#### **Survey: Rainfall Prediction Precipitation, Review of Statistical Methods**

SARAH BENZIANE Computer Science, USTO MB, ALGERIA

Abstract: - Rainfall precipitation prediction is the process of using various models and data sources to predict the amount and timing of precipitation, such as rain or snow, in a particular location. This is an important process because it can help us prepare for severe weather events, such as floods, droughts, and hurricanes, as well as plan our daily activities. Processing rainfall data typically involves several steps, which may vary depending on the specific data set and research question. Here is a general overview of the steps involved: (1) Collecting data: Rainfall data can be collected using various methods, including rain gauges, radar, and satellite imagery. The data can be obtained from public sources, such as government agencies or research institutions. (2) Quality control: Before using the data, it's important to check for errors or inconsistencies. This may involve identifying missing or incomplete data, outliers, or inconsistencies in measurement units. Quality control can be performed manually or using automated software. (3) Pre-processing: Once the data has been quality controlled, it may need to be pre-processed for analysis. This may involve aggregating the data to a specific temporal or spatial resolution, such as daily, monthly, or annual averages, or converting the data to a specific format. (4) Analysis: The processed data can be used for various types of analysis, such as trend analysis, frequency analysis, or spatial analysis. These analyses can help to identify patterns, changes, or relationships in the data. (5) Visualization: Finally, the results of the analysis can be visualized using graphs, maps, or other types of visualizations to help communicate the findings.

Overall, processing rainfall data requires careful attention to detail and a clear understanding of the research question and data sources.

*Key-Words:* - Rainfall, Prediction, Precipitation, survey, regression, techniques.

Received: May 21, 2023. Revised: November 6, 2023. Accepted: December 12, 2023. Published: January 8, 2024.

#### **1** Introduction

Rainfall prediction or precipitation prediction refers to the process of using various scientific techniques to estimate the amount of rainfall that is expected to occur in a particular location over a certain period of time. This information is essential for a range of activities, including agriculture, water management, disaster management, and climate research. Several methods can be used to predict rainfall precipitation, including:

- 1. Statistical models, [1]: These models use historical weather data to predict future weather patterns, [2]. They can be relatively simple, such as linear regression models, or more complex, such as machine learning algorithms, [3]. These involve analyzing historical rainfall data to identify patterns and trends that can be used to make predictions. For example, regression analysis can be used to determine the relationship between rainfall and different meteorological variables.
- Numerical weather prediction models: These models use mathematical equations to simulate the behavior of the atmosphere and predict future weather patterns. They require large amounts of data and computing power to run but can provide more accurate predictions than statistical models, [4]. These use mathematical equations and algorithms to simulate atmospheric conditions and predict weather patterns. NWP models can provide high-resolution forecasts for specific locations and timeframes.
- 3. Ensemble forecasting: This approach involves running multiple models and combining their predictions to create a more accurate forecast. This approach involves running multiple NWP models with slight variations to the input data and parameters to produce a range of possible outcomes.
- 4. Doppler radar: This technology uses radar to detect precipitation and analyze its movement and intensity. It can provide real-time information on precipitation, which is useful for short-term

forecasting, [5]. These involve using machine learning algorithms, [6], to analyze large amounts of weather data to identify patterns and make predictions. Overall, accurate rainfall precipitation prediction requires a combination of different models and data sources, as well as experienced meteorologists and weather forecasters to interpret and communicate the information effectively.

Overall, rainfall prediction is a complex process that involves combining multiple data sources, analytical methods, and models. While no prediction method can be 100% accurate, advances in technology and data analysis techniques have significantly improved the accuracy of rainfall predictions in recent years.

#### 2 What is the Impact of Rainfall Prediction Precipitation in Real Life?

The impact of accurate rainfall prediction and precipitation forecasting is significant and extends across various sectors, influencing daily life, economic activities, and environmental management. Here are some key areas where the impact is felt:

- 1. Agriculture: Crop Planning and Yield Prediction: Farmers rely on rainfall predictions to plan their planting and harvesting schedules. Accurate forecasts help optimize irrigation schedules, choose appropriate crops, and mitigate the risks of drought or excessive rainfall, ultimately influencing crop yields and food production.
- 2. Water Resource Management: Reservoir Management: Predictions of rainfall and precipitation patterns are essential for managing reservoirs and water storage facilities. Reservoir operators use forecasts to plan water releases and allocations, ensuring a balance between water supply and demand.
- 3. Flood Management: Early Warning Systems: Accurate rainfall predictions are crucial for the development of effective flood early warning systems. Timely alerts based on forecasted precipitation can help communities and authorities prepare for potential flooding events, minimizing damage to property and saving lives.
- 4. **Urban Planning: Infrastructure Design:** City planners and engineers use rainfall predictions to design stormwater drainage systems and infrastructure capable of handling varying

precipitation intensities. This is critical for preventing urban flooding and ensuring the resilience of cities to extreme weather events.

