A blue and white sign with white text

Description automatically generated

**Capstone Project Phase B**

**Weather Catastrophe Prediction**

**24-2-D-30**

**Students:**

**Nermeen Basila 209290048**

**Madi Swead 313310088**

**Supervisor:**

**Dr. Zakharia Frenkel**

**https://github.com/nermeenbasila**

**Table of Contents**

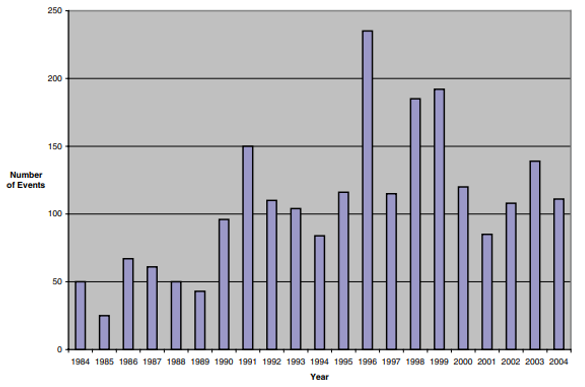
1. Abstract ………………………………………………………………………………3   
2.1 Introduction …………………………………………………………………………3  
 2.1. How historical data impacts predictions……………………………………… 4  
 2.2 short-Term Catastrophic Weather Prediction……………………………………5  
 2.3Long-Term Catastrophic Weather Prediction ……………………………………5  
2. Background ………………………………………………………………………….. 6  
 2.1 Temperature and Climate Change ………………………………………………..6  
 2.2 Humidity and Climate Change …………………………………………….…….7  
 2.3 Extreme Weather Events …………………………………………………………7  
 2.4 Climate Change Impacts ………………………………………………………….8  
 2.5 Definition of Catastrophic Weather Events ……………………………………… 9   
 2.5.1 Catastrophe in general ………………………………………………...……....9  
 2.5.2 Catastrophe in our concept …………………………………………………....9  
 2.6 Catastrophe reasons ……………………………………………………………... 9  
 2.7 Catastrophe Prediction …………………………………………………………..10  
 3 Review of previous Research ……………………………………….…………..11  
 3.1 Ways to calculate the parameters …………………………………………………15  
4. Available data ………………………………………………………………………...17  
 4.1 Weather Catastrophe Examples …………………………….………………...….. 18  
5. Work done ………………………………………………………………………….... 20  
 5.1 algorithm ……………………………………………………………….……….... 22  
 5.2 Use Case ………………………………………………………………………… ..23  
 5.2 Activity Diagram ……………………………………………………………...….. 23  
6. Project Review and description ……………………………………………………….24  
 6.1 Solution Description …………………………………………………………….…24  
 6.2 Data Description ……………………………………………………………………24  
 6.3 Data Processing ……………………………………………………………………..24  
 6.3.1 Data Ingestion and Parsing ……………………………………………………..24  
 6.3.2 Use Of Latitude and Longitude …………………………………………….…..24  
 6.3.3 Data Type Conversion …………………………………………………………..25  
 6.3.4 Data Cleaning ………………………………………………………………...…25  
 6.3.5 Data Resampling and Aggregation ……………………………………………...25  
 6.3.6 Data Normalization ………………………………………………………….….25  
 6.4 Model And Algorithm ………………………………………………………………26  
 6.4.1 Model Architecture ………………………………………………………………26  
 6.4.2 Model Implementation …………………………………………………………..26  
7. Tests ………………………………………………………………………………...….27  
8. Challenges ……………………………………………………………………..………29  
9. Code Explanation ……………………………………………………….……………..31  
 9.1 Formulas we use in our code ………………………………………...……………..34  
10. Results ………………………………………………………………………..………36  
 10.1 Catastrophes Points And Graphs ………………………………………………… 36  
11. Conclusion ……………………………………………………………………………..38  
12. References ……………………………………………………………………………. 39

**Abstract**

Our project presents a tool for detecting and analyzing catastrophic climatic events using geographic and temporal data. The Earth's surface is divided into a uniform grid of 200x200 km cells to facilitate the spatial organization of climate parameters. Thresholds for anomalies in metrics such as temperature, humidity, and pressure are defined to identify significant deviations from the norm. Daily data is compared to an ideal model, and flagged cells with deviations are analyzed to trace event trajectories, tracking their evolution over time and space to identify origin points. A backtracking approach lowers thresholds to uncover precursor signs, offering insights into early indicators of such events. Additionally, environmental conditions around origin cells are examined to establish “conditions for catastrophic production events.” A manual review of findings identifies patterns and recurring factors, enhancing predictive capabilities and understanding of event mechanisms. These insights aim to strengthen early warning systems and improve disaster preparedness.

**Introduction**

Catastrophic weather events have been on the rise over the years, reflecting an alarming trend. From 1984 to 2004, the number of state-level catastrophic events has shown a notable upward trajectory. For instance, in 1996, there were over 200 state-level catastrophic events. This increase is evident in the frequency and intensity of natural disasters such as hurricanes, storms, floods, and wildfires. Factors contributing to this rise include shifts in climatic conditions, such as global warming, which intensify weather patterns, and increased human activity in vulnerable areas, leading to greater exposure and impact. The trend suggests a growing vulnerability to natural catastrophes, posing significant challenges for communities, economies, and insurance markets alike.[12]

****

The frequency and intensity of catastrophic weather events have been on the rise in recent decades, posing significant challenges to communities worldwide. From devastating hurricanes to prolonged droughts and unprecedented flooding, these extreme weather phenomena not only disrupt daily life but also pose serious threats to infrastructure, agriculture, and human safety. Understanding the complex relationship between ordinary weather patterns and catastrophic events is crucial for effective mitigation and adaptation strategies in the face of climate change.

This project aims to delve into this intricate relationship by leveraging data-driven approaches to analyze historical weather patterns and predict catastrophic weather changes. By combining meteorological data, advanced analytics, and predictive modeling techniques, the study seeks to unravel the underlying factors contributing to the escalation of extreme weather events. Through comprehensive analysis and forecasting, the project endeavors to provide valuable insights into the dynamics of catastrophic weather changes and enhance preparedness measures for vulnerable regions.

The impacts of CC are far-reaching, affecting agriculture, public health, water resources, energy production, and biodiversity. Vulnerable populations, including marginalized communities and low-income individuals, are disproportionately affected. [1]

To systematically analyse catastrophic weather events, our project employs a data-driven approach grounded in a spatially consistent framework. The Earth's surface is divided into a uniform grid of 200x200 km cells, allowing precise organization of climate data and enabling region-specific analysis. For each cell, thresholds for catastrophic values and changes in climatic parameters—such as temperature, humidity, and pressure—are defined to identify significant anomalies. By calculating daily parameter changes and comparing them to ideal models, the project detects deviations that signal potential catastrophic events. These flagged events are tracked spatially and temporally, constructing trajectories to understand their evolution and pinpoint origin locations. Additionally, backtracking methods are employed to detect precursor signs by analysing less severe anomalies, while surrounding environmental conditions at event origins are evaluated to define the circumstances contributing to their formation. This comprehensive methodology enhances the ability to predict and mitigate catastrophic events, offering valuable insights for vulnerable regions.

**How historical data impacts predictions**

History can help to a limited extent because a climate situation never repeats itself.  
Observed data gives us a baseline to work from. We can see recent trends that help identify what to look for in the overall data. Take global temperatures, for example. We’ll have a simple metric of overall temperature changing over time. Sometimes that’s down to normal variability, and other times it’s more likely to be due to climate change. If we can isolate these from each other, we can make better predictions for the variable that we’re interested in. This could be temperatures at a specific location for example.

Sticking with the global temperature example, we know that temperatures are rising and that a warmer atmosphere can hold more moisture. This means there’s greater potential for more intense extreme rainfall events in future climates. But these events are likely to be less frequent because it takes longer for the moisture to build back up in the atmosphere. If we can predict global mean temperature, it will impact the predictions we make for mean rainfall in the future. It’s not just historical data of single weather phenomena that we use for our predictions. They’re often interlinked to consider things like carbon emissions, and climate change.

