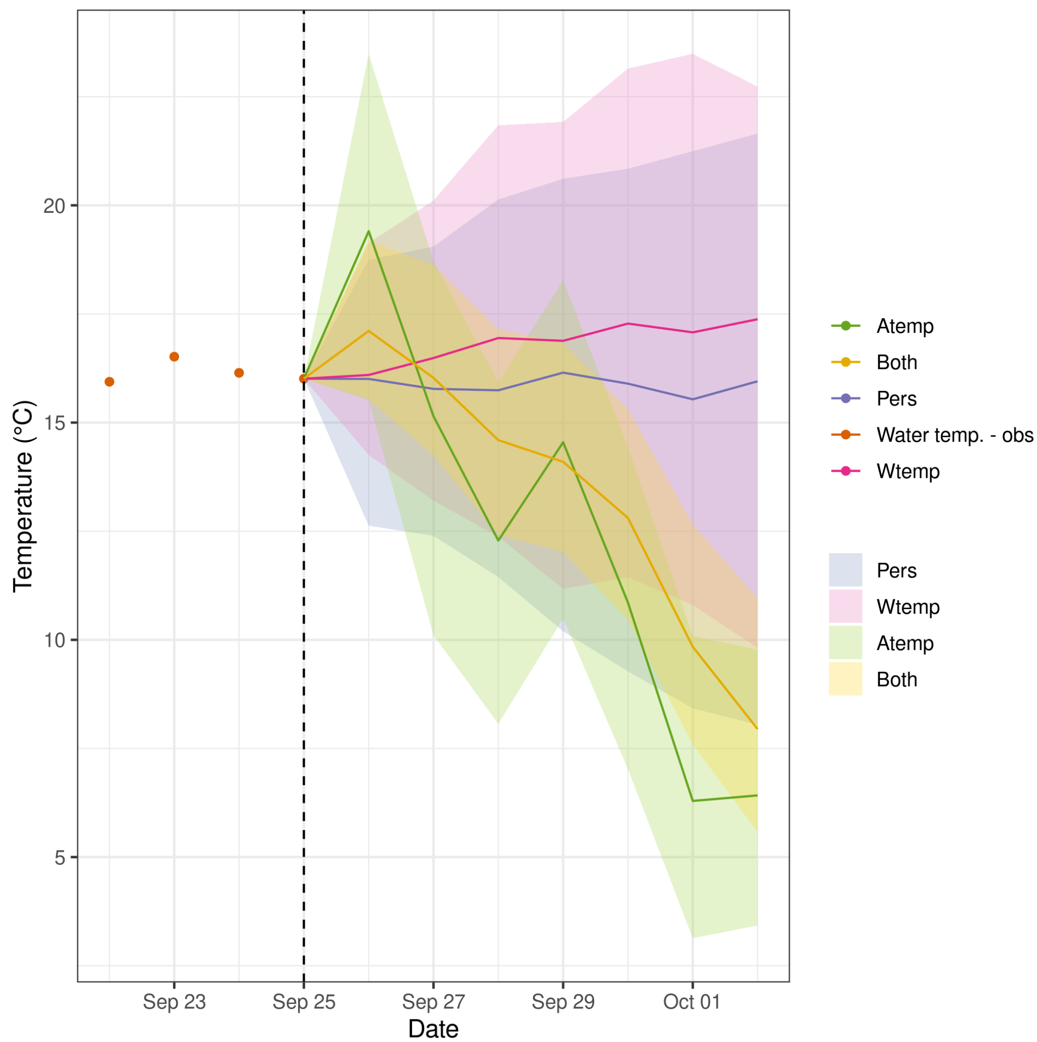


Figure 9. Water temperature forecasts with process uncertainty for your selected NEON lake.

b. Examine the plot of forecasts with process uncertainty. How does uncertainty change as the forecast goes farther into the future?:

Answer: Forecast uncertainty increases as the forecast goes farther into the future.

26. You have learned that process uncertainty arises when our models do not perfectly represent what is occurring in the real world. Knowing this, describe one way that you think process uncertainty in our water temperature forecasts could be reduced.

    Answer: You could revise the model structure to better capture real-world ecological processes that affect the variable you are trying to forecast.

6) Objective 6: Parameter Uncertainty

Explore how parameter uncertainty affects the different models when forecasting water temperature.

---

27. Compare the parameter values of the two models.

    a. Record the parameter values for the models fit to one and two years of data in the Q27a table your report.

