Water Stress And Drought Index Tracker

Project Report

Submitted by:

Thakur Raghwendra Singh Kunjam

(25300123)

Guided By:

Prof. Sonali Agarwal

Dr. SPM International Institute of Information technology DEPARTMENT OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

SESSION: 2025-26

TABLE OF CONTENTS

1. **Introduction** 1
    - 1.1 **Context** 1
    - 1.2 **The Challenge** 1
2. **Problem Statement and Objectives** 2
    - 2.1 **Problem Statement** 2
    - 2.2 **Objectives of the Project** 2
3. **Dataset Details** 3
    - 3.1 **Primary Dataset – NASA GRACE / GRACE-FO** 2
    - 3.2 **Secondary Dataset – FAO AQUASTAT** 3
    - 3.3 **Secondary Dataset – WRI Aqueduct 4.0** 3
    - 3.4 **Country Boundary Shapefile (Natural Earth)** 3
4. **Methodology and Implementation** 3
    - 4.1 **Overall Workflow** 4
    - 4.2 **Data Extraction Techniques** 4
    - 4.3 **GRACE Data Processing** 4
    - 4.4 **Data Integration and Feature Engineering** 4
    - 4.5 **Exploratory Data Analysis (EDA)** 4
    - 4.6 **Outlier Detection and Handling Temporal Mismatch** 5
5. **Modeling – Drought Risk Prediction** 5
    - 5.1 **Feature Selection** 5
    - 5.2 **Machine Learning Model Used** 5
    - 5.3 **Model Evaluation Metrics** 5
6. **Results and Discussion** 6
    - 6.1 **Dataset Summary** 6
    - 6.2 **Key Observations and Insights** 6
    - 6.3 **Interpretation of Results** 6
7. **Visualization and Dashboard** 7
    - 7.1 **Exploratory Plots and Graphs** 7
    - 7.2 **Streamlit Dashboard Overview** 7
8. **Conclusion** 7
9. **Future Scope** 7
10. **Relevance and Applications** 8
11. **Libraries and Tools Used** 8
12. **References** 8

Water Stress & Drought Index Tracker

---

1. Introduction

Water scarcity and drought are among the most critical global environmental challenges of the 21st century. Climate change, population growth, urbanization, and unsustainable water consumption have placed unprecedented pressure on freshwater resources. Many regions across the world experience seasonal or chronic water stress, threatening food security, economic stability, and ecosystem health.

With advancements in satellite remote sensing and open-access global datasets, it has become possible to monitor water availability and drought conditions at large spatial and temporal scales. This project leverages such datasets to build a data-driven framework for monitoring global water stress and drought risk.

1.1 Context

Water stress refers to the situation where water demand exceeds the available supply during a certain period or when water quality restricts its use. Prolonged water stress can lead to drought conditions, agricultural failure, groundwater depletion, and socio-economic instability.

Global initiatives such as NASA's GRACE satellite mission, FAO's AQUASTAT database, and WRI's Aqueduct framework provide complementary perspectives on water availability, usage, and risk. Integrating these datasets allows a more holistic understanding of global water stress patterns.

1.2 The Challenge

Despite the availability of high-quality global water datasets, they differ significantly in:

- Temporal coverage
- Spatial resolution
- Data formats
- Measurement methodologies

Satellite-based datasets provide continuous observations, whereas institutional datasets are often static or updated periodically. The key challenge is to harmonize these heterogeneous data sources into a unified dataset suitable for analysis, visualization, and predictive modeling.

2. Problem Statement and Objectives

2.1 Problem Statement

There is no single, integrated dataset that combines satellite-based groundwater observations with institutional water stress indicators in a form that supports drought monitoring and risk prediction. This fragmentation limits the ability to perform comprehensive drought analysis and decision-making.

2.2 Objectives of the Project

The objectives of this project are:

- To extract global water-related datasets from authoritative sources
- To preprocess and harmonize heterogeneous datasets
- To integrate groundwater anomalies with water stress indicators
- To perform exploratory data analysis and feature engineering
- To build a drought risk prediction model
- To visualize water stress and drought patterns interactively

---

3. Dataset Details

3.1 Primary Dataset – NASA GRACE / GRACE-FO

- **Source**: NASA Jet Propulsion Laboratory (JPL), PO.DAAC
- **Data Type**: Terrestrial Water Storage (TWS) anomalies
- **Time Period**: 2002–2025
- **Format**: NetCDF (.nc)

Extracted Features:

- Mean groundwater anomaly (`tws_mean_cm`)
- Median groundwater anomaly (`tws_median_cm`)
- Number of valid grid cells
- Monthly timestamps

GRACE data captures variations in groundwater, soil moisture, and surface water, making it the core dataset for drought detection.

3.2 Secondary Dataset – FAO AQUASTAT

- **Source**: Food and Agriculture Organization (FAO)
- **Data Type**: Country-level water use statistics
- **Time Period Used**: 2020–2022
- **Format**: CSV

AQUASTAT provides contextual information on water usage and efficiency, supporting interpretation of drought vulnerability.

