Anomaly Detection on Time Series Graphs
A.A 2024/2025

Riccardo Benedetti, Thomas De Palma, Mattia Pistoia

Abstract

In this paper we provide an analysis of the precipitation features in the region of Central Italy, based on the data from the NASA IMERG dataset. Our research comprehends a data obtained during a period of 6 years (2018-2023) in an area between longitudes 10°E to 15°E and latitudes of 41° N to 44° N. To describe the temporal characteristics of precipitations over the area we considered 9000 measurement points, allowing our study to provide a valuable resource for meteorological studies and climate analysis. Our methodology uses both temporal and spatial analysis with a statistical approach to identify trends and patterns in the precipitation registered data. The results show a decreasing frequency in precipitation from 2018 to 2023, with recurrent patterns, including areas of consistently higher rainfall and regions with significant temporal variability.

1 Introduction

The study of precipitation regimes is a crucial part of climate science today. It tells us how climate changes and develops. Earlier research [4] has shown that analysis of precipitation has a great bearing on the understanding of the behavior of local climates and their effects on the environment. The recently developed capabilities within satellite remote sensing, in particular with the NASA GPM mission, has revolutionized the study of rainfall patterns on scales of temporality and spatiality. While this technological leap forward has enabled high-resolution, multi-sensor precipitation data collection, there are also new challenges in effectively processing and analyzing such vast datasets to extract meaningful patterns and insights from them.
The objective of this paper is to study the precipitation pattern in Central Italy using the dataset provided by NASA's Integrated Multi-satellitE Retrievals for GPM (IMERG). This is a particularly interesting case study because of its intricate landforms and valuable position where continental and Mediterranean climates meet. Because of this, it is a valuable place to observe how various weather patterns intersect.
A thorough basis for our analysis is provided by the IMERG dataset, which aggregates observations from several satellites in the GPM constellation. This dataset provides high-resolution precipitation estimates for our study region (10°E-15°E longitude, 41°N-44°N latitude) over a six-year period (2018-2023) at both spatial ($0.1^\circ \times 0.1^\circ$ grid spacing) and temporal (30-minute intervals) scales. The dataset's fine-grained resolution allows for a thorough examination of precipitation dynamics across a range of spatial and temporal scales, and the integration of multiple satellite observations improves the dataset's reliability. We use these features to examine both short-term and long-term trends in precipitation, paying special attention to how precipitation interacts with Central Italy's intricate topography. The dataset is especially useful for researching the junction of the continental and Mediterranean climate systems due to its complex orography and strategic position at the confluence of different climate systems.
Our study's temporal scope allows for a thorough examination of potential longer-term precipitation trends as well as interannual variations, both of which are essential for comprehending regional climate patterns. In order to detect possible signs of climate change in the area, our method uses statistical techniques to measure changes in the temporal evolution and spatial distribution of precipitation patterns. This integrated analysis framework advances our knowledge of the dynamics of the Mediterranean climate by offering insightful information about the intricate relationship between precipitation and topography in Central Italy.
The importance of the research outgrows the boundaries of regional weather forecasting. In fact, with global warming continuing to alter precipitation patterns worldwide, the more detailed historical information there is on precipitation, the better future trends and anomalies can be predicted. Our findings also contribute to a broader body of knowledge in climate science while offering practical insights into regional meteorological applications. This study also signifies the full value of satellite-based multi-sensor data in environmental monitoring and climate research.

This paper is organized as follows: Sect. 2 presents the IMERG dataset, along with a detailed description of the spatial and temporal characteristics; Sect. 3 describes our approach to processing and analyzing data. Sect. 4 presents results related to rainfall distribution patterns and their temporal trends. Finally, in Sect. 5, implications of these findings are discussed together with some possible limitations of the study and ideas for further research.

2 Dataset Contents

2.1 IMERG Dataset Overview

The IMERG (Integrated Multi-satellitE Retrievals for GPM) system of NASA's Global Precipitation Measurement (GPM) mission is the source of the dataset [2]. In order to maximize accuracy across various climatic zones, this level 2 product synthesizes data from several satellites in the GPM constellation using retrieval algorithms.