- 5. Energy Production: Hydropower Generation: Hydropower plants depend on water availability, which is directly influenced by rainfall. Precise precipitation forecasts are crucial for optimizing energy production, managing reservoir levels, and ensuring a stable power supply.
- 6. Ecosystem Health: Ecological Balance: Rainfall patterns influence the health of ecosystems. Predicting precipitation helps in understanding and managing the impact on biodiversity, soil moisture, and overall ecosystem health.
- 7. **Disaster Preparedness: Natural Disasters:** In addition to floods, accurate rainfall predictions play a role in anticipating other natural disasters such as landslides and mudslides. Timely forecasts contribute to disaster preparedness and response efforts.
- 8. **Public Safety: Transportation Planning:** Rainfall predictions are crucial for planning transportation systems. Authorities can take preventive measures to ensure road safety, manage traffic, and reduce the risks of accidents during adverse weather conditions.
- 9. **Insurance and Risk Management: Risk Assessment:** Insurance companies use rainfall predictions to assess and manage risks associated with weather-related claims, such as those related to flooding or damage to crops.
- 10. Human Health: Disease Spread: Rainfall patterns can influence the spread of waterborne diseases and the breeding of disease vectors, such as mosquitoes. Accurate predictions allow public health authorities to implement preventive measures and respond to potential health risks.

In summary, accurate rainfall prediction and precipitation forecasting are foundational for informed decision-making in numerous sectors. The ability to anticipate and respond to changes in weather conditions has far-reaching implications for societal well-being, economic sustainability, and environmental conservation. Advances in forecasting technologies and continuous research in this field contribute to improving the reliability of predictions and, consequently, the effectiveness of related applications.

### **3** Which Model is more Effective at Forecasting Rainfall?

Researchers have utilized MLP and other machinelearning algorithms to forecast rainfall. Deep learning has a limited ability to predict rainfall, [7], especially when employing sensor-based datasets. According to recent polls, MLP is the most widely used neural network model for predicting rainfall, [8], [9], [10], [11]. Many researchers have attempted to integrate data-driven deep learning, [10], into weather forecasting at this time, and some tentative results have been obtained.

The effectiveness of rainfall forecasting models depends on various factors, including the specific characteristics of the region, the time scale of the forecast, and the available data. No single model is universally the most effective, and different models may perform better under different conditions. Here are some commonly used models for rainfall forecasting:

- 1 Numerical Weather Prediction (NWP) Models, [35]: NWP models, such as those developed by meteorological agencies like the European Centre for Medium-Range Weather Forecasts (ECMWF) or the National Centers for Environmental Prediction (NCEP). use mathematical equations to simulate the behavior of the atmosphere. These models consider various meteorological variables and can provide forecasts on different time scales, from short-term to medium-range.
- 2. Statistical Models, [36]: Statistical models use historical data and statistical techniques to make predictions. These models include techniques like regression analysis, autoregressive integrated moving averages (ARIMA), and other time series analysis methods. Statistical often used for models are short-term forecasting.
- 3. Machine Learning Models, [37]: Machine learning models, such as neural networks, support vector machines, and random forests, can capture complex patterns in large datasets. These models are particularly useful when dealing with non-linear relationships and can be effective for both short-term and long-term rainfall forecasting.
- 4. Ensemble Models, [38]: Ensemble models combine the predictions of multiple models to improve overall accuracy. This approach helps mitigate the weaknesses of individual models and provides more reliable forecasts.
- 5. **Hybrid Models, [39]:** Hybrid models combine elements of both physical and statistical models.

These models aim to leverage the strengths of each approach for improved forecasting accuracy.

6. **Regional Climate Models (RCMs), [40]:** RCMs focus on simulating climate at a regional scale. They provide detailed information about regional climate patterns and can be valuable for understanding local variations in rainfall.

The choice of the most effective model depends on the specific requirements of the forecasting task, the data available, and the desired forecast horizon. For example, NWP models are often used for medium-range forecasting, while statistical and machine-learning models may be more suitable for short-term predictions.

It's also important to note that continuous research and advancements in modeling techniques contribute to improvements in forecasting accuracy. The effectiveness of a model may vary over time as new methodologies and data sources are incorporated into the forecasting process. Validation against observed data for the specific region and time period of interest is crucial to assessing the performance of any forecasting model.

#### 4 Why Use Statistical Models?

A mathematical relationship exists between random and non-random variables in the statistical model. Data scientists can use statistical models to analyze raw data and generate intuitive visuals that help them discover correlations between variables and make predictions, [10].

Decision trees, time series, clustering, and logistic regression are a few examples of common statistical models, [11], [12], [13], [41].

### 5 Which Algorithms are Employed in the Forecasting of Rainfall?

Four classification methods-DT, [11], [14], NB, KNN, [3], [15], and SVM, [4], -are employed to make predictions. The accuracy of these categorization techniques is iteratively improved until it reaches its greatest potential, [14]. Rainfall prediction with maximum accuracy is a challenging task in the weather forecasting process. The use of machine learning techniques, [14], has increased the accuracy of rainfall prediction systems by exploring the hidden patterns of historical weather data. If due to any reason, the data which will be used for prediction is compromised, then the prediction cannot be trusted. Any type of malfunction in the weather sensor can also compromise the accuracy of the proposed rainfall prediction system. Therefore, a monitoring system to check the working of weather sensors has also been incorporated along with the information security system, which will ensure the integrity of the data until it is used for prediction.

Moreover, an appropriate feature selection technique would also be an effective addition to the system, which will ensure cost-effective prediction. Besides rainfall prediction, machine learning fusion will also be used for temperature prediction to efficiently utilize clean solar energy.

#### 6 Which Statistical Method is Suitable for Making Predictions?

Researchers utilize associational statistics as a technique to forecast outcomes and identify causes. They employ it to discover connections between various factors.

It is also used to assess whether researchers can draw conclusions and forecasts about one set of data from the features of another.

