5-5-2023

Quantifying the Nexus of Climate, Economy, and Health: A State-of-the-Art Time Series Approach

Kameron Blair Kinast
kbk7499@rit.edu

ROCHESTER INSTITUTE OF TECHNOLOGY

MASTER'S THESIS

Quantifying the Nexus of Climate, Economy, and Health: A State-of-the-Art Time Series Approach

---

by

KAMERON BLAIR KINAST

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in
Applied and Computational Mathematics
School of Mathematical Sciences, College of Science

May 5, 2023

R·I·T
College of
Science

Committee Approval:

Dr. Ernest Fokoué Date

Thesis Advisor, Professor, School of Mathematical Sciences

Dr. Tony Wong Date

Committee Member, Assistant Professor, School of Mathematical Sciences

---

Dr. Jason Quinones Date

Committee Member, Assistant Professor, School of Science, Technology, Accessibility, Mathematics, Public Health, and Psychology

---

Dr. Kara Maki Date

Director of Graduate Programs, School of Mathematical Sciences

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis advisor, Dr. Ernest Fokoué, for his invaluable guidance and support throughout the research process. His expertise and insight have played a noteworthy role in shaping my work and motivating me to strive for excellence. I am particularly grateful for Dr. Fokoué's patience and constructive feedback, which challenged me to refine my ideas and methodologies and deepen my understanding of my research topic. I am truly fortunate to have the opportunity to work with such an exceptional advisor who always brings joy to our weekly thesis meetings.

I also want to extend my gratitude to Dr. Tony Wong and Dr. Jason Quinones for their acceptance of being members of my thesis committee and for their critical advice on my research. Their willingness to contribute their time and knowledge has been deeply appreciated.

I wish to sincerely thank the Rochester Bridges to the Doctorate program for providing much-needed support during my two years of graduate school, including providing me with more research opportunities and paying my tuition. Their commitment to providing scientific mentorship to aspiring scholars like myself has been truly remarkable, and I am honored to have been a part of this program. The members of this program have provided me with guidance, encouragement, and feedback, which played an integral role in shaping my academic pursuits and future plans after graduation.

I am wholeheartedly grateful to my parents for their unwavering love, support, and encouragement throughout my academic journey. FaceTiming with my parents during my graduate school years, including on the darkest or brightest days, is a moment I will always cherish and remember. I am deeply indebted to them for the sacrifices that they have made to provide me with the opportunities that have allowed me to pursue my dreams.

Last but not least, I also am thankful for being able to see my dogs, Hurley, and Roxy, through video texts or FaceTime. Their wagging tails and goofy, playful energy have always managed to brighten my mood and lift my spirits without failing.

ROCHESTER INSTITUTE OF TECHNOLOGY

Abstract

Master of Science in Applied and Computational Mathematics

Quantifying the Nexus of Climate, Economy, and Health: A State-of-the-Art Time Series Approach

by KAMERON BLAIR KINAST

Extreme weather events pose significant threats to human life, the economy, agriculture, and various other socio-economic aspects. This thesis presents a comprehensive analysis of the patterns of climate factors and their impact on the economy and human health using state-of-the-art and emerging statistical machine learning techniques. This research consists of two parts: exploring and comparing the effectiveness of statistical models with respect to climate time series forecasting and analyzing the effects on the economy and human health. The study employs a predominantly computational approach, leveraging R, Python, and Julia to demonstrate the role of statistical computing in understanding climate change and its impacts. This thesis aims to construct powerful statistical models that establish a functional relationship between climate measurements, economic indicators, and human health. Furthermore, we speculate on potential causal relationships within the data to contribute to a deeper understanding of the causes and consequences of extreme weather events. By providing insights into the complex interplay of climate factors, economy, and health, this research seeks to inform evidence-based policy decisions that help mitigate the adverse effects of extreme weather events and foster resilience in the face of dangerous climate change.

Contents

List of Figures

List of Tables

List of Abbreviations

ADF	**A**ugmented **D**ickey-**F**uller
ACF	**A**uto-**C**orrelation **F**unction
AIC	**A**kaike **I**nformation **C**riterion
AR	**A**uto-**R**egressive
ARIMA	**A**uto-**R**egressive **I**ntegrated **M**oving **A**verage
BIC	**B**ayesian **I**nformation **C**riterion
CNN	**C**onvolutional **N**eural **N**etwork
CPI	**C**onsumer **P**rice **I**ndex
DES	**D**ouble **E**xponential **S**moothing
FEVD	**F**orecast **E**rrors **V**ariance **D**ecomposition
FPE	**F**inal **P**rediction **E**rror
HQ	**H**annan **Q**uinn Criterion
KPSS	**K**wiatkowski-**P**hillips-**S**chmidt-**S**hin
LSTM	**L**ong **S**hort-**T**erm **M**emory
MAE	**M**ean **A**bsolute **E**rror
MA	**M**oving **A**verage
MSE	**M**ean **S**quared **E**rror
MLP	**M**ulti-**L**ayer **P**erceptron
OLS	**O**rdinary **L**east **S**quares
PACF	**P**artial **A**uto-**C**orrelation **F**unction
ReLU	**Re**ctified **L**inear **U**nit
RMSE	**R**oot **M**ean **S**quared **E**rror
SC	**S**chwarz **C**riterion
SES	**S**ingle **E**xponential **S**moothing
TES	**T**riple **E**xponential **S**moothing
VAR	**V**ector **A**uto-**R**egressive

Dedicated to My Younger Self

Chapter 1

Introduction

1.1 Background

How has the climate changed over the past 100 years? Figure 1.1 presents the change in the average maximum and minimum temperatures in the United States. In both cases, you can see the average temperatures are increasing over time from 1950 to 2022. Within the figure, a binomial smoothing filter is a data smoothing technique that reveals the underlying trends within the data (Aubury and Luk, 1996). Rising average maximum and minimum temperatures can be associated with changes in weather patterns. For instance, heat waves and winter storms, known as extreme weather events, have become more severe, long-lasting, and frequent due to increasing temperatures.

