Studio 8 Significance testing and El Niño-Southern Oscillation
18.05, Spring 2025

Overview of the studio

Vocabulary

Here is the list of abbreviations and terms that are used in this studio:

- **ENSO**: El Niño-Southern Oscillation
- **MSEA**: Mainland Southeast Asian countries (Cambodia, Laos, Thailand, Myanmar, and Vietnam)
- **SST**: Sea surface temperature (measured in °K)

In this studio, we want to investigate the relationship between ENSO and precipitation in the MSEA region. The hydroclimate of MSEA is significantly affected by ENSO, a climate pattern known for its substantial impact on weather conditions globally. ENSO plays a crucial role in shaping the region's weather extremes, leading to severe droughts and floods that greatly affect both human societies and ecosystems. These climatic events disrupt agricultural productivity, influence water resources, and can lead to economic challenges.

What we are specifically interested in is to see if the patterns that existed in the historical data have changed recently.

Here is all the data that we have for this studio:

- monthly sea surface temperature (SST) data from 1940 till 2024.[1]
- precipitation data from 1891 till 2019.[2]

R introduced in this studio

Download the zip file

- You should have downloaded the studio8 zip file from our Canvas site.
- Unzip it in your 18.05 studio folder.
- You should see the following R files

  studio8.r, studio8-samplecode.r, studio8-test.r

  and the following other files

  studio8-instructions.pdf (this file), studio8-test-answers.html

[1] Provided by European Centre for Medium-Range Weather Forecasts fifth generation.
[2] Provided by Global Precipitation Climatology Centre.

Prepping R Studio

- In R studio, open studio8-samplecode.r and studio8.r
- Using the Session menu, set the working directory to source file location. (This is a good habit to develop!)
- Answer the questions in the detailed instructions just below. Your answers should be put in studio8.r
- Solution code will be posted on Saturday at 4 am

Detailed instructions for the studio

Problem 1

In this problem, our goal is to find in which months of the year the ENSO effect is at its peak.

Problem 1a. In this problem, you finish the code for the function

`studio8_problem_1a()`

This function plots the average SST data on a map, as well as the same data only shown for the El Niño region. Take a look at the El Niño region SST and try to reason about the way the temperature is distributed. Write your explanation briefly in the `cat` statement at the end of the function.

Problem 1b. In this problem, you finish the code for the function

`studio8_problem_1b()`

Next, we calculate the area-weighted mean of the El Niño region. Note that since latitude-longitude grid cells are *larger* at the equator than at the poles, it is important to weigh the data correctly (by the cosine of the latitude) to get a correct average. This is not important for the Niño3.4 index which only spans $-5°$S to $5°$S, but would be important if you took the global average for example.

The Niño3.4 index is typically defined relative to average conditions by removing the long-term mean for each month. This also de-seasonalizes the data by removing the seasonal cycle. The 5-month rolling mean is then typically applied to remove short-term variability.

Next, we plot the the Niño-3.4 index. Here we define a threshold of ± 0.4 K to define *El Niño events* (Niño3.4 index > 0.4 K, in red) and *La Niña events* (Niño3.4 index < -0.4 K, in blue).

Problem 1c. In this problem, you finish the code for the function

`studio8_problem_1c()`

We get to the point where we can ask "how does the ENSO amplitude vary seasonally".

First, calculate the ENSO amplitude for each month (absolute values of the Niño3.4 index) that you derived in **problem 1b** and then, calculate the mean and standard deviation of the ENSO amplitude for each month. Plot these seasonal variations in ENSO amplitude in a simple bar chart.

In the `cat` statement at the end of the function, specify the three months that have the largest mean ENSO amplitude. The three months you choose should be one of NDJ, FMA, MJJ, and ASO. We will work with these three months in the rest of the studio.

The fact that there is a peak in these three months is a well-known property of ENSO and is called "*phase locking*". However, the reasons for this is still an active area of research!

