Macrosystems EDDIE:

Using Data to Improve Ecological Forecasts

Instructor's Manual

Module Description

To be useful for management, ecological forecasts need to be both accurate enough for managers to be able to rely on them for decision-making and include a representation of forecast uncertainty, so managers can properly interpret the probability of future events. To improve forecast accuracy, we can update forecasts with observational data once they become available, a process known as data assimilation. Recent improvements in environmental sensor technology and an increase in the number of sensors deployed in ecosystems have resulted in an increase in the availability of data for assimilation to help develop and improve forecasts for natural resource management. In this module, students will develop an ecosystem model of primary productivity, use the model to generate forecasts, and then explore how assimilating different types of data at different temporal frequencies (e.g., daily, weekly) affects forecast accuracy. Finally, students will assimilate different types of data into forecasts and examine how data assimilation affects water resource management decisions.

Pedagogical Connections

Phase	Functions	Examples from this module
Engagement	Introduce topic, gauge students' preconceptions, call up students' schemata	Short introductory lecture introducing ecological forecasting; explaining the method of data assimilation to improve forecasts and answering discussion questions
Exploration	Engage students in inquiry, scientific discourse, evidence-based reasoning	Developing hypotheses about how best to use data to improve forecast accuracy; Testing these hypotheses by generating an ecological forecast and assessing forecast accuracy
Explanation	Engage students in scientific discourse, evidence-based reasoning	Comparing forecast accuracy between forecasts generated using different data assimilation methods; Analyzing how the forecast accuracy varies based on the sensor observation uncertainty and frequency of data assimilation
Expansion	Broaden students' schemata to account for more observations	Generating and comparing forecasts at different sites in different climatic regions; Applying their findings regarding how data assimilation affects forecast accuracy to a new lake in a management case study
Evaluation	Evaluate students' understanding, using formative and summative assessments	In-class discussion of data assimilation and how ecological forecasting can be used to improve ecosystem understanding and natural resource management decision making

This module was initially developed by: Lofton, M.E., Moore, T.N., Thomas, R.Q., and C.C. Carey. 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 was supported by NSF grants DEB-1926050 and DBI-1933016. This document last modified: 12 March 2024412 March 2024 by M.E. Lofton.

Learning Objectives

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

- Define data assimilation
- Generate an ecological forecast for chlorophyll-a in a lake
- Describe how to assess ecological forecast accuracy
- Describe how data assimilation affects forecast accuracy and uncertainty
- 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

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 6 "Understanding Uncertainty in 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, Objective 7 (15 min.), and Activity C (15 min.); wrap-up (20 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 – 20 mins

Quick overview of the activities in this module:

- 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.

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/module7/**
        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/module7**
        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, data assimilation, and the forecast model students 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 chlorophyll-a (Activity A Objective 3), and generate forecasts with uncertainty using their fitted 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 with and without data assimilation (Activity B Objective 5) as well as forecasts that assimilate data with different amounts of observation uncertainty (Activity B Objective 6) and at different frequencies (e.g., daily vs. weekly; Activity B Objective 7). Students should compare their forecasts with their partners and with students working on different lakes and work together to answer questions embedded throughout this activity about how the forecasts are affected by assimilating data with different levels of observation uncertainty and at different frequencies.
7. In Activity C, student pairs complete a management scenario individually and discuss with their partner how the forecasts with different methods of data assimilation provided in the scenario affected their management decisions (Activity C Objective 8).

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 perspective article about the importance of data assimilation for ecological forecasting, and answer questions about how data assimilation can improve ecological forecasts.

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, Objective 7 (15 min.), and Activity C (15 min.); wrap-up (20 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 – 20 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 set of possible alternate future conditions predicted by a model and the chance of observing each outcome
    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. 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, generating a forecast, assessing the forecast, and most importantly, updating our forecast models with new data as they become available.
7. 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 regression: “y = mx + b”; for example, we might use the amount of available phosphorus in a lake as the independent variable (x) to predict chlorophyll-a (y) in a lake.

8. Because they are a simplified representation of the real world, models for ecological forecasting will necessarily be imperfect representations of complex, real-world ecosystems as they represent ecosystem processes using mathematical equations, thus creating a simplified “digital lake” in your computer.
9. The key question then becomes: How do we make sure the digital lake resembles the actual ecosystem as closely as possible?
   a. For example, if this lake experiences an algal bloom event, how do we make sure the lake model is able to capture or forecast that event?
   b. (click for animation) One way is by using data collected in the real world to calibrate and update our models so they mirror real-world conditions as closely as possible
10. It is important to note that observational data are also imperfect
    a. While it can be tempting to view our observational data as “the truth” (particularly when we have worked really hard to collect it!), we know that observational data are also imperfect. This can be due to:
       i. Human error
       ii. Sensor malfunction
       iii. And limits to measurement accuracy, among other factors
    b. When producing ecological forecasts, we must account for uncertainty in both models and data
    c. Ultimately, mode predictions and data BOTH have value in producing accurate forecasts, as they both can provide helpful information about the ecological system that is being forecasted
11. This leads to our focal question for this module: how can we use data to improve ecological forecasts?
12. Let’s use a familiar example to start to answer this question. We know weather forecasts are updated regularly – how is this done?
    a. It is done through the process of data assimilation, which is 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.
    b. Here, you can see updated forecasts of Hurricane Ian, which occurred in 2022, as new data were used to update the forecast model over time. The Track Forecast Cone shows the most likely path of the center of a tropical cyclone. The center of a tropical cyclone historically moves out of the cone about a third of the time. The cone is formed by connecting a set of circles at each forecast time. Each circle is drawn so that it encloses 67% of the official track forecast errors over the last 5 years (hence the likelihood of the center to move out of the cone about a third of the time). Forecasts are updated over time using the most recent up-to-date position of the storm.

c. Image source: https://www.weather.gov/news/101722-jamie-rhome-cone.
d. Students may be aware that this forecast cone visualization has come under criticism for being difficult to interpret. The article associated with the image has further information on this topic if they are interested.
e. We'll now walk through a high-level overview of how data assimilation is done.

13. At a high level, the process of data assimilation involves three steps.
    a. First, you generate a forecast with uncertainty.
    b. Next, as the future arrives, you update the forecast model using new data when they become available
    c. Finally, you use the updated forecast model to generate a new forecast that reflects your updated data.
    d. We will now walk through each of these steps in more detail.
14. 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 is an example of a distribution that represents the uncertainty around an observation of chlorophyll-a in a lake.
15. In forecasting, we use the most recent observational data to inform the initial condition, or starting condition, of the forecast. When we account for uncertainty in our observations, the initial condition will be a distribution.
16. Then, instead of running just a single forecast, we take many draws from our initial conditions distribution (blue distribution) and run a model many times using slightly different conditions.
    a. The complete set of forecasts is referred to as the ensemble
    b. Individual forecasts within the ensemble are ensemble members (gray lines)
    c. Usually, each member of the ensemble is equally likely to occur
    d. This allows us to generate a forecast with uncertainty (white distribution)We can only update our models with data when data are available
17. Once we have generated a forecast, we wait for the future to arrive and then update our initial condition using new data as they become available
    a. This updating can be done using a technique called an ensemble Kalman filter, which accounts for uncertainty in both model predictions and new observations
    b. You will learn about the ensemble Kalman filter in more detail in Activity B
18. Finally, the new initial condition is used to generate a new forecast, once again accounting for uncertainty
19. The effect of data assimilation on our models and forecasts depends on our relative confidence in our observations and our model predictions
    a. If we observe a data point outside our predictive interval and we have **low observation uncertainty**, our model initial conditions and forecast will be adjusted quite a lot
    b. If we observe a data point outside our predictive interval and we have **high observation uncertainty**, our model initial conditions and forecast will not be adjusted as much, because we have a relatively greater degree of confidence in our model compared to observations

20. We can only update our models with data when data are available
    a. Data availability can also be affected both by **data frequency** and **data latency**
        i. **Data frequency:** how often data are collected (i.e., temporal frequency)
        ii. **Data latency:** the time between when an observation is collected and when it is ready to be used in a forecast model
    b. Data frequency and latency depend on the method of data collection
        i. For example, an in-situ sensor deployed on a buoy in a lake might have high data collection frequency (e.g., every 10 minutes) and low latency (if the data are wirelessly streaming to a computer)
        ii. In contrast, manual water samples are likely collected at a lower frequency because it requires someone traveling to the site to collect a sample, and higher latency because the samples must be analyzed in the lab and the data digitized before it can be used in a forecast model
21. 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.
    a. **Phytoplankton** are microscopic primary producers that live in the water. They are the dominant primary producers in many lakes and reservoirs.
    b. Ask the class: What are the drivers of phytoplankton biomass in a lake? Click to reveal the list of possible drivers; note this is not a comprehensive list and there may be others that the class correctly identifies!
22. Why forecast **phytoplankton biomass in lakes?**
    a. Excess phytoplankton biomass can lead to harmful algal blooms, which compromise water quality through:
        i. Production of toxins,
        ii. Production of taste and odor compounds,
        iii. Creation of anoxic zones, leading to fish kills
    b. Pre-emptive notice of a possible harmful algal bloom via an ecological forecast could allow managers to prevent or mitigate these adverse water quality impacts.
23. **Chlorophyll-a data** are a useful proxy for phytoplankton biomass in lakes.
    a. Chlorophyll-a is a molecule that phytoplankton use for photosynthesis.
    b. Measuring chlorophyll-a concentrations in water can give a bulk estimate of phytoplankton concentrations.
    c. Chlorophyll-a data can be collected at a high frequency by automated sensors; this means that chlorophyll-a data can be high frequency and low latency, which is often helpful for generating iterative, up-to-date forecasts
24. The type of ecological model we will use today is an autoregressive model
    a. We will fit an **autoregressive model** to chlorophyll-a data from a lake site of your choice.
    b. An **autoregressive model** uses past values of a variable to predict future values.