    Table 5. Parameters for Linear Regressions Fit to Different Datasets

Model	Slope (m)	Intercept (b)
1 year model	0.85	4.9
2 year model	0.91	4.06

    b. How do the parameter values of the models fit to one year vs. two years of data compare? Why do you think this might be?

    Answer: They are slightly different, and this is probably because the models were fit to different datasets (one dataset had more data than the other).

28. Assess the contribution of parameter uncertainty across models.

a. Click 'Download plot' to download the plot of forecasts with parameter uncertainty and copy-paste it into your report.

Please copy-paste your Q-28a-plot.png image here.

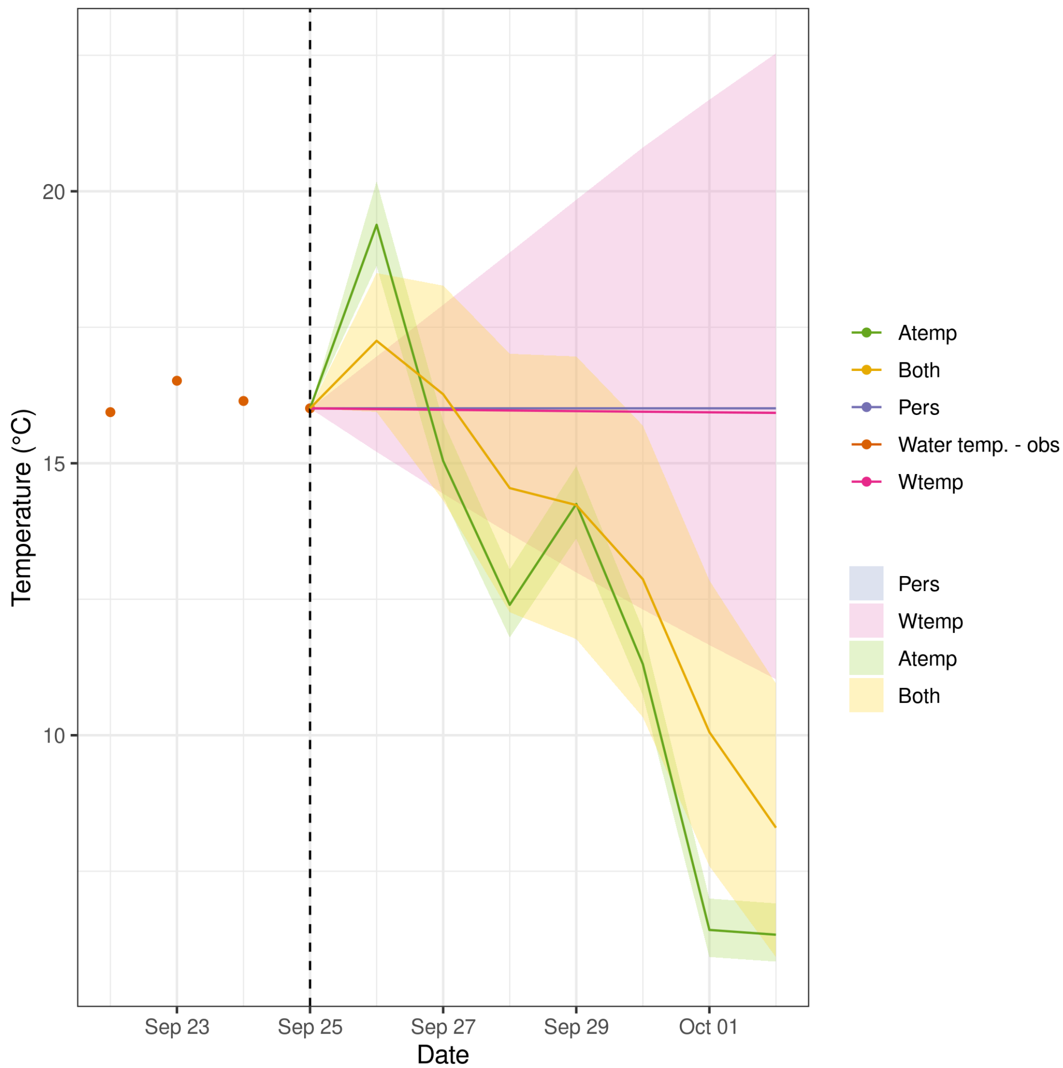


Figure 10. Water temperature forecasts with parameter uncertainty for your selected NEON lake.

b. Which model is affected most by the addition of parameter uncertainty? Why do you think that is?

