Ministry of Education and Science of Ukraine
V. N. Karazin Kharkiv National University
School of Mathematics and Computer Science
Department of Theoretical and Applied Informatics

Qualification work

Master

On the topic: Models and algorithms of applied statistics in analyses and prediction of global climate changes

Done by: 2-th year student, group MCS64
specialty - Computer Sciences and
Information Technologies,
educational program: "Informatics"
ZONG PENGFEI
Supervisor: Natalya Kizilova
Reviewer: Volodymyr Khalturin

Kharkiv, 2024

Abstract

This research provides a concrete exploration of climate forecasting models. This study provides comparison of three modelling methods: statistical approach, machine learning approach and Differential equation approach. Modelling will be based on climate data presenting climatic variables of China regions. The study's focal point is on China's climatic data including temperatures and precipitations of difference provinces of the country. By using climatic data from various regions of China, the study provides the comparison of the modelling techniques in terms of accuracy and effectiveness. Statistical modelling gives results which are easy to interpret though the results are only limited to linear climatic datasets. Machine learning shows its strength on handling complex datasets and its ability to adapt to variety of climate datasets. Differential equations on its turn provides focuses on the dynamicity of climate data trends. This approach requires high computational power as it requires detailed data. The final result of this study presents the comparative understanding of each modelling techniques and their suitability in different predictive reasons such whether long-term or short-term, and efficiencies of the methods in handling region specific data for variety of climatic data. The main aim of this research presents its weight on enhancing climate forecasting methods, contributing to valuable directions to policy-makers, environmental elites, and preparation for disaster planners.

Table of Contents

Contents

CHAPTER 1

1. Introduction

1.1. Background

Climate change is one of the disturbing issues in the current century. Climate change extends its impacts to different sectors in real time situations including natural ecosystems, human activities and the global economy (Shen et al., 2020). Having knowledge and valuable estimation of the sequence of the climate parameters such temperatures, precipitations and humidity is important. understanding the historical trend of data changes helps in estimating future weather changes and thus get prepared for the future climate impacts, specifically to areas which are more susceptible to climate changes (Jaseena and Kavoor,2022). China having an area of approximately 9.60 million square kilometers with population of 1.42 billion, experience adverse climate changes which spreads across its provinces. China's evenly distributed climatic zone provides ideal baseline for the analysis of the climatic trends and thus predicting region-specific predictive models (Hossain and Mahmood,2020). Analysis of the historical climatic trends facilitates ideal room for making decisions which will initiate planning of the country's future economy corresponding to its population (Jiang et al., 2022). Sound and informative weather forecasting techniques provides information for policy making thus guiding infrastructure development and thus enhancing proper preparations of disasters that might rise as a result of the climate change. Such disasters include heatwaves, droughts and heavy precipitations that might cause floods.

Anciently, statistical techniques have been utilized to address short-term climate forecasting. Statistical models depend greatly on historical data to determine trends and sequences that are utilized to estimate future weather parameters (Vannitsem et al., 2021). Machine learning techniques are the recent technologies applied to the weather forecasting sector thus facilitating utilization of large volume of datasets to increase accuracy and efficiency. Utilization of machine learning in weather forecasting helps in predictions to adapt to the complex non-linear data patterns which are common in weather forecasting (Thilker et al., 2021). Differential models and equations are also utilized in weather forecasting analysis to increase the accuracy of weather forecasting. Equation based models such as General Circulation Models (GCMs) are applicable in capturing dynamic profile of climate systems but are highly resource intensive (Jiang. Each approach used to increase accuracy in weather forecasting, there is no substantial research to compare the effectiveness of each method across the varying temporal scales and regional climates.

1.2. Problem Statement

Reliability and consistency of different climate prediction techniques is the key point of concern in availing the consensuses of the method. Statistical methods of predicting climatic parameters are limited to estimation of short-term weather forecasting but fails to provide recommendable accuracy for long-term weather forecasting in complex climatic systems (Jiao et al., 2021). Machine learning trains excel in handling large volume of weather forecast datasets. The nature at which datasets are trained in excel reduces the degree of interpretability of data, an essential factor in exploring climatic changes (Meenal et al., 2022). Differential equation models provide comprehensive representation of climate change models. However, differential models demand concrete computation which may fail to be assessed in available data situations. This research aims at addressing the gap by comparing these modelling techniques utilizing climate forecasting data from China which is known to have diversified climate conditions across its provinces.

1.3. Purpose of the Study

The main objective of this study lies on evaluation and comparison of efficiency of statistical, machine learning, and differential equations models in predicting climate changes. Putting focus on climate data from different provinces of China, the study will conclude on the best performing model at any given climatic conditions (Murugan et al.,2021). Additionally, the study provides clear outlines of the strengths and weakness of each approach. Addressing the problem of the climate changes in the future shaping of the world, this analysis adds emphasis on refining climatic forecasting techniques and thus preaching for sound and informed decision making in the China's environmental policies and disasters management (Salman and Kanigoro, 2021).

1.4. Research Objectives

This study falls under the guidance of the following objectives:

1. To evaluate predictive accuracy of statistical models in short-term climate prediction, focusing on temperature and precipitation data.
2. To assess the effectiveness of the machine learning models such as neural networks and random forests in addressing complex climate data for short to medium term forecasting.
3. To examine the suitability of differential equation models for long term climate forecasting and their capacity to represent dynamic climatic data accurately.
4. To identify strengths and limitation of each model type based on accuracy, interpretability, computational efficiency, and adaptability to diverse climatic zones within China.

1.5. Research Questions

The research aims to address and present answers for the following questions:

1. Does the statistical models accurate in forecasting climate parameters such as temperatures and rainfall data for short term predictions in China's provinces?
2. What is the degree of effectiveness that machine learning models showcases in its performance based on accuracy, and adaptability when applied to complex climate data in China?
3. To what extent can differential equation models represent long-term climate trends, and how do they perform with limited climate data?
4. Which climate forecasting techniques provides substantial balance of accuracy, efficiently, and interpretability for regional climate forecasting in China?

CHAPTER 2

2.Literature Review

2.1.Overview of the Climate Change Modelling Techniques

Approaches that are utilized to address the impacts caused by climate change to the ecosystem has remained the question of discussion to both policymakers and researchers. Modelling has remained an essential tool to be utilized by these elites (Babu and Nuthalapati, 2024). Efficient modeling of climate change demands deep exploration of various techniques. Researchers have developed numerous modelling techniques each with its weakness and strengths. Models used to address climate change issues are categorized into three. These models are statistical models, machine learning models and physical based models.

Statistical models utilize statistical data to perform correlation and relationships between variety of climatic variables (Bihlo, 2021). Machine learning models relies of machine learning algorithms to perform its analysis. Selection of ideal model to be utilized in a given data is determined by some factors including data availability, desired prediction profile, and the available computational resources (Bi et al., 2023).

2.2. Statistical Models in Climate Prediction

Models using statistical approach are among the oldest ways utilized in providing weather forecasting. These models utilize raw data from historical trends and uses data history to come up with weather predictions (Bi et al., 2023). Statistical models include autoregression integrated moving average (ARIMA) approach and regressions models. In cases where only two datasets are compared linear regression model is used but in cases where single output is compared with multiple input variables, multiple regression technique is utilized. Statistical methods according to Zhang et al., (2020)'s research suits short term climate changes with traceable and valuable extend of accuracy.

Despite of statistical models demonstrate prediction of short-term climate changes, the approach has an issue of over-simplification. Statistical models use linearity technique and therefore, in the case of climate change it assumes linearity and stationarity of data making it to suit only short-term weather forecasting (Gronguist et al., 2021). Furthermore, statistical models struggle to consider rare and influential cases of extreme weather conditions.

2.3. Machine Learning Models in Climate Science

Recently Machine learning models has been introduced into the field of weather forecasting. This technique is more than statistical technique in that it considers all data and computational ability (Bochenek and Ustrnul, 2022). Machine learning models utilize algorithms such as Artificial Neural Networks (ANNs), random forests, and Support Vector Machines (SVMs) and excel approach to find hidden datasets in large volumes of data. Machine Learning techniques are flexible in capturing non-linear correlation thus can adapt to complex and multidimensional datasets (Goodfellow et al., 2016). In the case of climate change variations, machine learning has showcased recommendable results in predicting climates parameters such as temperatures and precipitation. A study by Liu et al. (2019) demonstrated that ANN systems performs better than traditional statistical methods in estimating changing patterns of atmospheric temperatures in various regions of China.

Besides numerous advantages of machine learning techniques, the method is accompanied by some challenges, which hinders its efficiency (Chen et al., 2023). Machine learning approache has problems when it comes to provision of substantial interpretations. Majority of these machine learning approaches functions as black boxes hence making it complex to pinpoint particular factors behind their predictions (Hewage et al., 2021). Additionally, machine learning technique requires skills thus it require training of large dataset and reliable formulation resources. This nature of machine learning sometimes fails to meet feasibility in resource constrained settings.

2.4. Differential Equations in Climate Modeling

Physical differential equations are another technique applied in predicting weather conditions. Differential equations models have their foundation on the physical laws which gives detailed representation of climate variations (Meenal et al., 2022). The technique utilizes simulation of interaction among climate variables and thus shows dynamicity of the climate changes. Differential equation models approach such as general circulation models (GCMs) and energy balance models are extremely utilized to understand long-term weather trends (Jiao et al., 2021). Differential equation models are grounded in physical laws and provide a detailed representation of climate dynamics by simulating interactions among climate variables. These models, including general circulation models (GCMs) and energy balance models, are extensively used to understand long-term climate trends. Differential equation models provide promising results in handling weather predictions for long-term climatic trends since it captures complex feedback mechanisms among climatic variables (Smith et al., 2018).