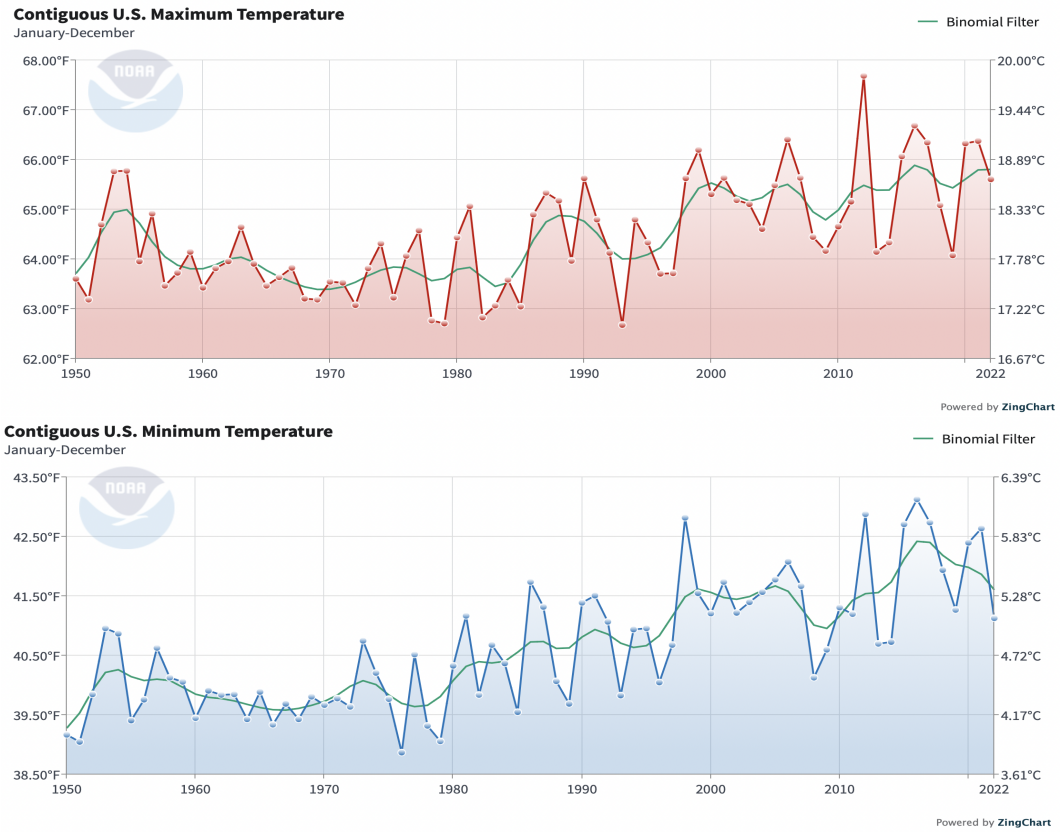


FIGURE 1.1: Line graphs of contiguous U.S. Maximum (top) and Minimum (bottom) temperatures from 1950 to 2022 (*Climate at a Glance: National Time Series* 2023).

Extreme weather events have become increasingly frequent and intense in recent years, posing significant risks to human life, the economy, agriculture, and various other aspects of society (Ebi et al., 2021). Extreme weather events are crucial weather phenomena that record above the thresholds of historical measurements (Seneviratne et al., 2021). According to National Weather Services, extreme weather events, known as Storm Data Events, can be defined in Table 1.1 and many more.

Events Type	Events Type	Events Type	Events Type
Avalanche	Astronomical Low Tide	Hail	Seiche
Blizzard	Debris Flow	Heavy Rain	Storm Surge/Tide
Coastal Flood	Dense Fog	Hurricane	Freezing Fog
Drought	Dense Smoke	Ice Storm	Heavy Snow
Dust Storm	Dust Devil	Lightning	High Wind
Excessive Heat	Flash Flood	Wildfire	Tropical Storm
Extreme Cold	Tornado	Tsunami	Volcanic Ash

TABLE 1.1: Types of Extreme Weather Events

Since 1980, the United States has experienced 341 extreme weather events that cost $2.48 trillion in damages. In 2022, the United States experienced 18 extreme weather events, including Hurricane Ian, which cost $112.9 billion in damages and 152 losses of human lives, and drought in Western and Southern Plains states that cost $22 billion in damages and 136 deaths (Smith, 2023). According to Figure 1.2, the number of severe storms and combined costs of disasters are increasing over time; it is concerning for the future state of this country's economy and human health. The growing impacts of extreme weather events have heightened the need for advanced methods to understand their patterns and relationships with the economy and human health and as well as to predict their consequences.

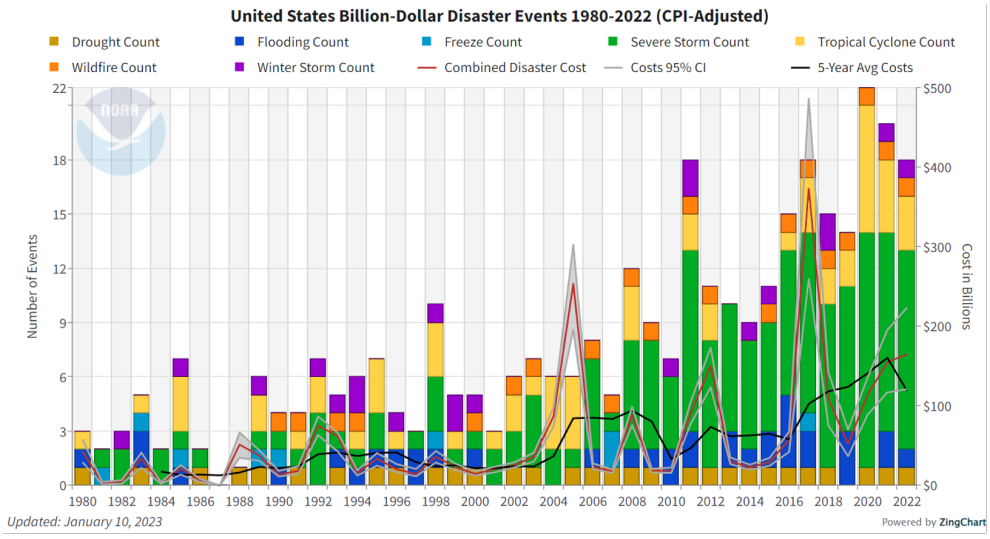


FIGURE 1.2: A bar chart of billion dollars climate disaster in the United States through years from 1980 to 2022 (Smith, 2023)

Climate change affects everyone, but specific communities or populations are affected more than others. Extreme weather events have different amounts of degrees of affecting particular populations or communities. For instance, people aged over 65 are more vulnerable to extreme weather events due to their limited physical ability or mobility. Extreme weather events have a concerning effect on particular social

determinants of health, including the communities of color or low-income populations, due to health disparities, inadequate infrastructures, fewer resources, and many more.