Answer: The Wtemp model, because it has parameters (unlike the Pers. model, which has no parameters and therefore no parameter uncertainty) and its predictions are not constrained by the air temperature forecast (unlike the Atemp and Both models).

29. You have learned that parameter uncertainty arises when we do not have enough data or when our data do not accurately represent the true relationship between environmental variables. Knowing this, describe one way that you think parameter uncertainty in our water temperature forecasts could be reduced.

    Answer: Improve the quality of collected data and/or collect more data.

7) Objective 7: Initial Conditions Uncertainty

Explore how initial conditions uncertainty affects the different models when forecasting water temperature.

---

30. What initial conditions from your chosen lake site are required to run the forecast models we are using today?

    Answer: Today's water temperature

31. One of the four models we are using for forecasting does not require any initial conditions. Which one is it? Explain how you know.

    Answer: The Atemp model, because the only independent (input) variable in that model is future (forecasted) air temperature.

32. Assess the contribution of initial conditions uncertainty across models.

a. Click ‘Download plot’ to download the plot of forecasts with initial conditions uncertainty and copy-paste it into your report.

Please copy-paste your Q-32a-plot.png image here.

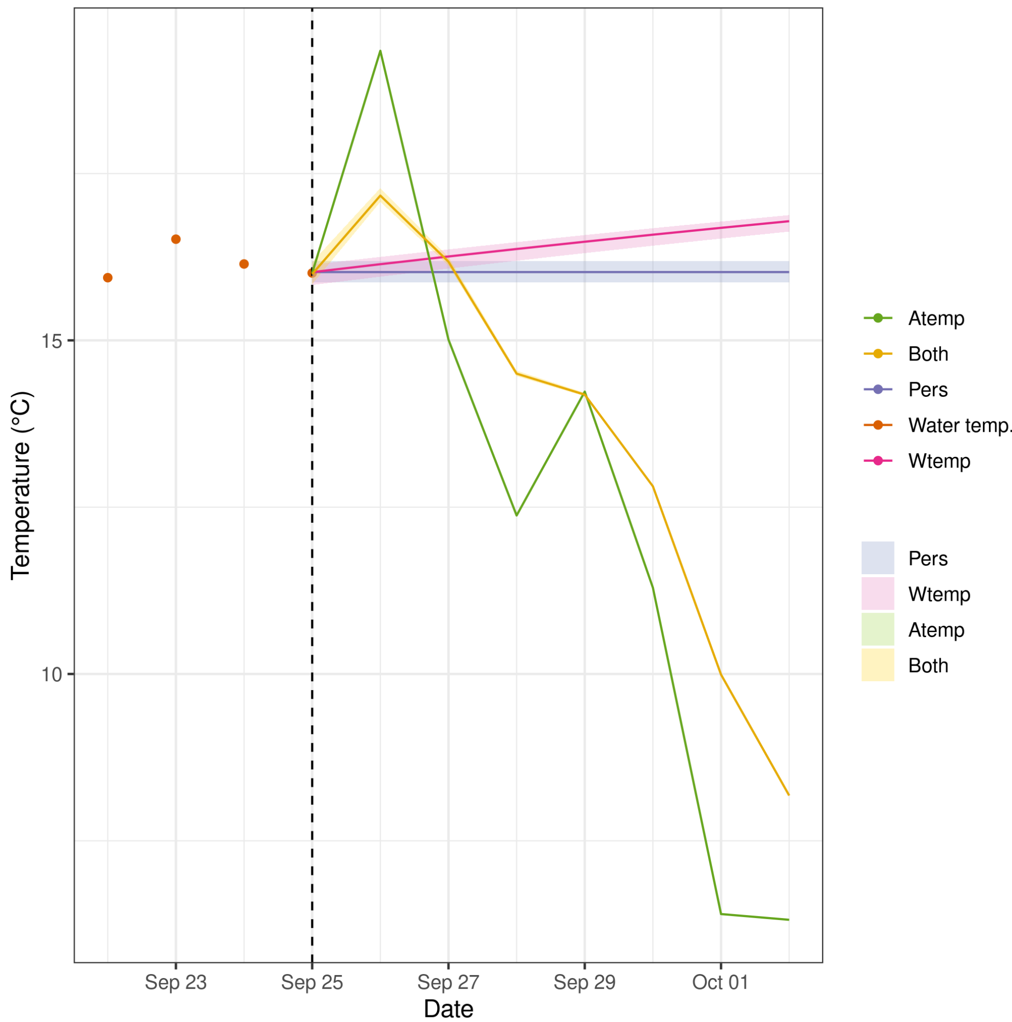


Figure 11. Water temperature forecasts with initial conditions uncertainty for your selected NEON lake.

b. Which model had the least initial conditions uncertainty, and why?