25. To calibrate our model, we will use data from the U.S. National Science Foundation's National Ecological Observatory Network, or NEON.
    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. We will be using data from five different lake sites within NEON.
26. Learning objectives!
    a. Talk through these with the students one by one: use the embedded animations to sequentially show each of the five bullet points.
27. Introduce Activity A, which has four objectives (have students work in pairs).
    a. **Objective 1:** Select and learn about a NEON lake site
    b. **Objective 2:** Explore chlorophyll-a data collected at your lake site
    c. **Objective 3:** Fit an autoregressive forecast model to chlorophyll-a data
    d. **Objective 4:** Generate a one-day-ahead forecast of chlorophyll-a with uncertainty
28. Activity B – continue working in pairs and answer the questions individually. Generate multiple forecasts while assimilating data at different temporal frequencies and with different amounts of observation uncertainty.
    a. Objective 5. Compare one-day-ahead forecasts generated with and without data assimilation.
    b. Objective 6. Compare one-day-ahead forecasts generated with data assimilation, using data with low vs. high observation uncertainty.
    c. Objective 7. Compare a series of one-day-ahead forecasts with no data assimilation, weekly data assimilation, and daily data assimilation.
    d. At the end of Activity B, regroup as a class and each of the groups present the results from their forecast and how their forecast got better or worse with different methods of data assimilation
29. Activity C – continue working in pairs and answer the questions individually.
    a. Objective 8. Make management decisions using forecasts generated with different frequencies of data assimilation.
    b. You may have the opportunity to present your choice and the resulting forecasts to the class.
30. Shiny App:
    a. The module can be accessed at: https://macrosystemseddie.shinyapps.io/module7/
    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
31. 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 30-32 in PowerPoint.
32. Saving & resuming progress in the Shiny app.
33. We recommend that you save your progress often!

34. 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.
35. Thank you for participating!

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).** We have observed a tendency to forget about these questions as students skip straight to Activity A!

Activity A: Build A Model With Uncertainty

Activity A challenges the students to build a model to simulate chlorophyll-a for their chosen NEON site and generate a forecast.

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.

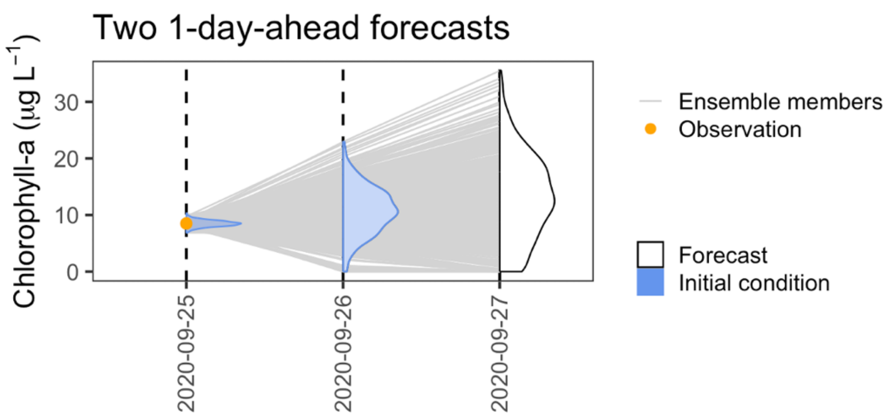
- Starting in Objective 3, many of the objectives have embedded carousels of scrolling PowerPoint slides which provide important information for the objective (as well as for some of the questions embedded throughout the app!). **Emphasize to students that scrolling through these slides can help them answer questions and understand the module.**
- Starting in Objective 3, students will be asked to interpret the goodness-of-fit of their models and forecasts using 1) visual assessment; 2) bias; and 3) RMSE. Students may lack confidence in their interpretation if they are not familiar with these metrics or with chlorophyll-a in lakes. Encourage students to compare their model fits, forecasts, and calculated bias and RMSE values with others in the class working on different lakes to gain a better understanding of how their models and forecasts perform.
- 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: Assimilate Data

At this point, students will generate multiple forecasts while assimilating data at different temporal frequencies and with different amounts of observation uncertainty. This will allow them to assess the effect of different methods of data assimilation on forecast accuracy.

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.
- In Objective 5, some students may notice that their forecast distributions appear truncated, like this:



This is because our very simple forecast model may sometimes predict negative chlorophyll-a, which is not physically possible. If such a prediction is generated, it is set to 0. Alternative approaches could include choosing a different forecast model or log-transforming chlorophyll-a data before fitting a model. However, for our learning purposes today, we will continue with this simple approach.

- Remind students to download plots and answer questions in their Word documents as they go. Also, encourage students to compare their forecasts plots with other students working on different lakes. One effective way of doing this is to ask students to briefly present one or more figures from Activity B to the class at the end of the lesson, and discuss differences in the effect of data assimilation on forecasts for different lakes.
- 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: Management Scenario

This is a relatively short activity with one objective. Students explore the effect of assimilating data from sensors with different levels of observation error on forecast accuracy and make management decisions using an ecological forecast. This will then provide discussion as a group for how data assimilation can affect decision making based on forecast output.

Tips for Activity C:

- **The management scenario in this activity has an unexpected outcome!** Students will likely expect that a forecasting system with a higher frequency of data assimilation will lead to a better management outcome. The scenario has been designed such that this is NOT the case. Instead, while the forecasting system with a higher frequency of data assimilation has a lower bias and RMSE, indicating superior forecast skill, the forecasting system with a lower frequency of data assimilation leads to a better forecast and management outcome in the end. We have intentionally designed the scenario in this way to encourage students to think critically about the value and limitations of data assimilation in forecasting. **We recommend that instructors emphasize to students that:**

- **Ecological forecasts must be assessed by how well they inform and improve management, in addition to skill metrics like bias and RMSE**
- **In this case, the underlying issue is that neither forecasting system is performing particularly well.** The low-frequency data assimilation misses the increase and decrease in chlorophyll-a over the course of the week, while the high-frequency data assimilation mis-predicts the chlorophyll-a concentration on Saturday, Oct. 11. This indicates that in order to improve the forecasting system, we may need to consider using a different forecasting model, rather than increasing the frequency of data assimilation.

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, Tadhg N., R. Quinn Thomas, Whitney M. Woelmer, and Cayelan C. Carey. 2022. Integrating ecological Fforecasting into undergraduate ecology curricula with an R Shiny application-based teaching module. *Forecasting* 4 (3): 604–33. DOI: 10.3390/forecast4030033
- Woelmer WM, Moore TN, Lofton ME, Thomas RQ, Carey CC. 2023a. Embedding communication concepts in forecasting training increases students' understanding of ecological uncertainty. Ecosphere 14: e4628.

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://module7.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.

By design, this module includes many formative questions which help ensure students follow along and engage with the module material, but do not provide a summative assessment of student understanding and may become repetitive from a grading perspective. Below, we give recommendations for which questions throughout the module provide a summative assessment of student understanding, and may be most valuable for instructors to focus on when grading.

Summative assessment questions:

Activity A: Q6, Q11, Q13, Q14, Q18, Q19, Q22, Q23c,d

Activity B: Q24, Q29c, Q31c, Q39

Activity C: Q40, Q43c, d, Q44, Q45

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

Module 7: Using Data to Improve Ecological Forecasts - Student Handout

Name: Key

Student ID: Key

Completed on: 2024-02-06

Copy-paste your save progress link from the Shiny app here for ease of reference:

Macrosystems EDDIE Module 7: Using Data to Improve Ecological Forecasts

Learning Objectives:

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

- Define data assimilation (Activity A)
- Generate an ecological forecast for primary productivity (Activity A)
- Describe how to assess ecological forecast accuracy (Activity A)
- Describe how data assimilation affects forecast accuracy and uncertainty (Activity B)
- 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 B, C)

Why macrosystems ecology and ecological forecasting?

Macrosystems ecology is the study of ecological dynamics at multiple interacting spatial and temporal scales (e.g., Heffernan et al. 2014). For example, *global* climate change can interact with *local* land-use activities to control how an ecosystem changes over the next decades. Macrosystems ecology recently emerged as a new sub-discipline of ecology to study ecosystems and ecological communities around the globe that are changing at an unprecedented rate because of human activities (IPCC 2013). The responses of ecosystems and communities are complex, non-linear, and driven by feedbacks across local, regional, and global scales (Heffernan et al. 2014). These characteristics necessitate novel approaches for making predictions about how systems may change to improve both our understanding of ecological phenomena as well as inform resource management.