Several statistical techniques can be used for prediction, depending on the nature of the data and the problem being addressed. Some commonly used techniques include:

- 1. Regression analysis, [16]: This involves using a mathematical model to describe the relationship between a dependent variable and one or more independent variables, and using the model to predict the value of the dependent variable for new observations.
- 2. Time series analysis, [13]: This involves modeling the patterns and trends in time series data, such as stock prices or weather patterns, and using the model to forecast future values.
- 3. Machine learning algorithms: These include techniques such as decision trees, random forests, support vector machines (SVM), [17], and artificial neural networks (ANN), [18], which can be used to classify or predict outcomes based on historical data.
- 4. Bayesian inference: This involves using Bayes' theorem to update prior beliefs about a parameter or event based on new data, and using the updated beliefs to make predictions.
- 5. Survival analysis: This involves modeling the time to an event of interest, such as disease progression or customer churn, and using the model to predict the likelihood of the event occurring in the future.
- 6. Ensemble methods: This involves combining multiple models to improve prediction accuracy, such as using a weighted average of the

predictions from different models. The choice of technique depends on the nature of the data, the problem being addressed, and the desired level of accuracy and interpretability. It's important to carefully evaluate the performance of the chosen technique and to assess its predictive accuracy before making any decisions based on the results.

Several statistical techniques can be used for rainfall prediction, depending on the nature of the data and the problem being addressed. Some commonly used techniques for rainfall prediction include:

- 1. Multiple Linear Regression: This involves modeling the relationship between rainfall and several independent variables such as temperature, humidity, wind speed, and atmospheric pressure. Multiple linear regression, [16], can be used to predict the amount of rainfall on a given day.
- 2. Autoregressive Integrated Moving Average (ARIMA) models: This is a time series, [13], analysis technique that can be used to model the temporal patterns in rainfall data, and forecast future values. ARIMA models can capture the trend, seasonality, and autocorrelation in the data.
- 3. Support Vector Machines (SVM), [17]: This is a machine learning algorithm that can be used to classify whether a particular day will have rainfall or not, or to predict the amount of rainfall on a given day. SVM can handle complex and nonlinear relationships between the features and the rainfall.
- 4. Artificial Neural Networks (ANN), [18]: These are machine learning models that can be used for rainfall prediction by modeling the nonlinear relationships between the input features and the rainfall.
- 5. Random Forest, [19]: This is a machine-learning algorithm that can be used to model the complex relationships between the input features and the rainfall. Random forests can handle a large number of input features and can provide variable importance measures.

The choice of technique depends on the nature of the data, the problem being addressed, and the desired level of accuracy and interpretability. It's important to carefully evaluate the performance of the chosen technique and to assess its predictive accuracy before making any decisions based on the results.

# 7 Which Regression Model Offers the Most Accurate Forecast?

- 1. Regular Regression: It ranks among the most used machine learning regression algorithms. The output variables are predicted by a significant variable from the data set (future values).
- 2. Ridge Regression, another widely used linear regression approach in machine learning is ridge regression. A linear regression ML technique is utilized when only one independent variable is used to predict the outcome.
- 3. Regression in a Neural Network: You must all be aware of how effective neural networks are at forming assumptions and predictions. In a neural network, each node has a unique activation function that determines the node's output based on a set of inputs. One can alter the last activation function to convert a neural network into a regression model. You can use "Keras," which is the ideal Python library for creating neural networks in machine learning.

# 8 Which Six Statistical Forecasting Techniques are There?

I look at six criteria for choosing forecasting methods: practicality, market acceptance, structured judgment, statistical standards, relative performance, and pre-existing research-based recommendations. These methods can be applied singly or in combination. There are many statistical forecasting methods, and the choice of method depends on the data being analyzed and the specific requirements of the forecasting problem. However, some of the commonly used statistical forecasting methods are:

- 1. Time-series forecasting: This method uses historical data to forecast future values of a variable based on patterns and trends observed in the past. Time series models may include components such as trend, seasonality, and cyclical fluctuations.
- 2. Exponential smoothing: This method is a time-series forecasting technique that calculates the weighted average of past observations to forecast future values. It is commonly used for short-term forecasting and can be used for both trended and non-trended time series.
- 3. Regression analysis: This method is used to forecast the relationship between a dependent variable and one or more independent variables. It is commonly used when there is a

known causal relationship between the variables.

- 4. ARIMA modeling: ARIMA (Autoregressive Integrated Moving Average) is a time-series forecasting technique that models the autocorrelation and trend in the data to forecast future values. It is commonly used for medium-term forecasting.
- 5. Neural networks, [20], [21]: This method uses a network of artificial neurons to learn the patterns and relationships in the data and make forecasts. It is commonly used for long-term forecasting and can handle nonlinear relationships between variables.
- 6. Bayesian forecasting: This method uses probability theory to forecast future values based on prior knowledge and new information. It is commonly used when there is uncertainty in the data and can be used for both short- and long-term forecasting.

There are several statistical methods, [22], used for forecasting rainfall. Here are six commonly used methods:

- 1. Persistence method: This method assumes that the current rainfall value will be the same as the previous value, which is suitable for short-term forecasts in areas with stable weather patterns.
- 2. Climatology method: This method uses longterm historical rainfall data to calculate the average rainfall for a specific location and time period. The forecast for the next period is then based on this historical average.
- 3. Autoregressive Integrated Moving Average (ARIMA) method: This method is used to model the temporal dependence of rainfall data and can forecast the next value based on a combination of past values and their differences.
- 4. Support Vector Machine (SVM) method, [23]: This method uses a nonlinear function to map input variables to the output variables and can predict rainfall based on past data.
- 5. Random Forest (RF) method, [19]: This method builds multiple decision trees from the historical rainfall data and can make predictions based on the collective results of these trees.
- 6. It is important to note that no single method is perfect for all situations and that the choice of method will depend on factors such as the availability and quality of historical data, the location, and the time period being forecasted.