2.2 Data Quality and Processing

The dataset incorporates both delayed-mode and near-real-time products that can be accessed via NASA's Giovanni portal [3], and it employs stringent quality control procedures:

- **Advanced Calibration:** Multi-stage sensor calibration process with continuous validation
- **Error Management:** Comprehensive error descriptions based on retrieval schemes
- **Quality Assurance:** Strict calibration procedures including:
  - Sensor-specific calibration coefficients
  - Cross-validation with ground-based measurements
  - Systematic bias correction
  - Uncertainty quantification for each measurement

In-depth studies of extreme weather events, seasonal cycles, and year-to-year variability are supported by this calibration framework, which guarantees accurate estimates even in complex meteorological scenarios.

The Framework for Calibration and Validation The IMERG system uses a thorough framework for calibration and validation:

- **Reality Checking:** Comparing weather radar data and rain gauge networks
- **Systematic Error Correction:** application of bias correction algorithms tailored to:
  - Various forms of precipitation (stratiform, convective)
  - Features of the regional climate
  - Seasonal fluctuations
- **Quality Metrics:**
  - Evaluation of relative bias
  - Calculation of root mean square error

- Probability of detection
- Analysis of the false alarm ratio

For applications involving climate analysis, this thorough validation guarantees excellent data quality and dependability.

2.3 Spatial and Temporal Coverage

Our dataset includes samples that cross the Central Italian region, which is distinguished by its diverse topography and distinct weather patterns. The samples span from 10E to 15E longitude and 41N to 44N latitude. With temporal granularity at 30-minute intervals and grid spacings of roughly 0.1×0.1, the IMERG dataset offers high-resolution precipitation estimates. In-depth analyses of both episodic and gradual changes in precipitation are made possible by this extended temporal record, which covers the years 2018–2023, which captures both seasonal variability and longer-term trends.

2.4 Data Characteristics

Additionally, the IMERG dataset—as accessed via NASA's [3] portal, these capabilities by offering enhanced spatial resolution with grid spacings of approximately $0.1^\circ \times 0.1^\circ$ and refined temporal granularity at 30-minute intervals. The dataset encapsulates both near-real-time and delayed-mode products, ensuring that the retrieved estimates remain reliable even under complex meteorological scenarios. This integrated approach facilitates a comprehensive exploration of precipitation patterns, allowing for in-depth investigations into extreme weather events, seasonal cycles, and year-to-year variability. Data quality is guaranteed by strict calibration processes and quality control. The data set contains near-real-time and delayed products, and detailed error descriptions based on different retrieval schemes and sensor calibration updates.

Table 1: IMERG Dataset Characteristics

Characteristic	**Description**
Temporal Coverage	2018-2023 (6 years)
Spatial Domain	10°E-15°E longitude, 41°N-44°N latitude
Spatial Resolution	$0.1^\circ \times 0.1^\circ$ grid spacing
Temporal Resolution	30-minute intervals
Number of Observations	9,000 (1,500 per year)
Data Format	GeoTIFF files
Variables Recorded	Precipitation (mm)

2.5 Data Access and Processing

The data was accessed through NASA's Giovanni portal [3], a data visualization and analysis portal. The system supported easy subsetting and processing of the data over our study region. The portal's capabilities are:

- Simple visualization and extraction of data

- Combination of complementary geophysical parameters
- Advanced quality control and validation tools

This data structure forms a sound foundation for precipitation pattern analysis, a valuable tool for in-depth analysis of the spatial distribution and temporal evolution of rainfall in the study region.

3 Data Analysis and Interpretation

We took an extensive IMERG dataset of 9,000 rainfall observations in Central Italy, which were gathered from 2018 to 2023. Our dataset analysis consists of different data types: geographic coordinates (10° E-15° E longitude, 41° N-44° N latitude), rainfall amounts (mm), time stamps (start_date, end_date), and source file names. During preprocessing, we removed all invalid or unavailable data and sorted the data points in time. We structured the dataset in a grid. This gave us 1,500 observations per year, which was enough to keep time balanced and still have good detail for the study area.