3.3 Secondary Dataset – WRI Aqueduct 4.0

- **Source**: World Resources Institute (WRI)
- **Data Type**: Baseline water stress indicators
- **Reference Year**: 2023
- **Key Variables**:
  - Water stress score
  - Risk category (Low to Extremely High)
  - Regional classifications

Aqueduct data is used as a static baseline risk indicator.

3.4 Country Boundary Shapefile (Natural Earth)

- **Source**: Natural Earth
- **Format**: Shapefile (.shp)
- **Purpose**: Spatial aggregation of GRACE grid data to country level

---

4. Methodology and Implementation

4.1 Overall Workflow

1. Data acquisition
2. GRACE data preprocessing
3. Country-level aggregation
4. Dataset harmonization
5. Feature engineering
6. Exploratory Data Analysis

7. Drought risk modeling
8. Visualization and dashboard development

4.2 Data Extraction Techniques

- **GRACE**: NetCDF processing using `xarray` and spatial aggregation via `geopandas`
- **AQUASTAT**: CSV extraction and pivot transformation
- **Aqueduct**: Excel-based tabular extraction
- **GRACE → Country:** Zonal statistics with shapefiles
- **Spatial Ops:** `geopandas, rasterstats`

4.3 GRACE Data Processing

GRACE data processing involved:

- Converted gridded satellite data to country-level monthly means
- Used zonal statistics
- Extracted:
  - Mean anomaly
  - Median anomaly
  - Valid pixel count

4.4 Data Integration and Feature Engineering

Datasets were merged using ISO A3 country codes and normalized country names.

Engineered Features:

- Rolling means (3, 6, 12 months)
- Groundwater anomaly trends
- Z-score normalization
- Binary drought flags
- Seasonal dummy variables

4.5 Exploratory Data Analysis (EDA)

EDA included:

- Distribution analysis

- Time-series analysis
- Correlation analysis
- Seasonal pattern detection

4.6 Outlier Detection and Handling Temporal Mismatch

Outliers were detected using:

- Interquartile Range (IQR)
- Z-score methods

Outliers were **flagged and capped**, not removed, preserving data integrity.

Handling temporal mismatch:

- AQUASTAT (2020–2022) used as static water-use context
- Aqueduct (2023) used as baseline vulnerability indicator
- GRACE (2002–2025) used for dynamic drought behavior

  This approach is scientifically valid and commonly used in climate risk studies.

---

5. Modeling – Drought Risk Prediction

5.1 Feature Selection

Key predictive features included:

- Groundwater anomalies
- Rolling averages
- Trend slopes
- Seasonal indicators

5.2 Machine Learning Model Used

A supervised machine learning classifier was used to predict drought risk. The model was selected for:

- Interpretability
- Robustness to noise
- Suitability for tabular data

Machine Learning Results :

- Model Used: Random Forest Classifier
- Target Variable: drought_flag_tws

Why Random Forest?

- Handles non-linearity
- Robust to outliers
- Performs well on environmental data
- Model Performance
- Accuracy: High (country-level drought classification)
- Feature importance highlights GRACE trends as dominant predictors

---

5.3 Model Evaluation Metrics

- Accuracy
- Precision
- Recall
- Confusion Matrix

---

6. Results and Discussion

6.1 Dataset Summary

- Final dataset spans over 20 years
- Monthly country-level observations
- Over 30 engineered features

6.2 Key Observations and Insights

- Negative GRACE anomalies strongly correlate with high water stress regions
- Seasonal patterns are evident across arid regions
- Persistent groundwater decline indicates chronic drought risk

---

6.3 Interpretation of Results

The integrated dataset successfully captures both short-term drought events and long-term water stress trends, validating the multi-source integration approach.

---

7. Visualization and Dashboard

7.1 Exploratory Plots and Graphs

- Time-series plots
- Distribution histograms
- Correlation heatmaps
- Drought frequency bar charts

7.2 Streamlit Dashboard Overview

The Streamlit dashboard provides:

- Interactive country selection
- Time-series visualization
- Drought risk indicators
- User-friendly analytics interface

---

8. Conclusion

This project successfully demonstrates that:

- Satellite-derived groundwater anomalies provide **critical drought insights.**
- Integrating GRACE, AQUASTAT, and Aqueduct enhances drought risk assessment.
- The developed framework supports **early drought detection.**
- The system is scalable, reproducible, and suitable for decision-making.

---

9. Future Scope

- Integrate precipitation data (CHIRPS / ERA5)
- Build a composite drought index (SPI + GRACE)
- Sub-national (basin/province) analysis
- Climate change scenario modeling

- Automated quarterly update pipeline
- Policy-focused dashboards

---

10. Relevance and Applications

End Users

- Government water agencies
- Disaster management authorities
- Climate researchers
- NGOs & policymakers

Practical Use

- Early drought warning
- Water resource planning
- Risk-based policy decisions

---

11. Libraries and Tools Used

- Python
- Pandas, NumPy
- Xarray, GeoPandas
- Rasterio, Rasterstats
- Scikit-learn
- Streamlit
- Plotly.express(choropleth map)

---

12. References

1. NASA JPL GRACE Tellus – https://podaac.jpl.nasa.gov
2. FAO AQUASTAT – https://data.apps.fao.org/aquastat
3. WRI Aqueduct – https://www.wri.org/aqueduct
4. Natural Earth – https://www.naturalearthdata.com