Progress Report: Climate Data Exploration

Full name: EL AMINE Soumya
Supervisor: BALHANE Saloua
Institution: International Water Research Institute
Project: Learning predictable and informative dynamical drivers of extreme precipitation usi

February 11, 2026

Abstract

This report summarizes the initial data exploration phase of the CMM-VAE project. Before proceeding with model implementation, I dedicated time to thoroughly understanding the multi-dimensional climate data used in this research. This report documents the development of a comprehensive tutorial notebook, key learning outcomes, and insights gained about ERA5 geopotential height and CHIRPS precipitation datasets. The exploration phase has established a solid foundation for the subsequent model training and evaluation phases.

Contents

1 Introduction

Before proceeding with the main phases of the CMM-VAE model implementation, I decided to dedicate time to thoroughly understanding the climate data used in this project. This is my first experience working with multi-dimensional climate data in NetCDF format, and I wanted to ensure a solid foundation before moving forward.

This report summarizes the data exploration work completed, including the development of an interactive tutorial notebook.

2 Motivation

Working with climate data presents unique challenges:

- **Multi-dimensional structure:** Data organized as time $\times$ latitude $\times$ longitude
- **NetCDF format:** Specialized scientific data format requiring specific libraries
- **Geospatial visualization:** Understanding how to create and interpret weather maps
- **Physical interpretation:** Connecting atmospheric variables to real-world weather patterns

To address these challenges systematically, I developed a comprehensive hands-on tutorial that allows me to explore and understand each aspect of the data.

3 Data Overview

3.1 ERA5 Geopotential Height (Z500)

Purpose: Represents atmospheric pressure patterns at 500 hPa ($\sim$5.5km altitude)

Parameter	**Value**
Temporal coverage	1940–2022 (83 years), daily data
Spatial coverage	North Atlantic (20°N–87.5°N, 50°W–42.5°E)
Resolution	28 lat $\times$ 38 lon = 1,064 grid points
Total observations	30,316 days

Table 1: ERA5 Z500 dataset specifications

Physical meaning:

- High values (5,800+ m) $\rightarrow$ High pressure systems (anticyclones) $\rightarrow$ Clear, dry weather
- Low values ($<$5,200 m) $\rightarrow$ Low pressure systems (cyclones) $\rightarrow$ Stormy, wet weather

3.2 CHIRPS Precipitation

Purpose: Satellite-based rainfall estimates over Morocco

Parameter	Value
Temporal coverage	1981–2023 (42 years), daily data
Spatial coverage	Morocco (27.7°N–35.9°N, 13.2°W–1.0°W)
Resolution	166 lat × 244 lon = 40,504 grid points (∼5km)
Total observations	6,353 days

Table 2: CHIRPS precipitation dataset specifications

Physical meaning:

- Values in mm/day (0 = no rain, 50+ mm = heavy rainfall)
- High spatial resolution captures local rainfall variability

4 Tutorial Development

4.1 Tutorial Structure

I created an interactive Jupyter notebook (`explore_climate_data.ipynb`) with 10 progressive sections:

1. **Import Libraries:** Introduction to xarray and visualization tools
2. **Load Z500 Data:** Opening NetCDF files and understanding structure
3. **Visualize Single Day:** Creating weather maps with proper projections
4. **Load Precipitation Data:** Loading CHIRPS dataset and comparing resolutions
5. **Side-by-Side Comparison:** Understanding spatial relationships
6. **Time Series Analysis:** Observing temporal evolution and correlations
7. **Calculate Anomalies:** Computing deviations from climatology
8. **Model Input Pipeline:** Understanding data flow into CMM-VAE
9. **Interactive Exploration:** Experimenting with different dates
10. **Summary:** Key takeaways

4.2 Tutorial Location

The tutorial is completely isolated from the main project:

Listing 1: Tutorial directory structure

```
CMM-VAE-Morocco/
          data_exploration_tutorial/
              explore_climate_data.ipynb  (main tutorial)
```

5 Key Learning Outcomes

5.1 Technical Skills Acquired

Technical Skills

- **NetCDF file handling:** Using xarray to open, explore, and manipulate climate data
- **Dimension selection:** Extracting specific time periods, spatial regions, and grid points
- **Geospatial visualization:** Creating professional weather maps with cartopy
- **Time series analysis:** Plotting and interpreting temporal evolution
- **Anomaly calculation:** Computing deviations from climatology
- **Data pipeline understanding:** How raw data becomes model input

5.2 Scientific Understanding

Scientific Insights

- **Z500 interpretation:** High/low pressure systems and their weather impacts
- **Precipitation patterns:** Spatial distribution and temporal variability
- **Atmospheric-rainfall connection:** How North Atlantic patterns affect Moroccan rainfall
- **Weather regime concept:** Persistent circulation patterns lasting several days
- **Anomaly importance:** Why deviations from normal matter more than absolute values