Using our close analysis, we were able to determine clear patterns both in the location and timing of local rainfall. Our method combines the study of trends over time with sensitivity to patterns in space so that we could better understand how rainfall patterns changed over the course of the study.

3.1 Temporal Evolution Analysis

The temporal analysis of precipitation patterns shows a significant variation in our year to year analysis. In our picture below we also include the Standard Deviation, which measures the amount of variation in our precipitation values for the mean.

- **Annual Patterns:** Mean annual precipitation shows variations across the study period (2018-2023), with observable differences between years. The highest precipitation mean registered was in 2018, with a value of 2.8, which was followed by a notable decrease of 1.8mm in 2020. A higher valued was recorded in 2021 (2.5mm) followed by a decline to 1.9mm in 2023 Figure 1. We calculate the mean annual precipitation with this formula:

  $$\bar{x}_y = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  Where $\bar{x}_y$ is the mean for the year **y**, n is the **number of observations**, and x_i the **precipitation values**.

- **Standard Deviation:** The annual standard deviation, calculated as:

  $$\sigma_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_y)^2}$$

  where σ_y is the standard deviation for year y, reflects varying degrees of precipitation variability, starting at 0.38 mm in 2018 and increasing to a high of 0.65 mm in 2021, then settling back to 0.52 mm in 2023. Increasing trend (Figure 2) in standard deviation reflects increased variability in precipitation in later years, as would be easily observed in the 2021 high.

- **Temporal Trend**: The precipitation trend showed a clear decrease over the period of study Figure 2. It began with a mean level of 2.8mm in 2018, then the rainfall levels fell to 1.8mm in 2020. It slightly picked up to 2.5mm in 2021 before falling again to 1.9mm in 2023. This overall decrease indicates a substantial change in the rainfall pattern of the region, with mean rainfall going down by about 32%.

  $$\Delta P = \frac{P_{2023} - P_{2018}}{P_{2018}} \times 100\% = -32\%$$

  with P_x being the **annual precipitation mean**.

