Project Guide: Classification Using CMIP6 Climate Variables

Andrés Navarro

General Objective

Develop a global classification based on selected climate variables from a set of 10 CMIP6 global climate models. The classification should be tailored to the research objective — such as identifying climate zones, land use categories, environmental typologies, or risk areas — using dimensionality reduction (PCA) and clustering techniques (k-means).

Tools to Use

- Anaconda (development environment)
- Python
- xarray
- matplotlib / seaborn
- CDO (Climate Data Operators)
- GitHub

Project Structure

`data/`	Processed NetCDF files
`notebooks/`	Jupyter Notebooks with analysis
`scripts/`	Reusable Python scripts
`figures/`	Generated plots
`report/`	Final report (PDF)
`README.md`	Project description

Project Steps

1. Data Selection and Download

- Select 10 CMIP6 climate models.

- Choose relevant variables (e.g., temperature, precipitation, wind, humidity).
- Period: 1850–2014, monthly data.

2. Preprocessing with CDO

- Compute monthly climatologies (mean over 1850–2014).
- Calculate anomalies.
- Optionally compute climatologies of anomalies.

3. Creation of Multi-Model Ensemble

- Compute the multi-model average for each variable (using CDO or xarray).
- Optionally include derived variables (e.g., seasonality index, aridity).

4. Dimensionality Reduction with PCA

- Create a feature matrix for each grid point.
- Standardize the data.
- Apply PCA and retain components explaining at least 90% of the variance.

5. Classification with k-means

- Apply k-means on the principal components.
- Determine the optimal number of clusters.
- Visualize clusters on a global map.
- Label clusters according to the chosen classification objective (e.g., land use types).

6. Interpretation of Results

- Compare with existing classifications (e.g., Köppen, FAO land use maps).
- Analyze spatial patterns and discuss limitations.

7. Final Report

- Write a PDF report including:
  - Introduction and objectives
  - Methodology
  - Results and visualizations
  - Interpretation and discussion
  - Conclusions

8. GitHub Repository

- Upload code, processed data, report, and README.

Deliverables

- GitHub repository with all content.
- Final report in PDF format.
- Clear and well-labeled visualizations.
- Clean and documented code.
- **Oral presentation (12 minutes)** summarizing objectives, methodology, results, and conclusions.

Evaluation Criteria

Criterion	**Weight**
Quality of analysis and code	30%
Scientific interpretation	25%
Visualizations and presentation	10%
Repository organization	10%
Final report (PDF)	10%
Oral presentation (12 minutes)	15%