Evolution of Europe’s Climate Monitoring Infrastructure

Anthony Gumucio, Dustin Enriquez, Jennifer Mazas,
Kathlyne Alilain, Nina Norseweather
Department of Information Systems, California State University
Los Angeles
E-mails: kalilai2@caltstatela.edu
jmazas@calstatela.edu
agumuci@calstatela.edu
denriq18@calstatela.edu

Abstract: This project explores how big data tools like Hadoop and Hive can be used to analyze the temporal-spatial characteristics of maximum temperature data collection across Europe. Using the E-OBS stations_info_tx_v31.0e.txt dataset, we conducted spatial queries and visualized weather station metadata, including location, elevation, and operational timelines. We built external Hive tables on IBM BigInsights, ran SQL-based queries to extract trends, and exported the data for interactive visualization in Excel 3D Maps. Although the dataset did not include actual temperature readings, our analysis highlights the expansion of infrastructure for climate data collection over time, providing a foundation for future integration with gridded temperature datasets.

1. Introduction

The intensification of global climate change has led researchers to increasingly rely on big data technologies to uncover tempo-spatial climate trends. Our project focused on analyzing Europe’s shifting maximum temperatures using the E-OBS dataset, a gridded observational dataset provided by the Copernicus Climate Change Service (C3S). We utilized Hadoop and Hive to process a large metadata file, stations_info_tx_v31.0e.txt, containing thousands of European weather station records. Our objective was to extract regional and temporal insights from this dataset, ultimately presenting these patterns through data visualizations in Excel 3D Maps. This paper presents our methodology, key findings, and how big data platforms enabled a deeper understanding of temperature changes across Europe.

2. Related Work

Previous studies have explored how temperatures across Europe have changed over time using complex statistical models and detailed climate datasets such as the use of generalized extreme value (GEV) distributions in ASCMO to detect shifts in maximum daily temperatures. [1]the European Environment Agency analyzed trends in average temperatures using long-term data from the European Climate Assessment & Dataset.[2] Another study by Cornes et al. improved temperature and rainfall estimates by creating an ensemble version of the E-OBS dataset, helping to better understand uncertainty in climate data.[3]

In contrast, our project takes a different approach by using big data tools and cloud computing to examine how and where climate data is being collected. Using Hadoop and Hive on IBM BigInsights, we analyzed the metadata from the *E-OBS stations_info_tx_v31.0e.txt* file, focusing on station locations, elevation, and how long each station has been operating. We ran SQL queries on large datasets and created 3D visualizations in Excel to show how weather station infrastructure has grown over time. While past studies focused on analyzing temperature trends, our work is unique in that it focuses on the data collection process itself and demonstrates how cloud based big data tools can manage and analyze large climate datasets efficiently.

3. Background/Existing Work

Our work is based on past research on European climate trends, which primarily used statistical models and gridded datasets like ECA&D and E-OBS to study climate changes. ECA&D provides a comprehensive archive of daily meteorological data from thousands of weather stations across Europe, offering critical insights into long-term climate patterns. Building on this, the E-OBS dataset delivers high-resolution gridded data on temperature, precipitation, and other climate variables, enabling spatial analysis and model validation. These datasets support a wide range of applications in climate monitoring, impact assessments, and policy planning, making them essential tools for understanding climate variability and change across Europe.

4. Data Source & Platform Specifications

This project utilizes the E_OBS gridded dataset maintained by the European Climate Assessment & Dataset (ECA&D) team. E_OBS provides highly accurate daily data for critical weather elements, including temperature, precipitation, pressure, wind speed, humidity, and radiation across Europe from 1950 to today.

A key aspect of the E_OBS dataset includes the use of a Digital Elevation Model (GTOP030) for elevation corrections to ensure statistical validity. The latest version, 31.0e (March 2025), expands coverage by including updates for Spain, Italy, and Poland, and provides continuous monthly, half-yearly, and yearly updates for previously established countries. The E-OBSpre1950 dataset offers valuable long-term climatic context despite limitations near domain edges and regions of sparse data.

This carefully curated dataset serves as a foundational source for evaluating long-term climate variability, validating climate models, and supporting climate impact assessments across Europe.

Below are the specifications of the dataset used, along with the platform specifications.

Table 1. Data Specifications

Dataset (stations_info_tx_v31.0e.txt)	Size (Total 4.6GB)

Table 2. H/W Specification

Platform	IBM Bluemix BigInsights
CPU Speed	2.4 GHz
Number of CPU Cores	4 cores per node
Number of Nodes	3 nodes
Total Memory Size	24 GB (8 GB per node x 3 nodes)

5. Implementation Flowchart

The implementation flowchart below (Figure 1) outlines the implementation process of the project, beginning with uploading the E-OBS metadata file into Hadoop's HDFS. Then we created an external Hive table using Beeline to structure and query the data. Key analyses included identifying station counts per country, elevation trends, and deployment timelines. Results were exported as CSV files and transferred to a local machine for visualization in Excel 3D Maps, where date formatting and animations illustrated how Europe's weather station network evolved over time.