Artificial Neural Networks (ANN) method, [8]: This method is based on a network of interconnected processing elements that can learn from historical rainfall data and make predictions based on the learned patterns.

51

#### 9 What are the 2 Most Common Models of Regression Analysis?

The two most common models of regression analysis for rainfall prediction are:

- 1. Multiple Linear Regression: This model is used when multiple independent variables can affect rainfall. The model assumes that the relationship between the dependent variable (rainfall) and multiple independent variables (such as temperature, humidity, wind speed, etc.) can be represented by a linear equation. The equation of the line is represented as rainfall =  $\beta 0 + \beta 1x1 + \beta 2x2 + \beta 3x3 + ... + \beta nxn + \varepsilon$ , where  $\beta 0$  is the intercept,  $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ , ...  $\beta n$  are the slopes, and  $\varepsilon$ is the error term.
- 2. Nonlinear Regression: This model is used when the relationship between the dependent variable (rainfall) and independent variables is not linear. Nonlinear regression models can capture more complex relationships between the variables, such as logarithmic, exponential, or polynomial relationships. These models are more flexible than linear regression models and can often provide better predictions.

Based on historical data and other environmental conditions, both models can be used to forecast rainfall. The type of data, the research issue, and the model's underlying assumptions all influence the model choice. It is crucial to remember that a variety of variables, such as the quality and quantity of data, the choice of predictors, and the modeling strategies employed, affect how accurately rainfall will be predicted.

#### 10 What is an Improvement over Linear Regression?

In general cases, **Decision trees** will have better average accuracy. For categorical independent variables, decision trees are better than linear regression.

The use of machine learning methods, such as Artificial Neural Networks (ANN), [8], Support Vector Machines (SVM), [23], and Random Forests, is an improvement over linear regression rainfall prediction (RF). In addition to simulating nonlinear correlations between rainfall and environmental parameters, these algorithms can capture more intricate patterns and interactions that linear regression models can miss.

For instance, ANNs, which draw their design cues from the structure and operation of the human brain, can recognize patterns in previous data and forecast the future using those patterns. SVMs, [24], can forecast rainfall based on historical data and employ a nonlinear function to transfer input factors to the output variable. Based on the findings of all the decision trees that are built by Random Forests, [25], using historical rainfall data, predictions can be made.

Overall, these machine learning methods are more adaptable and versatile than linear regression models, and they can frequently estimate rainfall more precisely and consistently. They might be harder to implement and interpret than linear regression models, and they might need more data and processing resources. The exact research issue, the type of data, and the trade-off between accuracy and complexity all play a role in the model selection process.

# 11 Is Linear Regression a Useful Tool for Forecasting?

Linear regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a longestablished statistical procedure, the properties of linear regression models are well understood and can be trained very quickly.

Linear regression can be useful for forecasting rainfall to some extent, but it may not be the best tool for this specific task.

Rainfall prediction involves many complex factors, including atmospheric pressure, humidity, temperature, wind patterns, and more. While some of these factors may have a linear relationship with rainfall, others may not, and the interactions between these factors may be highly nonlinear. Therefore, a more sophisticated model that takes into account the complex relationships between these variables may be necessary for accurate rainfall prediction.

Furthermore, the accuracy of linear regression models for rainfall prediction may be limited by the availability and quality of data. Rainfall data can be highly variable across different locations and time periods, and missing or incomplete data can further reduce the accuracy of linear regression models.

In summary, while linear regression may be a useful tool for some aspects of rainfall prediction, it may not be the most appropriate tool for accurately forecasting rainfall. More sophisticated statistical and machine learning models, such as time series analysis or artificial neural networks, may be more appropriate for this task.

### 12 Which Four Statistical Approaches are there?

It all boils down to applying the proper statistical analysis techniques, [26], when processing and gathering data samples to find patterns and trends.

There are five options available for this analysis: mean, standard deviation, regression, hypothesis testing, and sample size calculation.

There are several statistical approaches for rainfall precipitation prediction. Here are four common ones:

- 1. Autoregressive Integrated Moving Average (ARIMA): ARIMA is a time series forecasting method that can be used to model and predict the behavior of a variable over time. It involves analyzing past patterns in the data to identify trends, seasonality, and other patterns that can be used to make predictions about future rainfall.
- 2. Generalized Linear Models (GLMs): GLMs are a family of statistical models, [1], that can be used to model and predict the relationship between a dependent variable and one or more independent variables. They are commonly used in environmental science and hydrology to model rainfall as a function of factors such as temperature, humidity, and wind patterns.
- 3. Artificial Neural Networks (ANNs): ANNs are machine learning models that can be used to predict future rainfall based on historical data. They are particularly useful for modeling complex relationships between variables that may not be easily modeled using traditional statistical methods, [22].
- 4. Support Vector Machines (SVMs), [24]: SVMs are another machine learning technique that can be used to predict future rainfall based on historical data. They are particularly useful for modeling nonlinear relationships between variables and can be used to identify patterns and trends in the data that may not be apparent using traditional statistical methods.

It's worth noting that the effectiveness of each of these approaches may depend on the specific data set and the particular characteristics of the rainfall patterns being analyzed. Therefore, it may be useful to try multiple approaches and compare their performance before selecting a final model for precipitation prediction.

# 13 What are the Three Methods of Forecasting?

There are three basic types—qualitative techniques, time series analysis and projection, and causal models.