**short-Term Catastrophic Weather Prediction**

Short-term predictions, on the other hand, are critical for immediate weather events such as storms, floods, and heatwaves. Machine learning techniques, including deep learning and ensemble forecasting, are employed to analyze real-time data from various sources, such as satellite imagery and ground measurements. These methods can improve the accuracy of short-term forecasts by identifying patterns and anomalies in weather data that traditional models might miss.[19]

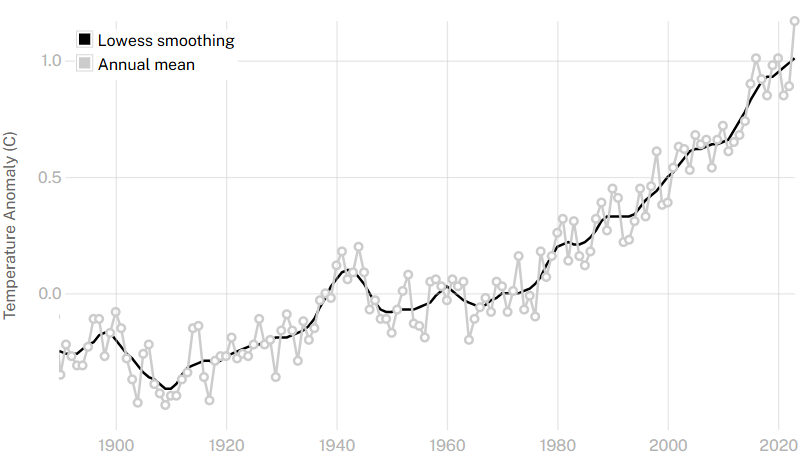
**Long-Term Catastrophic Weather Prediction**

Long-term predictions often focus on climate change impacts and trends over extended periods. Machine learning can enhance the accuracy of these predictions by analyzing vast datasets, including historical climate data and current environmental factors. The integration of machine learning with numerical weather prediction (NWP) models allows for better data assimilation and bias correction, which are crucial for understanding long-term climate patterns and extreme weather events.[19]

Machine learning is positioned as a transformative tool in both long-term and short-term catastrophic weather prediction, enhancing our ability to forecast and respond to extreme weather events.[19]

**Background**

Human civilization has evolved during the Holocene Era, the stability of which is now threatened by human-caused climate change. As a result, global catastrophic risk events from climate change are growing increasingly likely, there are many other potential global catastrophic risk events, both natural and human-caused, posing serious risks and warranting humanity’s careful consideration. there are cautions of “large uncertainty both for the likelihood of such events occurring and for their wider impact.” [2]



This graph shows the change in global surface temperature compared to the long-term average from 1951 to 1980. Earth’s average surface temperature in 2023 was the warmest on record since recordkeeping began in 1880 (source: [NASA/GISS](https://data.giss.nasa.gov/gistemp/graphs/graph_data/Global_Mean_Estimates_based_on_Land_and_Ocean_Data/graph.txt)). [3]

After researches we decided to explore How Humidity, Temperature affects global Climate and Extreme Weather.

Climate change is one of the most pressing challenges of our time, driven by a variety of factors, including temperature and humidity. Understanding how these elements interact and contribute to extreme weather events is crucial for developing effective strategies to mitigate and adapt to the changing climate.

* **Temperature and Climate Change:**

Temperature plays a fundamental role in regulating the Earth's climate. Over the past century, global temperatures have been rising at an unprecedented rate due to human activities, primarily the emission of greenhouse gases such as carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O). This phenomenon, known as global warming, has far-reaching consequences on the planet's weather patterns and ecosystems.  
We can measure the temperature in different heights.

**1. Global Warming**: The increase in average global temperatures leads to a variety of climatic changes. Warmer temperatures result in the melting of polar ice caps and glaciers, contributing to sea-level rise.

**2.** **Heatwaves**: Higher temperatures increase the likelihood and severity of heatwaves. These extreme heat events pose significant risks to human health, agriculture, and infrastructure.

* **Humidity and Climate Change:**

Humidity, the amount of water vapor present in the air, also plays a critical role in climate dynamics. It influences various atmospheric processes, including cloud formation, precipitation, and the overall energy balance of the planet.

**1.** **Water Vapor as a Greenhouse Gas**: Water vapor is the most abundant greenhouse gas in the atmosphere. It amplifies the warming effect of other greenhouse gases through a positive feedback loop. As the temperature rises, more water evaporates, increasing humidity. Higher humidity, in turn, traps more heat in the atmosphere, further accelerating global warming.

**2. Precipitation Patterns**: Changes in humidity levels directly impact precipitation patterns. Higher temperatures increase the rate of evaporation and the capacity of the air to hold moisture. This can lead to more intense and frequent rainfall events, contributing to flooding in some regions. Conversely, other areas may experience prolonged droughts due to altered precipitation patterns.

* **Extreme Weather Events:**

The interplay between temperature and humidity significantly influences the occurrence and intensity of extreme weather events.

**1. Hurricanes and Typhoons**: Warmer Ocean temperatures provide more energy for tropical storms, leading to stronger and more destructive hurricanes and typhoons. Increased humidity levels contribute to heavier rainfall associated with these storms, exacerbating flooding and storm surge impacts.

**2. Droughts and Wildfires**: Rising temperatures can lead to more frequent and severe droughts, particularly in regions already prone to dry conditions. Prolonged droughts reduce soil moisture and vegetation health, increasing the risk of wildfires. These fires can spread rapidly, fueled by dry conditions and high temperatures.

**3. Flooding and Heavy Rainfall**: As global temperatures rise, the atmosphere can hold more moisture, resulting in more intense and frequent heavy rainfall events. This can lead to severe flooding, particularly in areas with inadequate infrastructure to manage large volumes of water.

The intricate relationship between temperature, humidity, and extreme weather events highlights the complexity of climate change. Understanding these interactions is essential for predicting future climate scenarios and developing strategies to mitigate and adapt to their impacts. This project aims to explore the mechanisms through which temperature and humidity contribute to global climate change and extreme weather, providing insights into the challenges we face and potential solutions for a sustainable future. [4]

**Climate Change Impacts**:

1. **Agriculture**: CC affects crop yields, food security, and farmers' livelihoods due to changing climate patterns, precipitation levels, and extreme weather events. Sustainable agricultural practices and the development of resilient crop varieties are essential to mitigate these effects. The research highlights the importance of soil seed banks in maintaining plant community stability and resilience in the face of climate change, particularly in regions experiencing significant variations in rainfall and temperature.
2. **Sea Level Rise**: This poses significant risks to coastal regions and island nations. Strategies such as coastal protection measures and sustainable urban planning are crucial to protect these areas.
3. **Water Resources**: CC exacerbates water scarcity by affecting freshwater reserves and altering precipitation patterns. Sustainable water management practices and investment in water-efficient technologies are vital.
4. **Human Health**: CC increases health risks through elevated temperatures and changing patterns of infectious diseases. Vulnerable communities are especially at risk, necessitating health policies that address these emerging challenges.
5. **Biodiversity**: CC threatens biodiversity by altering or destroying habitats, pushing many species toward extinction. Conservation efforts and sustainable land use practices are critical.
6. **Forest Ecosystems**: Forests face threats from both deforestation and CC, impacting their role as carbon sinks. Mitigation measures like afforestation and improved forest management are important. [1]

**Catastrophe in general**

A catastrophic weather event is a severe and extreme weather phenomenon that causes significant damage, destruction, and loss of life. These events are often characterized by their intensity, scale, and impact on the affected area. Examples of catastrophic weather events include hurricanes, tornadoes, floods, wildfires, and severe storms. These events can result in widespread devastation, displacement of populations, and long-term consequences for the environment and infrastructure.[11]

**Catastrophe in our concept**

Catastrophic weather change refers to extreme and unexpected weather events that significantly deviate from the usual climatic patterns of a region and cause severe disruption and damage. These events are characterized by their rarity and the magnitude of their impact. For instance, while tornadoes are relatively common in certain parts of America and are expected occurrences, they would be considered catastrophic in regions where they rarely occur. Similarly, if a country like Sweden, which typically experiences mild summers, were to encounter temperatures soaring above 40°C, it would be deemed a catastrophic event due to its unprecedented nature and the potential for widespread harm to the environment, infrastructure, and public health.

**Catastrophe reasons**

catastrophes, particularly weather-related catastrophes, can occur due to various reasons. Some common factors contributing to catastrophes include:

**1.** **Severe Weather Events:** Catastrophes can result from extreme weather conditions such as hurricanes, tornadoes, floods, and wildfires, which can cause widespread destruction and significant property damage.

**2.** **Climate Change:** Changes in climate patterns, including global warming, can lead to more frequent and intense weather events, potentially increasing the likelihood of catastrophes.

**3.** **Infrastructure Vulnerability:** The susceptibility of infrastructure to damage from severe weather events can contribute to the occurrence of catastrophes, especially in areas with inadequate infrastructure resilience.[6]

**Catastrophe prediction**

Predicting catastrophic weather events involves a combination of advanced technology, scientific knowledge, and data analysis. Here are some key methods and tools used to forecast and predict catastrophic weather events:

**1. Meteorological Models:** Meteorologists use sophisticated computer models to simulate the atmosphere and predict weather patterns. These models analyze data on temperature, humidity, wind speed, and other factors to forecast the development of severe weather events.[5]

**2.Weather Observations:** Ground-based weather stations, weather balloons, and other observational tools collect data on temperature, pressure, humidity, and wind conditions. This data is essential for understanding current weather patterns and making accurate forecasts.[5]

**3. Climate Models:** Climate models analyze long-term trends and patterns to predict the likelihood of extreme weather events, such as heatwaves, droughts, and heavy rainfall. These models consider factors like sea surface temperatures, atmospheric circulation, and greenhouse gas concentrations.[5]

**4. Historical Data Analysis:** Studying historical trends of catastrophes, including frequency, severity, and impact, can help identify patterns and potential risk factors for future events. [6]

**5. Early Warning System:** Developing early warning systems for natural disasters like hurricanes, tsunamis, and floods can help alert populations in advance, giving them time to evacuate and take necessary precautions.