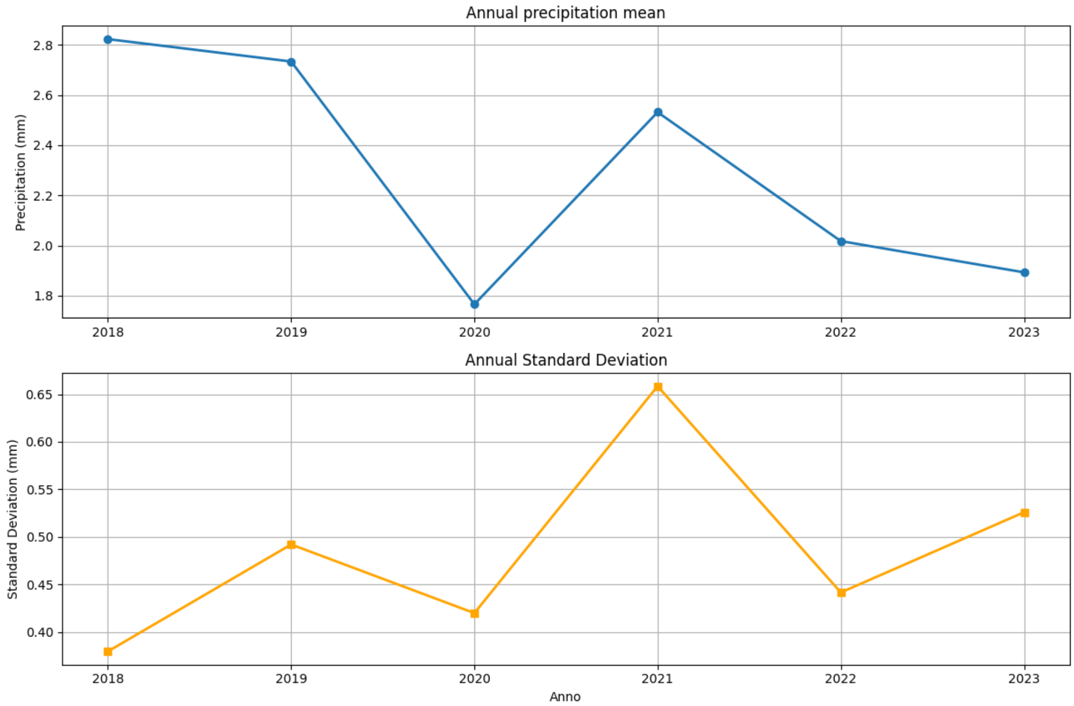

Figure 1: Annual Precipitation Mean Mesurements

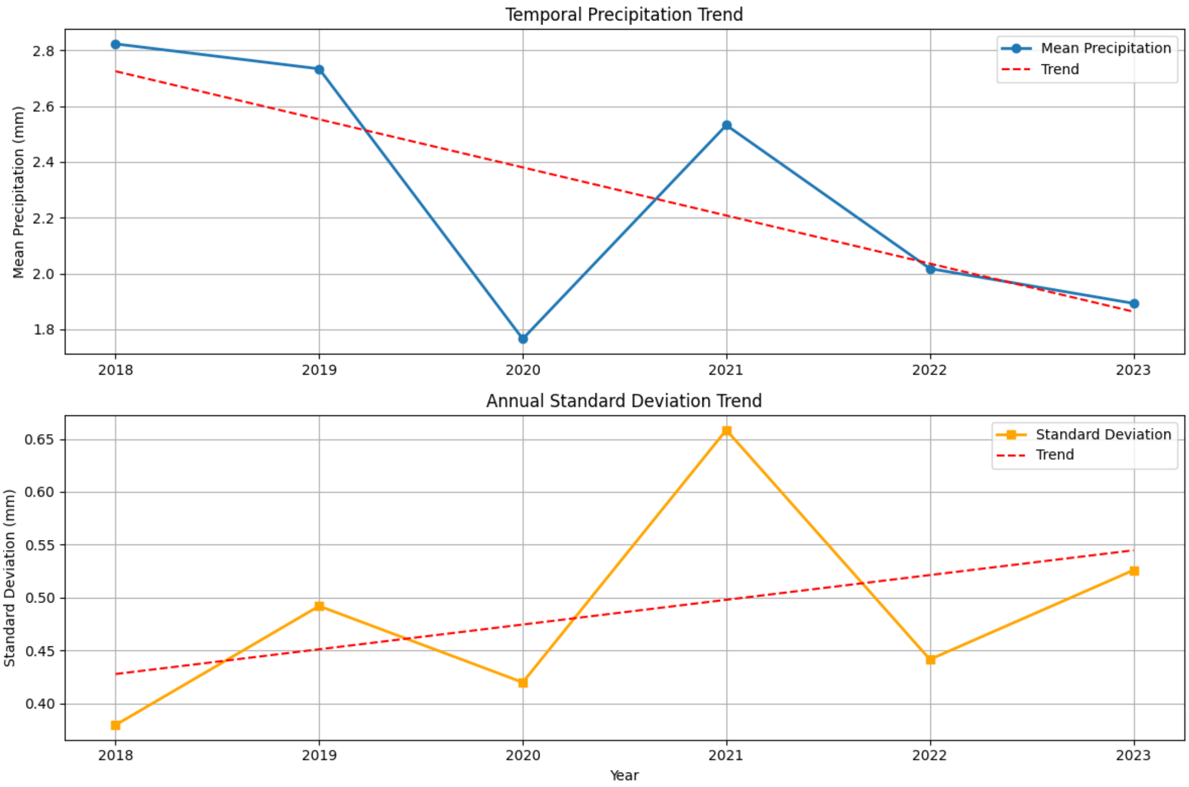

Figure 2: Temporal Trend Analysis of The Precipitation Mean

3.2 Spatial Analysis

We provide an overview of the spatial patterns of precipitation throughout the 2018–2023 research period, with an annual precipitation distribution analysis. The Study domain (D), is defined as it follows:

$$D = \{(lon, lat) \in R^2 : 10E \leq lon \leq 15E, 41N \leq lat \leq 44N\}$$

. The annual precipitation distribution analysis shows values going from approximately 0.75mm to over 3.0mm (Figure 3).

