Instructions

Each group should use the dataset provided for the project, and conduct a thorough analysis of the dataset choosing some of the tools that you learnt during the course.

Report

You are required to compile a **brief report** in the template provided in Overleaf without changing the font style and size, and the margins of the document. The report should be

- maximum 8-page long (excluding references);

- containing a maximum of 4 figures;

- containing a maximum of 50 references (note, references are not counted for the page limit).

The report is due **on Week 13**.
The Overleaf template can be found here: https://www.overleaf.com/read/xrnxwmdxqmkr
We ask you to copy the project, and start drafting your own project. For those who have no experience with Latex, the only files you need to modify are: ME5311-template.tex (for the report), and ME5311-references.bib (for the bibliography).

Grading

We will be ultimately grading all your projects, based on the report. However, we ask you to grade your project, as well as your group peers contributions.

Dataset ($\sim$ 2×1 GB): Spatio-temporal data

Description

The Earth's climate, widely known as a multi-scale, high dimensional, nonlinear and chaotic system, significantly influences ecosystems, societies, and economies globally. A better understanding of its mechanisms will bring significant benefits for humanity. By combining historical observations and numerical model outputs, researchers now are able to provide highly detailed understanding of past weather and climate using reanalysis datasets.

This dataset, with a spatial resolution of 0.5°×0.5° and daily temporal resolution, spans from 1979-12-31 to 2022-12-31 and covers a portion of the Indo-Pacific region (70°E-150°E, 10°S-40°N). It includes two key atmospheric variables: sea level pressure and two-meter temperature. The dataset comprises 16,071 daily snapshots, each containing 16,261 (161×101) grid points. One visualization example of time snapshot (2022-12-31) is provided in Fig. 1.

Sea Level Pressure and Two-meter Temperature are stored in "slp.nc" and "t2m.nc", respectively. If your laptop lacks sufficient memory during the project, please refer to the attached code for instructions on how to downsample the data further.
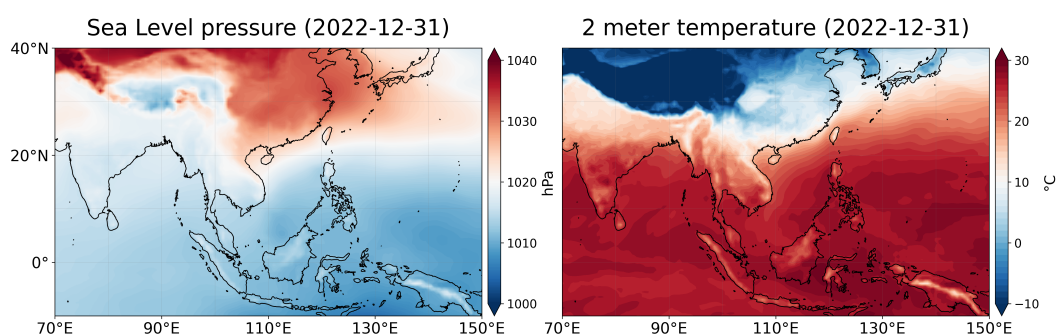

Figure 1: Daily mean sea level pressure and two-meter temperature in 2022-12-31. (The unit of T2m is converted from Kelvin to degree Celsius by minusing 273.15)

How to load the data

An example of how you can load the data in Python is reported below (file is in the same folder as the dataset):

```python
import os
import xarray as xr

# dimensions of data
n_samples = 16071
n_latitudes = 101
n_longitudes = 161
shape = (n_samples, n_latitudes, n_longitudes)

# load data
ds = xr.open_dataset('slp.nc') # change to 't2m.nc' for temperature data

# visualize dataset content
print(ds)

# get 'slp'/'t2m' values
# 'msl' is the variable name for slp data, change to 't2m' for temperature data
da = ds['msl']
x = da.values
print(x)
print(x.shape)

# get time snapshots
da = ds['time']
t = da.values
print(t)

# get longitude values
da = ds['longitude']
```

```
30  lon = da.values
31  print(lon)
32
33  # get latitude values
34  da = ds['latitude']
35  lat = da.values
36  print(lat)
37
38  # # ONLY if not enough memory
39  # low_res_ds = ds[{'longitude': slice(None, None, 2), 'latitude': slice(None, None, 2)}]
40  # low_res_ds.to_netcdf(path='slp_low_res.nc') # change to 't2m_low_res.nc' for temperature
        data
```

Listing 1: Loading the dataset in Python. See file load_data.py

Similarly, you can load the data in Matlab as follow (file is in the same folder as the dataset):

```
1   clc; clear; close all
2
3   % dimensions of data
4   n_samples = 16071;
5   n_latitudes = 101;
6   n_longitudes = 161;
7
8   % load data
9   ncfile = 'slp.nc'; % change to 't2m' for temperature data
10
11  % visualize dataset content
12  ncinfo(ncfile)
13  ncdisp(ncfile)
14
15  % get 'slp'/'t2m' values
16  % 'msl' is the variable name for slp data, change to 't2m' for temperature data
17  x = ncread(ncfile, 'msl');
18
19  % get time snapshots
20  t = ncread(ncfile, 'time');
21
22  % get longitude values
23  lon = ncread(ncfile, 'longitude');
24
25  % get latitude values
26  lat = ncread(ncfile, 'latitude');
```

Listing 2: Loading the dataset in Matlab. See file load_data.m

Modeling suggestions

Although the provided dataset is related to climate, no background in climate science is required for its analysis.

- What kind of information can you retrieve about this type of data?

- For example, can we model it via a reduced order model?

- Is the system predictable?

- How can you characterize this data from a dynamical systems point of view?

- Some methods introduced in the lectures seem only applicable to low dimensional systems. Is it possible to apply them to high dimensional system?

The above are just some suggestions.
Pick one or more topics from the course, to make a compelling analysis of the data!