**Review of Previous Research**

**Similar approaches have been previously undertaken by several researchers, including:**

1) In the document, the author utilized a Back Propagation Neural Network (BPNN) to classify and predict weather conditions. Historical data on temperature, humidity, and pressure were used to train the BPNN model, which was then employed to forecast the weather for the subsequent day. The performance of the BPNN model was compared with other machine learning techniques, including Decision Tree and Naive Bayes. The results indicated that the BPNN method performed well in weather prediction. The thesis concludes that the Backpropagation Neural Network is an effective method for weather forecasting, demonstrating high accuracy in classifying and predicting weather conditions.[7]

Here are potential parameters that used:[7]

1. **Temperature**: Extreme highs or lows that are significantly above or below the historical average.
2. **Humidity**: Extremely high humidity levels that can indicate the potential for storms or other weather events.
3. **Pressure**: Rapid drops in atmospheric pressure which can indicate the approach of a storm or other severe weather event.
4. **Wind Speed**: High wind speeds which can indicate storms, hurricanes, or tornadoes.
5. **Precipitation**: High levels of precipitation, such as heavy rain or snowfall, which can lead to flooding or other weather-related disasters.

2)This paper introduces a novel simulation algorithm that incrementally creates a new year loss table (YLT) by adjusting the original YLT with just enough events to capture the change in extreme weather hazard, specifically focusing on hurricanes and climate change. This incremental simulation method aims to reduce simulation noise and enhance the precision of estimates of change in catastrophe risk models. By copying the original YLT and selectively adding or removing events to reflect the adjusted hazard, this algorithm eliminates the need for approximations and ensures that the new YLT closely represents the desired changes. Through testing on various U.S. hurricane loss models with adjustments for climate change, the study demonstrates that the incremental simulation approach significantly improves the accuracy and precision of estimating the impact of climate change on extreme weather events, providing more reliable insights for risk assessment and decision-making processes.[8]

Here are potential parameters that used:[8]

1. **Frequency**: Frequency refers to how often a particular weather event occurs over a specific period.
2. **Intensity:** measures the strength or severity of a weather event.
3. **Rainfall**: measures the amount of precipitation (rain) that falls during a weather event. It is usually measured in millimeters or inches over a specific period.
4. **Speed**: is the speed at which a weather system, such as a storm or hurricane, moves across the surface of the Earth. It is usually measured in kilometers or miles per hour.

3)Deep learning algorithms have been increasingly employed for weather prediction due to their ability to handle complex and non-linear patterns in data. Convolutional Neural Networks (CNNs) are prominently used for extracting features from meteorological data, such as satellite images and radar data, to predict extreme weather events. CNNs excel in identifying spatial patterns and can effectively process large datasets, although they require significant computational resources​. Additionally, Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are utilized for their capability to capture temporal dependencies in sequential data, making them suitable for time-series forecasting of weather phenomena. These deep learning frameworks can automatically learn from vast amounts of historical weather data, enhancing the accuracy of predictions for events such as heavy rainfall, thunderstorms, and hurricanes. By integrating these advanced algorithms, meteorologists can improve the prediction of extreme weather conditions, potentially mitigating their adverse impacts on society.[9]

Here are potential parameters that used:[9]

**1.Temperature:** The ambient temperature, which can impact the likelihood of certain weather events, such as heatwaves or cold spells.

**2.Humidity:** The amount of moisture in the air, which is critical for predicting events like fog, rain, and storms.

**3.Rainfall:** The amount of precipitation, which is crucial for predicting flooding and other water-related disasters.

**4)**The prediction algorithm used in the article "Support Vector Machine Weather Prediction Technology Based on the Improved Quantum Optimization Algorithm" is the Support Vector Machine (SVM). The study focuses on utilizing SVM in conjunction with the Improved Quantum Genetic Algorithm (IQGA) to optimize parameters for weather prediction. Additionally, the study compares the performance of SVM models optimized by different algorithms such as Genetic Algorithm (GA-SVM) and Improved Ant Colony Optimization. [10]

Here are potential parameters that used:[10]

**1.** **Temperature:** Extreme highs or lows indicating potential heatwaves or cold spells.

**2.** **Humidity:** High humidity levels indicating potential for storms or heavy rainfall.

**3.** **Wind Speed:** High wind speeds indicating potential for storms, hurricanes, or tornadoes.

**4.** **Precipitation:** Amount of rainfall indicating potential for flooding or severe weather.

**5.** **Cloud Cover:** Dense cloud cover indicating potential for precipitation or severe weather.

5) The researchers utilized a genetic algorithm-based framework for automatically identifying weather systems from numerical weather prediction (NWP) data. Genetic algorithms are optimization algorithms inspired by the process of natural selection and genetics. In this context, general algorithms were employed to analyze multiple meteorological elements and identify patterns representing weather systems such as tropical cyclones, fronts, troughs, ridges, and pressure extrema. By formulating the weather system identification problem as a pattern recognition task and designing a generic model of weather systems, the researchers were able to use the genetic algorithm framework to automatically discover these patterns from NWP data. The framework allowed for the analysis of different meteorological elements and the evaluation of fitness functions to assess the effectiveness of the model and framework. Overall, the genetic algorithm-based approach provided a method for automatically identifying and locating weather systems with high precision, offering an independent and objective source of information to assist forecasters in their work.[16]

Here are potential parameters that used:[16]

1. Pressure: Pressure is the force exerted by the atmosphere at a given point. In meteorology, pressure is an essential parameter for understanding weather patterns and systems.

2. Temperature: Temperature is a measure of the average kinetic energy of particles in a substance. In meteorology, temperature plays a crucial role in determining weather conditions and atmospheric stability.

3. Dewpoint: Dewpoint is the temperature at which air becomes saturated with water vapor and dew begins to form. It is a key parameter for assessing humidity levels in the atmosphere.

6)The reviewed papers utilized various algorithms for prediction, including:

1. MetNet-2: A deep neural network-based weather model that outperforms existing physics-based models in predicting high-resolution precipitation up to 12 hours ahead.

2. MLP feedforward backpropagation ANN model: Employed for the retrieval algorithm in predicting Total Precipitable Water (TPW) and Convective Available Potential Energy (CAPE).

3. Artificial Neural Networks (MLP, RBF, GRNN): Used to predict Tropical Cyclone (TC) frequency based on large-scale climate variables.

4. Convolutional Neural Network (CNN): Utilized for predicting hailstorms and severe hail events, encoding spatial weather data for improved accuracy.

5. LSTM algorithm: Employed for nowcasting typhoon tracks and predicting tropical cyclone movement over a 24-hour timeframe.

6. Various modeling approaches (Linear Regression, Lasso Regression, Polynomial Regression, AdaBoost, Decision Trees, Random Forest, CNN, CNN with Recurrence Plots, RP+CNN with binary fusion): Explored for long-term air temperature prediction in summer using ECMWF's ERA5 reanalysis data.

These algorithms showcase the diverse applications of deep learning techniques in predicting various weather phenomena and improving forecast accuracy.[17]

Here are potential parameters that used:[17]

**1. Sea Surface Temperature (SST):** The temperature of the ocean surface, which plays a significant role in influencing weather patterns and climate variability. Changes in SST can impact atmospheric circulation, precipitation patterns, and the development of tropical cyclones

**2. Relative Humidity (RH):** The amount of water vapor present in the air relative to the maximum amount the air can hold at a specific temperature. RH is essential for understanding atmospheric moisture levels, cloud formation, and precipitation processes.

**3. Geopotential Height:** A measure of the height of a pressure surface in the atmosphere above mean sea level. Geopotential height is used in weather forecasting to analyse atmospheric stability, pressure systems, and the movement of air masses.

**4. Dewpoint:** The temperature at which air becomes saturated with water vapor, leading to the formation of dew or fog. Dewpoint is a critical parameter for assessing atmospheric moisture content and predicting the likelihood of precipitation.

**5. Zonal Wind**: The component of wind that blows parallel to lines of latitude, from west to east or east to west. Zonal wind patterns influence global circulation patterns, weather systems, and the movement of air masses across the Earth's surface.

**6. Meridional Wind:** The component of wind that blows perpendicular to lines of latitude, from north to south or south to north. Meridional wind patterns play a role in atmospheric circulation, weather fronts, and the transport of heat and moisture in the atmosphere.

7)This research references several algorithms used for weather prediction, particularly for extreme weather events. These algorithms include:[18]

**1. Artificial Neural Networks (ANN):** Used in various studies for weather forecasting by modelling nonlinear relationships between input data and weather conditions. For example, ANN with Error Backpropagation Algorithm has been employed for predicting weather patterns.

**2. Random Forest Algorithm:** This algorithm has been applied for controlling weather-dependent tasks and predicting specific weather-related events.

**3. Fireworks Algorithm:** This algorithm has been used for training neural networks specifically for weather forecasting purposes.

**4. Dynamic Neural Network Architecture with Immunology Inspired Optimization:** Used for weather data forecasting, integrating advanced optimization techniques to improve prediction accuracy.

These machine learning algorithms help improve the accuracy and reliability of weather forecasts, especially in predicting extreme weather events.

Here are potential parameters that used:[18]

The parameters used to predict extreme weather events in the referenced document include:

1. **Temperature**: Various temperature-related parameters are considered, such as surface temperature and temperature at different atmospheric levels.
2. **Humidity**: Humidity levels, including specific humidity and relative humidity, are used to predict weather conditions.
3. **Wind Speed and Direction**: These parameters help in understanding and predicting the movement of weather systems.
4. **Precipitation**: Amount and intensity of precipitation are critical for predicting extreme weather events like heavy rainfall and storms.
5. **Pressure**: Atmospheric pressure measurements are used to analyze and predict weather patterns.
6. **Solar Radiation**: This is considered in some models to understand its impact on temperature and weather dynamics.
7. **Sea Surface Temperature**: Particularly relevant for predicting phenomena like hurricanes and other ocean-related weather events.
8. **Cloud Cover**: The extent and type of cloud cover help in forecasting weather changes.