Forecasting is a tool that can be used for understanding and predicting macrosystems dynamics. To anticipate and prepare for increased variability in populations, communities, and ecosystems, there is a pressing need to know the future state of ecological systems across space and time (Dietze et al. 2018). 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. Ecological forecasts are a powerful test of the scientific method because ecologists make a hypothesis of how an ecological system works; embed their hypothesis in a model; use the model to make a forecast of future conditions; and then when observations become available, assess the accuracy of their forecast, which indicates if their hypothesis is supported or needs to be updated. Forecasts that are effectively communicated to the public and managers will be most useful for aiding decision-making. Consequently, macrosystems ecologists are increasingly using ecological forecasts to predict how ecosystems are changing over space and time (Dietze and Lynch 2019).

In this module, students will generate an ecological forecast for a NEON site and explore how to use ecological data to improve forecast accuracy. This module will introduce students to the concept of data assimilation within an ecological forecast; how data assimilation can be used to improve forecast accuracy; how the level of uncertainty and temporal frequency of observations affects forecast output; and how data assimilation can affect decision-making using ecological forecasts.

Module overview:

- Introductory presentation to 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: Update forecasts with data that have different levels of observation uncertainty for making management decisions

Today's focal question: *How can we use data to improve ecological forecasts?*

To be useful for management, ecological forecasts need to be both accurate enough for managers to be able to rely on them for decision-making and include a representation of forecast uncertainty, so managers can properly interpret the probability of future events. To improve forecast accuracy, we can update forecasts with observational data once they become available, a process known as data assimilation. Recent improvements in environmental sensor technology and an increase in the number of sensors deployed in ecosystems have increased the availability of data for assimilation to develop and improve forecasts for natural resource management.

In this module, you will explore how assimilating data with different amounts of observation uncertainty and at different temporal frequencies affects forecasts of lake water quality at an ecological site of your choice.

R Shiny App:

The lesson content is hosted on an R Shiny App at https://macrosystemseddie.shinyapps.io/module7/

This can be accessed via any internet browser and allows you to navigate through the lesson via this app. You will fill in the questions below on this handout as you complete the lesson activities.

Optional pre-class readings and video:

Webpages:

- NOAA Ecological Forecasts
- Ecological Forecasting Initiative

Articles:

- Silver, N. (2012) Chapter 6: How to drown in three feet of water.Pages 176-203 inThe Signal and the Noise: Why so many Predictions Fail – but some Don't. Penguin Books.
- Dietze, M. and 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., et al. 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

Videos:

- NEON's Ecological Forecast: The Science of Predicting Ecosystems
- Fundamentals of Ecological Forecasting Series: Why Forecast?
- Fundamentals of Ecological Forecasting Series: Forecast Analysis Cycle
- Fundamentals of Ecological Forecasting Series: Ensemble Kalman Filter

Pre-class activity: Explore how data assimilation can affect forecast accuracy

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

Niu S, Luo Y, Dietze MC, Keenan TF, Shi Z, Li J, Iii FSC. 2014. The role of data assimilation in predictive ecology. Ecosphere 5: 1–16. https://doi.org/10.1890/ES13-00273.1

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

A. Define 'data assimilation'.

   Answer: *Data assimilation is using data to inform the starting conditions and parameters of a model, with the aim of ensuring that model output resembles reality as closely as possible.*

B. Summarize why, in the authors' opinion, data assimilation is needed for ecological forecasting (which the authors refer to as "predictive ecology").

   Answer: *Data assimilation helps to constrain the starting conditions and parameters of a model, which may improve forecast accuracy. In addition, data assimilation can help quantify the uncertainty associated with a forecast, and it is important to have a good estimate of the uncertainty associated with a forecast when people use the forecast to make management decisions.*

C. The authors review four examples of how data assimilation has been applied in ecology (infectious disease, fisheries, wildfires, and the terrestrial carbon cycle). Choose ONE of these examples and explain 1) how data assimilation has been applied; and 2) how it has advanced research in this area.

   Answer: *There are multiple possible answers; see the section "Application of DA in Predictive Ecology" in the assigned paper. One example answer is that data assimilation has been applied in fisheries to predictive models for fish stocks and species distributions. In both cases, various methods for DA, including the Kalman filter, improved model predictions.*

D. The authors discuss several areas which present challenges and opportunities for application of data assimilation in ecology (models becoming more complex, data becoming more available, ecological issues becoming more complex, and real-time predictions). Choose ONE of these areas and explain, in your own words, the challenges and opportunities for applying data assimilation.

   Answer: *There are multiple possible answers; see the section "Future Challenges and Opportunities of DA" in the assigned paper. One example answer is that the greater availability of data in ecology provides more opportunity to assimilate data into models. However, a challenge is that it remains unknown which data streams will ultimately be most valuable for improving predictive skill.*

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 ‘data assimilation’ in the context of ecological forecasting?

   Answer: *In ecological forecasting, data assimilation is the process of updating ecological forecasting models with new environmental data as they become available.*

2. How do you think the process of integrating the most recently observed data into models can improve forecasts?

   Answer: *Updating the initial, or starting, conditions of a model with data will help the model better resemble current reality, thus potentially improving the accuracy of the forecast.*

Activity A - Build A Model and Generate A Forecast

Explore chlorophyll-a data at a NEON site of your choice, then fit a model to the data and generate a forecast with uncertainty.

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: *Lake Barco*

   b. Four letter site identifier: *BARC*

   c. Latitude: *29.675982*

   d. Longitude: *-82.008414*

   e. Lake area (km2): *0.13*

   f. Elevation (m): *27*

Objective 2: Explore water temperature

Explore the chlorophyll-a data measured at the selected site. This is data that has been downloaded from the NEON Data Portal. While we are using chlorophyll-a data in this module, there are a wide range of variables collected at each NEON site.

---

4. Why might a forecast of lake chlorophyll-a concentration days to weeks into the future be a useful tool for water managers?

   Answer: *If managers had advance notice of water quality changes in a lake, they might be able to take pre-emptive management actions, such as planning to withdraw water from a different depth in a reservoir, closing a recreational beach, or switching to another water supply.*

5. Describe chlorophyll-a data at your lake.

   a. Download the timeseries plot of chlorophyll-a data and copy-paste it into your report.

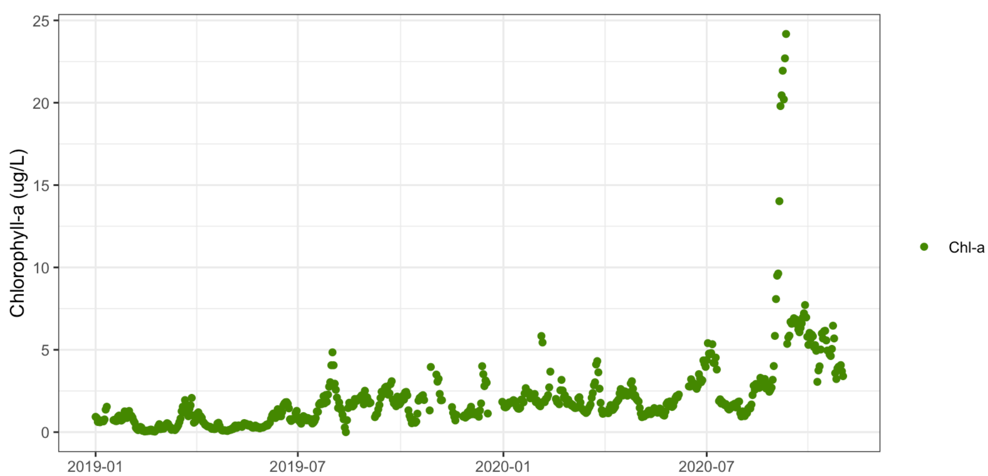


   Figure 1. Time series of chlorophyll-a data for your selected NEON lake.

   b. Describe how chlorophyll-a changes over time in your lake. Do you notice any patterns or trends?

   Answer: *Answers will vary. Most chlorophyll-a time series exhibit high variability.*

Objective 3: Fit model

We will explore autocorrelation in the chlorophyll-a data at your selected NEON site and then fit an autoregressive model for forecasting.

---

6. Explain, in your own words, how autocorrelation in a variable can help forecasters make predictions of the future.

Answer: *Autocorrelation is correspondence between previous and current/future values of a variable. If future values have a strong correspondence with previous values, knowing previous values of a variable can provide important information for forecasting future values.*

7. Describe what you observe on the timeseries plot of lagged chlorophyll-a. How do the two lines plotted on the timeseries (chlorophyll and 1 day lag of chlorophyll) relate to each other?

    a. Download the plot of lagged chlorophyll-a and copy-paste it into your report.

