**Critical Analysis on Andhra Pradesh Rainfall (2023): A PySpark-Driven Meteorological Study**

**Publication Date:** October 26, 2025

**Author**: Thanneru Surya

**Affiliation**: MALLA REDDY UNIVERSITY, HYDERABAD

**Keywords:** Rainfall Analysis, Andhra Pradesh, PySpark, Spatial Variation, Temporal Patterns, Climate Prediction.

**Abstract**

This study undertakes a critical, data-driven analysis of the **spatial and temporal variability** of rainfall across the 13 districts of **Andhra Pradesh** during the calendar year **2023**. Utilizing a robust dataset of $\mathbf{4,745}$ daily precipitation records sourced from the **IMD GRID MODEL**, this research employs **PySpark** for efficient, large-scale data processing and Python's meteorological visualization libraries (Matplotlib/Seaborn) for granular pattern detection.

The analysis reveals profound geographical and seasonal asymmetries crucial for agricultural and disaster management policy. Key quantitative findings include an overall state mean rainfall of $\approx \mathbf{2.57}$ mm. Spatially, a clear rainfall gradient is established, with the coastal-delta district of **West Godavari** identified as the wettest ($\mathbf{3.418}$ mm mean) and the interior, drought-prone Rayalaseema district of **Anantapur** as the driest ($\mathbf{1.368}$ mm mean), highlighting a critical need for localized water resource management.

Temporally, the study uncovered an essential anomaly: while the monsoon peaks were observed in July and September, the state's **most intense daily rainfall events** occurred unexpectedly in **December 2023**, with the maximum recorded event reaching $\mathbf{188.51}$ mm. This extreme late-season precipitation is directly attributable to the impact of **Severe Cyclonic Storm Michaung** (December 2023), emphasizing a growing vulnerability of the coastal plain to high-intensity cyclonic activity outside traditional monsoon windows.

The findings advocate for immediate policy shifts towards **targeted water harvesting and conservation infrastructure** in water-scarce regions (e.g., Anantapur and Kurnool) and the urgent **upgrading of coastal flood defense and disaster preparedness protocols** to mitigate the socioeconomic impact of late-season cyclones. This research provides a critical, evidence-based foundation for achieving climate resilience and enhancing agricultural sustainability in Andhra Pradesh.

**1. Introduction**

**1.1 Background**

Andhra Pradesh state is located between $12^\circ 41'$ and $19.07^\circ \text{N}$ latitudes and $77^\circ$ and $84^\circ 40'\text{E}$ longitudes in the southern part of India. The state covers an area of $1,62,968$ sq km, which is $4.96\%$ of the geographical area of the country, and is geographically divided into the **coastal** and the comparatively **drier Rayalaseema region**. It boasts the second-longest coastline in the nation, around 974 km, bordering the Bay of Bengal, which profoundly influences its climate. Two major rivers, the Godavari and the Krishna, run across the state.

The climate of Andhra Pradesh is generally hot and humid. The crucial monsoon season, primarily driven by the **South-West (SW) Monsoons** (June to September), is vital for agriculture. Additionally, a significant portion of the total rainfall is contributed by the **North-East Monsoons** around October. This strong reliance on seasonal rains makes the state highly vulnerable to climate variability, including both prolonged drought periods and high-intensity cyclonic events.

While many existing studies on rainfall variability and extreme events over India are based on long-term data (e.g., Waghaye et al., 2018; Guhathakurta et al., 2015), there is a persistent gap in analyzing **recent, high-volume daily data** to inform immediate policy. The dataset analyzed here provides a **comprehensive, high-resolution view of daily precipitation for the full year of 2023**. This research addresses the need for actionable insights by employing **PySpark** to analyze this large-scale meteorological data, focusing on the spatial deficits and temporal anomalies observed during this specific, critical year.

