Homework 3: Prediction over Time and Space

36-740, Fall 2026

Due at 6 pm on Thursday, 17 September 2026

Abstract

AGENDA: Practice with linear prediction of time series, especially over space. Practice with linear prediction of spatial data, especially over time. Practice with estimating covariance functions. Practice with extracting seasonal trends and de-trending.

The data set `wind` in the library `gstat` was introduced in Lecture 3. It consists of daily measurement of wind speed at twelve different locations across Ireland from 1961 through 1978. It is accompanied by the data set `wind.loc` which gives the full names and map coordinates for the stations. $X(r, t)$ will stand for the wind speed at station r on day t.

1. *Yesterday is like today*

   (a) (3) Calculate the slope and intercept for the optimal linear predictor of $X(\mathrm{DUB}, t)$ from $X(\mathrm{DUB}, t-1)$; be explicit about which variances and covariances you need to find this. What should the root-mean-squared (RMS) prediction error be?

   (b) (3) Use `lm` to linearly regress $X(\mathrm{DUB}, t)$ on $X(\mathrm{DUB}, t-1)$. What are the coefficients it reports? What is the connection between this and what you did in problem 1a? *Hint:* What would `lm(wind$DUB[2:11] ~ wind$DUB[1:10])` do?

   (c) (1) Suppose we want to retrodict $X(\mathrm{DUB}, t-1)$ from $X(\mathrm{DUB}, t)$. What are the optimal linear coefficients?

2. *Yesterday is like tomorrow*

   (a) (1) What is the optimal slope for linearly predicting $X(\mathrm{DUB}, t)$ from $X(\mathrm{DUB}, t-2)$?

   (b) (3) Suppose we want to linearly predict $X(\mathrm{DUB}, t)$ using both $X(\mathrm{DUB}, t-1)$ and $X(\mathrm{DUB}, t-2)$. Which variances and covariances we need to know in order to find the coefficients? What are all those variances and covariances? What are the resulting coefficients? (You will probably want to report things in tables.)

3. *Correlations over time*

(a) (2) Plot the autocorrelation function for $X(\mathrm{DUB}, t)$ out to a lag of 800 days. Describe the shape of the function. Can you explain the location of the peaks and troughs of the function?

(b) (2) What is the slope for linearly predicting $X(\mathrm{DUB}, t)$ from $X(\mathrm{DUB}, t-365)$? What are the slopes for linearly predicting $X(\mathrm{DUB}, t)$ from $X(\mathrm{DUB}, t-1)$ *and* $X(\mathrm{DUB}, t-365)$?

(c) (2) Plot the cross-correlation function between Dublin and Shannon out to a lag of 800. Describe the shape of the function. Can you explain the location of the peaks and troughs of the function?

4. *Our first spatial regression* Suppose we want to predict $X(\mathrm{DUB}, t)$ using all the $X(r, t)$, for all $r \neq \mathrm{DUB}$.

   (a) (3) What are the coefficients of the optimal linear predictor? (There should be twelve of them, so think carefully about how to report them.)

   (b) (2) What, theoretically, should the RMS prediction error be?

5. *Our first spatiotemporal regression* Suppose we want to predict $X(\mathrm{DUB}, t)$ using all the $X(r, t-1)$, for all $r \neq \mathrm{DUB}$.

   (a) (2) What are the coefficients of the optimal linear predictor? How do they compare to the coefficients that you found in problem 4a

   (b) (2) What, theoretically, should the RMS prediction errors be?

6. *In- and out- of sample comparisons*

   (a) (5) You have now fitted linear predictors in problems 1a, 2b, 4a and 5a. For each predictor, calculate its error in predicting $X(\mathrm{DUB}, t)$ for all t. Report the mean and RMS errors for each predictor in a table, along with their theoretical values. Which predictor seems to do best? Does theory match data?

   (b) (3) Re-estimate the coefficients of each of those four models, using only data from the first nine years of the data set (i.e., 1961–1969). What are the old and new coefficients for each model?

   (c) (3) For each of the four models, use the coefficients you found in problem 6b to calculate a prediction for each day in 1970–1978. Report the mean and RMS prediction errors for all four predictors.

   (d) (3) Why do the numbers in problem 6a differ from those in 6c? Which set of numbers gives a better idea of how well the different models predict?

7. *Ignoring inconvenient facts* (1) Four rows of the data set represent leap days. What are their row numbers? Create a new data set which removes those four rows, and use that data in all the rest of this homework assignment.