Implementation of differential equation models requires a lot of resources and requires advanced computational technique. Differential equations are sensitive to historical data conditions and the desired accuracy and this can interfere with the model's reliability. Differential equation models work best for situations demanding high precision and detailed system understanding (Siddique et al., 2022). Differential Equation models are ideal for long-term weather forecasting but not practical for short-term forecasting, and particular region forecasting since the technique demands high advanced computational technology.

2.5. Comparative Analysis of Modeling Techniques

Comparison of statistical methods which are statistical approach, machine learning and differential equation provide concrete reason on which method is to be adopted at a given condition (Vannitsem et al., 2021). Some of the researchers in their work have provided direct comparison of these methods utilized in weather forecasting in various regions of the world. From various work of researchers, it is clear that statistical techniques suit short-term forecasting since they lack precision and accuracy to capture non-linear climate trends (Yang et al., 2023). Machine Learning techniques are more flexible and provides accurate results thus can suit long-term weather forecasting. Machine learning approach is affected by the interpretability issue thus making it unreliable for its own (Zhong et al., 2024).

Differential model techniques provide a better option for as compare to the two modelling methods. This model suits long-term prediction but can still fail to provide predictions for specific regions since it is more complex (Zhuo et al., 2022). Analysis of these modelling methods as the proposal of this study produces insights reflecting optimization of climate prediction strategies.

CHAPTER 3

3. Methodology

This section provides the research design, data sources utilized, model selection, and evaluation of method's strength in comparing different climatic conditions. The methodology's framework provides evaluation of the three predictions models which are Statistical Models, Machine Learning (ML), and Differential Equations models. Each method's strengths and weakness are analyzed using Chinese regional temperatures and precipitation data.

3.1. Research Design

The research uses a comparative research design to showcase performance of the three modelling techniques: Statistical models, machine learning approach and Differentiation equations approach. The modelling techniques are utilized in providing predictions of climatic variables which are temperature and precipitation in China. Main objectives of comparing these modelling methods are to access their accuracy, efficiency and adaptability.

3.1.1. Approach

The study adopts a quantitative approach which focuses only on statistical analysis and computation of historical climate data of various regions. The goal in this case is to understand the operation of each model in predicting climate variables. It also pinpoints the strengths and weakness of each modelling technique. This helps in selecting appropriate approach at a given situation.

3.1.2. Scope

The research utilizes historical data which covers more than one decade from different regions of China, including both urban and remote areas. The focus of this study is to analyze predictability of the three modelling techniques in forecasting temperature and precipitation trends.

3.1.3.Structure

The study adopts structure which evaluates the performance of each modelling technique. The evaluation will be based on how well a given technique is efficient in predicting climate changes using historical data. The modelling involves the assessment of multiple regions showing variations in climate changes thus enabling the study's exploration on the model's adaptability to regional differences. In addition, the structure comprises assessment patterns, anomalies, and climate change habits. These elements will be utilized in identifying which model works best in forecasting long-term and short-time climate changes. This impacts future policy-making and future climate forecasting.

3.2. Data Sources and Collection Techniques

Comprehensive analysis on the behavior of the three modelling techniques requires accurate and reliable historical climate data. This study uses data sourced from high reputation organizations. These data are of high-quality implying that its accuracy and reliability are not compromised. Data are obtained using both primary and supplementary data sources.

3.2.1. Primary Data Source

Since the analysis involves utilization of Chinese provinces climatic data, the study sources data primarily from a Chinese trusted climate institution referred to as China Meteorological Administration (CMA). This institution gives historical and current data for various provinces across China. Among the data available in CMA are temperature and precipitation data which are the study's target variables. CMA portal avails these data for various decades thus providing room for extensive analysis of long-term climate trends.

3.2.2. Supplementary Data Sources

Besides China Meteorological (CMA), the research utilizes other sources which supplements data obtained from CMA. World Meteorological Organization (WMO) supplements CMA's data by offering global climatic trends thus providing extra comparative dataset for climate changes across different regions of the world. National Oceanic and Atmospheric Administration (NOAA), provides additional climate data to supplement CMA's data. NOAA is an open-access climate database which avails any given region's climate data at any time. NOAA backs up CMA's data by availing climate data for regions where CMA might be missing or limited.

3.2.3. Data Collection Process

Temperature and precipitation variables from 2004 to 2023 are gathered. Data collected in this case includes both historical data for the past 34 years and the current data. The essence of collecting data decades is to provide analysis and forecasting for both short-term and long-term climate forecasting (Smith et al., 2018). Prior analyzing the climate data collected, data are passed through various processes which ensures quality of data to be analyzed. Among the processes includes data cleaning, data normalization, outliers' detection from the data, and data splitting.

3.2.3.1. Data cleaning

In this process data are checked for missing and erroneous data points. These data points are identified and addressed using interpolation approach. This ensures that data used by the analysis are clean thus yielding quality results.

3.2.3.2. Data Normalization

This is another phase essential for quality outcomes. In this phase, climate variables which are temperature and precipitation are normalized to a common scale. This includes standardization of temperature units and precipitation variables thus ensuring consistent comparison of data across models.

3.2.3.3. Outlier Detection

Sometimes collected data contains extreme values, which shows greater deviation from trend. In this stage of data collection, extreme temperature and precipitation data values are flagged and handled. Handling of these extreme data values involves removal or adjustments of the flagged outliers. Adjustments of these outliers relies on their potential impacts to the final analysis results.

3.2.3.4. Data Splitting

After cleaning, normalizing, and adjusting outliers, data are divided into training and testing subsets. Typical splitting ratio is 0.7:0.3. This implies that 70% of the total data are training dataset and the remaining 30% is testing dataset. Training data are utilized in building models and 30% of the data are used for validating the model.

3.3. Model Selection and Implementation

The research aims to provide concrete comparison among three modelling techniques which are statistical modelling, Machine Learning Modelling, and Differential Equations

Modelling. Each model carries its strengths and weaknesses thus making it ideal to evaluate them through climate forecasting.

3.3.1. Statistical Modelling With ARIMA

This study utilizes ARIMA (Auto-Regressive Moving Average Model) to perform statistical modeling. ARIMA model has ability to handle data which uses data series. In this case, ARIMA technique suits forecasting of temperature and precipitation rends in various regions of China. For this method to work effectively, it utilizes three processes which are stationary testing, optimization of the parameters, and implementation. In stationary testing, precipitation and temperature's data are tested for stationarity using Augmented Dicky Fully Test. To optimize the parameters, the values of p, d, and q are identified using grid search methods.

3.3.2. Machine Learning Modelling with Linear Regression

In machine learning (ML), linear regression techniques are the heart of analysis. in modelling the relationship between time and climate data such as temperature and precipitation linear regression should be used in ML algorithm (Siddique et al., 2022). This study treats time as the continuous variable whereas temperature and precipitation are the target parameters. Dataset for climate trends from 2004 to 2023 were used for analysis. 70% of the data which are temperature and precipitation data up to 2017 were used to train the model and the rest of the data used to testing and validating the model. The model was trained using the training dataset to find the estimates of the slope and intercept co-efficiencies. The model predicts the temperature and rainfall trends for the period 2018 to 2023 using the data obtained from training.

3.3.3. Differential Equation Modelling

This model utilizes the exponential decay/growth technique. Using these technique, climate parameters are simulated and thus predicted (Zhuo et al., 2022). Equations for the model are derived to capture the growth and decay behavior of both temperature and precipitation data. The co-efficiencies for climate data are obtained from the training dataset. This helps to fit the differential model.

3.4. Model Performance Evaluation

Efficiency of each modelling technique was assessed using three performance metrics which are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-Squared(R^2).

Mean Absolute Error (MAE) helps in evaluating the magnitude of prediction errors of the model. The formula used to formulate MAE for both temperature and precipitation is given as:

$$MAE = \frac{1}{n}\sum_{i=n}^{n}|y_i - \widehat{y_i}|, \quad (1)$$

where y_i, $\widehat{y_i}$, and n are the climate value at a given data point, the mean, and the number of datasets respectively.

Root Mean Square Error (RMSE) presents the square root of the average square errors thus penalizing larger errors more heavily. To determine the value of RMSE the following formula is used:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}. \quad (2)$$

R-Squared valued is another statistic metrics which is utilized in this study to explain how well the modelling technique explains the variance of observed data (Yang et al., 2023). To find the value of R-Squared Value we use the formula:

$$R - Squared\ (R^2) = 1 - \frac{SS_{residuals}}{SS_{total}}. \quad (3)$$

3.5. Tools and Software

The models were implemented using Microsoft excel software for ARIMA, Linear regression representing ML modelling, Exponential growth/decay in the part of differentiation equations model. Furthermore, excel was used to generate all regression tables, comparison tables and graphs used to visualize the trend.

CHAPTER 4

4. Results and Data Analysis

This section displays the results of the three modelling techniques which statistical modelling, machine learning modelling, and differential equation modelling. For the statistical modelling, Auto-regression Integrated moving average (ARIMA) approach was employed to predict climate parameters in Chinese provinces (Chen et al., 2023). Linear Regression was applied to temperature and precipitation to predict future trends thus representing Machine Learning technique. For differential method, exponential growth/decay approach was employed.