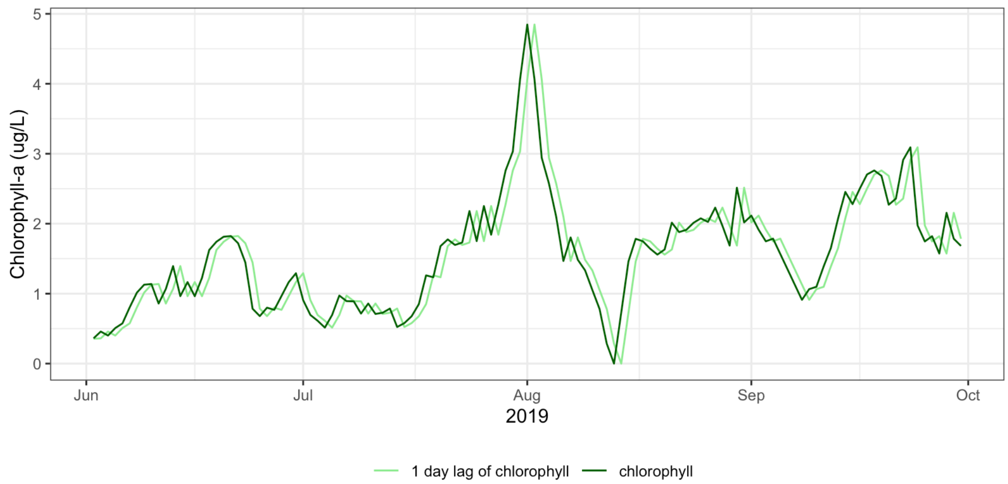


Figure 2. Time series of lagged chlorophyll-a data for your selected NEON lake.

    b. How do the two lines plotted on the timeseries (chlorophyll and 1 day lag of chlorophyll) relate to each other?

    Answer: *The 1-day lag is the same as the chlorophyll-a timeseries, just offset by one day.*

8. Describe what you observe on the scatterplot figure to the right. Do you think the chlorophyll-a data at your chosen lake site exhibit autocorrelation? Why or why not?

    a. Download the scatterplot and copy-paste it into your report.

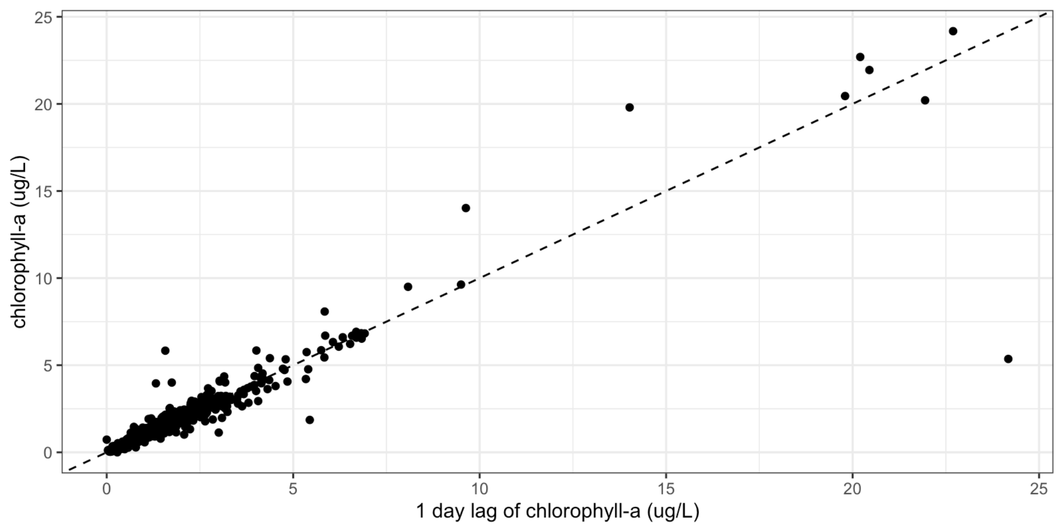


Figure 3. Scatterplot of chlorophyll-a data vs. lagged chlorophyll-a data for your selected NEON lake.

b. Do you think the chlorophyll-a data at your chosen lake site exhibit autocorrelation? Why or why not?

Answer: *Yes, because many of the points cluster along the 1:1 line in the plot, indicating a strong relationship between chlorophyll-a data and lagged chlorophyll-a data.*

9. Record the autocorrelation value you calculated. Does this value indicate low or high autocorrelation between chlorophyll-a and a 1-day lag of chlorophyll-a?

Answer: *0.932; this indicates high autocorrelation because the value is close to 1*

10. Describe autocorrelation across many lags at your lake site.

a. Download the autocorrelation plot and copy-paste it into your report.

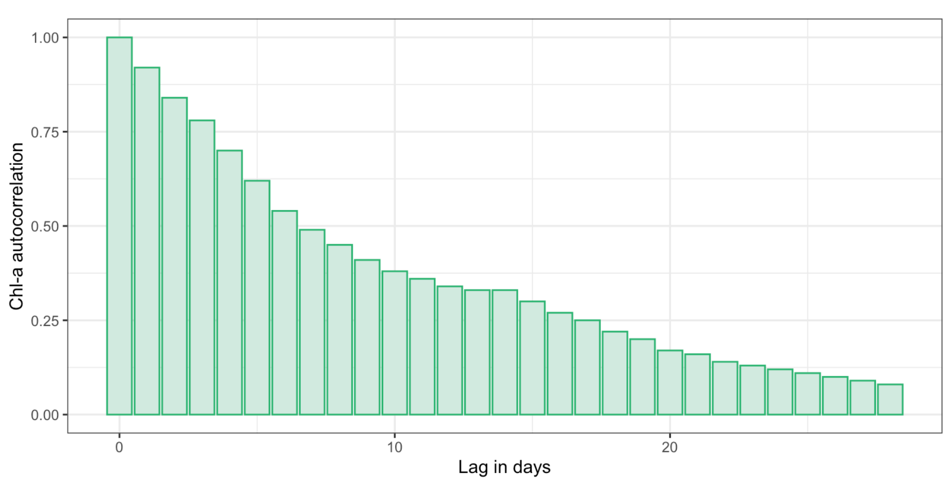


Figure 4. Autocorrelation of lagged chlorophyll-a at your lake site.

b. Describe how autocorrelation changes as the lag in days increases. Why do you think this pattern occurs?

Answer: *Autocorrelation decreases as the lag in days increases. This is likely because observations which have occurred farther in the past are less related to current values than values that have been observed more recently.*

11. Imagine you are asked to develop a forecasting model that uses lagged values of chlorophyll-a to predict future chlorophyll-a. Examining the autocorrelation plot above, how many lags of chlorophyll-a would you include in your forecasting model? Provide your answer in days (e.g., I would include up to a 3-day lag) and explain your reasoning.

    Answer: *Answers will vary.*

12. Examine the PACF plot.

    a. Download the PACF plot and copy-paste it into your report.

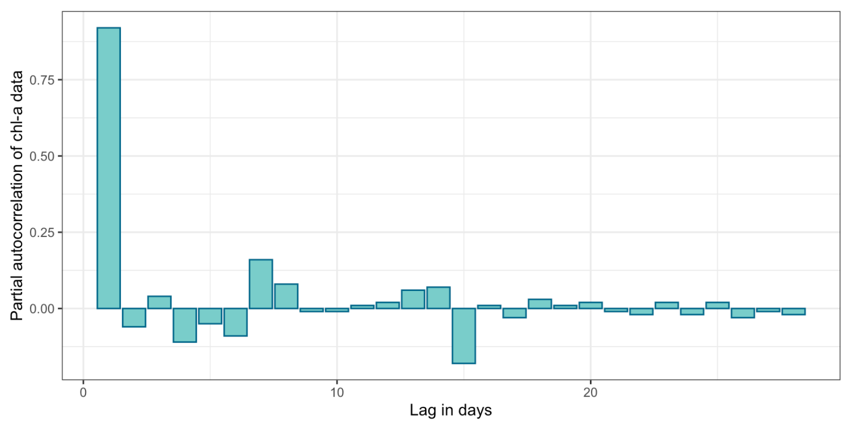


Figure 5. Partial autocorrelation of chlorophyll-a at your lake site.

b. Examine the PACF plot. Which lag contributes the most to autocorrelation in the chlorophyll-a data? Explain how you know.

Answer: *The one-day lag, because it has by far the highest PACF.*

13. Once again, imagine you are asked to develop a forecasting model that uses lagged values of chlorophyll-a to predict future chlorophyll-a. Examining the PACF plot above, how many lags of chlorophyll-a would you include in your forecasting model? Provide your answer in days (e.g., I would include up to a 3-day lag) and explain your reasoning.

    Answer: *Answers will vary.*

14. Did the number of lags you chose to include in your forecasting model change from Q11 to Q13? Why or why not?

    Answer: *Answers will vary.*

15. Use the plot above to assess the model fit to data.

    a. Download the plot of model predictions and observations and copy-paste it into your report.