Answer: The Atemp model, because the Atemp model does not require any initial conditions as input.

33. You have learned that initial conditions uncertainty arises from observation error and missing observations. Knowing this, describe one way that you think initial conditions uncertainty in our water temperature forecasts could be reduced.

    Answer: Collect more frequent data on initial conditions and/or improve data collection procedures to reduce the risk of missing data and observation error.

8) Objective 8: Driver Uncertainty

Explore how driver uncertainty affects the different models when forecasting water temperature.

34. Assess the contribution of driver uncertainty across models.

    a. Click 'Download plot' to download the plot of forecasts with driver uncertainty and copy-paste it into your report.

    Please copy-paste your Q-34a-plot.png image here.

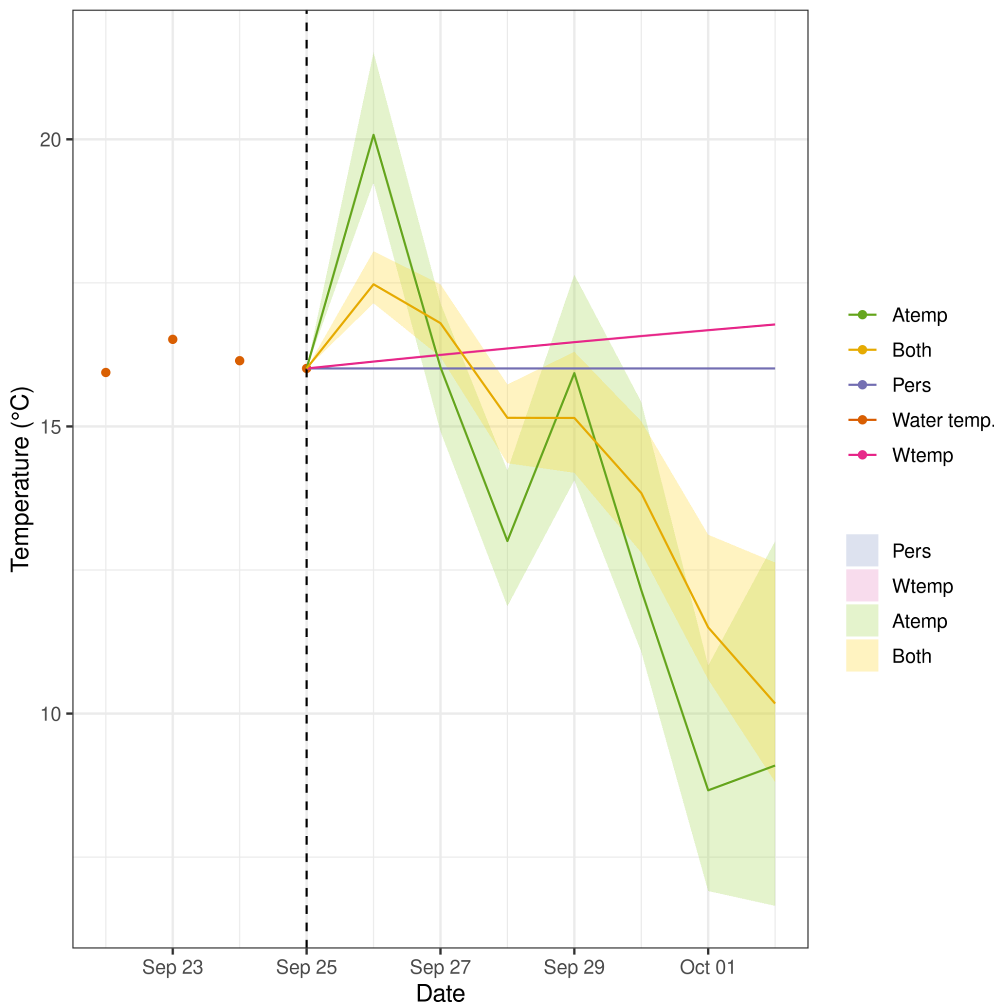


Figure 12. Water temperature forecasts with driver uncertainty for your selected NEON lake.

b. Why do only two of the models exhibit driver uncertainty in their water temperature forecasts?