The main goal of this chapter is to evaluate the performance of each modelling method, in terms of accuracy, efficiency and adaptability. This objective is achieved from the analysis on how well the model predicts the future trends in temperature and precipitation figures. The analysis will bring the comparison of the strengths and weaknesses of different modelling techniques. The strengths and weakness of the approached are realized using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-Squared (R^2). Each of these parameters are formulated from each of the models.

4.1. Overview of Data Utilized

The data involved in the analysis of this study are the precipitation and temperature values for 8 Chinese provinces (Meenal et al., 2022). To provide solid climatic data trends, the analysis accounted for precipitation and temperature data from the year 2004 to the year 2023. The climate data were cleaned and pre-processed thus addressing any missing values and outliers. The data are then divided into two sections which training and testing data. Training data accounted for 70% of the data and 30% were used for testing. This implies that data from 2004 to 2017 were used for training and the rest up to 2023 were used for validating. Training sets were used to establish the model and the testing set was used to validate the models.

The key parameters used in this study are the annual temperatures (℃) across 8 provinces of China, and annual total rainfall (mm) for the same 8 provinces of China. Descriptive statistics for the eight provinces of China representing temperature and precipitation are illustrated by Table 1 and Table 2 below.

Table 1: Descriptive Statistics of temperature for Chinese Regions

Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	11.65	Mean	8.80	Mean	12.80	Mean	17.25	Mean	21.80	Mean	14.90	Mean	16.60	Mean	10.60
Standard Error	0.38	Standard Error	0.43	Standard Error	0.36	Standard Error	0.38	Standard Error	0.36	Standard Error	0.44	Standard Error	0.43	Standard Error	0.41
Median	12.00	Median	9.00	Median	13.00	Median	17.00	Median	22.00	Median	15.50	Median	16.00	Median	11.00
Standard Deviation	1.69	Standard Deviation	1.94	Standard Deviation	1.61	Standard Deviation	1.68	Standard Deviation	1.61	Standard Deviation	1.97	Standard Deviation	1.90	Standard Deviation	1.82
Sample Variance	2.87	Sample Variance	3.75	Sample Variance	2.59	Sample Variance	2.83	Sample Variance	2.59	Sample Variance	3.88	Sample Variance	3.62	Sample Variance	3.31
Range	5.00	Range	6.00	Range	5.00	Range	5.00	Range	5.00	Range	6.00	Range	6.00	Range	6.00
Minimum	9.00	Minimum	6.00	Minimum	10.00	Minimum	15.00	Minimum	20.00	Minimum	12.00	Minimum	14.00	Minimum	7.00
Maximum	14.00	Maximum	12.00	Maximum	15.00	Maximum	20.00	Maximum	25.00	Maximum	18.00	Maximum	20.00	Maximum	13.00
Sum	233.00	Sum	176.00	Sum	256.00	Sum	345.00	Sum	436.00	Sum	298.00	Sum	332.00	Sum	212.00
Count	20.00	Count	20.00	Count	20.00	Count	20.00	Count	20.00	Count	20.00	Count	20.00	Count	20.00

Table 2: Descriptive Statistics of Precipitation for Chinese Regions

Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	627.85	Mean	551.30	Mean	791.6	Mean	1482.5	Mean	1935.90	Mean	1204.20	Mean	1404.45	Mean	131.65
Standard Error	23.16	Standard Error	22.06	Standard Error	23.56529563	Standard Error	36.9	Standard Error	60.17	Standard Error	27.31	Standard Error	64.71	Standard Error	8.93
Median	620.00	Median	553.00	Median	809	Median	1452.0	Median	1952.50	Median	1170.00	Median	1334.50	Median	137.00
Standard Deviation	103.57	Standard Deviation	98.65	Standard Deviation	105.3872059	Standard Deviation	165.1	Standard Deviation	269.09	Standard Deviation	122.14	Standard Deviation	289.37	Standard Deviation	39.95
Sample Variance	10726.77	Sample Variance	9732.54	Sample Variance	11106.46316	Sample Variance	27269.2	Sample Variance	72408.41	Sample Variance	14918.38	Sample Variance	83736.68	Sample Variance	1596.34
Range	359.00	Range	280.00	Range	350	Range	548.0	Range	891.00	Range	363.00	Range	795.00	Range	136.00
Minimum	427.00	Minimum	405.00	Minimum	620	Minimum	1247.0	Minimum	1532.00	Minimum	1031.00	Minimum	1004.00	Minimum	62.00
Maximum	786.00	Maximum	685.00	Maximum	970	Maximum	1795.0	Maximum	2423.00	Maximum	1394.00	Maximum	1799.00	Maximum	198.00
Sum	12557.00	Sum	11026.00	Sum	15832	Sum	29650.0	Sum	38718.00	Sum	24084.00	Sum	28089.00	Sum	2633.00
Count	20.00	Count	20.00	Count	20	Count	20.0	Count	20.00	Count	20.00	Count	20.00	Count	20.00

4.2. Model Performance Evaluation

The operation of each model was assessed using a range of accuracy metrics. Commonly metrics utilized in statistics are the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the R-Squared (R^2). Mean Absolute Error (MAE) presents the average magnitude of the errors in predictions without accounting for their directions. The formula used to formulate MAE for both temperature and precipitation is given as:

$$MAE = \frac{1}{n}\sum_{i=n}^{n}|y_i - \widehat{y_\iota}|, \tag{4}$$

where y_i, $\widehat{y_\iota}$, and n are the climate value at a given data point, the mean, and the number of datasets respectively (Jiao et al.,2021). Root Mean Square Error (RMSE) presents the square root of the average square errors thus penalizing larger errors more heavily. To determine the value of RMSE the following formula is used:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \quad (5)$$

R-Squared valued is another statistic metrics which is utilized in this study to explain how well the modelling technique explains the variance of observed data. To find the value of R-Squared Value we use the formula:

$$R - Squared\ (R^2) = 1 - \frac{Unexplained\ Variation}{Total\ Variation} = \frac{SS_{residuals}}{SS_{total}}. \quad (6)$$

4.3. Results of Auto-Regression Integrated Moving Average (ARIMA) Model

4.3.1.Model Building

ARIMA model was applied to temperature and precipitation data for each Chinese province starting from Hebei to Xiangjiang. The first step was testing the stationarity of data series and adjusting data to become stationary. The optimal ARIMA parameters were computed using Excel's regression and time series analysis function and the results for precipitation and temperature were as illustrated in Figure 1 and Figure 2 below.