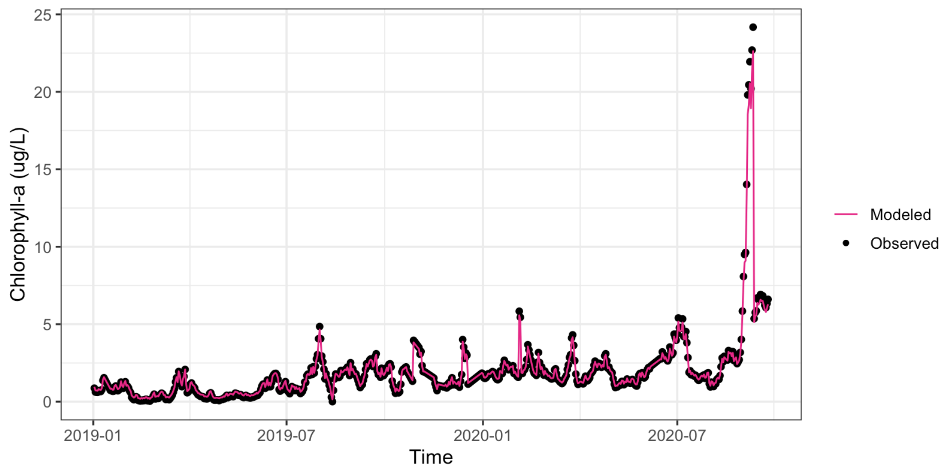


Figure 6. Autoregressive model predictions and chlorophyll-a observations at your lake site.

    b. How well do the predictions match the observations?

Answer: *The predictions match the observations very closely.*

Note: Answers to questions 16 & 17 hwill likely vary as students do not have much basis for comparison yet for the values of bias and RMSE. You might encourage the class to compare values among lakes to help them answer questions 16 & 17.

16. Record your model bias. Then, use the calculated bias to assess the model fit to data. How good is the model fit? Explain your reasoning.

Answer: *0.0103; this indicates a very good model fit because the smaller the absolute value of the bias, the better the fit*

17. Record your model RMSE. Then, use the calculated RMSE to assess the model fit to data. How good is the model fit? Explain your reasoning.

Answer: *0.17; this indicates a very good model fit because the smaller the RMSE, the better the fit*

Objective 4: Generate forecast

We will use the autoregressive model fit in the previous objective to generate a one-day-ahead forecast of chlorophyll-a with uncertainty for your chosen lake site.

---

18. Explain, in your own words, what forecast uncertainty is and why it is important to account for uncertainty in forecasts.

    Answer: *Forecast uncertainty is the range of possible future outcomes predicted by a model. Knowing the uncertainty in a forecast allows forecast users to make informed decisions based on the likelihood of various outcomes.*

19. Record the value of your process uncertainty standard deviation in ug/L, and explain in your own words how this value was calculated.

    Answer: *0.14; this value is the standard deviation of the model residuals, which is the difference between each observation and the model prediction for that observation*

20. What data from your chosen lake are needed to provide the initial condition for your forecast model?

    Answer: *chlorophyll-a data*

21. Examine the plot of high-frequency chlorophyll-a data.

    a. Download the plot of high-frequency chlorophyll-a data and copy-paste it into your report.

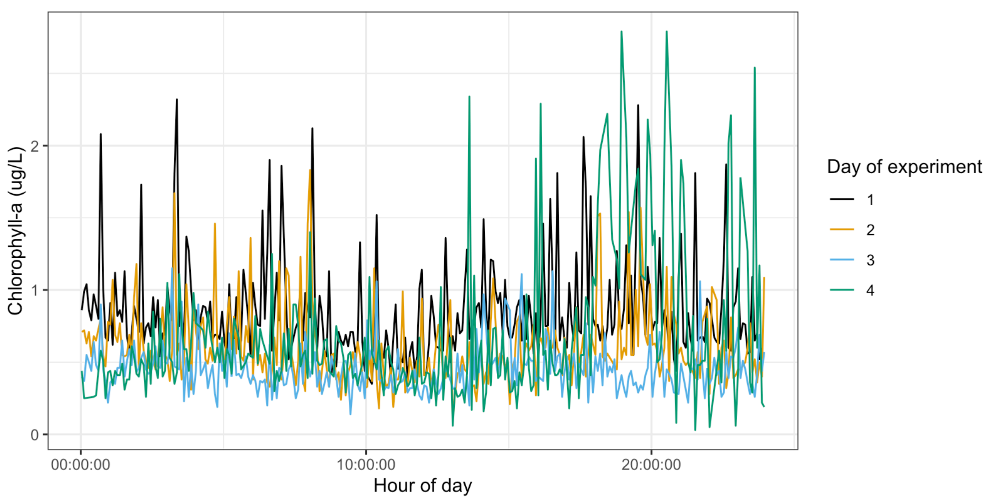


Figure 7. High-frequency chlorophyll-a data at your lake site.

    b. How variable is chlorophyll-a over the course of a day?

    Answer: *Answers here will vary according to which lake students have chosen; encourage students to compare figures across lakes to gain a better sense of how variable chlorophyll-a at their chosen lake site is.*

22. Record the value of your initial condition uncertainty standard deviation in ug/L, and explain in your own words how this value was calculated.

Answer: *0.32; this is the mean daily standard deviation of the high-frequency chlorophyll-a measurements at the chosen lake site*

23. Examine the chlorophyll-a forecast plot.

    a. Download the chlorophyll-a forecast plot and copy-paste it into your report.

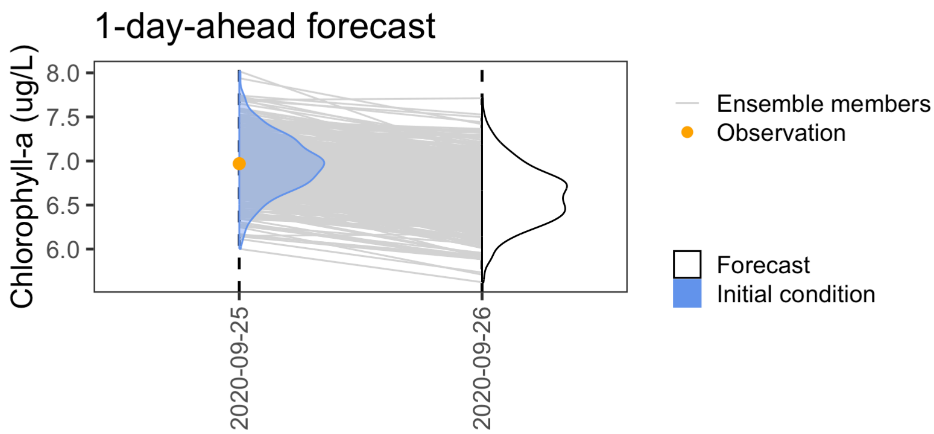


Figure 8. A one-day-ahead forecast of chlorophyll-a at your lake site.

    b. What is the forecasted chlorophyll-a concentration for 2020-09-26?

    Answer: *~6.7 ug/L is the most likely outcome, with a possible range of ~5.5-7.7 ug/L; students may have difficulty with this question if they are not used to interpreting figures with uncertainty; encourage them to describe the uncertainty in the forecast as part of their answer*

    c. What is the relationship between the observed chlorophyll-a for 2020-09-25, and the initial condition distribution (shown in blue)?

    Answer: *the observation is used as the mean to generate the initial condition distribution*

    d. Each one of the gray lines in the figure above represents an ensemble member. Explain what this means in your own words, with specific reference to the autoregressive model we are using for forecasting in this exercise.

    Answer: *each ensemble member is one prediction generated using our autoregressive forecast model; each ensemble member is generated using a slightly different initial condition and value of process uncertainty*

Activity B - Explore Data Assimilation

Now we will explore the effect of data assimilation on forecasts. First, we will compare forecasts generated with and without data assimilation. Then, we will investigate how the frequency of data assimilation (e.g., daily vs. weekly) and the amount of uncertainty associated with observations affects data assimilation and forecasts.

Objective 5: Assimilate data

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

---

24. Briefly describe in your own words how an ensemble Kalman filter can be used to assimilate data into an ecological forecast.

    Answer: *Ensemble Kalman filters are specifically used for ensemble forecasts. When new data are available, the ensemble Kalman filter statistically combines the new observation and the previous model prediction to generate a new initial condition which accounts for uncertainty in both model predictions and observations, which is the process of data assimilation. This new initial condition is then used to generate an updated forecast.*

25. Examine the forecast plot with an updated initial condition.

    a. Download the plot of a forecast with an updated initial condition and copy-paste it into your report.

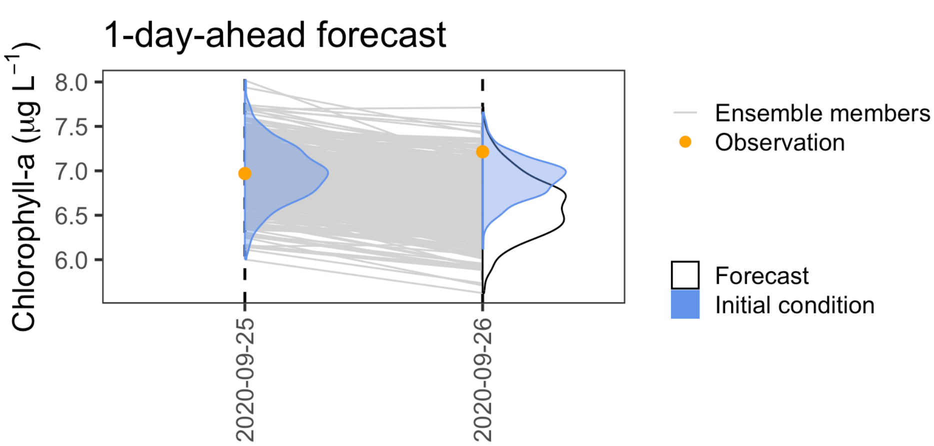


Figure 9. One-day-ahead forecast with updated initial condition using newly observed chlorophyll-a data.

    b. Compare the difference between the forecast distribution (white) for 2020-09-26 and the updated initial condition (blue distribution) for 2020-09-26. How are these two distributions different?