- **Spatial Consistency:**The data shows persistent patterns in the precipitation patterns across all of the region. Higher precipitation are consistently observed in the central-eastern portions of the area of study. with a gradient showing higher precipitation values (2.8-3.2mm) in particularly dense regions.

- **Temporal Evolution:** These spatial patterns' temporal evolution reveals significant annual variations. Although the spatial distribution was fairly uniform in 2018, the following years showed increasing heterogeneity, with 2020 and 2022 showing the most. This increasing spatial variability is consistent with the temporal trends found in our earlier analysis, indicating that there may be a connection between the region's increased spatial heterogeneity and declining mean precipitation.

- **Geographic Features:** The precipitation patterns appear to follow consistent geographic features, suggesting a strong relationship between topography and precipitation distribution. This is particularly evident in the higher precipitation zones that persist across multiple years.

- **Topographic Influence:** The three-dimensional view (Figure 4) further emphasizes the interaction between topography and precipitation patterns. There is an obvious diagonal trend in precipitation distribution throughout with higher values residing in areas with increasing latitude and east longitude. This visualization obviously demonstrates how the complex terrain in Central Italy controls precipitation accumulation, where elevation plays a key role in locally controlling the intensity of precipitation.

The spatial variety in precipitation intensity is adequately highlighted by the color scales in both images. Precipitation patterns across various elevations and geographic regions are clearly represented by the 3D visualization, which uses a complete color scheme from purple (lower values) to yellow (higher values), in contrast to the yellow-to-red gradient used in the annual distribution maps.