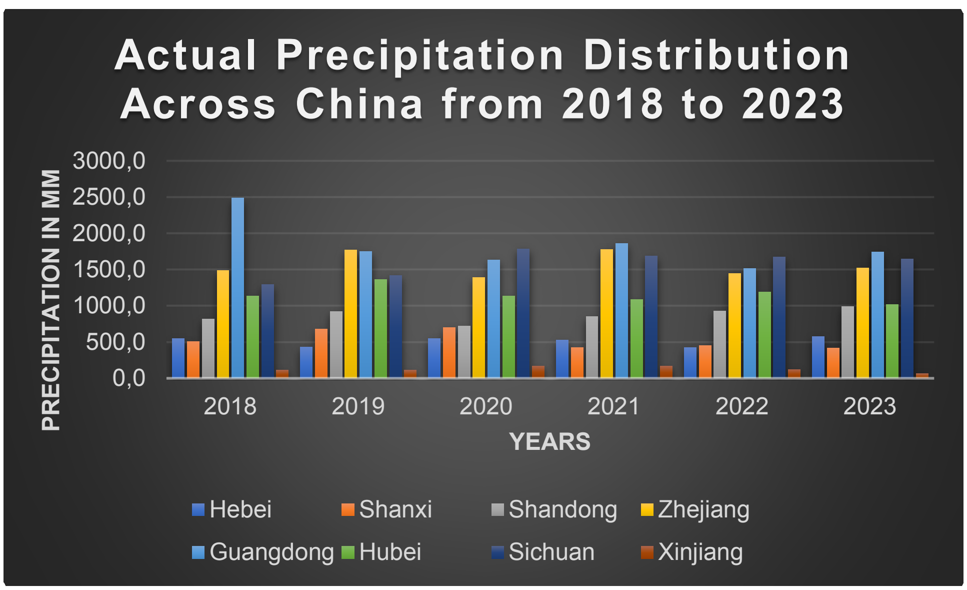

Figure 1: Actual Precipitation Distribution

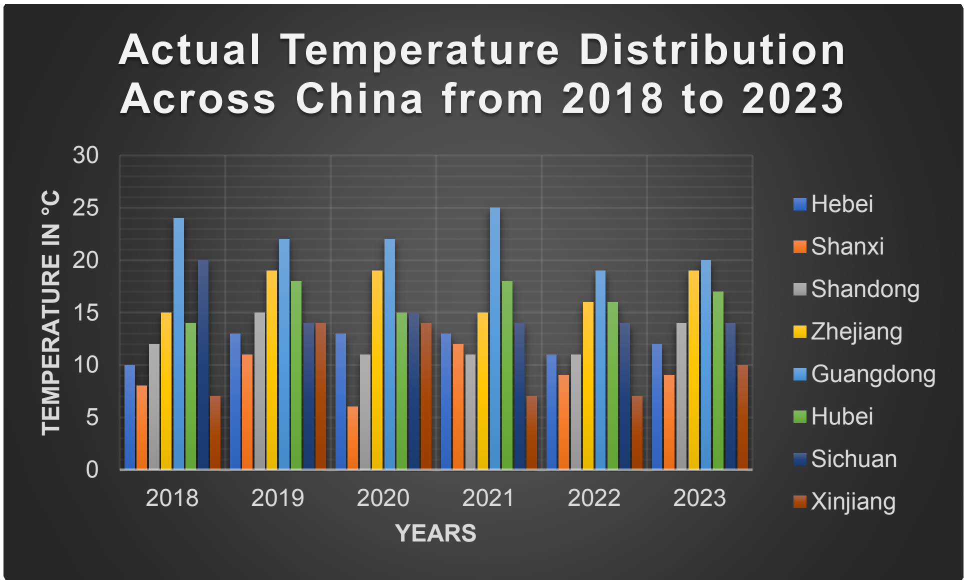

Figure 2: Actual Temperature Distribution

4.3.2. Forecasting Results

Upon application of the optimal ARIMA, the model was utilized to generate forecasts for future temperatures and precipitation values. The outcomes obtained for each province was compared to the actual observed values for both precipitation and temperature. Figure 3 and Figure 4 below illustrates forecasted climate trend for the 8 provinces for the forecasted values of temperature and precipitation.

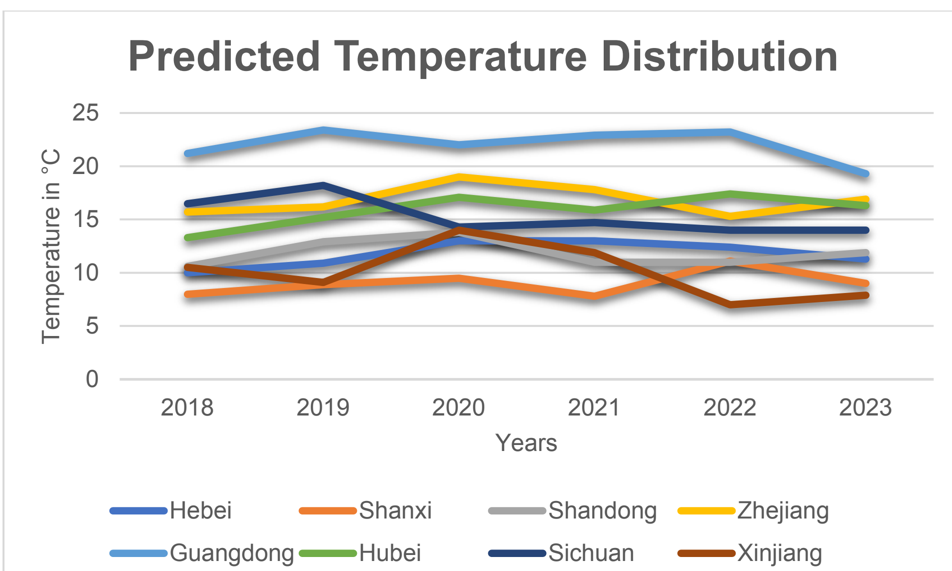

Figure 3: Forecasted Temperature Distribution from 2018 to 2023

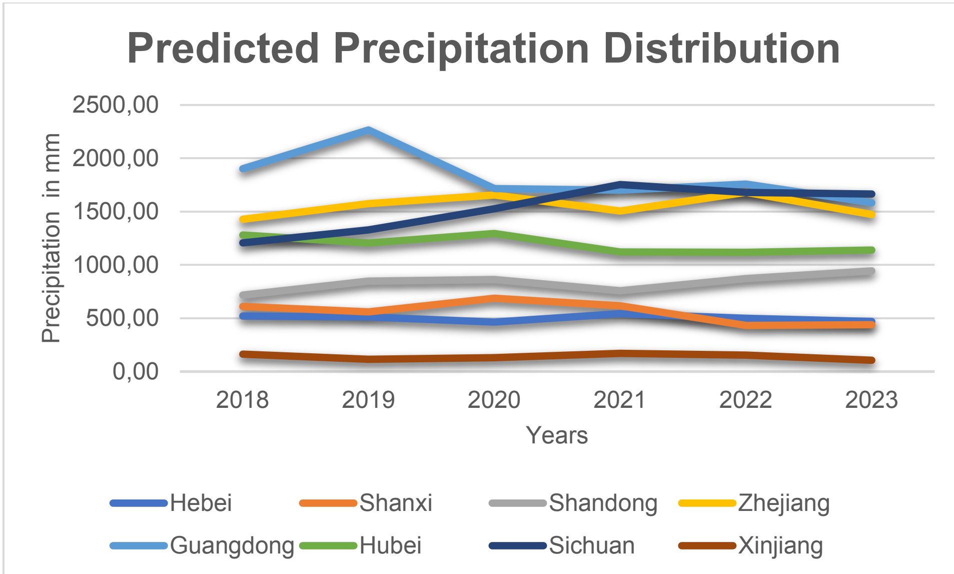

Figure 4: Forecasted Precipitation Distribution from 2018 to 2023

4.3.3. Evaluation Metrics of the ARIMA Model

ARIMA utilizes three metrics to showcase its performance, since is a prediction-based model is uses MAE, RMSE, and R^2. Using predicted values and actual values the formula for MAE and RMSE was formulated for each province. Formulation of MAE and RMSE uses validation set (Siddique et al., 2022). Computation of R^2 was also performed to disclose the efficiency of the model in terms of variations. After formulating the values of MAE and RMSE Table 3 and Table 4 shows the performance of ARIMA model for predicted temperature and precipitation.

Table 3: ARIMA Performance in Predicting Precipitation Across China

Province	RMSE	MAE	R-Squared
Hebei	21.5	21.5	1.010699
Shanxi	2.6	-2.6	1.000147
Shandong	28.8	-28.8	1.006032
Zhejiang	29.2	-29.2	1.00192
Guangdong	48.5	48.5	1.003881
Hubei	21.0	21.0	1.001957
Sichuan	74.8	-74.8	1.014951

Xinjiang	**13.1**	**-13.1**	**1.067458**

Table 4: ARIMA Performance in Predicting Temperature Across China

Province	RMSE	MAE	R-Squared
Hebei	**0.1167**	**0.1167**	**1.0007**
Shanxi	**0.1167**	**-0.1167**	**1.0012**
Shandong	**0.1167**	**-0.1167**	**1.0005**
Zhejiang	**0.1167**	**0.1167**	**1.0003**
Guangdong	**0.1167**	**-0.1167**	**1.0002**
Hubei	**0.3500**	**-0.3500**	**1.0043**
Sichuan	**0.2333**	**0.2333**	**1.0014**
Xinjiang	**0.1167**	**-0.1167**	**1.0009**

4.3.4.Interpretation of ARIMA Results

The model showcased concrete performance in predicting climatic trends across Chinese regions. For instance, using the model to predict precipitation Xinjiang showed a slightly higher R^2 of 1.0674. In terms of RMSE Sichuan displayed the highest value of RMSE 74.8 while other regions like Shanxi reported the lowest value of 2.6. In predicting temperature, the model performed well with almost the same values of R^2 implying that greater variance in terms of regional data. Difference in the values of RMSE and MAE indicates that the method works best on different regions thus the model is not region specific.

4.4. Results of Linear Regression Model

4.4.1. Model Construction

The linear regression model was applied for both temperature and precipitation as dependent variables and time years as the independent variables. Using excel Data Analysis Toolpak the regression model was generated. The linear regression model was fit to the training data and the parameters which are slope and intercept were estimated. The estimations of slope and intercepts are summarized in Table 5.

Table 5: Summary of Forecasted Temperature and Precipitation using ML Technique

Province	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Hebei	675.20	648.60	664.60	658.40	621.00	644.40
Shanxi	602.60	634.00	625.20	602.20	577.20	610.80
Shandong	740.20	779.20	786.00	800.80	812.60	825.80
Zhejiang	1577.80	1595.00	1596.80	1581.00	1517.00	1579.00
Guangdong	1957.20	2060.20	2074.40	2101.20	2106.20	2088.00
Hubei	1222.80	1259.80	1197.20	1231.20	1248.60	1245.80
Sichuan	1638.40	1528.80	1456.60	1421.40	1340.60	1323.40
Xinjiang	132.40	155.20	142.20	117.80	101.60	113.40

Province	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
Hebei	10.6	10.8	10.4	9.8	9.6	9.2
Shanxi	9.8	10	10.4	10	9.4	9.4
Shandong	12.6	12.2	12.6	11.8	12.4	12.2
Zhejiang	16	15.4	15.6	16.4	16.8	17.4
Guangdong	20.4	21	21.8	21.8	21.4	21.8
Hubei	13.4	14.2	15.2	15.4	15.8	16
Sichuan	18	17	16.2	15.6	15.2	16.2
Xinjiang	11.8	12.8	12.2	12.2	12	11.4

4.4.2. Forecasting Results

By using modelled linear regression as the representation of Machine Learning Technique, estimated values for precipitation and temperature for various provinces of China were formulated. Table 6 illustrates descriptive statistics for the predicted temperature and rainfall values across Chinese provinces.