Answer: *Answers will vary depending on which lake students have chosen; in general, while the updated initial condition distribution will likely be more closely centered on the observation than the forecast, we do not expect the updated initial condition distribution to be exactly centered on the observation. Because the ensemble Kalman filter accounts for uncertainty in both observations and model predictions, it does not consider the observation to be "truth".*

26. Examine the forecast plot with an updated initial condition when data are missing.

    a. Download the plot of the updated initial condition when data are missing and copy-paste it into your report.

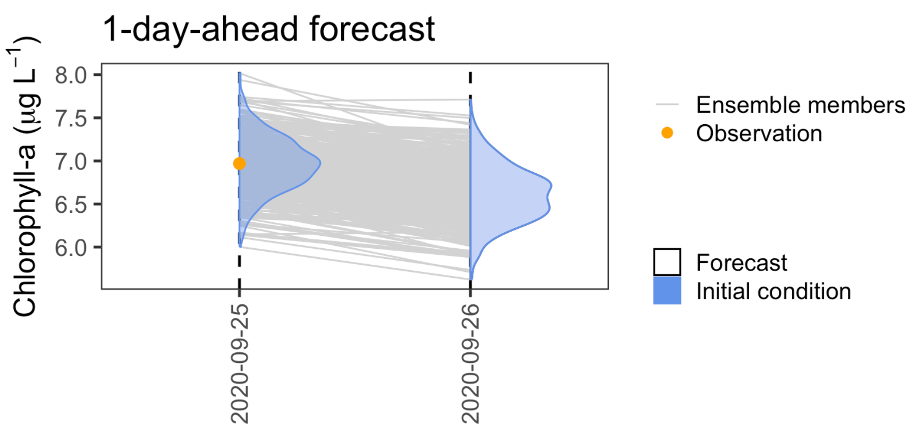


Figure 10. One-day-ahead forecast with updated initial condition when chlorophyll-a data are missing.

    b. Compare the forecast distribution (white) for 2020-09-26 and the updated initial condition (blue distribution) for 2020-09-26. What happened when there was no new observation to update the forecast?

Answer: *The forecast and the updated initial condition are the same. When there are no data available to update the forecast, the previous forecast is used as the new initial condition.*

27. Compare the two-forecast plot with no data assimilation (missing observation) to the two-forecast plot with data assimilation.

    a. Download the two 1-day-ahead forecasts plot with data assimilation and copy-paste it into your report.

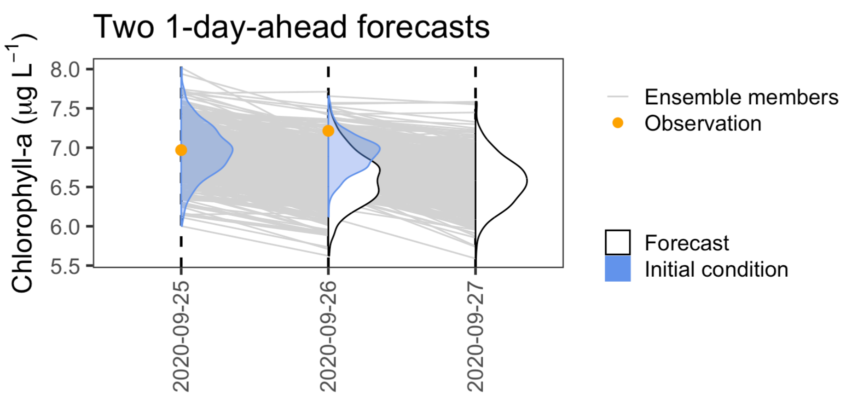


Figure 11. Two-forecast plot with data assimilation.

b. Download the two 1-day-ahead forecasts plot with no data assimilation (missing observation) and copy-paste it into your report.

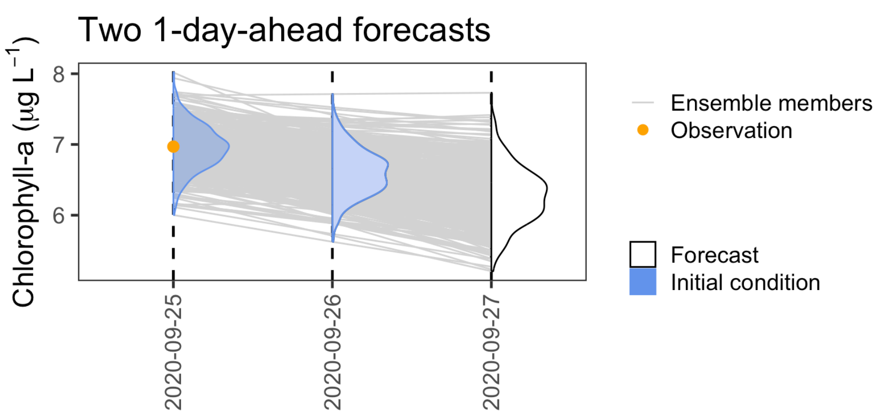


Figure 12. Two-forecast plot without data assimilation.

c. How do the forecasts for 2020-09-27 on each plot compare?

Answer: *Answers will vary depending on which lake students have chosen. In some lakes, data assimilation will result in a very different second forecast, while in others, the difference is slight. Encourage students to compare their results across lakes.*

Objective 6: Explore observation uncertainty

Compare one-day-ahead forecasts generated with data assimilation, using data with low vs. high observation uncertainty.

28. Make a prediction. How do you think a decrease in observation uncertainty will affect the forecasts?

Answer: *Answers will vary.*

29. Compare the initial conditions distributions (blue) in Figure A with those in Figure B.

    a. Download the plot of forecasts assimilating data with low observation uncertainty and copy-paste it into your report.

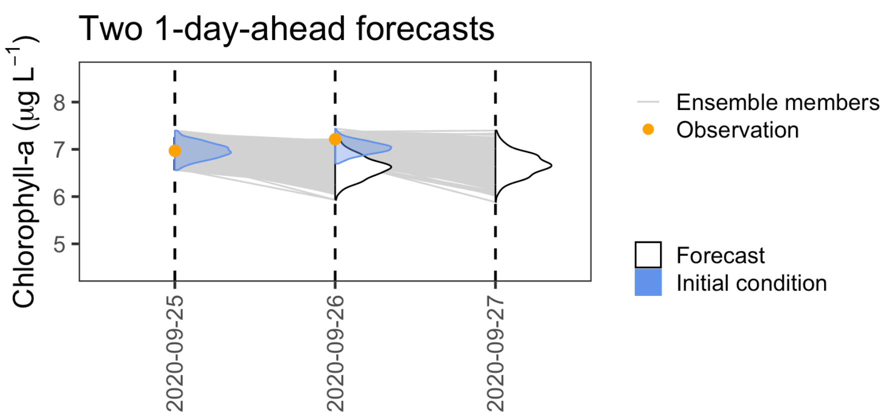


Figure 13. Plot of forecasts assimilating data with low observation uncertainty.

    b. Describe how the initial conditions distributions differ between the two figures.

    Answer: *The initial condition distributions in the figure with low observation uncertainty are narrower than the previous initial condition distributions.*

    c. What is the effect of a decrease in observation uncertainty on the forecasts? Does this match what you predicted in Q28?

    Answer: *Answers will vary; in general, the forecasts with lower observation uncertainty will be less uncertain (have tighter distributions) than the previous forecasts.*

30. Make a prediction. Using your experience from the previous example, how do you think an increase in observation uncertainty will affect the forecasts?

Answer: *Answers will vary.*

31. Compare the initial conditions distributions (blue) in Figure C above with those in Figure D.

    a. Download the plot of forecasts assimilating data with high observation uncertainty and copy-paste it into your report.

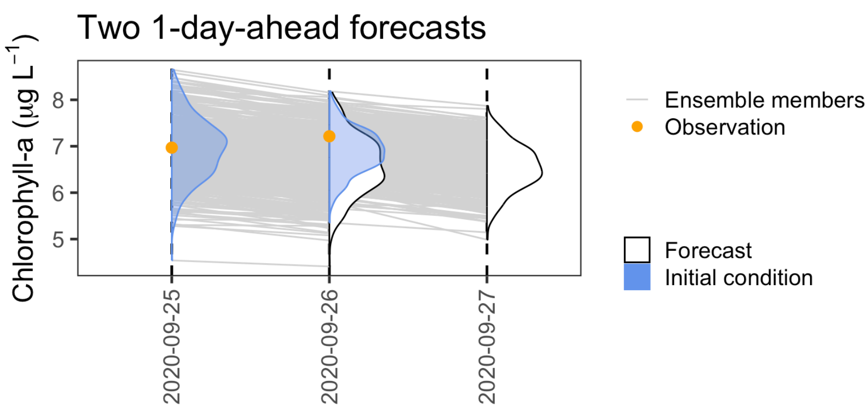


Figure 14. Plot of forecasts assimilating data with high observation uncertainty.

b. Describe how the initial conditions distributions differ between the two figures.