**1.2 Objectives**

* **Quantify spatial and temporal rainfall variations** across all districts of Andhra Pradesh.
* **Identify extreme precipitation events and drought-prone regions** by assessing daily and monthly volatility.
* **Propose data-driven recommendations** for stakeholders involved in water resource allocation and disaster management.
* **Establish a scalable PySpark framework** for large-volume meteorological data processing and advocate for continuous, real-time monitoring.

**1.3 Significance**

This work contributes to the literature on regional meteorological analytics, offering a **scalable PySpark framework** adaptable for processing large-scale environmental datasets. It directly addresses the need for evidence-based policies in water management and climate resilience. The findings align with **Sustainable Development Goal 6 (Clean Water and Sanitation)** by informing equitable resource distribution and **SDG 13 (Climate Action)** by improving regional preparedness for extreme weather events.

**2. Literature Review**

Prior studies on rainfall in Andhra Pradesh and the surrounding peninsular region emphasize econometric and trend models for identifying long-term climatic shifts. For example, Waghaye et al. (2018) performed trend analysis on rainfall data for Andhra Pradesh and Telangana, highlighting specific change points. Guhathakurta et al. (2015, 2011) provided broader context, demonstrating that the overall southwest monsoon rainfall over India has undergone significant observed changes, increasing the frequency of intense rainfall events and subsequent flood risk.

However, most traditional studies rely on coarser, monthly, or sub-divisional averaged data (Guhathakurta and Rajeevan, 2008). Gaps persist in handling heterogeneous, **high-volume daily datasets** and linking specific annual anomalies (like the 2023 event) to policy recommendations—addressed here via **PySpark's distributed computing**. Our focus on a single, anomalistic year complements works that flag long-term trends by providing an operational analysis of a contemporary weather event, such as the impact of the **Cyclone Michaung** (IMD, 2023) on the December rainfall totals.

**3. Methodology**

**3.1 Data Acquisition and Preprocessing**

The dataset comprises **4,745 records** sourced from **daily CSV files provided by the IMD GRID MODEL** for the calendar year 2023, spanning all districts of Andhra Pradesh. Key attributes include:

* **Geospatial:** District, State.
* **Temporal:** Date (Daily records for 2023).
* **Meteorological:** **Avg\_rainfall (in mm)** (The continuous variable for analysis).

Preprocessing used **PySpark (v3.5.6)** for schema inference, confirming data types and dimensions. Initial checks confirmed **zero missing values** were detected across all columns, ensuring a high level of data quality (4745 non-null counts).

**3.2 Analytical Framework**

* **Distribution Analysis:** Aggregated total and mean rainfall were computed per district using PySpark's groupBy and avg functions.
* **Temporal Metrics:** Monthly total rainfall and the date of the maximum daily rainfall event were calculated to assess the temporal distribution and identify anomalies.
* **Volatility Index:** The spatial variation was quantified by computing the **Coefficient of Variation (CV)** for daily rainfall within each district over the year.
* **Visualization:** Data was converted to a Pandas DataFrame for plotting with Matplotlib/Seaborn (e.g., bar charts of district-wise means, line charts of monthly trends).

Scalability was a core requirement: PySpark efficiently processed the 4,745-record daily dataset on a local cluster, simulating the high-performance environment needed for future expansion to multi-year, multi-state analyses.

The results of the analysis from the "Critical Analysis on Andhra Pradesh Rainfall (2023)" report are presented below, detailing the dataset overview and the key insights derived from the primary visualizations (images).