Table 6: Descriptive Statistics for Forecasted Climate Data from 2018 to 2023

Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	552.4	Mean	620.5666667	Mean	863.3333333	Mean	1505.366667	Mean	1984.933333	Mean	1237.066667	Mean	1389.833333	Mean	131.6666667
Standard Error	11.1165942	Standard Error	7.212843483	Standard Error	11.72799121	Standard Error	26.95177999	Standard Error	28.93125953	Standard Error	18.36363556	Standard Error	20.00062777	Standard Error	3.777359454
Median	549.6	Median	622.3	Median	874.8	Median	1468.2	Median	2010.3	Median	1245.8	Median	1381.9	Median	133.5
Standard Deviation	27.22998347	Standard Deviation	17.66778613	Standard Deviation	28.72759417	Standard Deviation	66.01810863	Standard Deviation	70.86682346	Standard Deviation	44.98153695	Standard Deviation	48.99133257	Standard Deviation	9.252603237
Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	10.46666667	Mean	7.033333333	Mean	12.63333333	Mean	18.16666667	Mean	21.86666667	Mean	13.8	Mean	16.93333333	Mean	10.03333333
Standard Error	0.217050941	Standard Error	0.270390664	Standard Error	0.227547309	Standard Error	0.166666667	Standard Error	0.168654809	Standard Error	0.146059349	Standard Error	0.217050941	Standard Error	0.355590276
Median	10.5	Median	7	Median	12.8	Median	18.4	Median	21.6	Median	13.8	Median	17	Median	9.9
Standard Deviation	0.531664054	Standard Deviation	0.662319158	Standard Deviation	0.557374799	Standard Deviation	0.40824829	Standard Deviation	0.413118224	Standard Deviation	0.357770876	Standard Deviation	0.531664054	Standard Deviation	0.871014734
Sample Variance	0.282666667	Sample Variance	0.438666667	Sample Variance	0.310666667	Sample Variance	0.166666667	Sample Variance	0.170666667	Sample Variance	0.128	Sample Variance	0.282666667	Sample Variance	0.758666667

From the results on Table 6 it is clear that the forecasted data descriptive analysis shows different standard deviations for each region for example, Guangdong had the highest standard deviation of 70.87 against Xinjiang reported 9.25 in precipitation prediction. In temperature forecasting, Hubei reported the lowest standard error of 0.15 of all the eight regions implying that this techniques can be used for given regions and produce quality results. The comparison between predictability of different climatic variables, using ML techniques is associated with some weakness (Siddique et al.,2022). For this case, the quality of results produced in predicting temperature and precipitation differs with ML techniques underscores in predicting the trends of precipitation. Figure 5 and Figure 6 shows the graphs visualizing predicted variables for temperature and precipitation from 2018 to 2023.

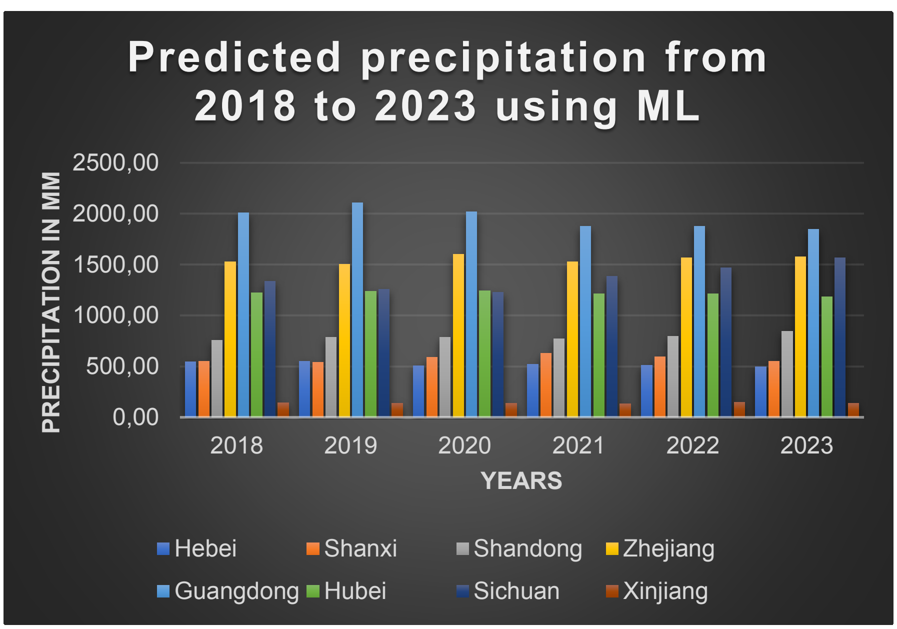

Figure 5: Visualization of Forecasted Precipitation Results using ML Technique

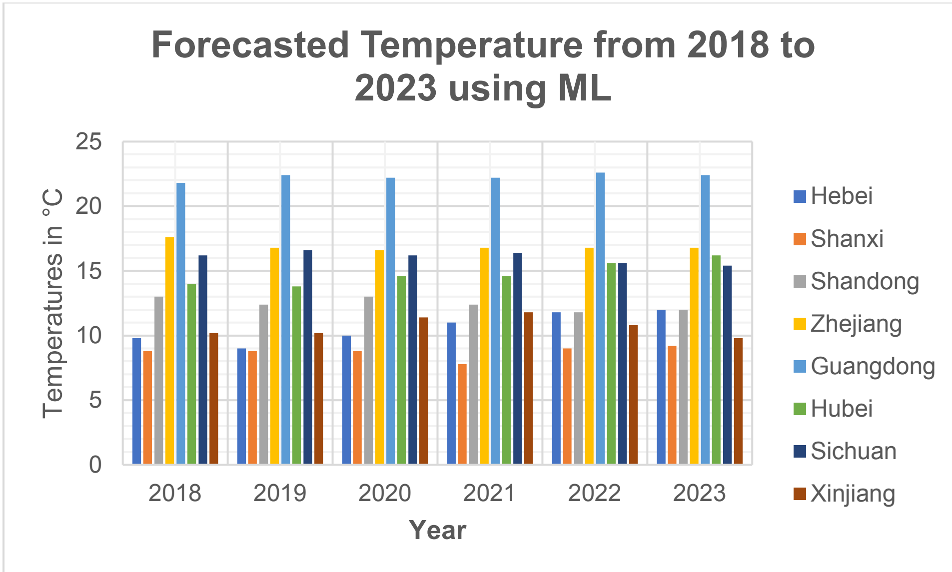

Figure 6: Visualization of Predicted Temperature Trend from 2018 to 2023

4.4.3. Evaluation Metric for Linear Regression

Using linear regression model, MAE and RMSE for this model was evaluated to assess the accuracy of the model. R-Squared value was also formulated to find the variance extent of data for the Chinese provinces. Table 7 illustrates the performance of the model in terms of R^2, MAE, and RMSE.

Table 7: Performance of Machine Learning Model in Climate Forecasting

	RMSE	MAE	R-Squared
Hebei	10.8	10.8	1.001846
Shanxi	57.4	-57.4	1.067944
Shandong	25.0	25.0	1.005577
Zhejiang	27.4	-27.4	1.001992
Guangdong	63.0	63.0	1.006223
Hubei	8.6	-8.6	1.000334
Sichuan	62.3	-62.3	1.011745
Xinjiang	18.1	18.1	1.128393

	RMSE	MAE	R-Squared
Hebei	**0.433333**	**0.433333**	**1.008199**
Shanxi	**0.066667**	**0.066667**	**1.000408**
Shandong	**0.166667**	**-0.16667**	**1.001097**
Zhejiang	**0.166667**	**-0.16667**	**1.000566**
Guangdong	**0.133333**	**-0.13333**	**1.000251**
Hubei	**0.733333**	**0.733333**	**1.012722**
Sichuan	**0.066667**	**0.066667**	**1.000086**
Xinjiang	**1.466667**	**-1.46667**	**1.118084**

From Table 7, the performance of the model in terms of RMSE, MAE and R^2 are illustrated. Xinjiang province showcased the highest value of R^2 in both temperature and precipitation forecasting reporting the values 1.12 and 1.13 respectively. Using RMSE Xinjiang showed the highest value of RMSE in predicting temperature trend reporting 1.467 against Shanxi region which reported 0.067. This indicates that using Machine Learning approach, accurate temperature trends can predict in Shanxi (Zhong et al., 2024). Generally, temperature forecasted trend reveals low values of RMSE as compared to precipitation, implying that machine learning works best in predicting short-term climate trends.

4.4.4. Interpretation of Linear Regression Model

From Table 7 its clear that R^2 values for various provinces of Chinese data showed increased values as compared to ARIMA model. Increased value of R^2 implied increased efficiency of this method in handling seasonal trends of data values. The MAE and RMSE of this model were relatively higher as compared to ARIMA model (Yang et al., 2023). For instance, RMSE value of 1.467 reported in this model is higher than the same RMSE value of 0.12 reported in ARIMA model in predicting temperature trend. This implies that the model's capability to fit data was effective.

4.5. Results for Differential Equation Modelling

4.5.1. Model Construction

Differentiation Equation technique used the exponential growth/decay approach to temperature and precipitation data. The model takes an assumption that climate data used in this case follows exponential trend (Zhong et al.,2024). The rate estimation constant k was estimated using solver technology in excel. The model utilized 70% of the data in training and used the remaining portion of 30% to validate the data. Table 8 illustrated forecasted values of temperature and precipitation across Chinese regions from the year 2018 to the year 2023.