Answer: Because only two of the models (Atemp, Both) use air temperature as an input, and thus require forecasted air temperature as a driver. The other two models (Pers, Wtemp) do not have drivers.

35. You have learned that driver uncertainty arises because we cannot perfectly know what the future values of our model inputs will be. Knowing this, describe one way that you think driver uncertainty in our water temperature forecasts could be reduced.

Answer: Improve forecasts of driver variables.

Activity C - Managing Uncertainty

Quantify and partition the uncertainty for forecasts made using different models and make management decisions using an ecological forecast.

9) Objective 9: Quantify Uncertainty

Generate forecasts of water temperature with total forecast uncertainty and quantify the uncertainty at each forecast horizon and its source.

---

36. Describe the sources of uncertainty for model 1.

    a. Record the name of the model you chose.

       Answer: Pers/Persistence (this will vary depending on student choice)

    b. Click ‘Download plot’ to download the plot of quantified uncertainty for model 1 and copy-paste it into your report.

       Please copy-paste your Q-36b-plot.png image here.

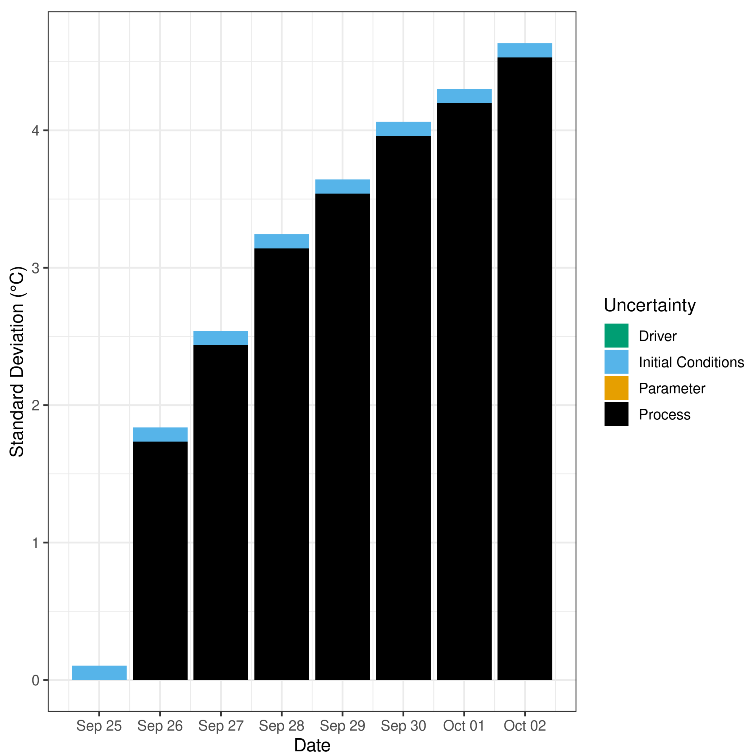


Figure 13. Total forecast uncertainty partitioned among process, parameter, initial conditions, and driver uncertainty for model 1.

c. Which source of uncertainty contributes the most to total forecast uncertainty for this model?

Answer: Process uncertainty (this will vary depending on student model choice)

37. Describe the sources of uncertainty for model 2.

a. Record the name of the model you chose.

Answer: Both (this will vary depending on student choice)

b. Click 'Download plot' to download the plot of quantified uncertainty for model 2 and copy-paste it into your report.

Please copy-paste your Q-37b-plot.png image here.