1.2 Motivation

For many decades, climate scientists have had concerns about the impact of climate change. People might have some uncertainty with the rising concerns about the effects of extreme weather events because they have not experienced them firsthand. This study aims to address this challenge by employing state-of-the-art statistical machine learning techniques to uncover hidden patterns within extreme weather event data and analyze their devastating effects on various aspects of human life, including health, economy, agriculture, and more.

Understanding, modeling, and predicting extreme weather events is an ongoing challenge. Scientists use different methods of machine learning or traditional approaches to analyze and predict these events. For instance, various studies have used deep learning techniques to analyze and forecast tropical cyclone intensity (Wang, Wang, and Yan, 2020; Chen, Kuo, and Huang, 2023), precipitation nowcasting (Beutler et al., 2022), and extreme precipitation events (Subrahmanyam et al., 2021). Moreover, Kumar and Middey (2023) create the projection of extreme climate indicators using the hybrid of random-forest and autoregressive integrated moving average model.

Furthermore, the economic and health consequences of extreme weather events have been investigated using machine learning techniques. Paudel et al. (2022) employed machine learning for regional crop yield forecasting in Europe, highlighting the importance of these techniques in assessing the impacts of climate change on agriculture. Nishimura et al. (2021) developed a framework to estimate the number of patients with heat-related illnesses based on average ambient temperatures using the Long Short-Term Memory model.

Thus, the specific questions that this thesis will address:

- How did the maximum and minimum temperatures change over time?
- Which statistical models are best suited for time series forecasting?
- What is the relationship/association between climate factors, economy, and human health considered?
- Do the economic indicators and human health have some dependency on climate factors?

1.3 Thesis Scope

My research will be predominantly computational, with approximately 70% of the work focusing on the development and implementation of statistical models. The remaining 30% of the work will be dedicated to gathering and refining data from various sources available on the internet. In order to construct powerful statistical models, I will establish a functional relationship between climate measurements, economic indicators, and human health. This research will utilize R, Python, and Julia to demonstrate the role of statistical computing in understanding climate change

and its impacts. Furthermore, I will delve into the complex topic of causal inference, speculating on potential causal relationships within the data.

This thesis explores various aspects of traditional approaches and machine learning methods for time series forecasting with climate time series data. Specifically, I will use two classical methods and two machine learning methods. I will only focus on maximum and minimum daily temperatures to explore those time series forecasting methods. The first objective of this research is to provide as thorough a review as possible of the effectiveness of some of the most commonly used classical and machine learning methods when it comes to time series forecasting.

Also, a part of this thesis is dedicated to the study of the relationship/association between climate factors, inflation rates, and fatalities and injuries. I will use a multivariate regression model to determine the co-movements between climate factors, consumer economy, and human health. The second objective of this research is to provide an understanding of the correlation between weather components, economic variables, and human health.

While causal inference remains a challenging endeavor, particularly in the context of complex systems such as climate and economic and health interactions, recent advancements in machine learning and statistical methods provide a foundation for exploring these relationships. By developing and implementing advanced algorithms that can unveil hidden patterns and relationships within the data, I aim to contribute to a deeper understanding of the causes and consequences of extreme weather events. This knowledge, in turn, will inform evidence-based policy decisions and help mitigate the adverse effects.

1.4 Thesis Organization

The thesis consists of 7 chapters, with this chapter being the first and presenting a brief background of the context of the research and motivation of this topic. This thesis is divided into two parts of time series analysis with climate factors: 1) the exploration of time series forecasting methods and 2) the evaluation of the effect on health or economic factors.

Chapter 2 reviews the relevant works of Part 1) different approaches for time series forecasting and Part 2) the impact on human health or economic variables. All previous work shares similar data collection, such as the climate time series dataset of different regions within the United States.

Chapter 3 is considered Part 1, which explores some concepts of time series decomposition and narrows down to four methods for time series forecasting. Chapter 4 is considered Part 2, which examines the context of a multivariate time series regression model. Within these two chapters, each method contains a comprehensive procedure for the development of the models.

Chapter 5 elaborates on the data collection of this research, which illustrates the study areas with detailed data descriptions. Then, this chapter explains the preliminary data preparation, which is data preprocessing, and wraps up with the data visualization. Chapter 6 presents the results of Part 1 and 2, and Chapter 7 summarizes the whole research work and discusses any potential future work.

Chapter 2

Related Work

A time series is data that involves a sequence of data points organized within an interval of time. Therefore, time is a crucial variable, where it can show the change of data points over time. Time series analysis is a method of analyzing historical data to understand the patterns and trends over time. Once the patterns and trends are collected, the time series can predict the likelihood of future events with time series forecasting techniques. Time series analysis splits into two branches, which are univariate and multivariate. Univariate time series consists of a single variable over time, while multivariate time series consists of multiple variables, or more than one variable, over time. Various fields rely on time series analysis for diverse purposes, such as examining financial stocks, tracking climate variables, or monitoring health indicators.

2.1 Univariate Time Series Forecasting

Time series forecasting is a technique for predicting future observations by analyzing past values, or past trends, with the assumption that future trends behave similarly to past trends. In the time series forecasting models (Figure 2.1), the input data, or independent variable, will be the past values of observations where it will forecast the dependent variables (Kumar, 2022). There are many different methods of univariate time series forecasting, and each method has strengths and weaknesses. Numerous comparison studies compare the classical and machine learning methods for univariate time series forecasting.