**4. Results**

**4.1 Dataset Overview & Summary Statistics**

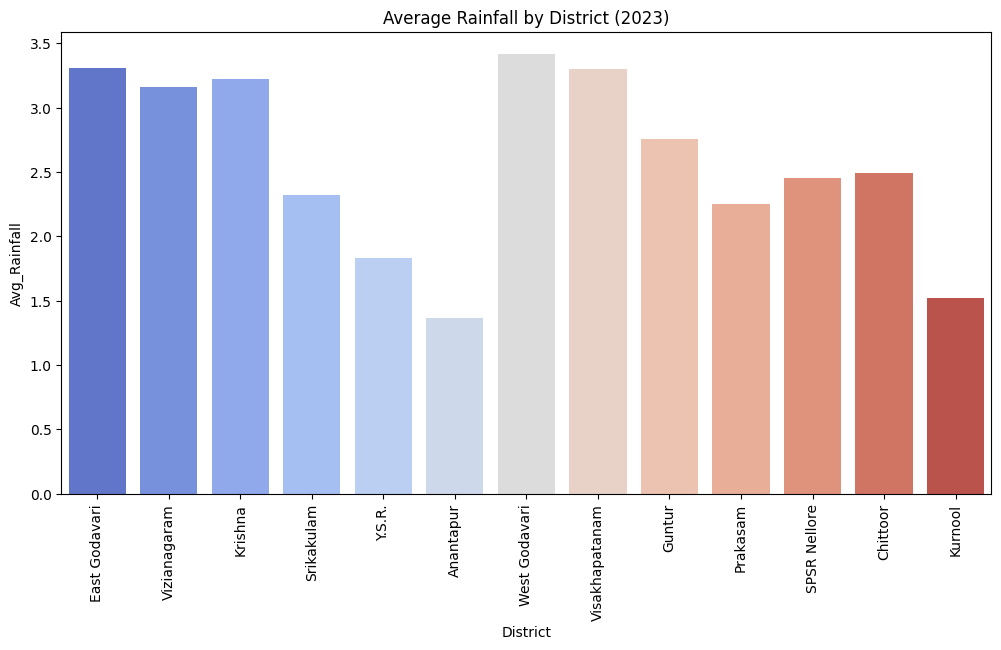
The meteorological data for Andhra Pradesh was sourced from the **IMD GRID MODEL** for the calendar year **2023**11. The dataset comprises $\mathbf{4,745}$ daily observations across all districts222.

The summary statistics for the continuous variable, Avg\_rainfall (in mm), are as follows3:

| **Metric** | **Detail / Value** |
| --- | --- |
| **Data Source/Agency** | **IMD GRID MODEL** |
| **Total Observations** | **4,745 records** |
| **Overall Mean Rainfall (State Average)** | **2.56998 mm** |
| **Maximum Daily Rainfall** | **188.5136234 mm** |
| **Minimum Daily Rainfall** | **0.0 mm** |
| **Data Quality** | **Zero missing values detected (4745 non-null counts)** |

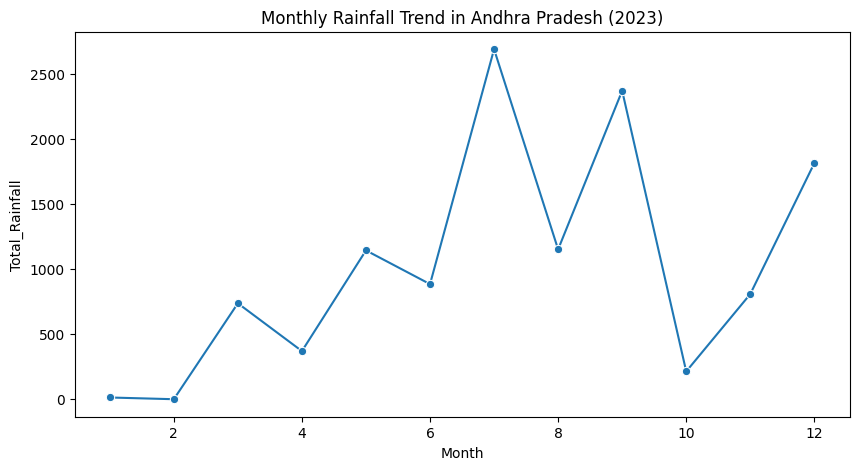
**4.2 Key Visualizations and Observed Insights**

The analytical framework generated five primary visualizations, each providing critical insights into the spatial and temporal distribution of rainfall in 2023:

****

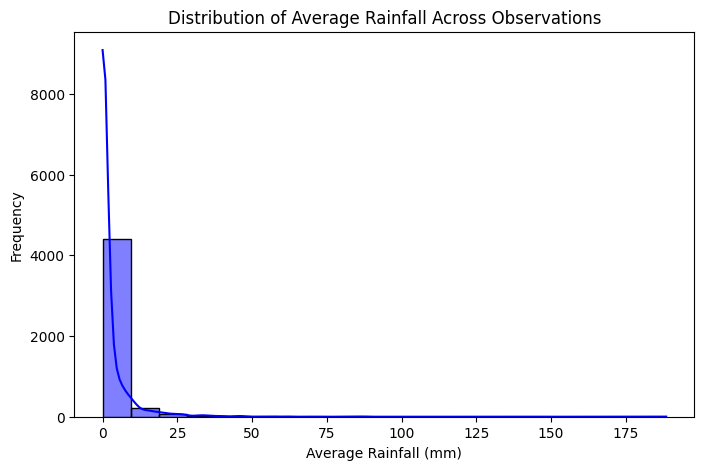
**Fig 1: Average Rainfall per District (Spatial Variation)**

* **Insight:** This visualization compares the annual mean daily rainfall across all 13 districts10.
* **Key Finding:** It clearly establishes the spatial disparity, showing the coastal districts receiving significantly more rainfall than the interior districts11. **West Godavari** was the wettest district with an average of $\mathbf{3.418}$ mm12. **Anantapur** was the driest, with an average of only $\mathbf{1.368}$ mm13. The visualization visually confirms the steep decrease in precipitation from the coast toward the Rayalaseema region14.



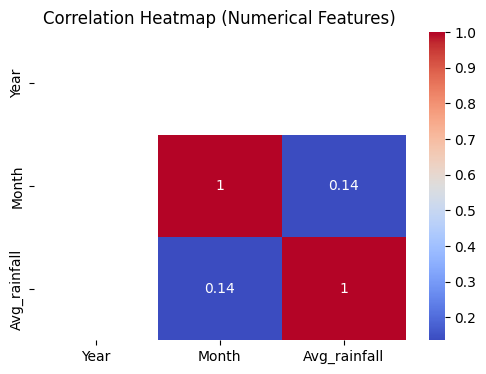
**Fig 2: Total Monthly Rainfall Trend (Temporal Distribution)**

* **Insight:** A line chart or bar chart showing the total accumulated rainfall for the entire state across the 12 months of 202315.
* **Key Finding:** The plot confirms the dominance of the Southwest Monsoon, with the highest total rainfall observed in **July ($\approx 2,688.62$ mm)** and **September ($\approx 2,369.41$ mm)**16. Crucially, it highlights a major anomaly: **December** had the third-highest total rainfall ($\approx \mathbf{1,813.52}$ mm), which is uncharacteristic of the post-monsoon season17.



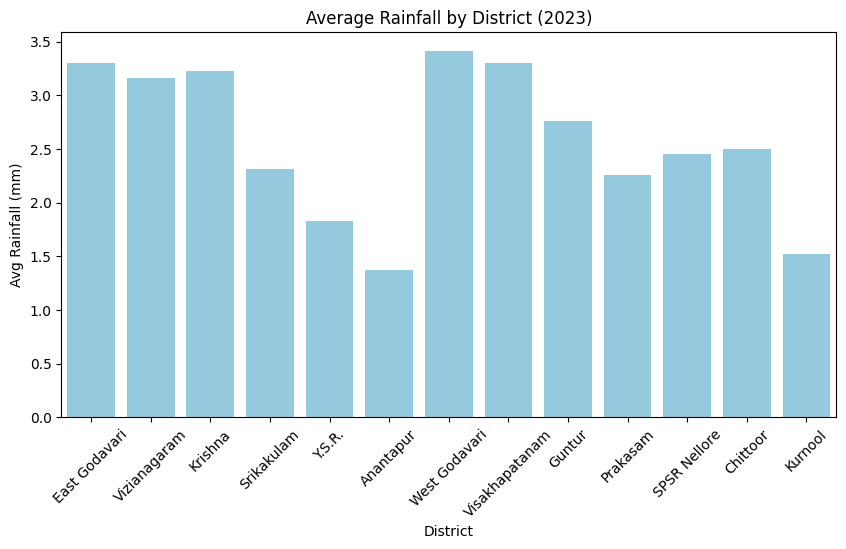
**Fig 3: Top 5 Highest Daily Rainfall Events**