8. *Annual patterns*

   (a) (1) What are the average wind speeds at all 12 stations on January 1st? You should get a vector of length 12, each element of which is an average of 18 measurements.

   (b) (3) Find the average wind speed at each day in the calendar year, at each of the 12 stations. Your answer should have a 365 rows and at least 12 columns. (You may find it convenient to add two columns for the month and the day of the month.) Display your results by plotting these twelve annual trends. (It is ideal if you can fit all 12 annual trends into the same plot.)

   (c) (1) Describe, in words, the two annual patterns at Dublin and at Valentina.

9. *Seasonally-adjusted values*

   (a) (3) Subtract the average wind speed at each station, on January 1st, from the actual wind speeds on January 1st. Display the results as time series. (You should have 12 series, each of 18 points.) If possible, show all the series in the same figure.

   R hint: The `scale()` function can be used to subtract a vector from each row of a data frame, without scaling. (There are other ways to approach this as well.)

   (b) (4) Create a new data frame where the appropriate trend has been subtracted from the observation for each station. Display the 12 new, de-trended time series (either in one plot or in a mosaic of plots, whichever you can make clearer).

10. *Covariance and seasons* In this problem, use the de-trended data you prepared in problem 9b. Assume that each of the time series is stationary.

    (a) (3) Report the autocovariance function for Dublin, and the cross-covariance between Dublin and Shannon, both out to a lag of 800 days. How do these functions compare to those you computed earlier?

    (b) (3) Assuming the data is now stationary, use the data from the first nine years to estimate a model which predicts Dublin on day t from Dublin on day $t-1$ and from Shannon on day t. This model should have an intercept and two slopes — what are they? Use this model to predict the seasonally-adjusted wind speed in Dublin for each day of 1975. Plot the predictions and the actual values as two time series (in the same plot). What is the RMS error?

    (c) (3) Using the same model, predict the *un*-adjusted wind speed in Dublin for each day in 1975. Plot the predictions and the actual values as two time series (in the same plot). What is the RMS error? Is the value you get for the RMS error surprising, or something you should anticipate, given your results in previous problems?

11. *Spatial kriging* In this problem, use the seasonally-adjusted data from problem 9b. *General hint:* The slides for lecture 6 (linear prediction over space) will be helpful; read all of them before plunging in to this problem.

    (a) (4) Using `lm`, and the first nine years of data, linearly regress the wind speed at Dublin on day t on the wind speed for the same day at the other 11 stations. What are the coefficients? (There should be 12 of them; use a table or figure.)

    (b) (4) For each pair of stations, calculate the correlation coefficients in wind speed on day t, using the first nine years of the data. Plot the correlation coefficients against the distance, in kilometers, between stations.

    (c) (5) Fit an exponential function to these correlations. What is the correlation length (in kilometers)?

    (d) (5) Using the estimated spatial correlation function, *not* the sample covariances, find the coefficients of the best linear predictor of the wind speed at Dublin on day t from the wind speed at the other stations on day t. What are the 11 slopes and the intercept? (Report your answer in the form of a table or figure.) Compare these values to those in problem 11a, and explain why they are the same (if they *should* be the same) or explain the differences (if they should be different). *Hint:* Make sure Dublin isn't used to predict itself.

    (e) (4) Use your model from problem 11d to give predictions for the wind speed in Dublin on each day of 1975. Plot these predictions as a time series, along with the actual (seasonally-adjusted) speeds. What is the RMS error?

    (f) (4) Using the estimated spatial correlation function an assumption of spatial stationarity, find a prediction for the wind speed at Belfast for every day of 1975. Plot this as a time series.

    (g) (1) Explain why you cannot find the RMS error for this last set of predictions.

12. (1) How much time did you spend on this problem set?

RUBRIC (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision. Any code displayed in the write-up is relevant, without dangling or unused fragments, and is indented, commented, and uses meaningful names. All parts of all problems are answered with coherent sentences, and raw computer code or output are only shown when explicitly asked for.

Extra Credit (5): Using the model from problem 11d, create maps of (seasonally adjusted) wind speed that cover the whole of Ireland for the first day of each month in 1975. (Why is it complicated to un-do the seasonal adjustment here?) Provide a mosaic of 12 maps, clearly labeled by date. *Hints:* You will want probably want to make predictions on a grid of evenly-spaced points; try to avoid inverting a matrix for each grid point; display predictions with something like `contour()`.