These parameters are fed into various machine learning algorithms to model and predict extreme weather events accurately.

Ways to calculate these parameters:

* **Temperature:**

TST: is the true solar time in decimal hours since sunrise.  
 T-max and T-min: are the maximum and minimum ambient temperature during the day. [14]

**Or**

T=T0+ΔT

where T0​ is the initial temperature and ΔT is the change in temperature over time, which can be modeled using differential equations.[18]

**2.Humidity:** RH=100X

Rh: relative humidity percent.  
e:is the current vapor pressure.  
 : is the saturated vapor pressure at the given temperature. [13]

**3. wind Speed:** V=d/t

V: Wind speed (e.g., in m/s)  
d: Distance travelled by the air (e.g., in meters)  
t: Time taken to travel that distance (e.g., in seconds)

**Or**

The wind speed can be calculated using the formula: W=U2+V2 where:   
a. W is the wind speed.  
b. U is the zonal wind component.   
c. V is the meridional wind component

**4.Precipitation:**

a. Simple Linear Regression Formula P=a⋅T+b

P: Predicted precipitation (e.g., in mm)  
T: Temperature (e.g., in °C)  
a, b: Coefficients determined through regression analysis of historical data

b. Multiple Linear Regression Formula P=a0+a1⋅T+a2⋅H+a3⋅W

P: Predicted precipitation  
T: Temperature  
H: Humidity (e.g., in %)  
W: Wind speed (e.g., in m/s)  
a\_0, a\_1, a\_2, a\_3: Coefficients determined through regression

**5. Intensity:** I= **I:** is the intensity.  
  **P**: is the power.  
 **A:** is the area of cross-section. [15]

**6. Pressure (P):**

a. The pressure at a specific location can be calculated using the ideal gas law: P=ρ⋅R⋅T, where:   
P is the pressure.  
ρ is the density of the air.  
R is the specific gas constant for dry air.   
T is the temperature.

b. Calculated using the ideal gas law: P= ​, where P is pressure, n is the number of moles of gas, R is the gas constant, T is temperature, and V is volume.

**7.Dewpoint (D):**

a. The dewpoint can be calculated using the formula: D=T−(5100−RH) where:  
D is the dewpoint temperature.  
T is the temperature.  
RH is the relative humidity.

b. The dewpoint temperature can be calculated using the formula: Td = T - ((100 - RH) / 5), where Td is the dewpoint temperature,

T is the air temperature,

RH is the relative humidity.

**8. Wind Speed and Direction**: Calculated using vector components of wind velocity, u (zonal wind) and v (meridional wind): Speed= ​, and direction θ=arctan ).[18]

The current algorithms for predicting extreme weather changes, including neural networks, simulation algorithms, CNNs, and SVMs, are not achieving the desired level of accuracy. To address this, we are exploring the new modern Prophet algorithm. Prophet, developed by Facebook, is particularly suited for time series forecasting due to its ability to handle missing data, seasonal variations, and trend changes more effectively than the other methods. Unlike traditional neural networks and CNNs that require extensive tuning and large datasets, or SVMs that struggle with large-scale data, Prophet offers a robust and user-friendly approach. Its flexibility and strong performance with historical data make it a promising alternative for improving the accuracy and reliability of extreme weather predictions.

**Available data**

There are several types of catastrophes and climatic parameters associated with them. In our project we concentrate majorly on the temperature, because such data is more available. Generally, we reviewed two types of data: the first is directly associated with catastrophes. Such data is required for criteria for automatic detection of the events. The second type is historical climatic data for “reverse engineering” and testing of our predictive algorithms.

**Weather Catastrophe Examples:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **When it happens** | **What happens** | **Damages** |
| **Hurricane Katrina [20]** | August 23, 2005 | 1) Hurricane Katrina was a Category 5 Atlantic hurricane that caused severe destruction along the Gulf Coast of the United States.  2) The storm surge overwhelmed the levees in New Orleans, causing widespread flooding. | 1) Over 1,800 deaths.  2) $125 billion in damages.  3)Displacement of hundreds of thousands of residents.  4)Extensive destruction of homes, infrastructure, and the local economy. |
| **Indian Ocean earthquake and tsunami [21][22]** | December 26, 2004   at 07:58:53 | 1)Triggered by a massive undersea earthquake off the coast of Sumatra, Indonesia.  2) Generated a series of devastating tsunamis that affected 14 countries**.** | 1) Approximately 230,000 to 280,000 deaths.  2) Millions of people displaced.  3) Widespread destruction of coastal communities, infrastructure, and economies.  4) Long-term environmental and social impacts. |
| **Typhoon Haiyan [23][24]** | November 3, 2013 | 1) Known as Yolanda in the Philippines, it was one of the strongest tropical cyclones ever recorded.  2) Produced winds of up to 195 mph and massive storm surges. | 1) Over 6,300 deaths in the Philippines.  2) $2.98 billion in damages.  3) Widespread destruction of homes, infrastructure, and livelihoods.  4) Severe humanitarian crisis, with millions requiring aid and support. |
| **European Heatwave**  **[25]** | Summer of 2003 | 1)An intense heatwave affected much of Europe, with temperatures reaching unprecedented levels.  2) Particularly severe in France, Italy, Spain, and Portugal**.** | 1) Estimated 70,000 deaths due to heat-related illnesses.  2) Agricultural losses estimated at €13 billion.  3) Increased risk of wildfires and damage to infrastructure. |
| **Pakistan Floods [26]** | July 26,2010 | 1)Unprecedented monsoon rains led to widespread flooding across Pakistan.  2) The Indus River overflowed, affecting large swathes of the country. | 1) Approximately 1,985 deaths.  2) Over 20 million people affected.  3) $43 billion in damages.  4)Extensive damage to infrastructure, homes, and agriculture. |
| **Hurricane Sandy [27]** | **October 24, 2012** | 1)A post-tropical cyclone that affected the Caribbean and the Eastern United States.  2)Caused widespread flooding and power outages, particularly in New York and New Jersey. | 1) 233 deaths.  2)$70 billion in damages.  3)Extensive damage to homes, infrastructure, and the economy.  4) Long-term displacement and recovery efforts. |
| **Hurricane Maria [28]** | September 16, 2017 | 1)Hurricane Maria was a Category 5 hurricane that devastated Puerto Rico and other parts of the northeastern Caribbean.  2) It brought intense winds, heavy rainfall, and massive storm surges. | 1) Approximately 3,000 deaths, making it one of the deadliest hurricanes to hit the region.  2) Over $90 billion in damages.  3) Extensive destruction of infrastructure, homes, and power grids, leaving many residents without electricity for months. |

**Work done**

**Part 1:** System Setup and Geographical Division

**1.** System Setup: Read data from CSV files containing geographical stations , historical events and associated parameters.

Components Involved:

- CSV files: The database containing the required data.  
- Connection Interface: Use a programming language such as C++ and Python to establish the connection and drawing the graphs .  
- Data retrieval: Run queries to retrieve complete data related to geographic stations, historical events, and associated parameters.

**2.** Geographical Division: we divide the Earth into Regions

Objective: Divide the earth into distinct regions based on station data correlations to facilitate localized analysis.

Steps Involved:

**-** Region Definition: Define regions by analyzing correlations between stations. Stations with strong data correlations will be grouped into the same region. **-** Formulas: Develop formulas to calculate the inter-station correlations and define regional boundaries.

**Part 2:** Parameter Assumption and Catastrophe Identification

**1.** Parameter Assumption: Defining Parameters for Catastrophe Detection

Objective: Analyze historical data to establish parameters for detecting catastrophes in each region.

Steps Involved:

**-** Historical Data Analysis: Analyze past catastrophic events (e.g., earthquakes, floods, hurricanes) to identify patterns and thresholds for key parameters such as temperature, seismic activity, and pressure. **-** Parameter Definition: Define specific parameters that indicate potential catastrophes, such as: **-** Thresholds: Set upper and lower limits for each parameter. **-** Standard Deviations: Calculate the deviation of current data from the historical mean. Large deviations may signal unusual activity. **-** Region-Specific Customization: Adjust the parameters for each region based on its unique historical and environmental characteristics.

**2.** Catastrophe Identification: Detecting Catastrophes

Objective: Detect potential catastrophes by comparing real-time data against the predefined parameters.

Steps Involved:

- Data Monitoring: Continuously monitor data from stations in each region.  
**-** Anomaly Detection: Compare real-time data to the defined thresholds and standard deviations to detect anomalies that may indicate a catastrophe.

**Part 3:** Collection and Analysis of Catastrophic Events

**1.** Collection and Analysis

Objective: After identifying a potential catastrophe, collect relevant data and perform an in-depth analysis to understand its onset and causes.

Steps Involved:

- Data Compilation: Gather data from the stations during the period of anomaly detection.  
- Onset Analysis: Determine the starting point of the unusual event by identifying the first significant deviation.  
**-** Reason Documentation: Analyze potential causes of the catastrophe, including environmental conditions, human activities, and other contributing factors.  
- Report Generation: Document the event details, including the onset, the parameters that triggered the alert, and the findings from the analysis.