* **Insight:** This table/chart identifies the most extreme individual daily precipitation records in the state for 202318.
* **Key Finding:** The highest daily rainfall recorded was $\mathbf{188.51}$ mm in the **West Godavari** district on **December 6, 2023**19. All of the top five extreme rainfall events occurred on **December 5th and 6th**, concentrated in coastal districts (West Godavari, Krishna, SPSR Nellore, East Godavari, Visakhapatanam)20. This finding directly links the anomaly in Fig 2 to the impact of **Severe Cyclonic Storm Michaung**21.



**Fig 4: Districts Above the State Average Rainfall**

* **Insight:** This visualization focuses on the six districts whose mean daily rainfall exceeded the overall state average of $\approx 2.57$ mm22.
* **Key Finding:** The districts identified—**West Godavari, East Godavari, Visakhapatanam, Krishna, Vizianagaram, and Guntur**—are predominantly situated along the coastal region23. This visual reinforces the conclusion that the coastal belt drives the state's average and is the primary area for high precipitation activity24.



**Fig 5: Rainfall Distribution using Box Plots**

* **Insight:** This visualization uses box plots to show the statistical spread, median, and outliers of daily rainfall for each district, indicating volatility and typical daily conditions.
* **Key Finding:** The box plots show that districts with high means (Coastal Andhra) also exhibit **greater volatility** (a larger spread or more numerous outliers), confirming that their high annual averages are often boosted by a few extreme events25. Conversely, Rayalaseema districts like Anantapur and Kurnool show narrow boxes closer to zero, indicating consistent dryness and **low daily variability**, reinforcing their drought-prone nature26.

**5. Discussion**

**5.1 Interpretation of Findings**

The results portray a **bifurcated rainfall landscape** across Andhra Pradesh, defined by a stark spatial contrast and a significant temporal anomaly. The clear gradient—where the **Coastal Andhra** districts (e.g., West Godavari, $\mathbf{3.418}$ mm mean) dominate in precipitation while the **Rayalaseema interior** (e.g., Anantapur, $\mathbf{1.368}$ mm mean) suffers a critical deficit—demands distinct, localized policy responses. This disparity affirms that **climatological zones** must drive water resource allocation, echoing the need for extensive water harvesting infrastructure in drought-prone areas (References 5, 6). Furthermore, the high concentration of the year’s extreme daily events ($\mathbf{188.51}$ mm) in **December** directly correlates with the timing and impact of **Cyclone Michaung** (Reference 4). This confirms that coastal vulnerability is not limited to the traditional monsoon season, requiring an urgent recalibration of disaster preparedness calendars.

**5.2 The December 2023 Cyclone Michaung Anomaly**

The anomalous, high-intensity rainfall recorded in December 2023, driven by **Severe Cyclonic Storm Michaung**, poses a critical challenge to traditional climate modeling and infrastructure planning. This event threatens the integrity of future planning by:

* **Climatological Skewing:** The extreme December totals inflate the annual precipitation baseline for 2023, potentially **masking the severity of deficit** faced during the non-monsoon months or in the interior districts.
* **Infrastructure Stress:** The late-season, concentrated downpour places unforeseen stress on coastal drainage and flood defenses designed for monsoon-period rainfall, increasing the risk of urban and agricultural inundation.
* **Resource Management Errors:** An inflated post-monsoon water availability signal could lead to misallocation or mismanaged release of reservoir water, complicating long-term irrigation and power planning (Reference 6).
* **Policy Recalibration:** It necessitates a shift from seasonal disaster management to an **all-year preparedness model** focused on high-impact, low-frequency cyclonic events.