There are several methods for forecasting rainfall, but here are three common methods:

- 1. Statistical Methods, [4]: These methods use historical data on rainfall patterns to develop statistical models that can forecast future rainfall. Statistical methods include regression analysis, time-series analysis, and artificial neural networks.
- 2. Dynamic Methods: These methods use mathematical models to simulate atmospheric processes and predict future weather patterns. Dynamic methods include numerical weather prediction, and weather forecasting models like the Global Forecast System (GFS), and the European Centre for Medium-Range Weather Forecasts (ECMWF) model.
- 3. Remote Sensing Methods: These methods use remote sensing data from satellites, radars, and other sensors to estimate rainfall. Remote sensing methods include using satellite imagery to track cloud formations, microwave radiometers to estimate the amount of water vapor in the atmosphere, and Doppler radar to track precipitation.

It's worth noting that these methods are often used in combination to improve the accuracy of rainfall forecasts.

### 14 What are the 4 Basic Types of Forecasting?

Four common types of forecasting models, [27]

- Time series model.
- Econometric model.
- Judgmental forecasting model.
- The Delphi method.

There are several types of forecasting methods for rainfall prediction, but here are four basic types:

- 1. Short-term Forecasting, [5]: This type of forecasting predicts rainfall over a period of a few hours up to three days in advance. Short-term forecasts are based on current weather conditions and use models such as radar and satellite data, as well as numerical weather prediction models.
- 2. Medium-term Forecasting, [28]: This type of forecasting predicts rainfall over a period of three to ten days in advance. Medium-term forecasts use models that are based on climate

and weather patterns, as well as the analysis of atmospheric conditions and sea surface temperatures.

- 3. Long-term Forecasting, [29]: This type of forecasting predicts rainfall over a period of several weeks to months in advance. Long-term forecasts are based on climate models that take into account the factors that influence climate patterns, such as El Niño and La Niña.
- 4. Seasonal Forecasting, [30], [31]: This type of forecasting predicts rainfall patterns over a period of three months to a year in advance. Seasonal forecasts are based on climate models that take into account the interactions between the atmosphere, ocean, and land surface, as well as the factors that influence climate patterns.

It's worth noting that different forecasting methods may be more appropriate for different regions and time scales and that the accuracy of these forecasts can vary depending on a range of factors, such as the quality and quantity of available data, and the complexity of the atmospheric and climate systems.

### 15 What are the three Most Common Types of Regression Models?

The different types of regression in machine learning, [3], techniques are explained below in detail:

- Linear Regression. Linear regression is one of the most basic types of regression in machine learning. ...
- Logistic Regression. ...
- Ridge Regression. ...
- Lasso Regression.

Regression analysis is a statistical method commonly used in rainfall precipitation prediction. Here are three types of regression models that are commonly used:

- 1. Linear regression: This model assumes a linear relationship between the independent variables and the dependent variable. In the context of rainfall prediction, linear regression may be used to model the relationship between rainfall and variables such as temperature, humidity, and wind speed.
- 2. Multiple regression: This model involves more than one independent variable and assumes a linear relationship between the independent variables and the dependent variable. In rainfall prediction, multiple regression may be used to model the relationship between rainfall and a

combination of variables such as temperature, humidity, wind speed, and atmospheric pressure.

3. Logistic regression: This model is used when the dependent variable is binary or categorical. In the context of rainfall prediction, logistic regression may be used to predict the probability of rainfall occurrence or non-occurrence based on a combination of variables such as temperature, humidity, and wind speed.

It's worth noting that regression models are just one type of statistical method used in rainfall prediction, and they are often used in combination with other methods to improve the accuracy of the forecasts.

## 16 Do Predictions Well using Logistic Regression?

Logistic regression is commonly used for prediction and classification problems. Some of these use cases include: Fraud detection: Logistic regression models can help teams identify data anomalies, which are predictive of fraud.

If you want to predict whether it will rain or not based on a mix of factors like temperature, humidity, and wind speed, logistic regression can be a valuable tool.

The quality and quantity of the data, the selection of variables, and the assumptions and constraints of the model, however, all affect how accurate the forecasts are.

In some cases, logistic regression may not be the best approach for rainfall prediction. For example, if the dependent variable is continuous (e.g., the amount of rainfall), a different type of regression model such as linear or multiple regression may be more appropriate. Additionally, there may be other statistical methods such as time series analysis, machine learning algorithms, [29], or ensemble modeling techniques that can provide more accurate rainfall predictions depending on the specific context and data available.

Generally, the selection of a regression model for predicting rainfall is influenced by the particular research topic, the quantity and quality of the data, and the assumptions and constraints of the model. Before choosing the best strategy, it is crucial to carefully assess several approaches and compare their effectiveness.

### 17 What are the 5 Basic Methods of Statistical Analysis?

The five basic methods are mean, standard deviation, regression, hypothesis testing, and sample size determination.

To predict rainfall, many statistical analytic techniques are applied. Here are five simple approaches:

- 1. Descriptive statistics: This method involves summarizing and describing the properties of a dataset, such as the mean, median, mode, standard deviation, and range. Descriptive statistics can be used to explore the distribution and variability of rainfall data.
- 2. Correlation analysis: This method involves examining the relationship between two variables, such as rainfall and temperature, and determining the strength and direction of the relationship. Correlation analysis can be used to identify patterns and associations in rainfall data.
- 3. Regression analysis: This method involves modeling the relationship between a dependent variable, such as rainfall, and one or more independent variables, such as temperature and humidity. Regression analysis can be used to predict rainfall based on other weather variables.
- 4. Time series analysis: This method involves analyzing a sequence of observations over time, such as daily or monthly rainfall data, [32], to identify trends, patterns, and seasonal variations. Time series analysis can be used to forecast future rainfall patterns.
- 5. Spatial analysis: This method involves analyzing rainfall data across a geographic region, such as a country or continent, to identify patterns and variations in rainfall amounts and distributions. Spatial analysis can be used to identify areas that are more susceptible to drought or flooding.