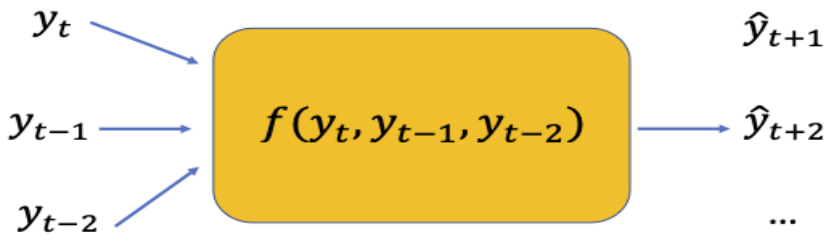


FIGURE 2.1: General Univariate Time Series Forecasting Model

2.1.1 Classical Methods

The classical approach to time series forecasting involves human analysis of the data, while classical methods depend on this analysis to identify patterns within the data. Classical methods typically use predefined techniques and statistical models that include confidence intervals to present a level of certainty. The objective of classical methods is to estimate a forecast value based on the analysis of historical data. The main advantage of classical forecasting methods is the level of transparency of how these methods function.

Naive Forecasting

Naive forecasting appeared in ancient times when mathematicians relied on observed patterns and sequences of events in different scenarios to forecast. Naive forecasting is considered one of the simplest forecasting methods where the forecasts are the value of past observations without any adjustments. In other words, this method generates predictions that are equal to the past observations, or mathematically, $y_{n+1} = y_n$, where y_n is the past observation (Ahuja and Kumar, 2022). Despite its potential for being imprecise at times, the Naive forecasting model is a benchmark for most time series forecasting methods.

Auto-Regressive Integrated Moving Average

In 1970, George Box and Gwilym Jenkins developed Autoregressive Integrated Moving Average (ARIMA) model, also known as the Box-Jenkins model, which uses the data-generating process to identify the patterns that generate and influence past observations (Jenkins, Box, and Reinsel, 2011). For decades, researchers have used the ARIMA model in the application of climate projections. ARIMA model includes three different parameters, ARIMA[p,d,q]. For instance, Figure 2.2 presents the difference in fluctuations of annual average temperature forecasts in Phoenix due to the parameters in the ARIMA model, where the left image illustrates the model of ARIMA [2,1,3] model compared to the model of ARIMA[0,1,1] in the right image (Lai and Dzombak, 2020). ARIMA forecast values are typically influenced by recent values and long-term historical trends.

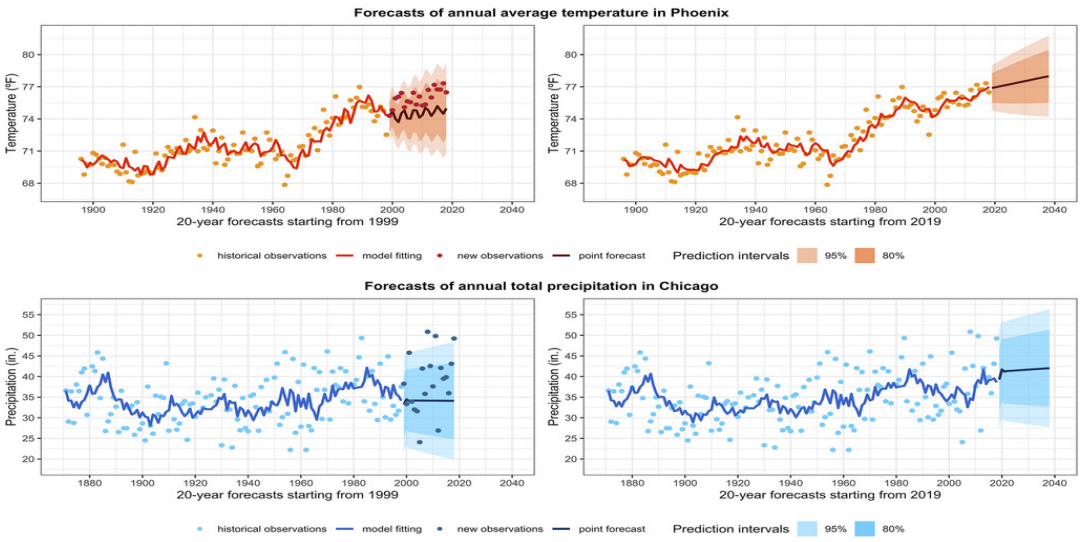


FIGURE 2.2: 20 years ARIMA Projection of (top) annual average temperature in Phoenix and (bottom) total precipitations in Chicago (Lai and Dzombak, 2020)

Exponential Smoothing

Exponential smoothing was first suggested in statistical literature by Robert Brown in 1956, where he used the application of exponential smoothing to forecast the demands in the Navy system. Also, Charles Holt has developed a similar methodology of exponential smoothing independently from Brown; however, Holt has a different approach to smoothing seasonal data than Brown's. Holt's methodology became popular in 1960 when Winters amended a new method to Holt's linear method, which considers all components of time series. Exponential smoothing generates a forecast based on the weighted average of past observations. A study compares three different exponential smoothing with rainfall data (Figure 2.3) and claims that Holt-Winters method is the best exponential smoothing for rainfall data (Dhamodharavadhani and Rathipriya, 2019).

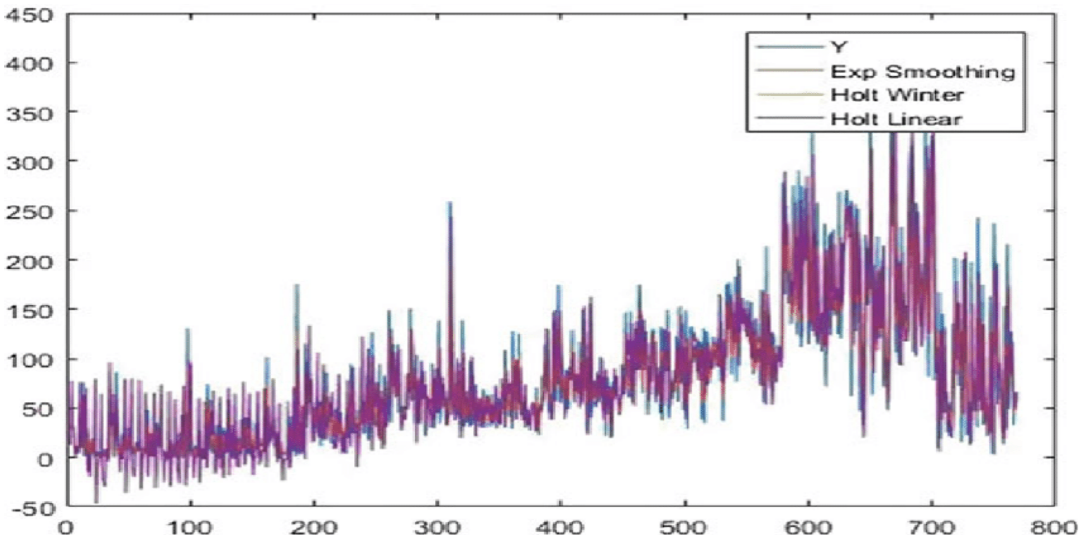


FIGURE 2.3: Rainfall Predictions using Exponential Smoothing methods (Dhamodharavadhani and Rathipriya, 2019).

