Data Loading Exercise

Dataset link :
https://drive.google.com/drive/folders/176Pb1jZBeEYbNZcaKxIM1WTZMRCf7oyP?usp=sharing

The goal of this exercise is to work with xarray and geophysical data. The dataset consists of 3 files from an idealized oceanic simulation:

- `toce`: 3D + Time temperature fields
- `soce`: 3D + Time salinity fields
- `ssh`: 2D + Time sea surface height (SSH) fields

This exercise has 3 main parts:

1. Dataloader

- Write a PyTorch dataloader that can sample from these files in a batched manner.
- Since the dataset may be large in realistic applications, your method should not require loading the full dataset into memory.
- The focus is on handling the peculiarities of the data, such as dealing with boundary pixels and normalizing the data, rather than building production-ready code.
- Bonus points if you can sample randomly from the time instants.

You can transform the files to another format if needed.

2. Visualization

- Write a Python script to generate visualizations from these fields.
- The goal is to create interpretable and insightful visualizations that provide understanding of the data (e.g., global statistics over the volumes).

3. Presentation

- Build a 3-5 slide presentation to showcase the key elements of your code and visualizations.
- The focus should be on providing clean insights, as you would in a weekly meeting, rather than creating a fancy presentation.