It's important to note that these strategies are just a few of the numerous statistical ones used in rainfall prediction; they are frequently combined with other techniques to increase forecast accuracy.

The research objective, the quantity and quality of the data, as well as the presumptions and constraints of each approach, all influence the method that is chosen.

## 18 What is the Statistical Approach to Weather Prediction?

However, in this part, we use the term statistical weather forecasting to mean **forecasting through the use of a formal statistical analysis of the data**,

#### [26], with the results of that analysis being clearly stated.

The statistical method for forecasting weather conditions makes use of statistical models and historical weather data, [12].

This strategy assumes that correlations between weather variables and historical weather patterns can be used to forecast future weather patterns.

Regression analysis, time series analysis, machine learning algorithms, [29], and other statistical methods that find patterns and trends in previous weather data and use them to forecast future weather conditions are examples of statistical models.

One advantage of the statistical approach to weather prediction is that it can be relatively simple and easy to implement. It also works well for shortterm weather forecasting, such as predicting weather conditions for the next few days.

However, the accuracy of the statistical approach depends on the quality and quantity of the data used to train the models, as well as the assumptions and limitations of the models themselves. The statistical approach may not be as effective for long-term weather forecasting or for predicting extreme weather events, which may require more complex and sophisticated models. Therefore, the statistical approach is often used in combination with other methods, such as numerical weather prediction models, to provide more accurate and reliable weather forecasts.

## 19 What is the Most Accurate Forecasting Method?

In contrast to the moving averages method, where the weight for each period is fixed, exponential smoothing gives the most weight to the most recent projection, which should make it the most accurate in forecasting demand.

There is no one "most accurate" forecasting technique for predicting rainfall because a forecasting technique's performance might vary depending on the geography, the season, the length of the forecast, and the kind of precipitation.

There are, however, several widely used forecasting techniques that have been proven to be successful in predicting rainfall.

Numerical weather prediction (NWP), which simulates the behavior of the atmosphere and forecasts future weather patterns, is one frequently employed technique. To create predictions, NWP models include physics equations, atmospheric measurements, and data assimilation methods.

One commonly used method is numerical weather prediction (NWP), which involves using computer models to simulate the behavior of the atmosphere and predict future weather patterns. NWP models use a combination of physical equations, atmospheric observations, and data assimilation techniques to make predictions.

Another commonly used method is statistical forecasting, which involves analyzing historical data to identify patterns and trends that can be used to make predictions. Statistical forecasting methods include regression analysis, time-series analysis, and artificial neural networks, [33].

Alternative techniques for predicting rainfall include ensemble forecasting (combining numerous models or predictions to boost accuracy), remote sensing techniques (such as using satellite data to estimate precipitation), and hybrid methods that combine NWP and statistical approaches.

The most accurate forecasting technique may ultimately depend on the specifics of the forecast; hence it is frequently helpful to employ several techniques and compare the outcomes to increase prediction accuracy.

## 20 Which Statistical Test Should I Use?

It is safer to apply non-parametric tests if the data distribution is not normal or if one is unsure of the distribution. A multiple-group comparison test, such as one-way analysis of variance (ANOVA) or the Kruskal-Wallis test, should be employed first when comparing more than two sets of numerical data.

Different statistical tests can be used for rainfall precipitation prediction, and the choice of test will depend on the specific question and data you have. Here are a few common statistical tests that can be used for rainfall precipitation prediction:

- 1. Correlation analysis: This test is used to examine the relationship between two variables, in this case, the relationship between rainfall and other factors that might affect it, such as temperature, humidity, or pressure.
- 2. Regression analysis: This test is used to model the relationship between two or more variables, to predict one variable based on the values of other variables. In this case, you might use regression analysis to predict rainfall based on other weather-related factors.

- 3. Time series analysis: This test is used to analyze data that is collected over time, to identify patterns or trends in the data. In this case, you might use time series analysis to predict rainfall based on historical rainfall data and other weather-related factors.
- 4. Machine learning algorithms, [34]: Machine learning algorithms, such as random forests or support vector machines, can also be used for rainfall precipitation prediction. These algorithms can learn complex relationships between variables and make predictions based on those relationships.

Ultimately, the choice of statistical test will depend on the specific question you are trying to answer and the data you have available.

#### 21 Conclusion

This paper presents a survey about the precipitation prediction rainfall with different types of questions and answers. We can conclude that there is no one optimal model for predicting rainfall because the model chosen will vary depending on the region, the forecast's time frame, the data that is available, and the application.

In the real world, meteorologists and hydrologists frequently combine a variety of models and data sources to produce the most precise forecasts.

Having stated that some of the most popular models for predicting rainfall include:

- 1. The Global Forecast System (GFS), the European Centre for Medium-Range Weather Forecasts (ECMWF), and the North American Mesoscale (NAM) model are examples of numerical weather prediction (NWP) models.
- 2. Statistical models include multiple linear regression, support vector machines, and autoregressive integrated moving averages (ARIMA).
- 3. Artificial Neural Networks (ANNs), a class of machine learning algorithms, [34], can recognize intricate patterns in data and make precise predictions.
- 4. Many models are used in ensemble forecasting, which combines their projections to make a more precise prediction.