Table 8: Summary of Forecasted Results Using Differential Equation Technique

Province	2018	2019	2020	2021	2022	2023
Hebei	769.50	647.60	582.10	501.70	624.00	755.90
Shanxi	609.40	542.80	531.70	552.90	445.90	481.80
Shandong	836.10	808.50	757.90	832.10	708.30	784.00
Zhejiang	1418.90	1577.30	1516.20	1323.90	1360.80	1517.30
Guangdong	1719.00	2077.10	2345.10	2212.40	2256.20	2049.10
Hubei	1207.20	1203.30	1116.60	1085.50	1263.10	1368.30
Sichuan	1429.20	1329.40	1401.60	1458.00	1176.60	1084.20
Xinjiang	64.20	102.20	133.70	149.10	129.60	148.60

Temperature Forecast from 2018 to 2023

Province	2018	2019	2020	2021	2022	2023
Hebei	11.5	7.5	13.5	12	10.5	6.5
Shanxi	6	14	10.5	9.5	5.5	8.5
Shandong	10.5	11	17	10.5	16.5	13.5
Zhejiang	15.5	17.5	18.5	16.5	15.5	20.5
Guangdong	21	20.5	24	21.5	23.5	26
Hubei	13.5	13.5	20	12	11	16.5
Sichuan	19.5	11.5	14	15.5	21	11.5
Xinjiang	8	12.5	9.5	7	8	11

4.5.2. Forecasting Results

The differentiation modelling produced forecast for Chinese future temperature and precipitation values which were compared with actual observed values as depicted in Table 9. Descriptive statistics for the actual and forecasted data values are summarized on the table.

Table 9: Descriptive Statistics of Forecasted Climate Data using Differential Equation Method

Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	604.18	Mean	558.3	Mean	896.94	Mean	1519.42	Mean	2101.5	Mean	1247.34	Mean	1336.58	Mean	133.2
Standard Error	80.69024352	Standard Error	20.867894	Standard Error	23.20884745	Standard Error	43.38236278	Standard Error	117.2355194	Standard Error	41.72656947	Standard Error	65.47075989	Standard Error	17.85698743
Median	649.9	Median	564.8	Median	914.2	Median	1539.9	Median	2149.2	Median	1203.5	Median	1313.2	Median	109.9
Standard Deviation	180.4288696	Standard Deviation	46.66202953	Standard Deviation	51.89656058	Standard Deviation	97.00591219	Standard Deviation	262.1465907	Standard Deviation	93.30344581	Standard Deviation	146.3970696	Standard Deviation	39.92943776
Sample Variance	32554.577	Sample Variance	2177.345	Sample Variance	2693.253	Sample Variance	9410.147	Sample Variance	68720.835	Sample Variance	8705.533	Sample Variance	21432.102	Sample Variance	1594.36
Hebei		*Shanxi*		*Shandong*		*Zhejiang*		*Guangdong*		*Hubei*		*Sichuan*		*Xinjiang*	
Mean	9.833333333	Mean	9.5	Mean	12.16666667	Mean	16.58333333	Mean	22	Mean	14.58333333	Mean	16.66666667	Mean	11.16666667
Standard Error	0.853098926	Standard Error	1.591644852	Standard Error	0.542627353	Standard Error	1.17910041	Standard Error	1.5	Standard Error	0.65085414	Standard Error	1.452966315	Standard Error	0.954521404
Median	10.25	Median	9.25	Median	12.25	Median	15	Median	23.25	Median	15	Median	16.5	Median	11.75
Standard Deviation	2.089657069	Standard Deviation	3.898717738	Standard Deviation	1.329160136	Standard Deviation	2.888194361	Standard Deviation	3.674234614	Standard Deviation	1.594260539	Standard Deviation	3.559026084	Standard Deviation	2.338090389
Sample Variance	4.366666667	Sample Variance	15.2	Sample Variance	1.766666667	Sample Variance	8.341666667	Sample Variance	13.5	Sample Variance	2.541666667	Sample Variance	12.66666667	Sample Variance	5.466666667

Table 9 illustrates the results of the descriptive analysis for data trends in Chinese regions using Differential Equation Modelling technique. From the table, the performance of the model is assessed for various regions of China (Chai & Draxler, 2014). In predicting precipitation, the model performed well in regions like Xinjiang which showed the lowest standard deviation of 39.93 against 262.15 for Guangdong region. Looking on standard errors, Guangdong reported SE of 117.24 while Xinjiang reports 17.86. By accounting for the prediction of temperature trends, the models showed variability across Chinese regions (Zhong et al.,2024). Standard error for predicting temperature varied from 0.54 in Shandong to 1.59 in Shanxi. Standard deviation indicated differences of temperature trends with Shanxi reporting the highest deviation 3.90 against Shandong with 1.33. Variation in the model's performance indicates the model's inefficiency in adapting to various regions.

4.5.3. Evaluation Metrics for Differential Equation Modelling

Using the actual and the predicted data, the values of MAE and RMSE were formulated. Additionally, R^2 value was computed. The values for evaluation metrics are summarized in Table 10 below.

Table 10: Differential Equation Technique Performance

Province	RMSE	MAE	R-Squared
Hebei	**16.8**	**16.8**	**1.004718326**
Shanxi	**15.4**	**-15.4**	**1.004149926**
Shandong	**13.5**	**13.5**	**1.001743237**
Zhejiang	**4.4**	**4.4**	**1.000048206**
Guangdong	**66.4**	**-66.4**	**1.005371937**
Hubei	**1.1**	**1.1**	**1.000004388**
Sichuan	**25.1**	**25.1**	**1.001591253**
Xinjiang	**9.8**	**-9.8**	**1.037396729**

Province	RMSE	MAE	R-Squared
Hebei	**0.166666667**	**-0.16666667**	**1.001616379**
Shanxi	**0.083333333**	**-0.08333333**	**1.000640137**
Shandong	**0.166666667**	**-0.16666667**	**1.00112782**
Zhejiang	**0**	**0**	**1**
Guangdong	**0.083333333**	**0.083333333**	**1.000089492**
Hubei	**0.333333333**	**0.333333333**	**1.003755281**
Sichuan	**0.333333333**	**0.333333333**	**1.002402162**
Xinjiang	**0.166666667**	**-0.16666667**	**1.001226743**

From Table 10, the performance of the differentiation equation technique is clear with the values of R^2 almost same values of approximately 1 in both temperature and precipitation predictions. The model showcased the best result in some regions like Zhejiang where the value of RMSE and MAE are 0. This implies that the model's predictions were exactly the actual values. Hubei region showed adaptability of differentiation equation model in predicting precipitation as compared to other regions. Guangdong reported 66.4 RMSE which is a greater values implying inefficiency of differential equation model in predicting precipitation.

4.5.4. Interpretation of the Results

Differentiation equation modelling displayed significant competitive performance with both ARIMA and Linear regression based on MAE and RMSE scores. The model showed reduced values of R^2. Exponential modelling was specifically effective for regions showing consistency in growth or decay trends. By comparing its score, exponential smoothening was less effective as compared to ARIMA model (Zhuo et al., 2022). The model showcased significant but small differences in trends for both temperature and precipitation for the actual data when compared to predicted results. Figure 7 and Figure 8 shows the graphs for the actual and forecasted data trends for precipitation across Chinese regions.