Answer: *The initial condition distributions in the figure with high observation uncertainty are broader than the previous initial condition distributions.*

c. What is the effect of an increase in observation uncertainty on the forecasts? Does this match what you predicted in Q30?

Answer: *Answers will vary; in general, the forecasts with higher observation uncertainty will be more uncertain (have broader distributions) than the previous forecasts.*

Objective 7: Explore data assimilation frequency

Compare a series of one-day-ahead forecasts with no data assimilation, weekly data assimilation, and daily data assimilation.

---

32. Make a prediction by filling in the blank: as the frequency of data assimilation increases, forecast accuracy ____________. Choose from 'increases', 'decreases', or 'stays the same.'

    Answer: *Answers will vary.*

33. Describe the series of one-day-ahead forecasts with no data assimilation.

    a. Download the plot of a series of one-day-ahead forecasts with no data assimilation and copy-paste it into your report.

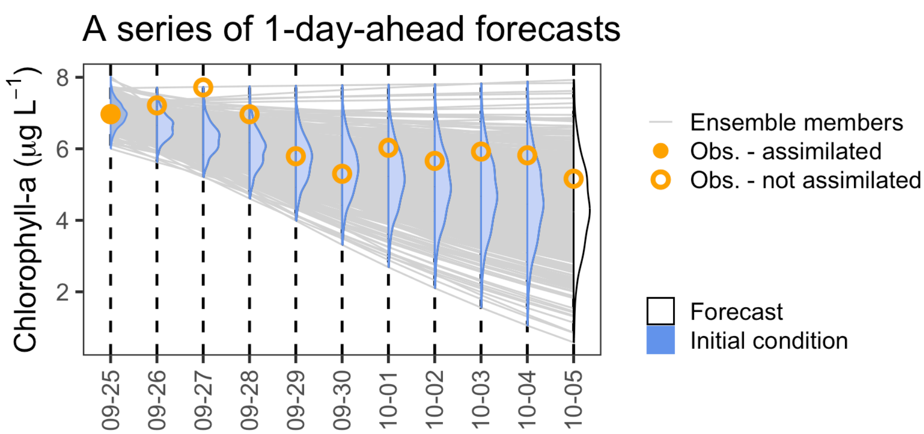


Figure 15. A series of one-day-ahead forecasts with no data assimilation.

b. Describe how the central tendency (most likely outcome predicted for each day) of the 1-day-ahead forecasts changes over time when no data are assimilated.

Answer: *Answers will vary depending on which lake students have chosen.*

c. Describe how the uncertainty distribution of the 1-day-ahead forecasts changes over time when no data are assimilated.

Answer: *Uncertainty increases over time.*

34. Assess the series of one-day-ahead forecasts with no data assimilation.

a. Click the "show observations" checkbox under the plot above, and then visually assess the forecasts and describe their accuracy. How well do they match observations?

Answer: *Answers will vary depending on which lake students have chosen. Encourage students to compare their series of forecasts with others working on different lakes.*

b. Click the buttons to calculate bias and RMSE, and then use the values of bias and RMSE to assess the forecasts with no data assimilation. How well do you think the forecasts are performing?

Answer: *Bias is -0.78 and RMSE is 0.91 (for Lake Barco). Answers will vary depending on which lake students have chosen. Encourage students to compare their values of bias and RMSE with others working on different lakes.*

35. Describe the series of one-day-ahead forecasts with weekly data assimilation.

a. Download the plot of a series of one-day-ahead forecasts with weekly data assimilation and copy-paste it into your report.

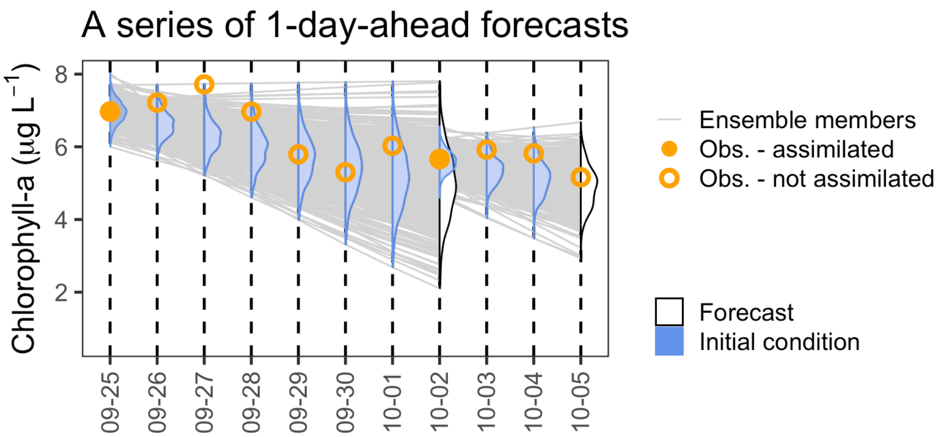


Figure 16. A series of one-day-ahead forecasts with weekly data assimilation.

b. Describe how the central tendency (most likely outcome predicted for each day) of the 1-day-ahead forecasts changes over time when weekly data are assimilated.

Answer: *Answers will vary depending on which lake students have chosen.*

c. Describe how the uncertainty distribution of the 1-day-ahead forecasts changes over time when weekly data are assimilated.

Answer: *Uncertainty increases until a data point is assimilated; then it decreases substantially. After the data point is assimilated, uncertainty begins increasing again.*

36. Assess the series of one-day-ahead forecasts with weekly data assimilation.

a. Click the "show observations" checkbox under the plot above, and then visually assess the forecasts and describe their accuracy. How well do they match observations?

Answer: *Answers will vary depending on which lake students have chosen. Encourage students to compare their series of forecasts with others working on different lakes.*

b. Click the buttons to calculate bias and RMSE, and then use the values of bias and RMSE to assess the forecasts with weekly data assimilation. How well do you think the forecasts are performing?

Answer: *Bias is -0.62 and RMSE is 0.75 (for Lake Barco). Answers will vary depending on which lake students have chosen. Encourage students to compare their values of bias and RMSE with others working on different lakes.*

37. Describe the series of one-day-ahead forecasts with daily data assimilation.

a. Download the plot of a series of one-day-ahead forecasts with daily data assimilation and copy-paste it into your report.

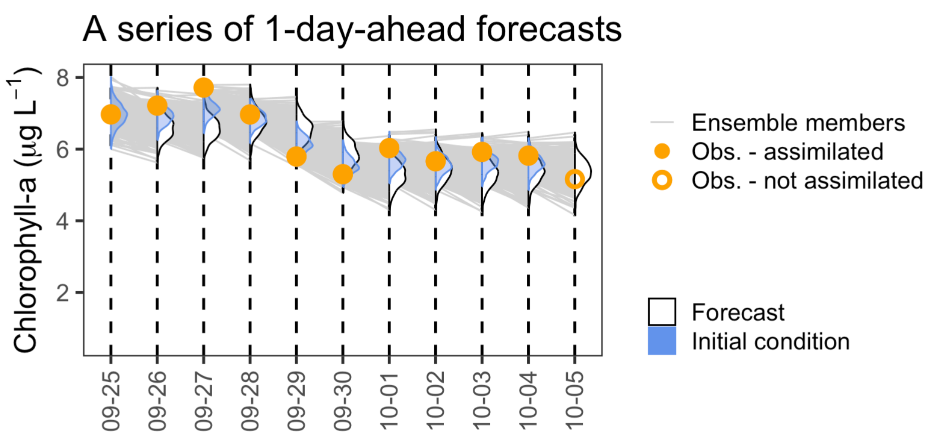


Figure 17. A series of one-day-ahead forecasts with daily data assimilation.

b. Describe how the central tendency (most likely outcome predicted for each day) of the 1-day-ahead forecasts changes over time when daily data are assimilated.

Answer: *Answers will vary depending on which lake students have chosen*

c. Describe how the uncertainty distribution of the 1-day-ahead forecasts changes over time when daily data are assimilated.

Answer: *Uncertainty in the forecasts is greater than uncertainty in the initial condition distributions. Every day, uncertainty in the forecasts is constrained (decreases) when data are assimilated and used to update the initial conditions.*

38. Assess the series of one-day-ahead forecasts with daily data assimilation.

a. Click the "show observations" checkbox under the plot above, and then visually assess the forecasts and describe their accuracy. How well do they match observations?

Answer: *Answers will vary depending on which lake students have chosen. Encourage students to compare their series of forecasts with others working on different lakes.*

b. Click the buttons to calculate bias and RMSE, and then use the values of bias and RMSE to assess the forecasts with daily data assimilation. How well do you think the forecasts are performing?

Answer: *Bias is -0.25 and RMSE is 0.61 (for Lake Barco). Answers will vary depending on which lake students have chosen. Encourage students to compare their values of bias and RMSE with others working on different lakes.*

39. Fill in the blank: as the frequency of data assimilation increases, forecast accuracy ____________. Choose from 'increases', 'decreases', or 'stays the same.' How does your answer now compare to what your answer in Q40?