**Algorithm:**

1. Grid Creation: Begin by constructing a geographic grid with a constant cell size of 200x200 kilometers. This grid will serve as the framework for spatially organizing climate data, ensuring consistent analysis of parameters across the Earth’s surface.

2. Definition of Catastrophic Metrics: For each defined climatic parameter (such as temperature, humidity, or pressure), we determine thresholds for what constitutes a “catastrophic value” and a “catastrophic change.” These thresholds will help identify significant deviations from normal conditions that may signal the onset of catastrophic events.

3. Daily Parameter Calculation and Comparison: Daily, calculate various climate-related parameters (e.g., temperature, pressure, wind speed, precipitation) and their differences between consecutive days for each cell in the grid. For each calculated parameter,we compare the daily average with an “ideal model,” which represents the expected or normal state. Deviations beyond predefined thresholds will help flag potentially catastrophic events.

4. Identification of Catastrophic Cells: Identify all cells within the grid that exhibit either a “catastrophic value” or a “catastrophic change” based on the threshold definitions. These cells are flagged for further analysis as they may indicate the presence of a catastrophic event.

5. Trajectory Building for Catastrophic Events: For each identified catastrophic event, trace its trajectory by locating neighboring cells in both space and time that also exhibit catastrophic characteristics. This step involves connecting neighboring events to form a spatial and temporal path, which will reveal how the event evolved over time and across regions.

6. Identification of Trajectory Starting Points: Once trajectories are constructed, identify the initial points—cells where the event originated. These starting points represent the locations where the catastrophic events were “born,” providing crucial information about the origins of the events.

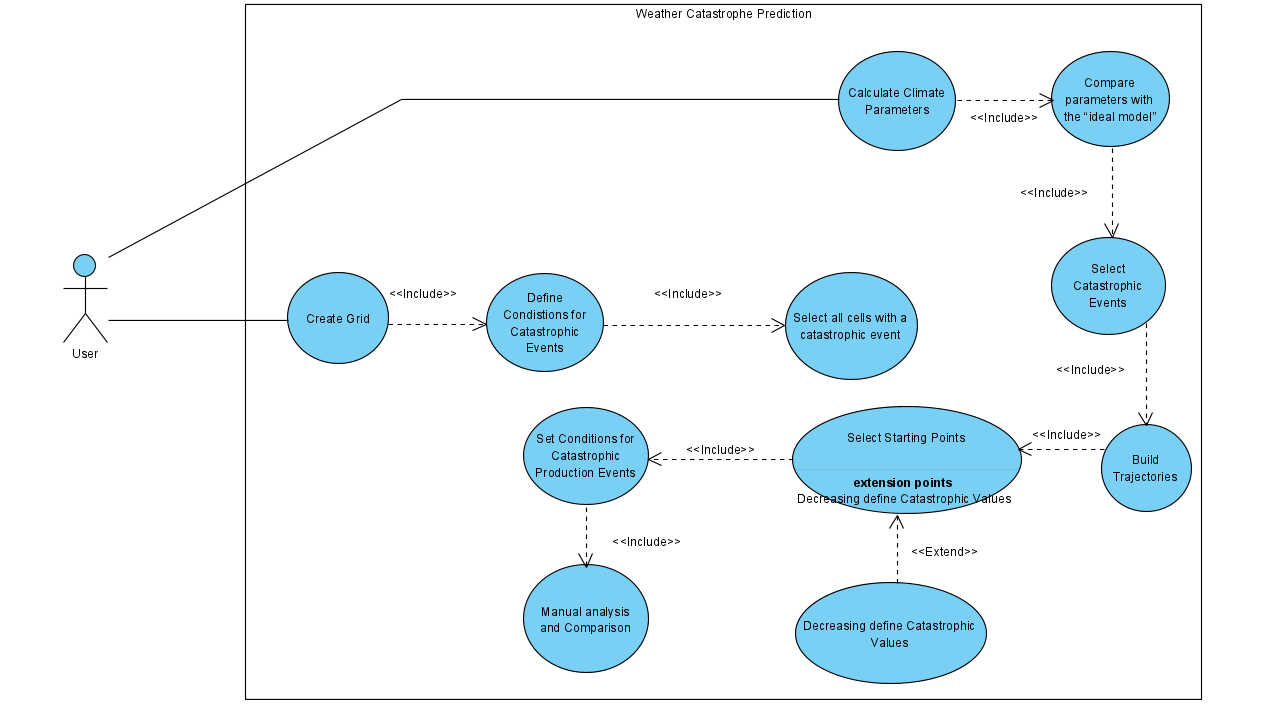
7. Backtracking to Earlier Points: To further understand the development of the event, we attempt to trace the trajectory backward by lowering the threshold for what is considered a “catastrophic value.” By identifying earlier points with less severe conditions, it may be possible to detect precursor signs of the event and extend the understanding of its origin.

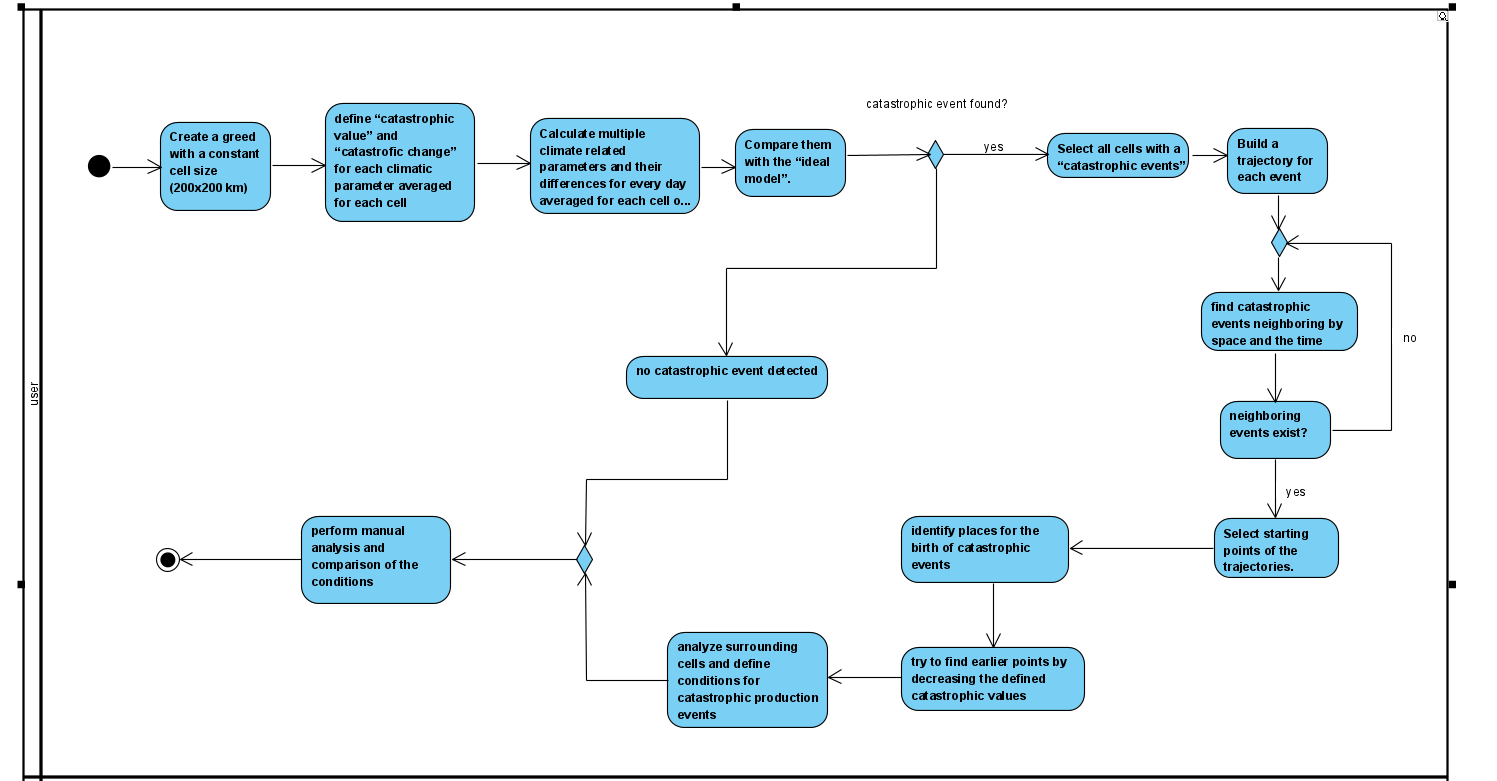
8. Conditions for Catastrophic Event Generation: Analyze the states of surrounding cells at the event’s starting point to determine the environmental conditions that contributed to the event's formation. These surrounding cells' climatic states will be compiled into a set of “conditions for catastrophic production events,” which describe the circumstances that led to the event's birth.

9. Manual Analysis and Comparison: Finally,we conduct a manual analysis of the identified conditions. This step involves comparing the catastrophic event conditions across multiple trajectories and events, searching for patterns or recurring factors that may provide insights into the underlying causes of catastrophic event generation. Manual analysis allows for nuanced interpretation and the integration of expert judgment to enhance the algorithm’s results.