6 Sample Visualizations

6.1 Z500 Atmospheric Pressure Map

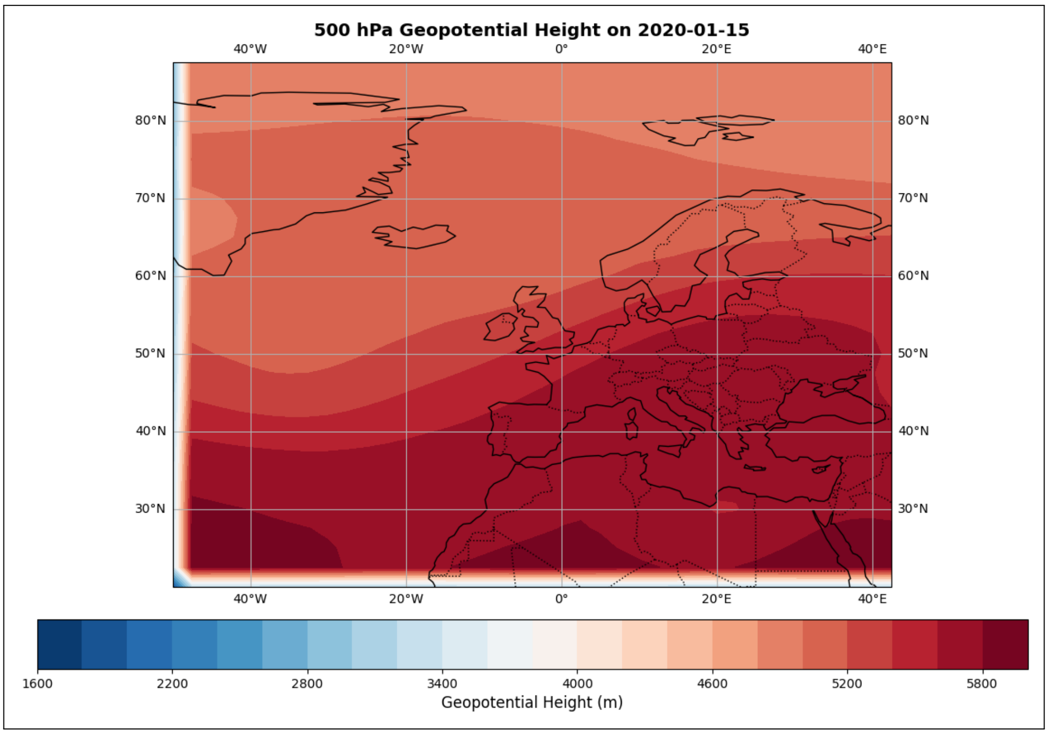

Figure 1: Z500 geopotential height on January 15, 2020. Red areas indicate high pressure while blue areas show low pressure.

6.2 Precipitation Distribution

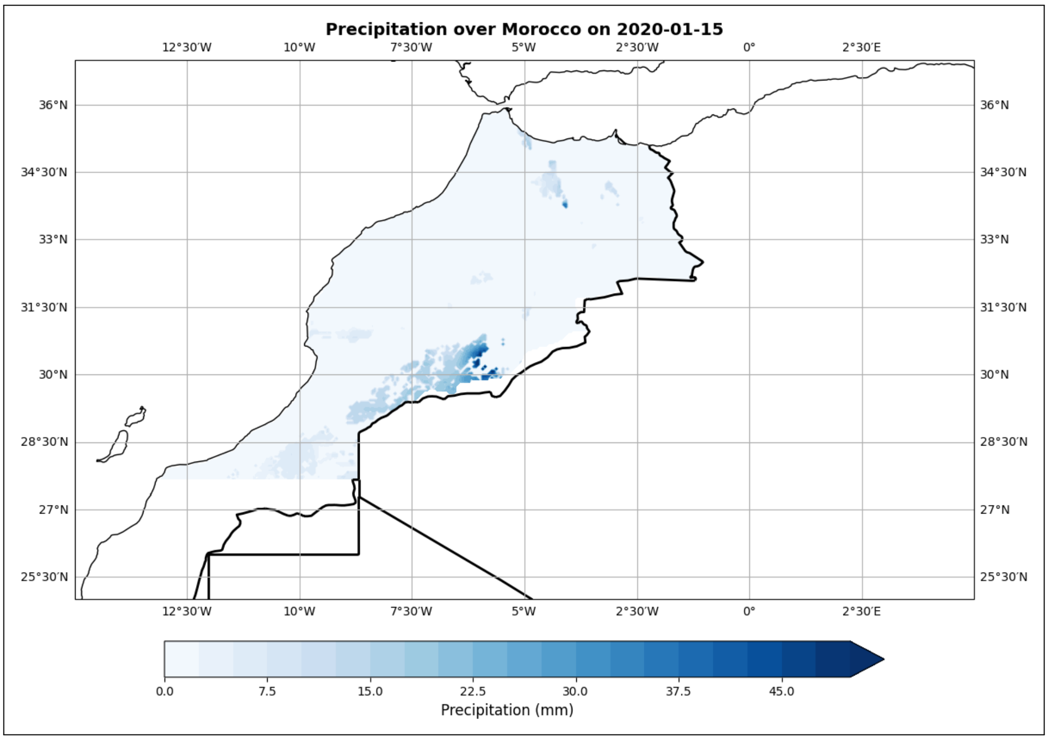

Figure 2: Precipitation over Morocco on January 15, 2020. Dark blue indicates heavy rainfall while white shows little to no precipitation.

7 Insights Gained

7.1 Data Characteristics

Z500 Data:

- Smooth spatial patterns (large-scale circulation)
- Temporal persistence (patterns last several days)
- Clear seasonal cycle (higher values in summer, lower in winter)

Precipitation Data:

- High spatial variability (local effects)
- Intermittent events (many days with zero rainfall)
- Strong seasonal signal (wet winters, dry summers)

8 Conclusions

Summary

During this phase, I developed a comprehensive data exploration tutorial, mastered NetCDF data handling, created professional geospatial visualizations, and gained understanding of the relationship between Z500 and precipitation.
Next, I will focus on gaining a thorough understanding of the CMM-VAE method and its implementation.

Attachments

1. **explore_climate_data.ipynb** – Complete interactive tutorial notebook