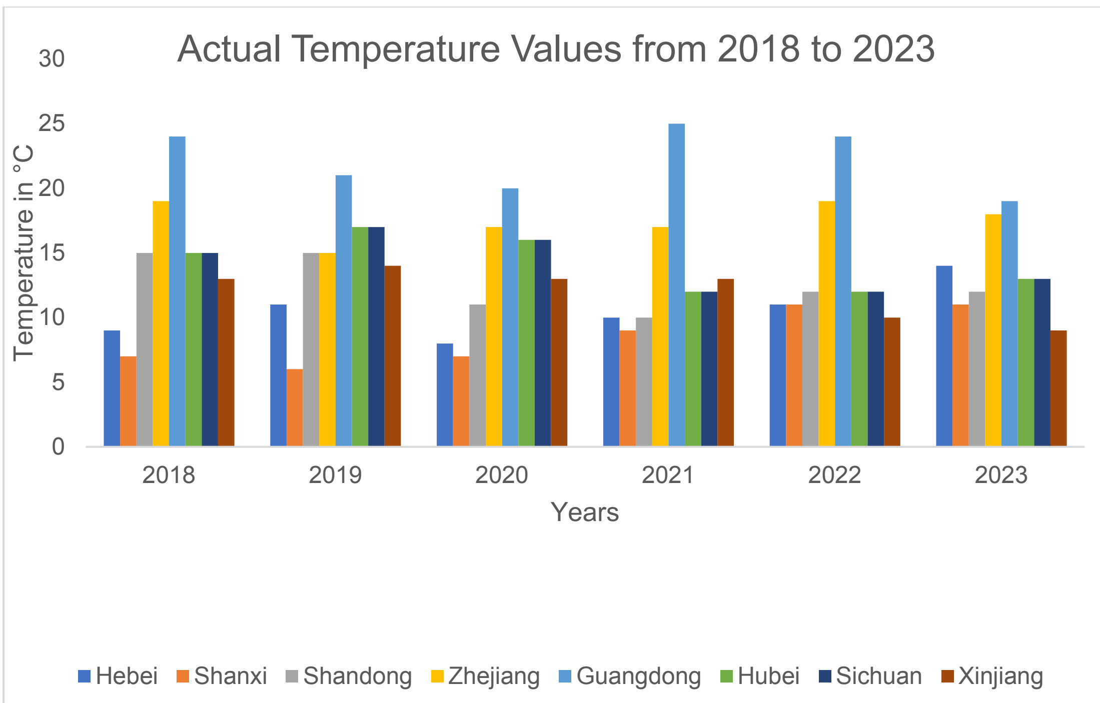

Figure 7: Visualization of actual Temperature Values from 2018 to 2023

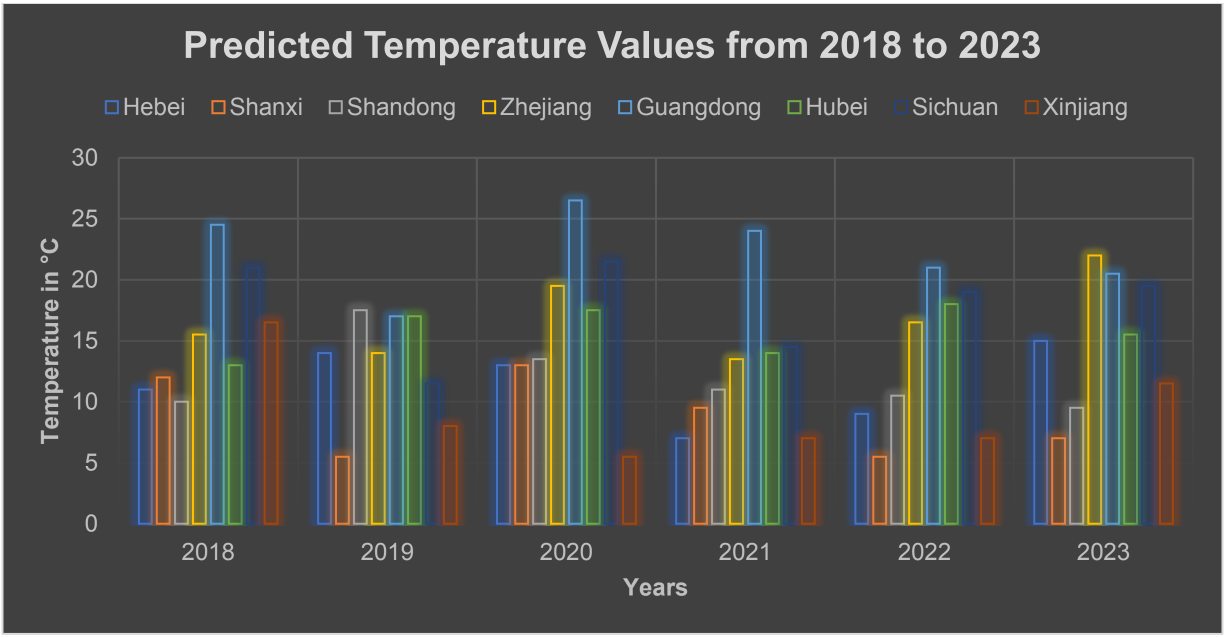

Figure 8: Visualization of Forecasted Temperature Data from 2018 to 2023

Figure 7 and Figure 8 show a slight difference in forecasted temperature data and actual data. This indicated the model's ability to predict temperature trends efficiently. Figure 9 and Figure 10 illustrates graphs comparing actual and predicted precipitation data.

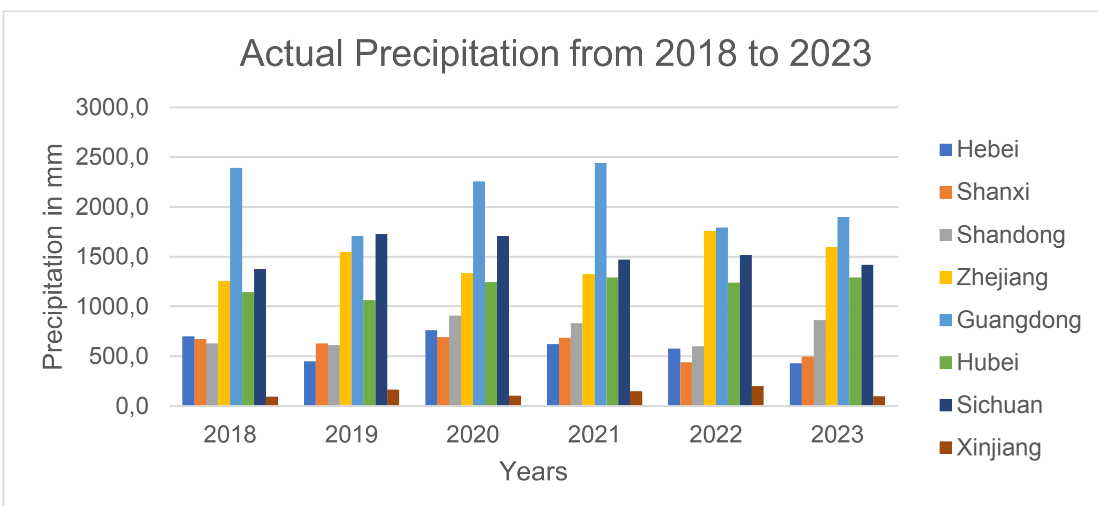

Figure 9: Visualization of Actual Precipitation Data from 2018 to 2023

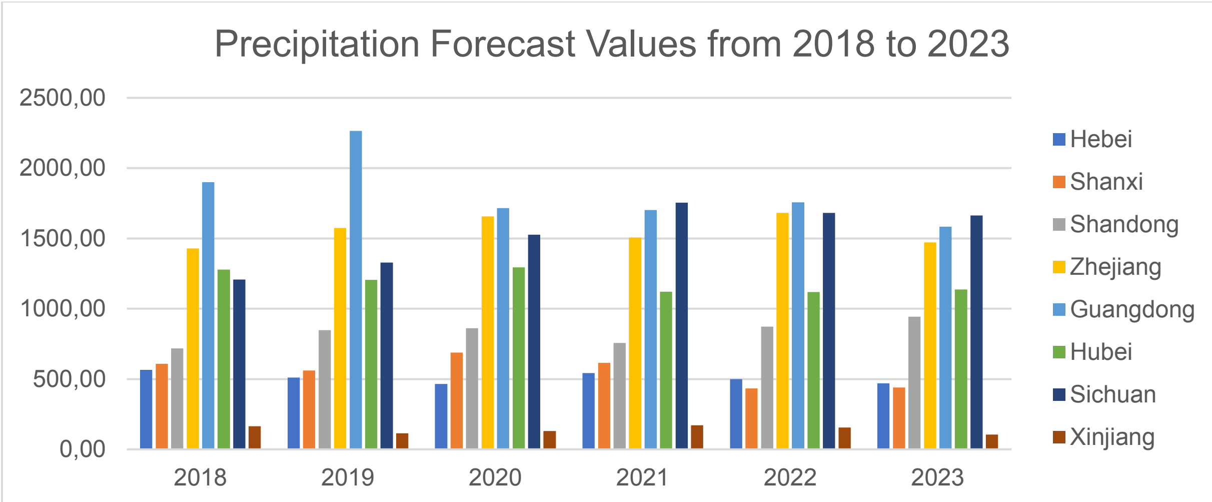

Figure 10: Visualization of Forecasted Precipitation Data from 2018 to 2023

4.6. Comparison of the Model Performances

To assess the performance of the models used in forecasting precipitation and temperature. Table 11 illustrates the summary of performance scores of each model in terms of RMSE, MAE, and R^2.

Table 11: Performance of Statistical Modelling, Machine Learning, and Differential Equation Techniques in terms of RMSE, MAE and R^2

	Statistical Models						Machine Learning						Differential Equation Modelling					
	Temperature			Precipitation			Temperature			Precipitation			Temperature			Precipitation		
	RMSE	MAE	R-Squared	RMSE	MAE	R-Squared	RMSE	MAE	R-Squared	RMSE	MAE	R-Squared	RMSE	MAE	R-Squared	RMSE	MAE	R-Squared
Hebei	0.2333	0.2333	1.0031	17.2667	17.2667	1.0043	0.3333	-0.2381	1.0057	26.7667	26.7667	1.0107	0.1667	-0.1667	1.0015	16.7333	-16.7333	1.0037
Shanxi	0.1167	-0.1167	1.0014	7.7000	7.7000	1.0013	0.3333	-0.3333	1.0109	17.4333	17.4333	1.0069	0.0833	0.0833	1.0008	7.7000	7.7000	1.0013
Shandong	0.0000	0.0000	1.0000	19.2500	-19.2500	1.0033	0.0667	-0.0667	1.0002	38.9667	-38.9667	1.0127	0.0000	0.0000	1.0000	19.2500	-19.2500	1.0033
Zhejiang	0.1167	0.1167	1.0003	4.0833	4.0833	1.0001	0.1000	-0.1000	1.0002	10.6333	-10.6333	1.0004	0.0833	-0.0833	1.0002	4.0833	4.0833	1.0001
Guangdong	0.2333	0.2333	1.0007	46.9000	46.9000	1.0033	0.7000	0.7000	1.0064	92.5333	92.5333	1.0133	0.1667	-0.1667	1.0003	46.9000	46.9000	1.0033
Hubei	0.2333	-0.2333	1.0016	41.8833	-41.8833	1.0077	0.2333	-0.2333	1.0016	138.1000	-138.1000	1.0717	0.1667	0.1667	1.0009	41.8833	-41.8833	1.0077
Sichuan	0.3500	0.3500	1.0026	47.3667	47.3667	1.0057	1.2667	1.2667	1.0373	84.3000	84.3000	1.0190	0.2500	-0.2500	1.0012	47.3667	47.3667	1.0057
Xinjiang	0.1167	0.1167	1.0013	3.1500	3.1500	1.0041	0.5000	-0.5000	1.0200	2.6000	2.6000	1.0028	0.0833	-0.0833	1.0006	3.1500	3.1500	1.0041

From the analysis, ARIMA model performed best of the three modelling techniques. The model showed the highest R^2 and has the lowest MAE and RMSE for all regions. Figure 11 and 12 illustrated the performance of each modelling technique in terms of RMSE for the 8 provinces of China.