**UML Diagrams:**

**Use Case:**

**Activity diagram:**

****

**Project Review and description**

**Solution Description**

Our solution analysing weather data to predict catastrophic events. It involves processing temperature data from a CSV file, splitting it into smaller subsets, and calculating statistical metrics such as averages, differences, and ratios for a 14-day period across multiple areas. Outliers with a ratio are flagged and stored separately. A second phase joins the flagged data with geographical coordinates from another file to enhance spatial analysis. The final stage calculates distances between affected areas and a target location, filtering results within a 500 km radius for further analysis. The modular design integrates preprocessing, statistical computation, geospatial analysis, and data filtering to support predictions effectively, ensuring scalability and clarity by dividing tasks across multiple functions and files.

**Data description**

The dataset consists of temperature measurements recorded across multiple geographic areas over time, stored in a CSV file where each row represents a day and columns correspond to area identifiers and their respective temperature readings. The first column contains non-numeric day labels, while the remaining columns store numeric temperature data. Additionally, a supplementary file, greed\_coord.dat, provides geographic coordinates (latitude and longitude) for each area, enabling spatial analysis. The dataset supports statistical computations over 14-day periods and facilitates anomaly detection and geospatial filtering to analyse catastrophic weather patterns effectively.

**Data processing**

**Data Ingestion and parsing**

The data ingestion and parsing process involves reading temperature measurements and metadata from a CSV file, skipping the header row, and extracting relevant temperature values while ignoring non-numeric identifiers such as the day index column. Geographic coordinates for the areas are integrated by joining the dataset with a separate file (greed\_coord.dat) using area identifiers as keys. Invalid or missing values are managed to ensure the dataset remains consistent and usable. Parsed data is organized into structured records, capturing details such as day, area, temperature, and statistical metrics. This comprehensive process ensures clean, structured data is prepared for anomaly detection, statistical analysis, and geospatial filtering.

**Use of Latitude and Longitude in Our Research**

In our research, latitude and longitude play a crucial role in analyzing catastrophic weather patterns by allowing us to pinpoint the exact locations of extreme temperature anomalies. Each measurement in our dataset is associated with geographic coordinates, enabling us to track the spatial distribution of temperature fluctuations and detect patterns over different regions.

By integrating latitude and longitude with temperature data, we can:

* Identify affected regions where significant temperature deviations occur.
* Analyze proximity relations between anomalous events to detect clusters of extreme weather activity.
* Filter relevant data by selecting only locations within a defined radius (e.g., 500 km) of an anomaly.
* Correlate climate changes geographically, helping us understand whether temperature anomalies are localized or part of a larger atmospheric phenomenon.

We use this formula to calculates the distance between two geographic points based on their latitude and longitude. It is derived from the Haversine formula, which determines the great-circle distance (shortest distance over the Earth's surface) between two points on a sphere.

These coordinates are essential for mapping and visualizing temperature variations across different areas, providing valuable insights into weather prediction and disaster preparedness.

**Data Type Conversion**

Data type conversion is essential throughout the process to ensure consistent and accurate handling of the dataset. Temperature values, initially read as strings, are converted to double-precision floating-point numbers for mathematical operations. Day labels, stored as strings with a "Day" prefix, are parsed into integers for easier comparison and filtering. In cases where invalid or non-numeric data is encountered, appropriate default values are assigned to maintain data integrity. These conversions ensure that the data is in the correct format for statistical calculations, filtering, and geospatial analysis.

**Data Cleaning**

There were points in the data file on certain days that didn't show what the temperature was, so we skipped them.

**Data Resampling And Aggregation**

After cleaning , data resampling and aggregation involve organizing temperature data into 14-day periods for each area to calculate statistical metrics. For each period, the average temperature is computed by summing valid entries and dividing by the count of non-missing values. The standard deviation is calculated to quantify variability within the period. These metrics are then used to determine daily temperature differences and their ratios to the standard deviation. This aggregation process enables identifying anomalies, as high ratios indicate significant deviations from the norm. Additionally, the data is divided into smaller subsets (e.g., by area or time period) for efficient processing and storage.

**Data Normalization**

Data normalization in the codes is achieved by calculating the ratio of daily temperature differences to the standard deviation for a 14-day period. This process scales the deviations relative to the data's variability, allowing for meaningful comparisons across different areas and time periods. By standardizing these differences, the method highlights significant anomalies, as ratios that defined threshold indicate extreme deviations. This normalized metric ensures that areas with naturally higher temperature ranges are treated fairly in the anomaly detection process.

**Model And Algorithm**

**Model Architecture**

The system follows a modular architecture designed for efficient processing, analysis, and integration of weather data to detect anomalies. It consists of several key components:

1. Data Ingestion and Preprocessing: Reads raw temperature data, cleans missing or invalid values, and joins it with geographic coordinates for spatial analysis.
2. Statistical Computation: Aggregates data into 14-day periods to calculate averages, differences, and standard deviations, identifying anomalies based on normalized ratios.
3. Geospatial Analysis: Integrates geographic coordinates and calculates distances between areas, filtering results within a defined radius (e.g., 500 km) for localized analysis.
4. Anomaly Detection: Flags areas with significant deviations for further investigation, storing these anomalies for downstream processes.
5. Output Generation: Organizes results into CSV files, enabling visualization and integration with external tools for trend analysis and geographic mapping.

This architecture ensures scalability, clear separation of tasks, and robust handling of large datasets for weather anomaly detection and spatial analysis**.**

**Model Implementation**

The model implementation is a structured pipeline that processes weather data to identify anomalies and conduct spatial analysis. It begins with reading and parsing temperature data from a CSV file, ensuring invalid or missing values are handled gracefully. Data is aggregated into 14-day periods, where averages, differences, and standard deviations are calculated to normalize temperature deviations. Anomalies are flagged based on ratios exceeding a threshold and stored for further analysis. Geographic data is integrated using coordinates from a supplementary file, enabling the calculation of distances between areas. The model filters results based on proximity criteria (e.g., within 500 km) and outputs structured CSV files for downstream use. This implementation ensures efficiency, accuracy, and readiness for visualization or further exploration.

**Tests**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test No.** | **Test case** | **Input** | **Expected output** | **Observed Results** |
| **1** | Ensure that the grid is correctly created with cells of size 200x200 km | Boundary coordinates, cell size (200x200 km) | Correct grid dimensions with no errors in size or positioning of cells | The grid was created with accurate dimensions and 200x200 km cells, correctly aligned to the boundary coordinates. No errors or overlaps were observed, ensuring seamless positioning and adherence to specifications. |
| **2** | Ensure that the entire area of interest is covered by the grid | Region boundary coordinates | No gaps in the grid for the defined region | The grid fully covered the region within the specified boundary coordinates, with no gaps or uncovered areas. All cells aligned seamlessly, ensuring complete spatial coverage of the area of interest. |
| **3** | Verify that each climatic parameter for a cell has its catastrophic value defined correctly | Climatic parameter averages for a cell | Defined catastrophic values based on predefined thresholds | Each climatic parameter for the cell was correctly assigned its catastrophic value based on the predefined thresholds. No discrepancies were observed, ensuring accurate identification of extreme conditions. |
| **4** | Ensure that significant changes in climatic parameters over time trigger a "catastrophic change" flag. | Climatic parameter time series. | Correct identification of catastrophic changes based on defined thresholds. | Significant changes in climatic parameters over time were accurately flagged as "catastrophic changes" when exceeding predefined thresholds. The results showed correct and timely identification of such events. |
| **5** | Ensure that climate-related parameters are averaged for each cell correctly on a daily basis. | Climate parameter data for a day | Correct daily average for each cell. | Daily averages for climate-related parameters were calculated correctly for each cell based on the input data. No errors or inconsistencies were observed, ensuring accurate aggregation of daily values. |
| **6** | Verify the comparison between the calculated parameters and the "ideal model. | Averaged parameters, ideal model data. | Correctly flagged cells where deviations from the model occur | The comparison accurately flagged cells with significant deviations from the ideal model based on the averaged parameters. All deviations were correctly identified, ensuring reliable anomaly detection. |
| **7** | Ensure that the correct cells are identified as having catastrophic events | Climatic parameter values for each cell. | Correct selection of cells that meet the catastrophic threshold | Cells meeting the catastrophic threshold were identified; however, some cases showed incorrect flagging or missed detections, suggesting potential issues with threshold definitions or parameter calculations. Debugging is needed to ensure accurate and reliable detection. |
| **8** | Verify that catastrophic events neighboring in both space and time are correctly linked | Spatial and temporal data of events | Correctly identified trajectories of connected catastrophic events | Catastrophic events within a 14-day window and 500 km range are correctly linked. |
| **9** | Ensure that the trajectory of events is accurately built based on proximity and sequence of occurrences. | Set of catastrophic events with timestamps | Correctly built trajectory of events | successfully links catastrophic events within 14 days and 500 km, identifying clusters of anomalies. |
| **10** | Verify that the correct starting points (where catastrophic events originate) are identified | Trajectories of catastrophic events | Correct identification of starting points | identifies anomalous events but does not explicitly determine the starting point of a trajectory. It filters events within 14 days and 500 km, but lacks logic to trace back to the initial event in a sequence. |
| **11** | Ensure that reducing the catastrophic value threshold leads to the identification of earlier starting points. | Trajectory data and lowered catastrophic value threshold. | Correct identification of earlier points in the trajectory | Lowering the catastrophic value threshold results in detecting more early-stage anomalies, potentially identifying earlier starting points. However, the code does not explicitly track the first occurrence in a trajectory, only filtering based on the new threshold. |
| **12** | Verify that the states of surrounding cells are correctly stored and analyzed | Climatic data for surrounding cells | Correct set of surrounding cell conditions is established | The code correctly filters and processes climatic data for surrounding cells within 500 km and 14 days, storing relevant conditions. However, it does not explicitly track the evolution of surrounding cells over time or establish state changes beyond filtering. |
| **13** | Ensure that the manual comparison of conditions between different catastrophic events is feasible and that conditions are clearly presented for analysis | Set of conditions for different catastrophic events | Clear and accurate presentation for manual analysis | The code outputs filtered catastrophic event data with relevant conditions, making manual comparison feasible. |

**Challenges**

Challenges Faced During Code Development

1. Data Quality and Preprocessing

* Handling Missing or Corrupted Data:

The raw data contained missing or invalid temperature values, which required handling ( skipping them during calculations) to avoid computational errors.

