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



### Name:

### Student ID:

#### Completed on:

#### 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](https://oceanservice.noaa.gov/ecoforecasting/noaa.html#:~:text=What%20is%20ecological%20forecasting%3F,%2C%20pollution%2C%20or%20habitat%20change.)
* [Ecological Forecasting Initiative](https://ecoforecast.org/about/)

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](https://www.youtube.com/watch?v=Lgi_e7N-C8E&t=196s&pbjreload=101)
* Fundamentals of Ecological Forecasting Series: [Why Forecast?](https://www.youtube.com/watch?v=kq0DTcotpA0&list=PLLWiknuNGd50Lc3rft4kFPc_oxAhiQ-6s&index=1)
* Fundamentals of Ecological Forecasting Series: [Forecast Analysis Cycle](https://www.youtube.com/watch?v=zBsqjmdLYfk&list=PLLWiknuNGd50Lc3rft4kFPc_oxAhiQ-6s&index=1)
* Fundamentals of Ecological Forecasting Series: [Ensemble Kalman Filter](https://www.youtube.com/watch?v=UeFsEOGz_cc&list=PLLWiknuNGd50Lc3rft4kFPc_oxAhiQ-6s&index=1)

## 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:

*Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J.S. and Schimel, D.S. (2011), Ecological forecasting and data assimilation in a data-rich era. Ecological Applications, 21: 1429-1442.* [*https://doi.org/10.1890/09-1275.1*](https://doi.org/10.1890/09-1275.1)

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

1. Define ‘data assimilation’.

* **Answer:**

1. Summarize why, in the authors’ opinion, data assimilation is needed for ecological forecasting.

* **Answer:**

1. Name one limitation of data assimilation in ecological forecasting.

* **Answer:**

1. The authors argue that the current, data-rich era in ecology is the perfect time to implement data assimilation more often in ecological modeling and forecasting. Why is the data-rich era an especially promising time for data assimilation efforts?

* **Answer:**

1. Provide one question you have after reading the paper. This could be, for example, a point you found confusing and would like to clarify, or a follow-up question on the topic.

* **Answer:**

Now navigate to the [Shiny interface](https://macrosystemseddie.shinyapps.io/module7) 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:**

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

* **Answer:**

# 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.

1. Fill out information about your selected NEON site.
   1. Name of selected site:
   2. Four letter site identifier:
   3. Latitude:
   4. Longitude:
   5. Lake area (km2):
   6. Elevation (m):

## 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.

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

* **Answer:**

1. Describe chlorophyll-a data at your lake.
   1. Download the timeseries plot of chlorophyll-a data and copy-paste it into your report.
   * *Please copy-paste your Q5a-plot.png image here.*
   * *Figure 1. Time series of chlorophyll-a data for your selected NEON lake.*
   1. Describe how chlorophyll-a changes over time in your lake. Do you notice any patterns or trends?
   * **Answer:**

## 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.

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

* **Answer:**

1. 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?
   1. Download the plot of lagged chlorophyll-a and copy-paste it into your report.
   * *Please copy-paste your Q7a-plot.png image here.*
   * *Figure 2. Time series of lagged chlorophyll-a data for your selected NEON lake.*
   1. How do the two lines plotted on the timeseries (chlorophyll and 1 day lag of chlorophyll) relate to each other?
   * **Answer:**
2. 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?
   1. Download the scatterplot and copy-paste it into your report.
   * *Please copy-paste your Q8a-plot.png image here.*
   * *Figure 3. Scatterplot of chlorophyll-a data vs. lagged chlorophyll-a data for your selected NEON lake.*
   1. Do you think the chlorophyll-a data at your chosen lake site exhibit autocorrelation? Why or why not?
   * **Answer:**
3. 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:**

1. Describe autocorrelation across many lags at your lake site.
   1. Download the autocorrelation plot and copy-paste it into your report.
   * *Please copy-paste your Q10a-plot.png image here.*
   * *Figure 4. Autocorrelation of lagged chlorophyll-a at your lake site.*
   1. Describe how autocorrelation changes as the lag in days increases. Why do you think this pattern occurs?
   * **Answer:**
2. 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:**

1. Examine the PACF plot.
   1. Download the PACF plot and copy-paste it into your report.
   * *Please copy-paste your Q12a-plot.png image here.*
   * *Figure 4. Partial autocorrelation of chlorophyll-a at your lake site.*
   1. Examine the PACF plot. Which lag contributes the most to autocorrelation in the chlorophyll-a data? Explain how you know.
   * **Answer:**
2. 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:**

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

* **Answer:**

1. Use the plot above to assess the model fit to data.
   1. Download the plot of model predictions and observations and copy-paste it into your report.
   * *Please copy-paste your Q15a-plot.png image here.*
   * *Figure 5. Autoregressive model predictions and chlorophyll-a observations at your lake site.*
   1. How well do the predictions match the observations?
   * **Answer:**
2. 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:**

1. 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:**

## End of module - nice work!

This module was initially developed by: Lofton, M.E., T.N. Moore, R.Q. Thomas, and C.C. Carey. 13 May 2022. Macrosystems EDDIE: Using Data to Improve Ecological Forecasts. Macrosystems EDDIE Module 7, Version 7. <https://serc.carleton.edu/eddie/teaching_materials/modules/module7.html>. Module development was supported by NSF grants DEB-1926050 and DBI-1933016. This app was last updated on: 2024-01-30