The best model for predicting rainfall ultimately depends on the particular application and the data at hand; to provide the most precise forecast, it may be necessary to combine many models and methodologies. The choice of regression model depends on the nature of the data, the number and type of independent variables, and the goals of the analysis. The most commonly used methods for processing data rainfall are for Spatial Averaging of Precipitation Data: (a) Arithmetic average, (b) Normal ratio method, (c) Distance power method, (d) Thiessen polygon method, and (e) Isohyetal method.

#### 22 Perspectives

Predicting rainfall and precipitation is a crucial aspect of weather forecasting, agriculture planning, water resource management, and disaster preparedness. There are various methods, including statistical approaches, that researchers can explore for improving rainfall prediction. Here are some potential future research directions:

- 1. **Machine Learning Techniques:** Explore advanced machine learning algorithms, such as deep learning models (e.g., neural networks) for rainfall prediction. These models can capture complex patterns and relationships in large datasets, potentially improving prediction accuracy.
- 2. Ensemble Methods: Investigate ensemble methods that combine multiple models to enhance prediction performance. Ensemble techniques, such as bagging and boosting, can help mitigate the weaknesses of individual models and provide more robust predictions.
- 3. **Hybrid Models:** Develop hybrid models that integrate statistical methods with machine learning approaches. Combining the strengths of both approaches may lead to more accurate and reliable predictions.
- 4. **Spatial-Temporal Modeling:** Focus on developing models that can capture the spatial and temporal dynamics of rainfall patterns. This involves considering not only the quantity of precipitation but also its distribution across different geographical locations and over time.
- 5. **Remote Sensing Data:** Integrate remote sensing data, such as satellite imagery and radar observations, into rainfall prediction models. These sources can provide valuable information on atmospheric conditions, cloud cover, and precipitation that may improve the accuracy of predictions.
- 6. **Feature Engineering:** Explore novel approaches to feature engineering, extracting meaningful information from various meteorological and environmental variables.

Feature selection and extraction techniques can help identify the most relevant factors influencing rainfall.

- 7. **Uncertainty Quantification:** Research methods for quantifying uncertainty in rainfall predictions. Uncertainty estimates are crucial for decision-makers, as they provide insights into the reliability of the predictions and help in making informed decisions in the face of uncertainty.
- 8. **Long-Term Predictions:** Extend the prediction horizon for rainfall forecasts, moving beyond short-term predictions to longer-term forecasts. This could be particularly valuable for agricultural planning and water resource management.
- 9. Climate Change Impact: Investigate the impact of climate change on rainfall patterns and develop models that can account for changing climate conditions. Understanding how climate change influences precipitation is essential for adapting to future environmental conditions.
- 10. **Citizen Science and Crowdsourcing:** Explore the integration of citizen science and crowdsourced data for rainfall prediction. Engaging the public in data collection can provide additional information and improve the coverage of meteorological observations.

Continued collaboration between meteorologists, climatologists, statisticians, and data scientists is essential for advancing research in rainfall prediction and precipitation modeling. Additionally, open data initiatives and the sharing of datasets can contribute to the development and validation of more robust models.

References:

- [1] Ashby, S.A., M.A. Taylor, and A.A. Chen, *Statistical models for predicting rainfall in the Caribbean.* Theoretical and Applied Climatology, 2005. 82: p. 65-80, 0177-798X.
- [2] Glahn, H.R., Statistical weather forecasting, in Probability, statistics, and decision making in the atmospheric sciences. 2019, CRC Press. p. 289-335, 0429303084.
- [3] Patil, R. and G. Bedekar, *Comparative Analysis of Machine Learning Algorithms for Rainfall Prediction*, in *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021*. 2022, Springer. p. 833-842.

- [4] Niu, J. and W. Zhang. *Comparative analysis* of statistical models in rainfall prediction. 2015. IEEE.
- [5] Tran Anh, D., Downscaling rainfall using deep learning long short- term memory and feedforward neural network. International Journal of Climatology, 2019. 39(10): p. 4170-4188 0899-8418.
- [6] Praveen, B., Analyzing trend and forecasting of rainfall changes in India using nonparametrical and machine learning approaches. Scientific reports, 2020. 10(1): p. 1-21 20452322.
- [7] Van, S.P., *Deep learning convolutional neural network in rainfall–runoff modelling*. Journal of Hydroinformatics, 2020. 22(3): p. 541-561 1464-7141.
- [8] Nayak, D.R., A. Mahapatra, and P. Mishra, *A* survey on rainfall prediction using artificial neural network. International journal of computer applications, 2013. 72(16 0975-8887).
- [9] Sundaravalli, N. and A. Geetha, A study & survey on rainfall prediction and production of crops using data mining techniques. International Research Journal of Engineering and Technology (IRJET), 2016. 3(12): p. 1269-1274.
- [10] Ren, X., Deep learning-based weather prediction: a survey. Big Data Research, 2021. 23: p. 100178, 2214-5796.
- [11] Endalie, D., G. Haile, and W. Taye, *Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia.* Water Supply, 2022. 22(3): p. 3448-3461, 1606-9749.
- [12] Zaw, W.T. and T.T. Naing, *Empirical statistical modeling of rainfall prediction over Myanmar*. World Academy of Science, Engineering and Technology, 2008. 2(10): p. 500-504.
- [13] Mishra, N., A comprehensive survey of data mining techniques on time series data for rainfall prediction. Journal of Ict Research & Applications, 2017. 11(2 2337-5787).
- [14] Rahman, A.-u., Rainfall prediction system using machine learning fusion for smart cities. Sensors, 2022. 22(9): p. 3504 1424-8220.
- [15] Huang, M., A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. Advanced Engineering Informatics, 2017. 33: p. 89-95 1474-0346.
- [16] Anshuka, A., F.F. van Ogtrop, and R. Willem Vervoort, *Drought forecasting through statistical models using standardised precipitation index: a systematic review and*

*meta-regression analysis.* Natural Hazards, 2019. 97: p. 955-977 0921-030X.