* Inconsistent Formats:

Different input files, such as data\_integr\_2024\_13.csv and greed\_coord.dat, had varying formats and structures, which necessitated careful parsing and trimming of strings to ensure consistency.

2. Scalability

* Large Dataset Handling:

With thousands of areas and multiple days of data, the sheer size of the dataset posed challenges for memory and file management. Splitting the data into manageable chunks (e.g., processing 3000 areas per file) was necessary to optimize performance.

* Dynamic File Generation:

Creating multiple output files dynamically for different areas and time periods while maintaining readability and traceability added complexity.

3. Statistical Computations

* Ratio and Anomaly Detection:

Calculating the standard deviation and ratios for identifying anomalies required precise mathematical operations. Small errors could lead to incorrect anomaly detection, affecting the reliability of results.

* Thresholds:

Deciding on thresholds for anomalies required balancing sensitivity and specificity to ensure meaningful results.

4. Geographical Integration

* Distance Calculations:

Calculating distances between areas based on latitude and longitude. Ensuring these calculations were accurate and efficient for thousands of areas required careful implementation.

5. Filtering and Proximity Analysis

* Time and Spatial Constraints:

Filtering data based on both temporal (14-day range) and spatial (500 km proximity) constraints required combining multiple datasets and creating efficient filtering algorithms.

6. Points Detected in the Sea

* Unexpected Anomalies:

During the analysis, many temperature anomalies were detected in areas located over the sea. This was unexpected, as most analysis focused on land-based regions.

**Code explanation**

Overview

This code suite is designed to analyze temperature data to predict weather-related catastrophic events. It processes raw data, identifies anomalies, combines results with geographical information, and filters data based on proximity to specific regions. The outputs allow spatial and temporal analysis of temperature anomalies.

**First Code File: Analysing Temperature Data**

1. Purpose:

Reads raw temperature data from data\_integr\_2024\_13.csv.

Divides data into 14-day periods for each area and calculates:

* + - Average temperature over the period.
    - Standard deviation of temperatures in the period.
    - Difference between each temperature and the average.
    - Ratio of the difference to the standard deviation.

1. Anomalies:

Identifies temperature anomalies where the ratio exceeds 3.

Saves anomalies in a separate file: ratios\_above\_3.csv.

1. Output Files:

daily\_differences\_X.csv: Contains detailed temperature analysis for subsets of areas.

ratios\_above\_3.csv: Records areas with significant temperature anomalies for further processing.

1. Key Logic:

Chunks Data: Processes a maximum of 3000 areas per file for scalability.

Anomaly Detection: High ratios indicate areas with significant deviations, potentially signalling weather anomalies.

**Second Code File: Combining Data with Geographical Information**

1. Purpose:

Combines temperature anomaly data with geographical coordinates from greed\_coord.dat.

Adds latitude and longitude for each area, enabling spatial analysis of anomalies.

1. Functions:

removeAreaPrefix: Ensures consistent area names by removing the "Area" prefix.

trim: Cleans strings by removing leading and trailing spaces.

printPreview: Displays a preview of file contents for quick verification.

joinTables:

* + - Matches areas in ratios\_above\_3.csv with coordinates in greed\_coord.dat.
    - Outputs the combined data as joined\_table.csv.

1. Output File:

joined\_table.csv: Contains temperature anomalies with corresponding geographical locations (latitude and longitude).

**Third Code File: Filtering and Proximity Analysis**

1. Purpose:

Filters and analyzes anomalies based on proximity to a specific area and day range.

Calculates distances between areas using GPS coordinates.

1. Functions:

DataRow: Represents a row of temperature data with fields for day, area, temperature, average, difference, and ratio.

parseRow: Parses CSV rows into DataRow objects for processing.

filterAndSave:

* + - Filters data for a specific area and 14-day period.
    - Saves filtered results in dynamically named output files (e.g., filtered\_data\_DayX-AreaY.csv).

calcGPSDistance: Calculates the distance between two geographical points using the Haversine formula.

calculateDistances:

* + - Reads joined\_table.csv and calculates distances from a target area.
    - Filters results based on proximity (e.g., within 500 km).
    - Outputs a filtered table: table\_with\_distances\_filtered.csv.

1. Key Logic:

Filtering by Time and Space: Identifies anomalies within 14 days of a target day and areas within 500 km of a target region.

Dynamic File Generation: Ensures outputs are specific to the area and time range under analysis.

1. Output Files:

table\_with\_distances\_filtered.csv: Contains anomalies filtered by proximity.

Dynamically named files (e.g., filtered\_data\_DayX-AreaY.csv): Detail anomalies for specific areas and time ranges.

**Formulas we use in our code:**

**1. Mean (Average) Temperature Calculation:**

We use this formula to compute the average temperature over a 14-day period.

Where:

= Mean temperature

N = Number of valid temperature readings (excluding missing data)

Ti = Individual temperature readings over the period

**2. Standard Deviation Calculation**

We use this formula to measure the spread of temperature values around the mean.

Where:

σ= Standard deviation

Ti= Individual temperature reading

= Mean temperature

N = Number of valid readings

**3. Temperature Difference**

We use this formula to calculate how much the current temperature differs from the average.

**D=T−**

Where:

D = Difference between the actual temperature and the average

T = Current temperature

= Mean temperature

**4. Temperature Anomaly Ratio**

We use this formula to determine whether a temperature deviation is statistically significant.

**R=**

Where:

R = Ratio (how extreme the deviation is)

D = Temperature difference

σ = Standard deviation

**5. Haversine Formula (Distance Between Two Locations)**

Used to calculate the distance between two points (latitude & longitude) on Earth's surface.

**d= 2 x R x atan2 ()**

Where:

a=

d = Distance between two points in meters.

R = Earth's radius (~6,372,797.56 meters).

ϕ1,ϕ2 = Latitudes of the two points (in radians).

λ1,λ2= Longitudes of the two points (in radians).

Δϕ=ϕ2−ϕ1 (difference in latitude).

Δλ=λ2−λ1(difference in longitude).

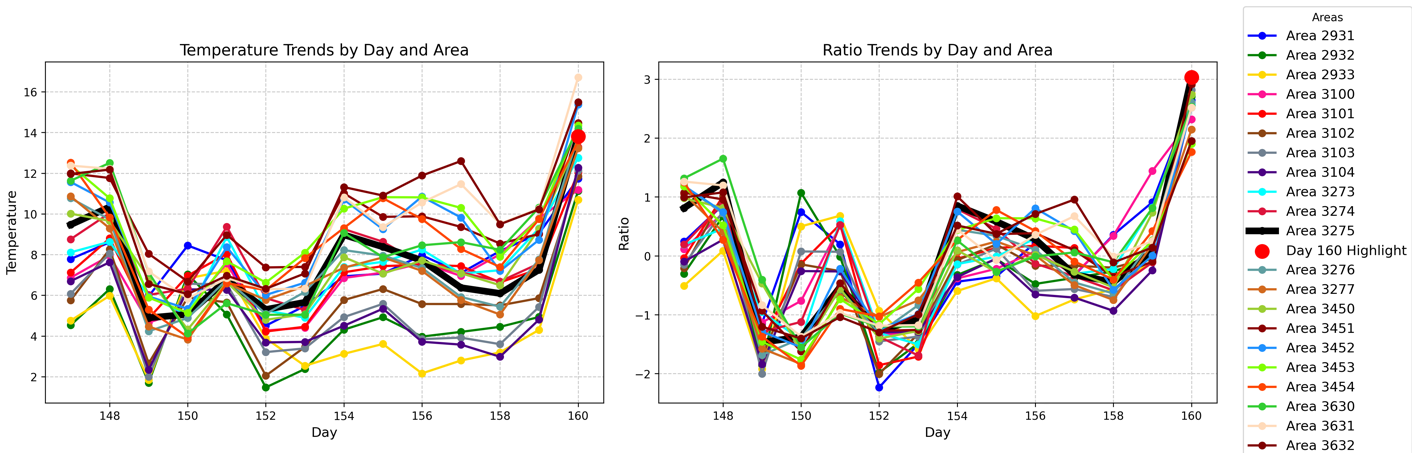
asin = Arcsine function, which helps calculate the angle from the sine value.