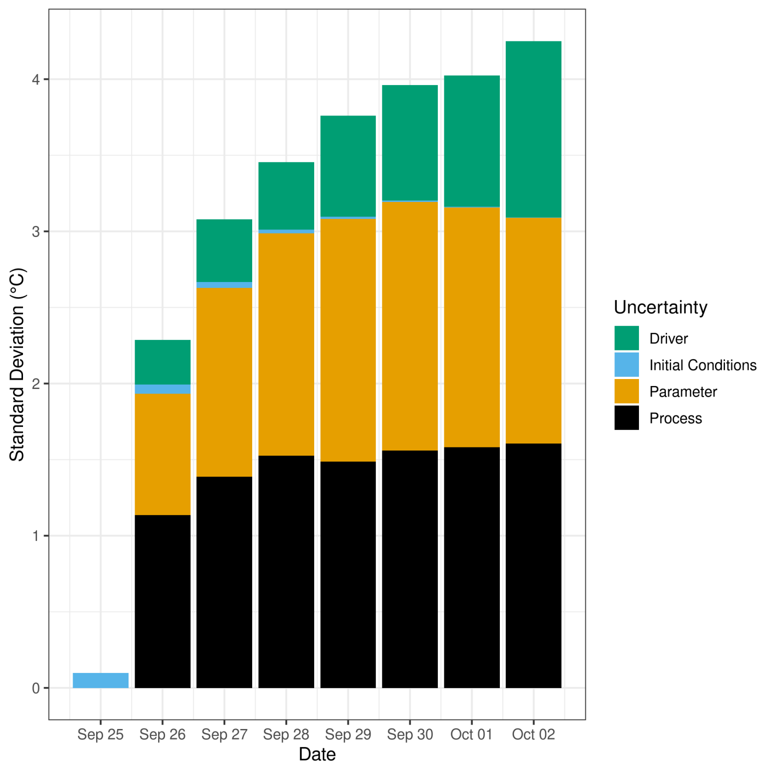


Figure 14. Total forecast uncertainty partitioned among process, parameter, initial conditions, and driver uncertainty for model 2.

c. Which source of uncertainty contributes the most to total forecast uncertainty for this model?

Answer: Both process and parameter uncertainty contribute similar amounts of uncertainty; based on this visualization, students may think that either of these sources contributes the most (this will vary depending on student model choice and lake).

38. Compare how the sources of uncertainty differed between your forecasts. Explain why each of the uncertainties (process, parameter, initial conditions & driver) is different between your models.

    Answer: Process: Process uncertainty contributed a lot more to total forecast uncertainty for the Pers than the Both model. This may be because the Pers model structure does not represent some of the real-world processes that affect water temperature which are represented in the Both model (e.g., the effect of air temperature).

    Parameter: Parameter uncertainty contributed more to total forecast uncertainty for the Both model than the Pers model, because the Pers model has no parameters.

    Initial conditions: Initial condition uncertainty contributes slightly more to the Pers model total uncertainty than the Both model; the influence of initial conditions may be relatively less in the Both model due to the existence of parameters and additional inputs (driver) to constrain model predictions.

    Driver: Driver uncertainty contributed more to total forecast uncertainty for the Both model than the Pers model, because the Pers model has no drivers.

10) Objective 10: Management Scenario

As a water resource manager you will use forecasts to make decisions about water releases from a reservoir. With your partner you will explore how forecast uncertainty affects your decisions.

---

39. Did the uncertainty visualization affect your decision? How?

    Answer: Answers will vary; it is likely that many students will have switched from choosing to release water from the Bottom in Decision #1 (because it appears there is no chance that the Bottom water temperature will exceed the threshold for salmon survival) to choosing to release water from the Surface in Decision #2 (because when all sources of uncertainty are included, it's clear that there is a substantial chance that Bottom water temperature will exceed the salmon survival threshold, even more so than for Surface water temperature).

40. Can you think of any potential risks when using a forecast without all sources of uncertainty included?

    Answer: Answers will vary; students should mention that failing to include all sources of uncertainty may provide an inaccurate or incomplete picture of the probability of future outcomes to decision-makers, thereby leading to potentially non-optimal decisions.

This module was developed by: Moore, T.N., M.E. Lofton, C.C. Carey, and R.Q. Thomas. 12 December 2023. Macrosystems EDDIE: Ecological Forecast Uncertainty. Macrosystems EDDIE Module 6, Version 2. http://module6.macrosystemseddie.org*. Module development was supported by NSF grants DEB-1926050 and DBI-1933016. This app was last updated on: 2023-12-12.*

Authorship contributions:

- *CCC and RQT conceived the idea of this module and acquired funding for this project.*
- *TNM, CCC, and RQT developed the learning objectives and website text.*
- *MEL developed RMarkdown activities with feedback from CCC and RQT.*
- *TNM and MEL developed the module activities and code for the module with feedback from CCC and RQT.*
- *TNM, MEL, and CCC developed student assessment questions and led module testing and collection of student assessment data.*
- *MEL and TNM developed the student handout, instructor PowerPoint, and instructor manual with feedback from CCC and RQT.*
- *MEL, TNM, CCC, and RQT worked with instructors of the module and integrated feedback into improving the module.*