Various types of applications often utilize these three classical methods for time series forecasting. An empirical study found that the ARIMA model had better accuracy than Holt-Winters exponential smoothing, where they used 106 different time series and measured with average squared error (Newbold and Granger, 1974). A recent study also claimed that the ARIMA model yields higher accuracy than exponential smoothing with weather time series forecasting (Chatwaranon and Tunyasrirut, 2022). The exponential smoothing method considers the time series components, while the ARIMA model considers the autocorrelations. Also, the ARIMA model has the assumption that the time series data is stationary, while exponential smoothing does not. In summary, all models might not give the same results from the same dataset, but it is critical to determine which one works best based on the univariate time series.

2.1.2 Machine Learning Methods

Over the last few decades, machine learning models have made significant strides, utilizing mathematical and statistical models as a framework to conduct data analysis. Unlike classical forecasting methods, machine learning methods rely on artificial intelligence to analyze time series data, rather than human analysis. These methods are more demanding and complex since they require multiple steps to select the

model with the smallest amount of loss. The main goal of machine learning methods is to minimize the losses between actual and predicted observations.

Multilayer Perceptron

Multilayer Perceptron (MLP) model dates back to 1986 when Rummelhart, Hilton, and Williams developed a backpropagation algorithm with a non-linear activation function (Albrecht, 2023). It is believed the researchers use the MLP model for the time series forecasting method because the MLP model can capture the correlations between patterns. In other words, the MLP model can learn a function that computes a forecast value with the input as a sequence of past observations. Like classical methods, the MLP model also has parameters researchers can alter for optimal model predictions. One study developed an optimization method to find an optimal number of hidden layers and hidden neurons for the MLP model with air temperature data in Padang (Kurniawan, Silaban, and Munandar, 2020). For instance, Figure 2.4 shows the optimal MLP model of two hidden layers and five hidden neurons for the air temperature data.

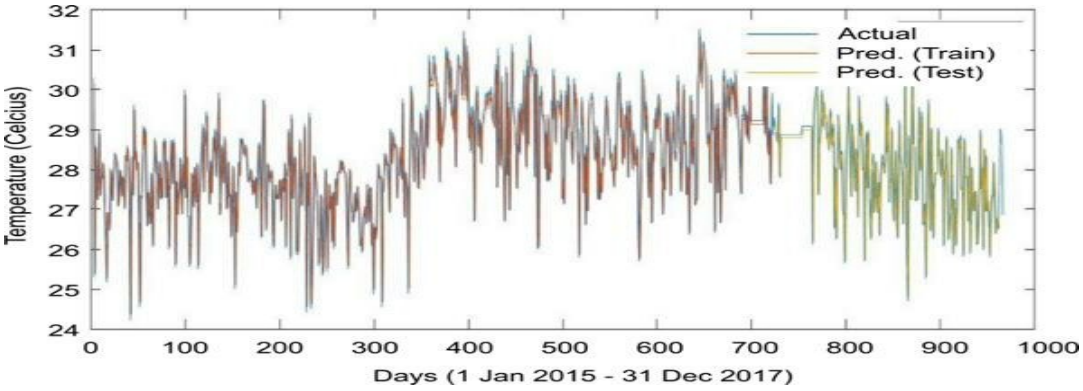


FIGURE 2.4: 2 years MLP Air Temperatures Projection of Actual versus Predicted Values (Kurniawan, Silaban, and Munandar, 2020)

Convolutional Neural Networks

The first appearance of Convolutional Neural Networks (CNNs) was in the 1990s by Yann LeCun and his colleagues. They train CNN with the well-known MNIST database of handwritten digits by introducing the assembly of patterns or dot products of matrices. The ability of the CNN model to recognize the patterns from input data can be useful for time series forecasting problems. Like the MLP model, CNN can learn a function that computes forecast values with input as a sequence of past observations, but CNN will treat a sequence as a one-dimensional image to learn its features. Kurniawan and their colleagues also used the CNN model to predict the air temperature in Padang, and the CNN's prediction was pretty similar to the MLP's prediction in Figure 2.4. However, this study presented CNN as the best performance in predicting the time series data of temperatures.

Long Short-Term Memory

In 1997, the Long Short-Term Memory (LSTM) network was introduced by Hochreiter and Schmidhuber. LSTM is a type of recurrent neural network with the capability of learning long-term dependencies in sequential data (Albrecht, 2023). The LSTM model is considered one of the most advanced and complex time series forecasting models. This model contains four components where a sequence of past observations will go through several layers, and then the model keeps important values before creating forecast values. Like MLP and CNN, LSTM consists of hidden layers and hidden neurons with activation functions in each layer. Inyoung Park and their colleagues studied daily temperatures for two months in Seoul, Korea, and compared the model performance by measuring root mean squared error (Park et al., 2019). Figure 2.5 illustrates a time series plot of observed and predicted temperatures from four models with different parameters.