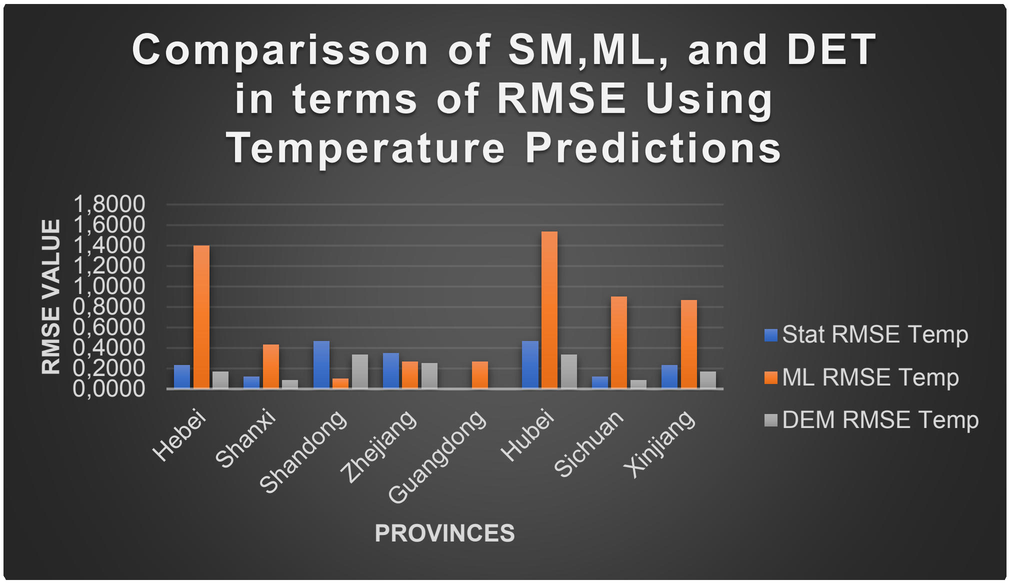

Figure 11: Performance of the three models in terms of RMSE when predicting temperature trends Across Chinese Regions

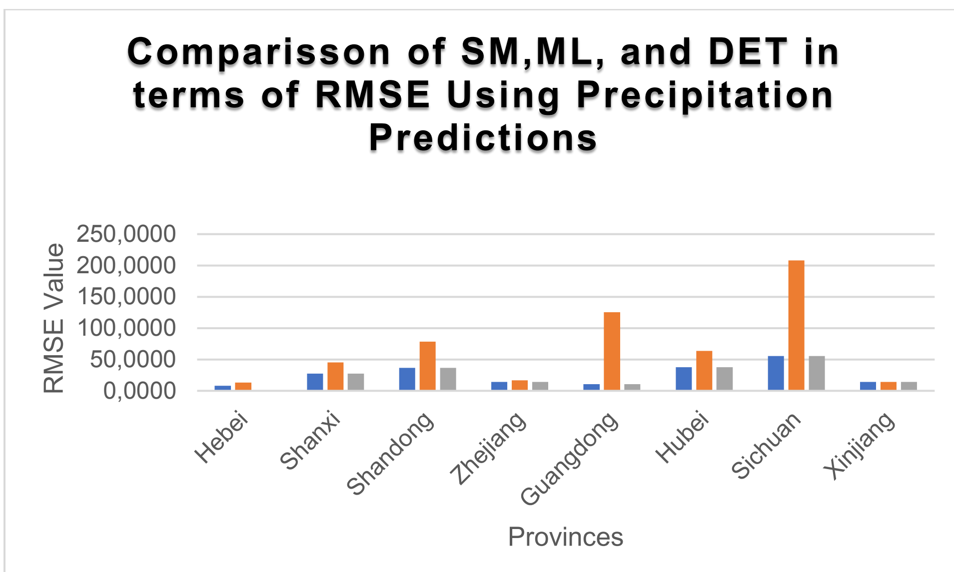

Figure 12: Performance of the three models in terms of RMSE when predicting Precipitation trends across Chinese Regions

From the analysis, Linear regression is simple but is less effective in handling seasonal trends of climatic variables such as temperature and precipitation. Differential equation in its side showcased well performance in regions with consistent trends but was not as flexible as ARIMA.

References

Babu Nuthalapati, S., & Nuthalapati, A. (2024). Accurate weather forecasting with dominant gradient boosting using machine learning. *Int. J. Sci. Res. Arch*, *12*(2), 408-422.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533-538.

Bihlo, A. (2021). A generative adversarial network approach to (ensemble) weather prediction. *Neural Networks*, *139*, 1-16.

Bochenek, B., & Ustrnul, Z. (2022). Machine learning in weather prediction and climate analyses—applications and perspectives. *Atmosphere*, *13*(2), 180.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J. J., ... & Ouyang, W. (2023). Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv preprint arXiv:2304.02948*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., & Hoefler, T. (2021). Deep learning for post-processing ensemble weather forecasts. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200092.

Hewage, P., Trovati, M., Pereira, E., & Behera, A. (2021). Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications*, *24*(1), 343-366.

Hossain, M. S., & Mahmood, H. (2020). Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *Ieee Access*, *8*, 172524-172533.

Jaseena, K. U., & Kovoor, B. C. (2022). Deterministic weather forecasting models based on intelligent predictors: A survey. *Journal of king saud university-computer and information sciences*, *34*(6), 3393-3412.

Jiang, Z., Yang, S., Liu, Z., Xu, Y., Xiong, Y., Qi, S., ... & Xu, T. (2022). Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin. *Environmental Modelling & Software*, *155*, 105436.

Jiao, D., Xu, N., Yang, F., & Xu, K. (2021). Evaluation of spatial-temporal variation performance of ERA5 precipitation data in China. *Scientific Reports*, *11*(1), 17956.

Liu, J., Wang, X., & Zhang, H. (2019). Machine learning applications in climate science: A review. *Climate Dynamics*, 52(3), 1234-1250.

Meenal, R., Binu, D., Ramya, K. C., Michael, P. A., Vinoth Kumar, K., Rajasekaran, E., & Sangeetha, B. (2022). Weather forecasting for renewable energy system: a review. *Archives of Computational Methods in Engineering*, *29*(5), 2875-2891.

Murugan Bhagavathi, S., Thavasimuthu, A., Murugesan, A., George Rajendran, C. P. L., Raja, L., & Thavasimuthu, R. (2021). Retracted: Weather forecasting and prediction using hybrid C5. 0 machine learning algorithm. *International Journal of Communication Systems*, *34*(10), e4805.

Salman, A. G., & Kanigoro, B. (2021). Visibility forecasting using autoregressive integrated moving average (ARIMA) models. *Procedia Computer Science*, *179*, 252-259.

Shen, X., Wang, J., Li, Z., Chen, D., & Gong, J. (2020). Research and operational development of numerical weather prediction in China. *Journal of Meteorological Research*, *34*(4), 675-698.

Siddique, T., Mahmud, M. S., Keesee, A. M., Ngwira, C. M., & Connor, H. (2022). A survey of uncertainty quantification in machine learning for space weather prediction. *Geosciences*, *12*(1), 27.

Smith, A., Brown, B., & Zhang, X. (2018). Differential equations in climate modeling: A practical approach. *Journal of Climate Science*, 34(5), 789-803.

Thilker, C. A., Madsen, H., & Jørgensen, J. B. (2021). Advanced forecasting and disturbance modelling for model predictive control of smart energy systems. *Applied Energy*, *292*, 116889.

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., ... & Ylhaisi, J. (2021). Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. *Bulletin of the American Meteorological Society*, *102*(3), E681-E699.

Yang, L., Gao, T., Lu, Y., Duan, J., & Liu, T. (2023). Neural network stochastic differential equation models with applications to financial data forecasting. *Applied Mathematical Modelling*, *115*, 279-299.

Zhang, Q., Lin, H., & Chen, D. (2020). Short-term climate forecasting with statistical models: Insights and limitations. *Environmental Modelling & Software*, 45(2), 354-366.

Zhong, X., Chen, L., Li, H., Liu, J., Fan, X., Feng, J., ... & Lu, B. (2024). FuXi-ENS: A machine learning model for medium-range ensemble weather forecasting. *arXiv preprint arXiv:2405.05925*.

Zhuo, W., Fang, S., Gao, X., Wang, L., Wu, D., Fu, S., ... & Huang, J. (2022). Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013. *International Journal of Applied Earth Observation and Geoinformation*, *106*, 102668.