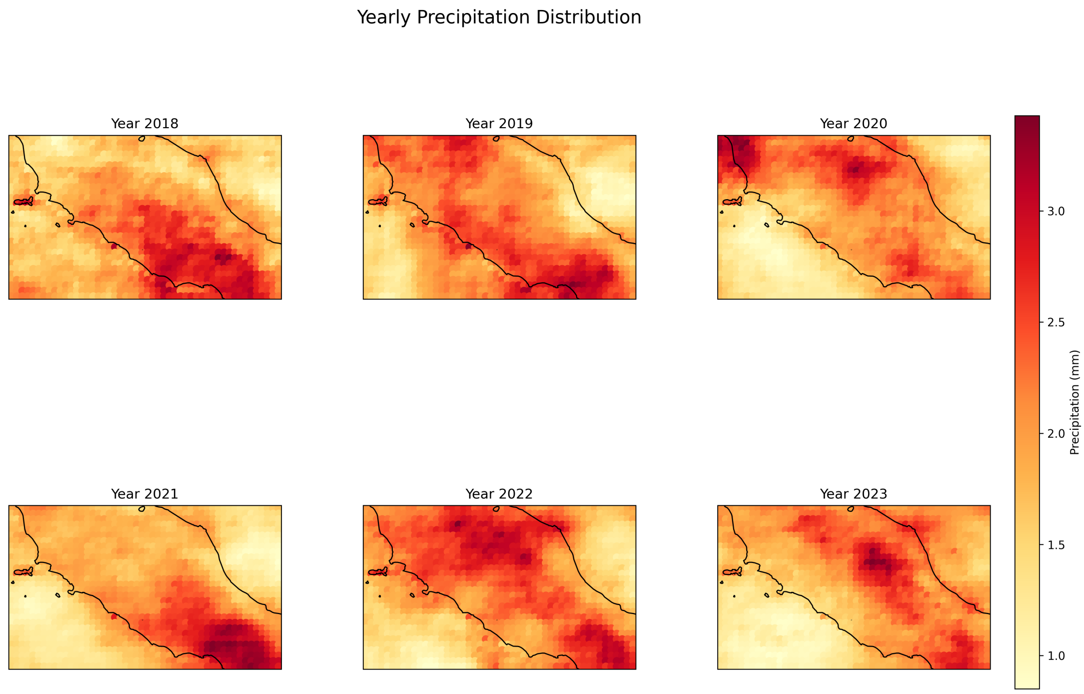

Figure 3: Spatial Distribution

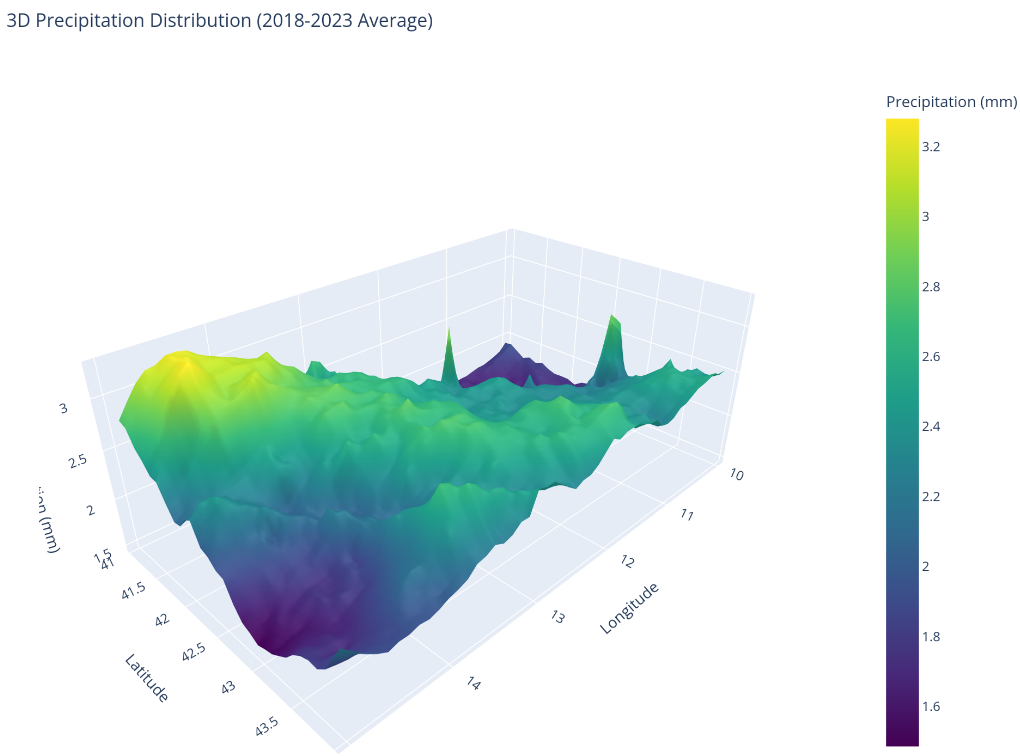

Figure 4: 3d Spatial Distribution

4 Results

Our analysis of precipitation patterns in Central Italy revealed significant temporal and spatial variations. The temporal analysis showed a clear declining trend in precipitation levels, characterized by notable year-to-year variability. Spatially, we identified consistent patterns strongly influenced by regional topography, with higher precipitation consistently observed in central-eastern areas. The combined temporal-spatial analysis suggests an evolving precipitation regime that could have significant implications for regional climate patterns.

5 Future Work

Several directions for future research emerge from this study:

- **Extended Temporal Analysis:** Expanding the study period beyond six years could reveal longer-term trends and cyclical patterns in precipitation.
- **Comparative Analysis:** Extending the study to neighboring regions could help understand broader precipitation patterns across the Mediterranean.
- **Anomaly Detection Systems:** Implementing an Anomaly Detection System could be useful in detecting unusual precipitation patterns and identifying temporal anomalies based on the spatial and temporal data of our dataset. This may find usage in the machine learning field to predict extreme precipitation events and weather alerts [1].

References

[1] *AnomalyDetection*. https://www.vmware.com/topics/anomaly-detection.

[2] *Data Processing Levels*. https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/data-processing-levels.

[3] *Giovanni Website*. https://giovanni.gsfc.nasa.gov/giovanni.

[4] S. Scarsoglio, F. Laio, and L. Ridolfi. "Climate Dynamics: A Network-Based Approach for the Analysis of Global Precipitation". In: *PLOS ONE* 8.8 (2013). DOI: 10.1371/journal.pone.0071129. URL: https://doi.org/10.1371/journal.pone.0071129.