- [17] Hamidi, O., A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. Theoretical and applied climatology, 2015. 119: p. 723-731, 0177-798X.
- [18] Kumar, S., T. Roshni, and D. Himayoun, A comparison of emotional neural network (ENN) and artificial neural network (ANN) approach for rainfall-runoff modelling. Civil Engineering Journal, 2019. 5(10): p. 2120-2130 2476-3055.
- [19] Li, M., Estimating annual runoff in response to forest change: A statistical method based on random forest. Journal of Hydrology, 2020. 589: p. 125168 0022-1694.
- [20] Benevides, P., J. Catalao, and G. Nico, Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. Remote Sensing, 2019. 11(8): p. 966 2072-4292.
- [21] Poornima, S. and M. Pushpalatha, Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. Atmosphere, 2019. 10(11): p. 668 2073-4433.
- [22] Barlow, M., North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. Climate Dynamics, 2019. 53: p. 6835-6875 0930-7575.
- [23] Hussein, E., M. Ghaziasgar, and C. Thron. Regional rainfall prediction using support vector machine classification of large-scale precipitation maps. 2020. IEEE.
- [24] Neelakandan, S. and D. Paulraj, An automated exploring and learning model for data prediction using balanced CA-SVM. Journal of Ambient Intelligence and Humanized Computing, 2021. 12: p. 4979-4990 1868-5137.
- [25] Liu, Y., Short-term rainfall forecast model based on the improved BP–NN algorithm. Scientific reports, 2019. 9(1): p. 1-12 2045-2322.
- [26] Sharif, H.O., Statistical analysis of radar rainfall error propagation. Journal of Hydrometeorology, 2004. 5(1): p. 199-212 1525-7541.
- [27] Zhao, Q., An improved rainfall forecasting model based on GNSS observations. IEEE Transactions on Geoscience and Remote

Sensing, 2020. 58(7): p. 4891-4900 0196-2892.

- [28] Gebremichael, M., The skills of medium-range precipitation forecasts in the Senegal River basin. Sustainability, 2022. 14(6): p. 3349 2071-1050.
- [29] Diez-Sierra, J. and M. Del Jesus, *Long-term* rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. Journal of Hydrology, 2020. 586: p. 124789 0022-1694.
- [30] Sattari, M.T., Estimation of monthly and seasonal precipitation: A comparative study using data-driven methods versus hybrid approach. Measurement, 2021. 173: p. 108512 0263-2241.
- [31] Badr, H.S., B.F. Zaitchik, and S.D. Guikema, *Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel.* Journal of Applied meteorology and climatology, 2014. 53(3): p. 614-636 1558-8424.
- [32] Pham, B.T., Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 2020. 237: p. 104845 0169-8095.
- [33] Bukhari, A.H., *Neuro-fuzzy modeling and prediction of summer precipitation with application to different meteorological stations*. Alexandria Engineering Journal, 2020. 59(1): p. 101-116 1110-0168.
- [34] Chowdhuri, I., Torrential rainfall-induced landslide susceptibility assessment using machine learning and statistical methods of eastern Himalaya. Natural Hazards, 2021. 107: p. 697-722 0921-030X.
- [35] Brotzge, J. A., Berchoff, D., Carlis, D. L., Carr, F. H., Carr, R. H., Gerth, J. J., & Wang, X. (2023). Challenges and Opportunities in Numerical Weather Prediction. Bulletin of the American Meteorological Society, 104(3), E698-E705.
- [36] Finkel, J., Gerber, E. P., Abbot, D. S., & Weare, J. (2023). Revealing the statistics of extreme events hidden in short weather forecast data. AGU Advances, 4(2), e2023AV000881.
- [37] Rakholia, R., Le, Q., Ho, B. Q., Vu, K., & Carbajo, R. S. (2023). Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. Environment International, 173, 107848.
- [38] Mayer, M. J., & Yang, D. (2023). Pairing ensemble numerical weather prediction with ensemble physical model chain for

probabilistic photovoltaic power forecasting. Renewable and Sustainable Energy Reviews, 175, 113171.

- [39] Venkatachalam, K., Trojovský, P., Pamucar, D., Bacanin, N., & Simic, V. (2023). DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM). Expert Systems with Applications, 213, 119270.
- [40] Barde, V., Nageswararao, M. M., Mohanty, U. C., & Panda, R. K. (2023). Performance of the CORDEX-SA regional climate models in simulating summer monsoon rainfall and future projections over East India. Pure and Applied Geophysics, 180(3), 1121-1142.
- [41] Kanchai, T., Tepkasetkul, N., Pongsart, T., & Klongdee, W. (2023). Rainfall Data Fitting based on An Improved Mixture Cosine Model with Markov Chain. networks, 11(12), 13.

#### Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

From the drafting of the problem to the final conclusions and solution, the author participated in every step of the current investigation.

#### Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

#### **Conflict of Interest**

The authors have no conflicts of interest to declare.

#### Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en \_US