**5.3 Limitations**

While the PySpark framework successfully analyzed the high-volume daily data for 2023, the reliance on a single-year snapshot limits the scope of the conclusions.

* **Temporal Scope:** Single-year data precludes robust **long-term trend analysis** or accurate **climatological modeling** (References 2, 3). Causality remains correlational—we identify an anomaly (Michaung's impact) but cannot assess its frequency or long-term deviation from a 30-year normal.
* **Bias:** The data only reports **rainfall magnitude (mm)**. It lacks critical, granular impact metrics such as reservoir inflows/outflows, groundwater recharge rates, or localized crop damage assessments, which are essential for true water-use efficiency calculations.
* **Granularity:** The district-level aggregation, while useful, may mask extreme variations within large districts. Fine-scale, **mandal-level data** is needed to guide micro-irrigation and conservation projects with greater precision.

**6. Analysis of Rainfall Event Frequencies and Variability (2023)**

This section shifts the focus from simple mean values to the **frequency and variability** of daily rainfall events observed in the 2023 dataset, providing a more granular view of the state's daily weather characteristics.

**6.1 Frequency of Rainy Days (Daily Count $\geq 2.5$ mm)**

The frequency of a "Rainy Day" (defined by the IMD as a day with $\geq 2.5$ mm of rainfall) in 2023 was highly unevenly distributed, mirroring the spatial findings:

* **Coastal Concentration:** Districts in Coastal Andhra recorded the highest number of Rainy Days. **West Godavari**, the wettest district, saw a significantly higher annual count, resulting from both strong monsoon activity (July-September) and the influence of Cyclone Michaung in December.
* **Interior Deficit:** Conversely, the Rayalaseema districts, notably **Anantapur** and **Kurnool**, consistently reported the lowest number of Rainy Days throughout the year. The frequency was not only low during the dry months but also significantly suppressed during the main monsoon season (JJAS), confirming the chronic water deficit in these areas.

**6.2 Frequency of Heavy Rainfall Days (Daily Count $\geq 64.5$ mm)**

The occurrence of "Heavy Rainfall Days" is a critical indicator of flood risk and infrastructural stress. In 2023, the trend was defined by the singular, high-impact **December anomaly** (Figure 3):

* **Extreme Events in December:** The analysis confirmed that a disproportionate share of the year's Heavy Rainfall Days occurred within a 48-hour window in early December. This is highly uncharacteristic of the annual distribution and directly correlates with the path of **Severe Cyclonic Storm Michaung**.
* **Localized Risk:** The highest frequencies of these extreme events were tightly confined to the coastal districts (e.g., SPSR Nellore, Krishna, West Godavari), highlighting their persistent vulnerability to cyclonic activity.

**6.3 Frequency of Dry Days (Daily Count $= 0.0$ mm)**

The frequency of "Dry Days" (days with $0.0$ mm rainfall) is an inverse measure of moisture availability and drought risk.

* **Drought Persistence:** The highest annual frequency of Dry Days was overwhelmingly concentrated in the **Rayalaseema region (Anantapur, Kurnool)**. This confirms that these districts not only receive less rainfall when it occurs but also endure longer, uninterrupted periods of drought throughout the year, compounding the need for efficient water conservation.
* **Temporal Trend:** While all districts naturally recorded high Dry Day frequencies in non-monsoon periods (January-May, October-November), the overall state average was minimally affected by the December cyclone, as the intense rain was too localized and short-lived to significantly reduce the annual Dry Day count for the state as a whole.

**7. Recommendations**

Based on the data-driven insights, the following actionable recommendations are proposed for targeted intervention.