Answer: *Students should select 'increases'. Answers about how their current choice compared to their choice in Q32 will vary.*

Activity C - Management Scenario

Make management decisions using ecological forecasts generated with different levels of observation uncertainty and different frequencies of data assimilation.

Objective 8: Management scenario

Make management decisions using ecological forecasts generated with different levels of observation uncertainty and different frequencies of data assimilation.

---

40. Make a preliminary recommendation. Based on what you have learned in Activities A and B, do you recommend that the Green Reservoir management authority invest in a new high-frequency sensor to inform their forecasts?

    Answer: *Answers will vary.*

41. Make a management decision.

    a. Download the plot of the series of forecasts generated using the current forecasting system and copy-paste it into your report.

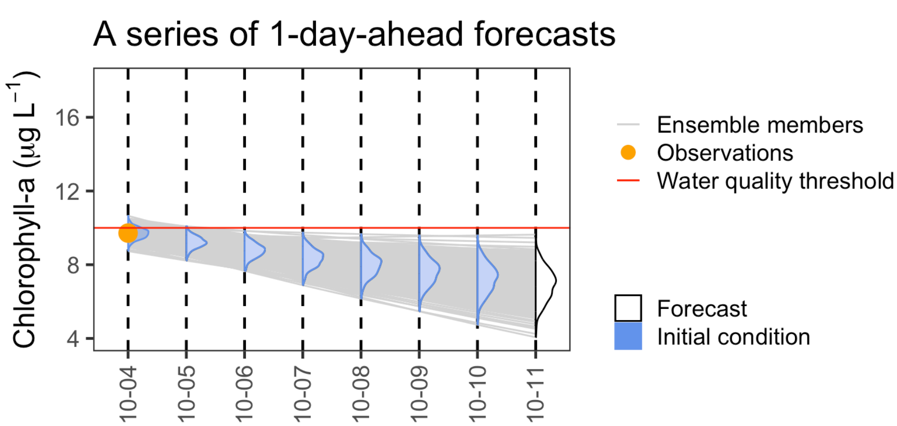


Figure 18. A series of of one-day-ahead forecasts generated using the current forecasting system and data collection method.

    b. Based on the forecasts presented here, as a manager, would you recommend a beach closure for Saturday, October 11?

    Answer: *Answers will vary; most students will probably not recommend a beach closure as the entire range of the forecast is below the water quality threshold.*

42. Make a management decision.

a. Download the plot of the series of forecasts generated using the borrowed high-frequency sensor and copy-paste it into your report.

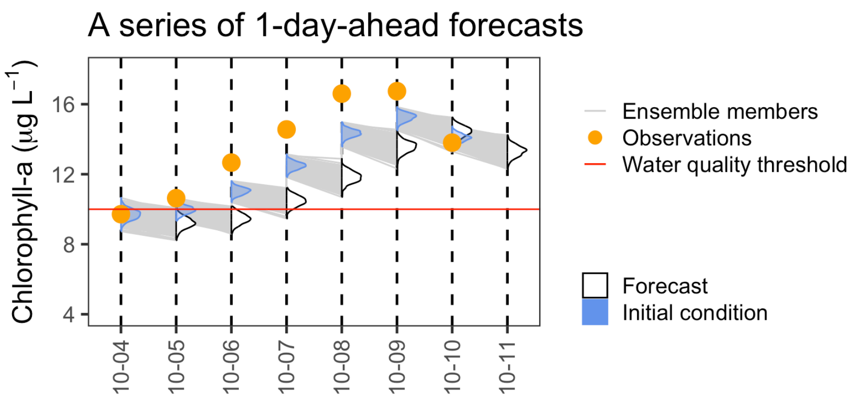


Figure 19. A series of of one-day-ahead forecasts generated using the borrowed high-frequency sensor.

b. Based on the forecasts presented here, as a manager, would you recommend a beach closure for Saturday, October 11?

Answer: *Answers will vary; most students will probably recommend a beach closure as the entire range of the forecast is well above the water quality threshold.*

43. Compare the two series of forecasts.

a. Download the plot of the series of forecasts generated using the current forecasting system AND the observation for Saturday, Oct. 11 and copy-paste it into your report.

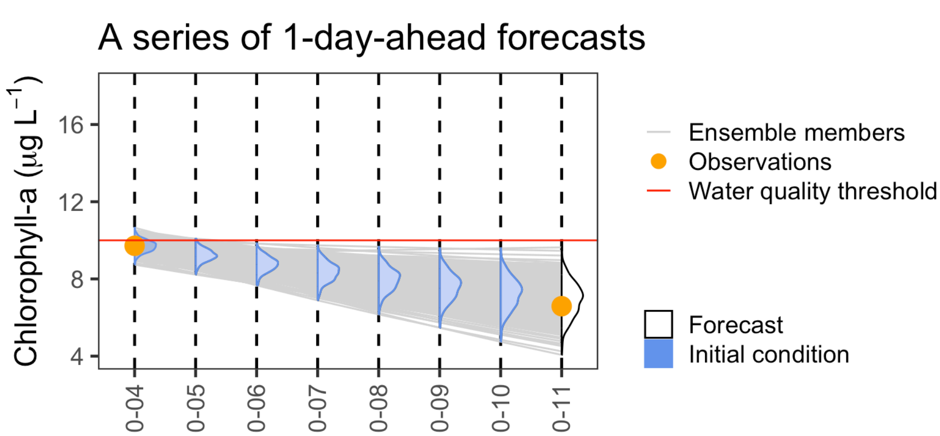


Figure 20. A series of of one-day-ahead forecasts generated using the current forecasting system and data collection method and showing the observation for Saturday, Oct. 11.

b. Download the plot of the series of forecasts generated using the borrowed high-frequency sensor AND the observation for Saturday, Oct. 11 and copy-paste it into your report.

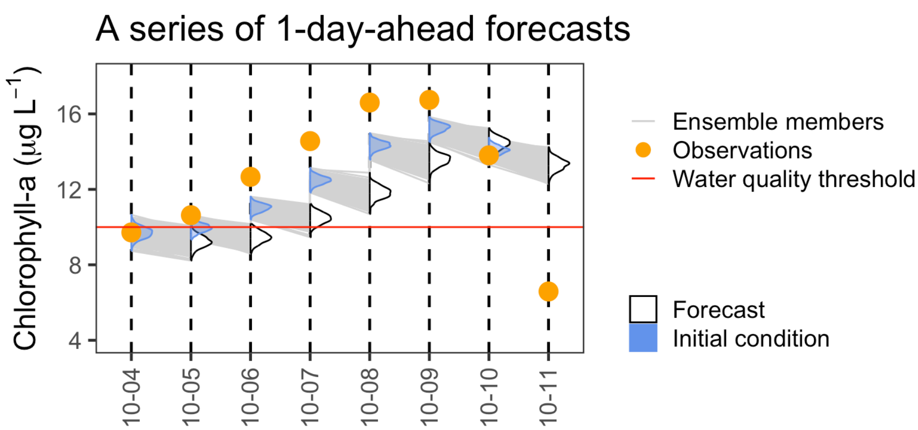


Figure 21. A series of of one-day-ahead forecasts generated using the borrowed high-frequency sensor and showing the observation for Saturday, Oct. 11.

c. Which forecast series - the one with or without the high-frequency sensor data - provided a more accurate forecast for Saturday, Oct. 11?

Answer: *The one without the high frequency sensor.*

d. How did your lake closure decision made using the forecast series with weekly, manual data assimilation compare to the decision made using the forecast series with daily high-frequency sensor data assimilation?

Answer: *Answers will vary; most students will likely have made the "wrong" decision to close the beach when using the forecasts with the high-frequency sensor.*

44. Compare the values of bias and RMSE for the forecast series with data assimilation of weekly manual data vs. the forecast series with data assimilation of daily sensor data. Overall, which forecast series produces more accurate forecasts?

Answer: *The forecasts with the high-frequency sensor are more accurate, as they have a lower absolute value of bias and a lower RMSE.*

45. Make a final recommendation.

a. Based on what you have learned from your forecast trials with the high-frequency sensor, do you recommend that the Green Reservoir management authority invest in a new high-frequency sensor to inform their forecasts?

Answer: *Answers will vary; students will be weighing the tradeoff between having more accurate forecasts using the high-frequency sensor with the increased cost of the sensor* *and the fact that in the end, the forecasts still weren't accurate enough to result in correct management decisions.*

b. Briefly explain the reasoning behind your final recommendation. Did it change from the preliminary recommendation that you made in Q.40?

Answer: *Answers will vary; students will be weighing the tradeoff between having more accurate forecasts using the high-frequency sensor with the increased cost of the sensor and the fact that in the end, the forecasts still weren't accurate enough to result in correct management decisions.*

This module was initially developed by: Lofton, M.E., T.N. Moore, R.Q. Thomas, and C.C. Carey. 07 March 2024. Macrosystems EDDIE: Using Data to Improve Ecological Forecasts. Macrosystems EDDIE Module 7, Version 1. https://serc.carleton.edu/eddie/teaching_materials/modules/module7.html. Module development was supported by NSF grants DEB-1926050 and DBI-1933016.

Authorship contributions:

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