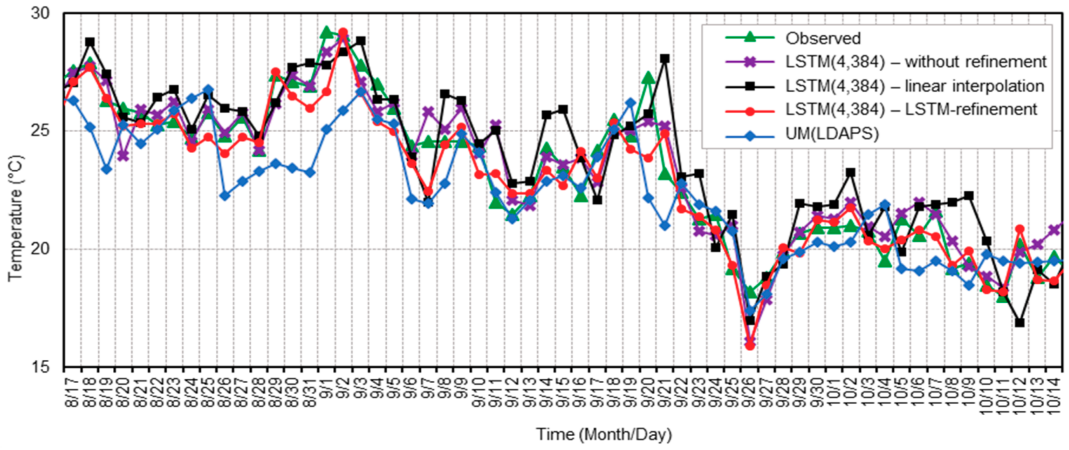


FIGURE 2.5: Daily time series plot of the observed and predicted temperature data for two months from August to October in Seoul, Korea. (Park et al., 2019)

Gaussian Processes

The earliest form of Bayesian optimization was introduced as Wiener processes by Norbert Wiener, which is a continuous-time stochastic process. However, this process is also often referred to as Brownian motion processes due to a physical phenomenon observed by Robert Brown (Keeler, 2022). Wiener processes are considered one of the Gaussian Process (GP), although there is no specific information about who introduced the GP since the concept of Gaussian distribution appeared in the 1900s. GP is a powerful model for time series forecasting that considers multivariate Gaussian distributions, with its covariance being defined by a kernel combination.

A research article studies different applications of Gaussian processes to various types of data, and one of these types is multidimensional weather sensor data. For this data case, Roberts and their colleagues use the sum of a periodic kernel and a disturbance term in the form of Matérn as a covariance for the Gaussian Processes (Roberts et al., 2013). Figure 2.6 illustrates the prediction of the Gaussian process with four different sensor measurements of the tide heights.

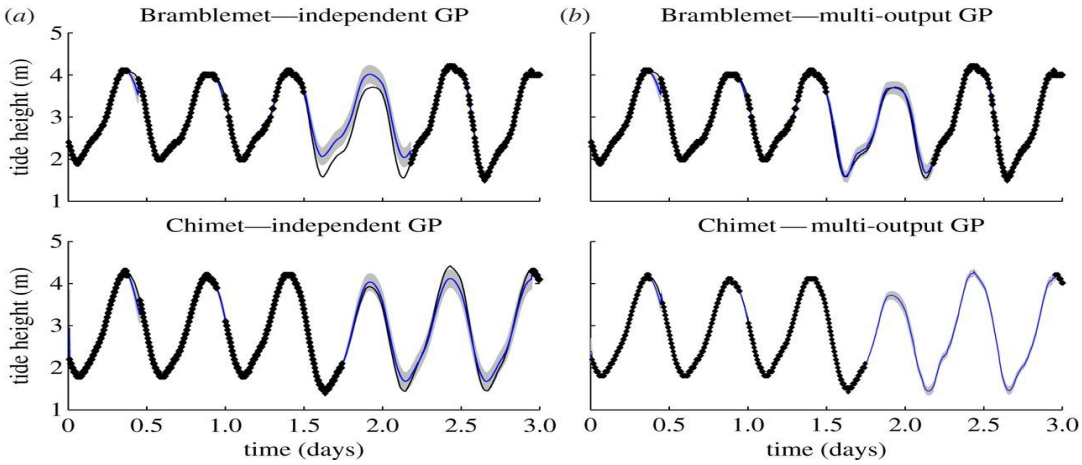


FIGURE 2.6: Gaussian Process's prediction of tide height data for (a) independent and (b) multi-outputs (Roberts et al., 2013)

Many researchers study an empirical comparison of classical and machine learning methods for time series forecasting. Most studies focus on the comparison of ARIMA and neural network models with econometric time series data. There is insufficient evidence of empirical comparison with classical and machine learning methods for time series forecasting with climate time series data. Clearly, econometric and climate time series data have different patterns and behaviors over time. Therefore, this thesis provides an empirical comparison of classical methods and machine learning methods for time series forecasting with climate time series data.

2.2 Multivariate Time Series Regression

Multivariate time series regression is a technique used to analyze the interaction among a group of time series variables. It is more complicated than univariate time series analysis since it involves multiple variables. Several variables that share the same time domain are typically stacked together, enabling the multivariate time series regression to analyze the dynamic relationship between several variables over time, as illustrated in Figure 2.7.

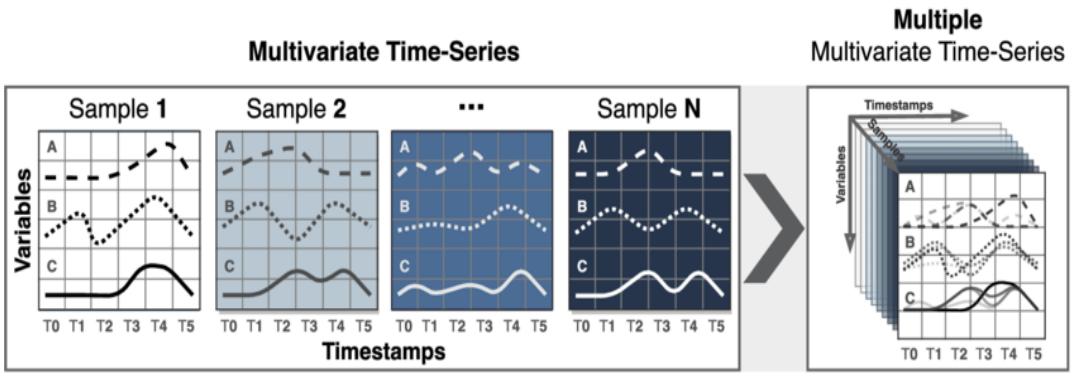


FIGURE 2.7: A multiple variables that shares the same domain of time series stacks together to create a multivariate time series data (Spadon et al., 2020)

Vector Autoregression (VAR) model is one of the most common multivariate time series regression models. The VAR model is one of the most successful and flexible models for the analysis of multivariate time series. Although VAR models are usually used in finance or econometrics, many researchers now apply this model to climate time series along with econometric and public health time series.