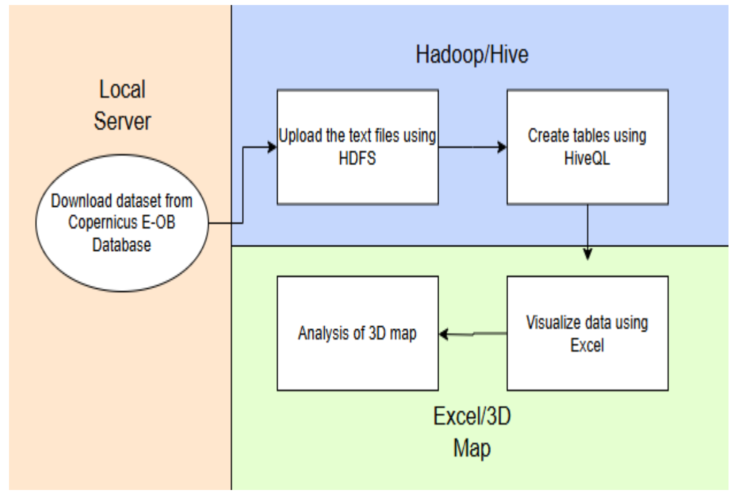

Figure 1: Implementation Flowchart

6. Our Work

The dataset is sourced from the official European Climate Assessment and Dataset download page. Our data preparation began with loading the metadata file into Hadoop's HDFS. Using Beeline, we created a Hive external table and validated its structure. We ran tempo-spatial queries to:

- Determine the number of weather stations per country
- Identify which countries had the highest average elevations
- Analyze station installation trends by decade
- Locate stations with long-term historical records (pre-1950 to present)

We then exported the cleaned data into CSV format and downloaded it to a personal machine using scp. In Excel, we created 3D Maps to visualize station density, elevation, and installation patterns over time. We converted start and end dates into proper date formats using Excel functions, such as DATEVALUE(), and built time-based animations that show the regional development of the weather station network.

Although the file we analyzed did not include temperature readings, our visualizations help contextualize where temperature data is being collected, its geographical reach, and how collection infrastructure has evolved. In future iterations, this metadata can be joined with the full gridded NetCDF temperature dataset to assess regional warming trends directly.

4.1 Graphics

country	avg_elevation
ARMENIA	1955.075
KYRGYZSTAN	1398.5
IRAN, ISLAMIC REPUBLIC OF	1355.142857142857
BOSNIA AND HERZEGOVINA	1348.5
SWITZERLAND	1018.972972972973
TAJIKISTAN	934.0
NORTH MACEDONIA	779.5
AUSTRIA	758.0833333333334
SAUDI ARABIA	689.0
TÜRKIYE	568.6090909090908

10 rows selected (8.929 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai>

Figure 2: Top 10 Countries with the Highest Average Elevations of Weather Stations

This Hive query ranks countries by the average elevation of their weather stations. Armenia, Kyrgyzstan, and Iran top the list, indicating a higher concentration of stations in mountainous or high-altitude regions within these countries.

country	num_stations
GERMANY	1013
RUSSIAN FEDERATION	318
SWEDEN	275
ITALY	204
NORWAY	183

5 rows selected (8.858 seconds)
0: jdbc:hive2://bigdaiun0.sub03291929060.trai>

Figure 3: Countries with the Most Weather Stations

Germany has the highest number of weather stations, at 1013. This reflects the broader spatial coverage of the temperature observation network in these regions, supporting detailed regional climate analysis.

station_id	country	elevation
2073	ITALY	3480.0
2941	ARMENIA	3223.0
15	AUSTRIA	3109.0
58	GERMANY	2964.0
878	ITALY	2600.0

Figure 4: Top 5 Highest-Elevation Weather Stations

Italy hosts two of the five highest-elevation weather stations. These high-altitude stations are crucial for capturing temperature extremes in mountainous regions.

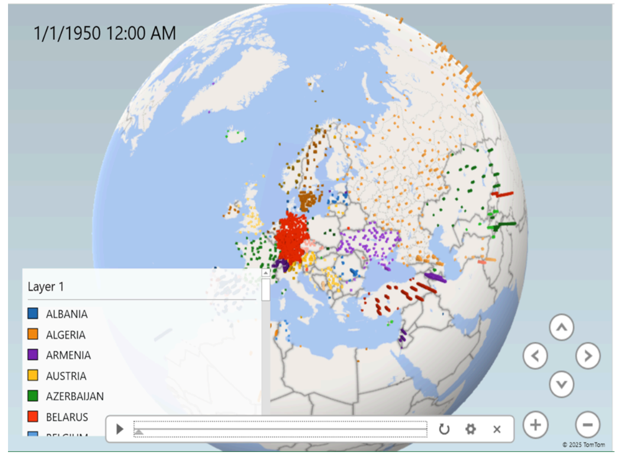

Figure 5: Animated 3D Map of Station Deployment by Country

This Excel 3D map displays the spatial and temporal distribution of weather stations across Europe, colored by country. The timeline indicates that station deployment commenced in the 1950s and continued to expand over time, particularly across Central and Eastern Europe.

7. Conclusion

This project demonstrates the power of Hadoop and Hive in managing and analyzing large-scale climate datasets. By focusing on the metadata structure of the E-OBS maximum temperature dataset, we laid the groundwork for future integrations and data analysis. We identified spatial-temporal patterns in the deployment of European weather stations. Our work provides insight into the underlying structure that supports climate data collection, setting the stage for an extended analysis of climate change impacts across different regions.

References

[1]Auld, G., Hegerl, G. C., & Papastathopoulos, I. (2023). Changes in the distribution of annual maximum temperatures in Europe. *Advances in Statistical Climatology, Meteorology and Oceanography, 9*, 45–66. https://doi.org/10.5194/ascmo-9-45-2023

[2]European Environment Agency. (2023). *Observed annual mean temperature trend (1960–2023)*. https://www.eea.europa.eu/en/analysis/maps-and-charts/observed-annual-mean-temperature-trend-3

[3]Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). *An ensemble version of the E-OBS temperature and precipitation datasets*. Journal of Geophysical Research: Atmospheres, 123(17), 9391–9409. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2017JD028200