**Results**

**Catastrophe points and graphs:**

* Day 160 Area 3275 this catastrophe happens in Beltana Station SA 5730 ,Australia

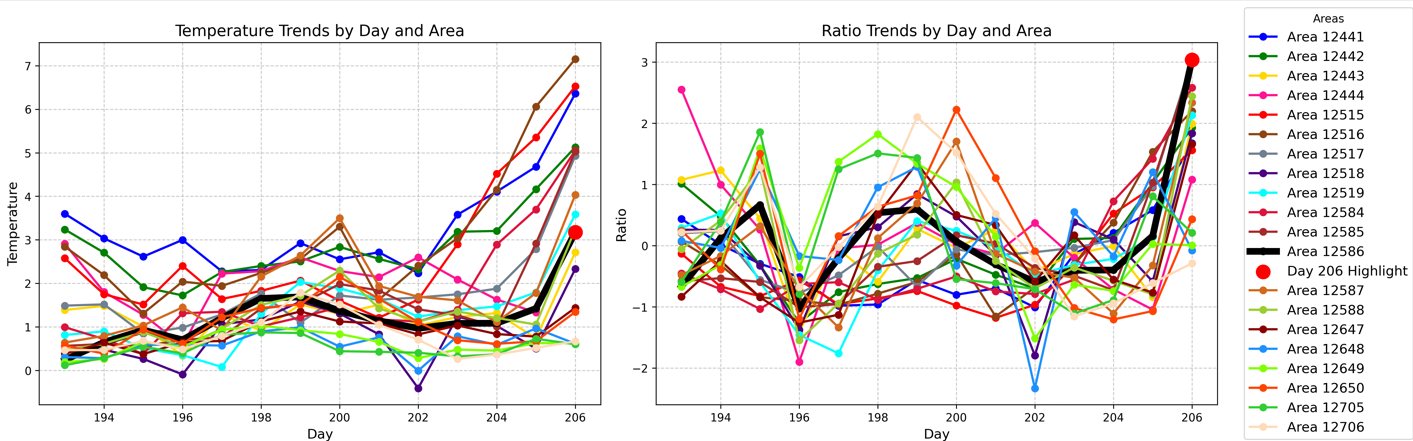
In early June 2024, New South Wales (NSW) in Australia. experienced a significant severe weather event characterized by heavy rainfall and widespread flooding. The Bureau of Meteorology issued flood warnings as parts of Sydney and southeast NSW faced substantial rainfall, with some areas receiving up to 250 millimeters within a 24-hour period.[30]

****

* Day 206 Area 12586 this catastrophe happens in Kitikmeot Region , Nunavut ,Canada.

In July 2024, the Kitikmeot Region of Nunavut, Canada, experienced an unprecedented heatwave. Where temperatures soared above 30°C, significantly higher than the typical average for that time of year . These temperatures were approximately twice the usual average for these regions in July.

This extreme heat prompted Environment Canada to issue heat warnings for the affected communities. The heatwave was attributed to a mass of hot air that typically remains over the prairie provinces but had extended much further north than usual.Such events are becoming more frequent, with the Arctic warming faster than other parts of the world due to climate change. [29]



* Day 208 Area 2515 this catastrophe happens in “ Uanamed Road ,Departamento Hucal ,Municipio de General San Martin , Argentina” .

In June 2024, Argentina's Patagonia region experienced an extreme cold snap, with temperatures plunging below those recorded in Antarctica. This unusual deep freeze was attributed to a high-pressure system situated between Patagonia and the Antarctic Peninsula, which funneled large amounts of cold air from the South Pole northwards. The severe cold led to road blockages, food shortages, and posed risks to livestock, among other challenges for the local inhabitants.

The cold snap continued into early July, with temperatures in Buenos Aires dropping to 0°C (32°F) on July 6, and the extreme cold persisting throughout the week.[31]

A graph of different colored lines

Description automatically generated

**Conclusion**

Our research successfully developed a tool for detecting and analyzing catastrophic climatic events using geographic and temporal data. By organizing the Earth's surface into a structured 200x200 km grid and defining anomaly thresholds for key climate parameters, we were able to track event trajectories and identify their origins. The backtracking approach provided insights into precursor signs, enhancing early warning capabilities.

Unfortunately, the data collection and preparation required to much time, and only manual review was done. It helps to establish key environmental conditions contributing to catastrophic event formation, supporting disaster preparedness and mitigation efforts.

**References**

**[1] :** Rawat, A., Kumar, D., & Khati, B. S. (2024). A review on climate change impacts, models, and its consequences on different sectors: a systematic approach. *Journal of Water and Climate Change*, *15*(1), 104-126.

**[2]:** [The future of global catastrophic risk events from climate change » Yale Climate Connections](https://yaleclimateconnections.org/2022/07/the-future-of-global-catastrophic-risk-events-from-climate-change/)

**[3]:** [Global Temperature | Vital Signs – Climate Change: Vital Signs of the Planet (nasa.gov)](https://climate.nasa.gov/vital-signs/global-temperature/?intent=111)

**[4]:** [National Oceanic and Atmospheric Administration (noaa.gov)](https://www.noaa.gov/)

**[5]:** Lloyd, E. A., & Shepherd, T. G. (2020). Environmental catastrophes, climate change, and attribution. *Annals of the New York Academy of Sciences*, *1469*(1), 105-124.

**[6]:** Changnon, D., & Changnon Jr, S. A. (1998). Evaluation of weather catastrophe data for use in climate change investigations. *Climatic Change*, *38*(4), 435-445.

**[7]:** <http://www.ijeert.org/pdf/v2-i2/35.pdf>

**[8]:**Jewson, S. (2023). A new simulation algorithm for more precise estimates of change in catastrophe risk models, with application to hurricanes and climate change. *Stochastic Environmental Research and Risk Assessment*, *37*(7), 2631-2650.‏

**[9]:**Fang, W., Xue, Q., Shen, L., & Sheng, V. S. (2021). Survey on the application of deep learning in extreme weather prediction. *Atmosphere*, *12*(6), 661.‏

**[10]:** Zhang, J., Qiu, X., Li, X., Huang, Z., Wu, M., & Dong, Y. (2021). Support vector machine weather prediction technology based on the improved quantum optimization algorithm. *Computational Intelligence and Neuroscience*, *2021*(1), 6653659.

**[11]:** Singh, M. S., & O’Neill, M. E. (2022). Thermodynamics of the climate system. *Physics Today*, *75*(7), 30-37.

**[12]:** Born, P., & Viscusi, W. K. (2006). The catastrophic effects of natural disasters on insurance markets. *Journal of risk and Uncertainty*, *33*, 55-72.‏

**[13]:** Lawrence, M. G. (2005). The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, *86*(2), 225-234.

**[14]:**Aoun, N. (2022). Methodology for predicting the PV module temperature based on actual and estimated weather data. *Energy Conversion and Management: X*, *14*, 100182.‏

**[15]:** [Intensity Formula -Equation and Solved Examples (byjus.com)](https://byjus.com/intensity-formula/)

**[16]**: Wong, K. Y., Yip, C. L., & Li, P. W. (2008). Automatic identification of weather systems from numerical weather prediction data using genetic algorithm. Expert systems with Applications, 35(1-2), 542-555.

**[17]:** <https://arxiv.org/pdf/2308.10995>

**[18]:** <https://www.researchgate.net/publication/376521669_Review_of_AI-Enhanced_Weather_Forecasting_Application_for_Communication_Networks>

**[19]**: Bochenek, B., & Ustrnul, Z. (2022). Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. Atmosphere, 13(2), 180.

**[20]**: [Hurricane Katrina - Wikipedia](https://en.wikipedia.org/wiki/Hurricane_Katrina)

**[21]**: [2004 Indian Ocean earthquake and tsunami: Facts and FAQs | World Vision](https://www.worldvision.org/disaster-relief-news-stories/2004-indian-ocean-earthquake-tsunami-facts)

**[22]**: [2004 Indian Ocean earthquake and tsunami - Wikipedia](https://en.wikipedia.org/wiki/2004_Indian_Ocean_earthquake_and_tsunami)

**[23]**: [Typhoon Haiyan: Facts, FAQs, and how to help | World Vision](https://www.worldvision.org/disaster-relief-news-stories/2013-typhoon-haiyan-facts)

**[24]**: [Typhoon Haiyan - Wikipedia](https://en.wikipedia.org/wiki/Typhoon_Haiyan)

**[25]**: [2003 European heatwave - Wikipedia](https://en.wikipedia.org/wiki/2003_European_heatwave)

**[26]**: [2010 Pakistan floods - Wikipedia](https://en.wikipedia.org/wiki/2010_Pakistan_floods)

**[27]**: [Hurricane Sandy - Wikipedia](https://en.wikipedia.org/wiki/Hurricane_Sandy)

**[28]**: [Hurricane Maria - Wikipedia](https://en.wikipedia.org/wiki/Hurricane_Maria)

**[29]**: [Heatwave hits Nunavut with ‘astounding’ temperatures](https://nunatsiaq.com/stories/article/heatwave-hits-nunavut-with-astounding-temperatures/?utm_source=chatgpt.com)

**[30]**: [BOM issues flood warnings as Sydney and south-east NSW faces another major wet weather event, heavy rain expected - ABC News](https://www.abc.net.au/news/2024-06-06/nsw-major-rain-event-snow-season-delayed/103941002?utm_source=chatgpt.com)

**[31]**: [Patagonia cold snap sees parts of Argentina colder than Antarctica - Buenos Aires Herald](https://buenosairesherald.com/environment/patagonia-cold-snap-sees-parts-of-argentina-colder-than-antarctica?utm_source=chatgpt.com)