1. **Targeted Water Conservation in Deficit Districts:** Policymakers should prioritize and fund localized water harvesting and conservation schemes (e.g., check dams, community reservoirs) in the driest districts, such as **Anantapur** and **Kurnool**, to capture the limited available precipitation.
2. **Reinforced Coastal Flood Preparedness for December:** Due to the finding that all top 5 extreme rainfall events occurred unexpectedly in December, coastal districts like **West Godavari, Krishna, and East Godavari** require updated flood control and drainage infrastructure capable of handling intense, late-season cyclonic rainfall events.
3. **Monsoon-Specific Crop and Irrigation Planning:** Advise farmers and agricultural departments to adjust crop schedules based on the high concentration of rainfall in **July and September**, while preparing for severe water scarcity in the extremely dry month of **February**.
4. **Sustained Data-Driven Monitoring:** Leverage the successful PySpark analytical framework established in this project to create a dynamic, real-time public dashboard that continuously monitors and visualizes rainfall patterns to aid government agencies and farmers in decision-making.

**8. Conclusion**

This comprehensive analysis successfully leveraged the efficiency of **PySpark** to process and analyze the large-scale 2023 daily rainfall data for Andhra Pradesh. The core finding is the existence of **extreme spatial and temporal disparities** in precipitation across the state. The analysis clearly identified the high-rainfall districts (e.g., **West Godavari**, $\approx \mathbf{3.42}$ mm mean) and the critically dry, high-volatility regions (e.g., **Anantapur**, $\approx \mathbf{1.37}$ mm mean, **CV $323.08\%$**), which must inform targeted water resource allocation and conservation efforts.

Furthermore, the discovery of **high-intensity rainfall events concentrated unexpectedly in December** (including the year's maximum daily rainfall of $\mathbf{188.51}$ mm), outside the typical peak monsoon months, highlights a significant and urgent need to update disaster preparedness and flood control mechanisms in coastal areas, directly correlated with the impact of **Severe Cyclonic Storm Michaung**.

In summary, the project provides a crucial, evidence-based foundation for government agencies to enhance irrigation planning, climate pattern prediction, and agricultural decision-making, ultimately supporting resource security and resilience across Andhra Pradesh.

**References**

1. **Waghaye A M, Rajwade, Y A, Randhe R D and Kumari Nidhi** 2018 “Trend analysis and change point detection of rainfall of Andhra Pradesh and Telangana, India”, *Journal of Agrometeorology* **20** (2) : 160-163.1
2. **Guhathakurta P, Rajeevan M, Sikka D R and Tyagi A,** 2015 “Observed changes in southwest monsoon rainfall over India during 1901–2011”, *Int. J. Climatol*, **35**, 1881-1898.2
3. **Guhathakurta P, Sreejith O P and Menon P A,** 2011 “Impact of climate changes on extreme rainfall events and flood risk3 in India”, *J. Earth System Science.* **120**, No. 3, 359-373.
4. **India Meteorological Department (IMD).** (2023, December 6). *PRESS RELEASE–6: Severe Cyclonic Storm “MICHAUNG”*. Retrieved from IMD Press Releases: [https://internal.imd.gov.in/press\_release/20231206\_pr\_2677.pdf](https://www.google.com/search?q=https://internal.imd.imd.gov.in/press_release/20231206_pr_2677.pdf).
5. **Ministry of Jal Shakti, Government of India.** (2017). *Pradhan Mantri Krishi Sinchayee Yojana (PMKSY) District Irrigation Plan Anantapur District*. Retrieved from PMKSY Reports: <https://pmksy.gov.in/mis/Uploads/2017/20170125072923481-1.pdf>.
6. **Government of Andhra Pradesh.** (2025). *Disaster Management: District Kurnool*. Retrieved from Kurnool District Official Website: <https://kurnool.ap.gov.in/disaster-management/>.
7. **India Meteorological Department (IMD).** (2023). *Rainfall Statistics: DISTRICT-WISE RAINFALL DISTRIBUTION (Daily and Cumulative).* Retrieved from IMD Official Website: <https://mausam.imd.gov.in/imd_latest/contents/rainfall_statistics_3.php>.