Problem 2

We now start working with the precipitation data. You do not have to write any code for this problem. Only run the function

`studio8_problem_2()`

to see the plots.

The timeseries of MSEA precipitation looks a lot noisier than the timeseries of SSTs in the Niño3.4 region! This is expected since SSTs are strongly governed by large-scale ocean dynamics whereas precipitation is affected by small-scale atmospheric processes like convection, cloud microphysics, and moisture transport, which introduce more variability.

Since the goal is to (eventually) analyze how the relationship between Niño3.4 SSTs and MSEA precipitation has changed over time, it is crucial to de-seasonalize the data! Many climate variables (e.g., SSTs, precip) co-vary seasonaly. Without removing the seasonality, the correlations will be dominated by the seasonal cycle rather than more meaningful climate relationships.

Problem 3

In this problem, you finish the code for the function

`studio8_problem_3()`

We now investigate the relationship between ENSO and precipitation in the MSEA region. To make our life easier, we downsample the monthly time-series data to seasonal data for the SST and precipitation data.

For ENSO, we focus on the three months that you identified in **problem 1c**. For MSEA precipitation, we do not really know what season(s) to focus on yet. Therefore, we explore how the ENSO-rainfall correlation changes with season! We can do this by plotting a *lead-lag correlation plot*. This plot refers to the correlation between two time series shifted in time relative to one another, and is useful for studying whether a lagged time series can be viewed as a good predictor for another one.

For each year shift in $\{-1, 0, 1\}$ and season (NDJ, FMA, MJJ, or ASO), find the correlation of the season you found for ENSO and the corresponding year-season for MSEA. For example, if the three ENSO months you found are NDJ, you shall compute the correlations between ENSO at NDJ and MSEA at the following seasons (± 1 or 0 denote the year shift):

• ND(-2)J(-1) • FMA(-1) • MJJ(-1) • ASO(-1) • **ND(-1)J(0)** (current season) • FMA(0) • MJJ(0) • ASO(0) • ND(0)J(+1) • FMA(+1) • MJJ(+1) • ASO(+1)

Find the year-season that has the highest (in absolute value) correlation.

Problem 4

Now that you have identified which season of ENSO SST anomalies correlate most strongly with which season of MSEA precipitation, let's see how this relationship has changed over time.

To visually see this relation over time, we can plot the timeseries of the running correlation coefficient between the Niño3.4 index and MSEA precipitation for these months. This means that we choose a window size (13 years in this example), and for each window, we calculate the correlation between the ENSO and MSEA precipitation in the specified seasons.

After this visualization, let us divide the whole data into two parts: pre-1980 (1940–1979) and post-1980 (1980–2019). For each of the parts, compute the running correlations and plot a histogram of these values on top of each other.

Lastly, do a two-sample t-test to see if these data have the same mean or not. Report the statistic and the p-value in the end.

Testing your code

For each problem, we ran the problem function with certain parameters. You can see the function call and the output in studio8-test-answers.html. If you call the same function with the same parameters, you should get the same results as in studio8-test-answers.html – if there is randomness involved the answers should be close but not identical.

For your convenience, the file studio8-test.r contains all the function calls used to make studio8-test-answers.html.

Before uploading your code

1. Make sure all your code is in studio8.r. Also make sure it is all inside the functions for the problems.
2. Clean the environment and plots window.
3. Source the file.
4. Call each of the problem functions with the same parameters as the test file studio8-test-answers.html.
5. Make sure it runs without error and outputs just the answers asked for in the questions.
6. Compare the output to the answers given in studio8-test-answers.html.

Upload your code

Upload your code to Gradescope.

Leave the file name as studio8.r.

You can upload more than once. We will grade the last file you upload.

Due date

Due date: The goal is to upload your work by the end of class. If you need extra time, you can upload your work any time before 6 PM ET on the day of the studio (Friday).

Solutions uploaded: Solution code will be posted on Canvas at 4 AM the day after the studio.