2.2.1 Influence on Public Health

Climate change and unpredictable weather events can have a serious effect on human health, especially in terms of the number of fatalities and injuries. Extreme weather events are weather events that are destructive and unpredictable. Fatalities and injuries are the direct health impacts most often associated with extreme weather events. Different parts of the United States have different extreme weather, which is a major cause of fatalities and injuries.

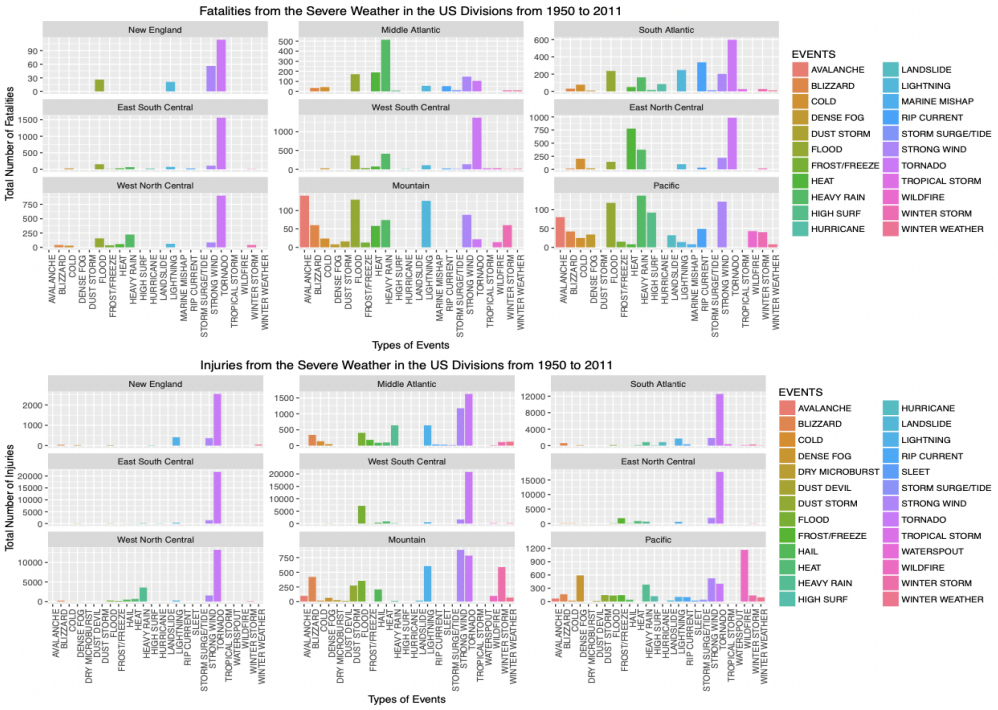


FIGURE 2.8: The number of fatalities and injuries from different extreme events among different regions of the United States from 1950 to 2011. (Kirakosyan, 2016)

Figure 2.8 presents a strong correlation between the number of extreme events and the geographic locations of the United States. For instance, avalanches are a major cause of fatalities in the mountain region, and most injuries within the same region come from strong winds. As expected, wildfires are a major cause of injuries in the Pacific region, and most deaths in the same region are due to heavy rains. All central regions, south Atlantic, and New England regions have major extreme events of a tornado that causes most of the fatalities and injuries.

Many studies use time series regression methods and claim that there is an association between temperatures and the mortality rate. For example, a research study on health vulnerability to extremely cold temperatures in London were conducted for the purpose of informing the Cold Weather Plan for England (Hajat et al., 2016). This

study uses Poisson regression, assuming the outcomes of each variable, including daily mean temperatures, daily all-cause mortality, and daily all-cause emergency admission, follow a Poisson distribution. The result of this study claims that each drop in temperature below 1°C increases the number of deaths by 3.44%. There are other studies that analyze the relationship between extreme heat and mortality rates.

As mentioned, many different studies examine the relationship between extreme weather events and moralities and injuries. However, few to no research articles use the VAR model to analyze the dynamic relationship between extreme weather events and public health. Additionally, not enough studies analyze different climate factors and their effects on human health. Therefore, this thesis analyzes the dynamic behaviors between climate factors and the number of fatalities and injuries by severe storms or extreme weather events.

2.2.2 Impacts on Inflation

Climate change has proven as an environmental issue that has a potentially significant impact on the economy in different locations. As mentioned, climate change influences rising sea levels, extreme rainfalls, and drought in areas where it impacts agricultural and crop production. Among all socio-economic sectors, the agricultural sector has the greatest impact on climate change. There are numerous studies performing analysis to establish the major cause of price spikes.

A study aims to estimate and trace the historical effects of extreme weather events on the change in maize prices. This study uses the structural VAR model to analyze the relationship between monthly data on maize prices in different regions of Kenya and the monthly climate data of maximum and minimum temperatures and rainfalls (Muriuki, Mung'atu, and Waititu, 2018). The result of this study claims that there is empirical evidence of co-movement between maize prices and climate factors in Kenya.

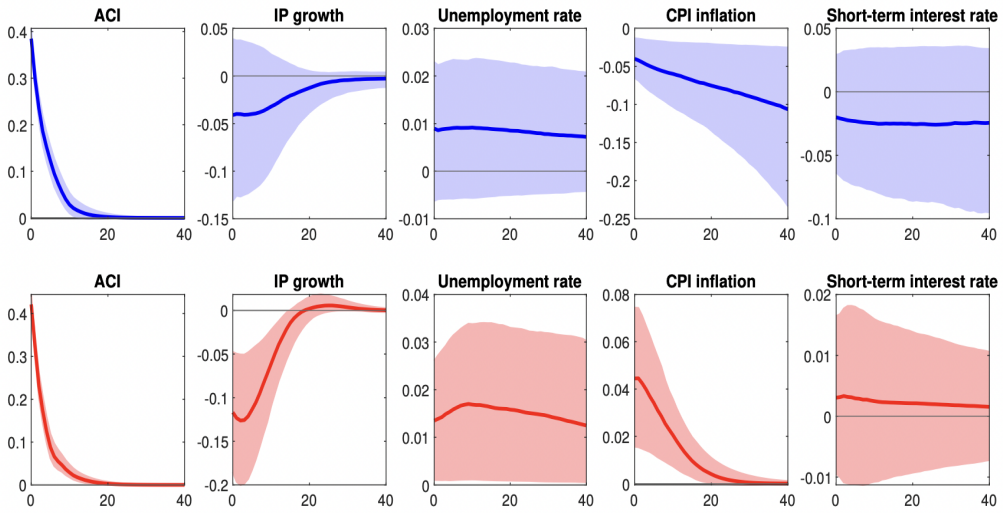


FIGURE 2.9: Impulse response functions of macro variables to severe weather data of the United States. Top panels represent the beginning of the samples, while bottom panels represent the end of the samples. (Kim, Matthes, and Phan, 2011)

