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
- Ecological Forecasting Initiative

Articles:

- Silver, N. 2012. The 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:

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

   Answer:

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:

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:

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.

Introduction

Think about it!

Answer the following questions:

1. What is meant by the term ‘data assimilation’ in the context of ecological forecasting?

   Answer:

2. 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.

Be sure you have answered questions 1 and 2 in the previous Introduction section before you begin Activity A!

3. Fill out information about your selected NEON site.

   a. Name of selected site:

   b. Four letter site identifier:

   c. Latitude:

   d. Longitude:

   e. Lake area (km2):

   f. 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 primarily focusing on 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:

5. Describe chlorophyll-a data at your lake.

   a. 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.

   b. 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 historical trends 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:

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.

      Please copy-paste your Q7a-plot.png image here.

      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:

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.

      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.

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

      Answer:

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:

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

    a. 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.

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

       Answer:

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:

12. Examine the PACF plot.

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

       Please copy-paste your Q12a-plot.png image here.

       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:

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:

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

    Answer:

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.

       Please copy-paste your Q15a-plot.png image here.

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

    b. How well do the predictions match the observations?

       Answer:

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:

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:

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:

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:

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

    Answer:

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.

       Please copy-paste your Q21a-plot.png image here.

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

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

       Answer:

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:

23. Examine the chlorophyll-a forecast plot.

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

       Please copy-paste your Q23a-plot.png image here.

       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:

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

       Answer:

    d. Each one of the gray lines in the figure above represents an ensemble member. Describe the sources of uncertainty in the ensemble forecast for 2020-09-26 in your own words.

       Answer:

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:

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.

       Please copy-paste your Q25a-plot.png image here.

       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:

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.

       Please copy-paste your Q26a-plot.png image here.

       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:

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.

       Please copy-paste your Q27a-plot.png image here.

       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.

       Please copy-paste your Q27b-plot.png image here.

       Figure 12. Two-forecast plot without data assimilation.

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

       Answer:

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:

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.

       Please copy-paste your Q29a-plot.png image here.

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

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

       Answer:

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

       Answer:

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:

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.

       Please copy-paste your Q31a-plot.png image here.

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

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

       Answer:

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

       Answer:

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:

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.

       Please copy-paste your Q33a-plot.png image here.

       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:

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

Answer:

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:

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:

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.

Please copy-paste your Q35a-plot.png image here.

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:

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

Answer:

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:

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:

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.

       Please copy-paste your Q37a-plot.png image here.

       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:

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

       Answer:

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:

    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:

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 Q32?

    Answer:

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:

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.

       Please copy-paste your Q41a-plot.png image here.

       Figure 18. A series 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:

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.

       Please copy-paste your Q42a-plot.png image here.

       Figure 19. A series 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:

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.

Please copy-paste your Q43a-plot.png image here.

Figure 20. A series 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.

Please copy-paste your Q43b-plot.png image here.

Figure 21. A series 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:

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:

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 over the trial forecasting period?

Answer:

45. Make a final recommendation.

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

Answer:

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

Answer:

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. This app was last updated on: 2024-03-07