Most recent research studies the relationship between severe weather and the macroeconomics of the United States. Their data collection focuses on extreme temperatures, rainfalls, and other extreme weather indicators with four different macroeconomic measures, including the consumer price index, short-term interest rate, and unemployment rate. This research uses the methodology of a smooth transition VAR model, where it adds the weight of a convex combination to the VAR model (Kim, Matthes, and Phan, 2011). Figure 2.9 presents the result of impulse response functions of four macro-variables to the severe weather data. According to the figure, severe weather has a positive effect on industrial production growth and unemployment rate, a negative effect on the consumer price index, and no effect on the short-term interest rate. In conclusion, an increase in severe weather leads to a persistent decrease in the growth rate of industrial production and a persistent increase in the unemployment rate and consumer price index inflation.

Many research studies use a specific type of VAR model to examine the relationship between climate factors and either economic factors or human health, such as mortality and injuries. Most of this research analyzes the impact of extreme weather events on home sales, crop yields, electricity costs, or stock prices. However, the recent study examines this relationship between severe weather and macroeconomic variables, including the consumer price index, within the United States, but it focuses on the entire country. Therefore, this thesis only focuses on the five most populated cities in different regions of the United States. Also, this thesis analyzes the dynamic behaviors between inflation rates and the climate factors among those cities.

Chapter 3

Forecasting Techniques for Time Series

Time series data usually display a variety of patterns, and a time series decomposition can help identify specific characteristics or patterns within time series data. The decomposition of time series data gives a better understanding of time series analysis. The time series decomposition typically focuses on three main components as follows:

- **Trend Component:** A change in data values over a length of time. In other words, it shows the tendency of the data to increase or decrease during a long period of time.
- **Seasonal Component:** Repeated level shifts within the same period of time, indicating the presence of rhythmic patterns in a regular and periodic manner within a time series data
- **Random Noise Component:** Random variations or irregular patterns within the time series data, also known as residuals after all other components are removed from the time series data.

Moreover, the time series decomposition can determine if the model is additive or multiplicative, which is significant for time series forecasting models. If the seasonality exhibits similar patterns of frequency and amplitude, then the time series is an additive model. Otherwise, it is a multiplicative model. In the case of studying maximum and minimum temperatures, the time series model is more likely to be an additive model. Therefore, this thesis will focus solely on additive models.

There are different decomposition techniques to split the components of time series data. This thesis will approach the time series decomposition with the procedure of a seasonal-trend decomposition procedure based on loess (SLT) (Cleveland et al., 1990). Figure 3.1 shows a time series decomposition using the SLT technique. SLT is a simple design that decomposes three components: seasonal, trend, and remainder. If the time series is an additive model, SLT proposes:

$$Y_t = T_t + S_t + R_t, \tag{3.1}$$

where T_t is a trend component in time, S_t is a seasonal component in time, and R_t is a remainder component in time.

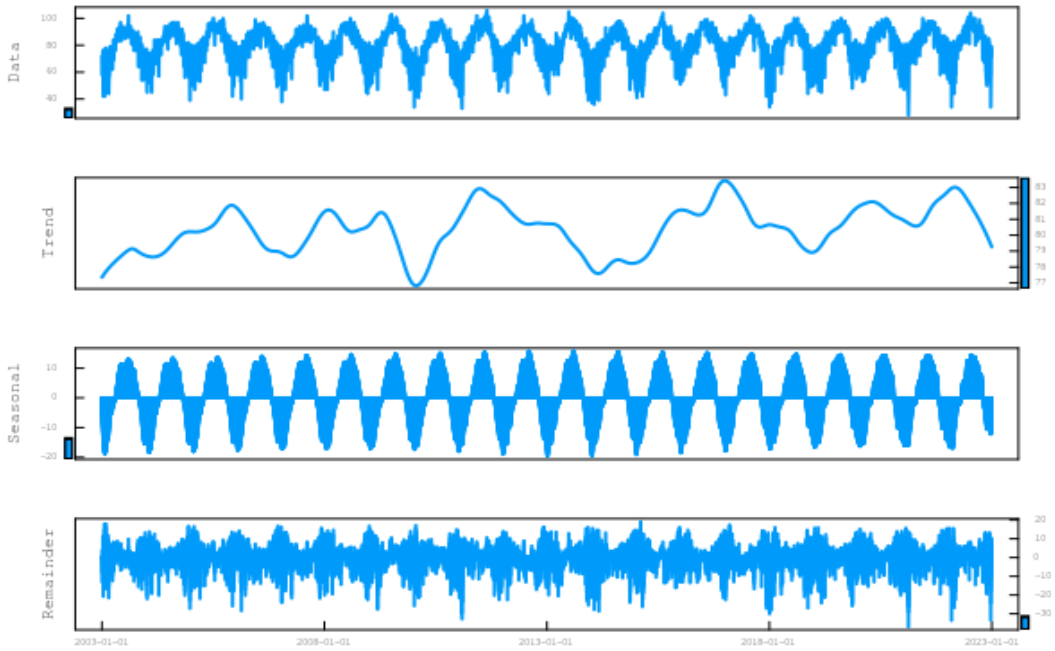


FIGURE 3.1: Time Series Decomposition of Houston's Maximum Temperatures Data

3.1 Selecting Time Series Forecasting Methods

Numerous comparison studies compare the classical and machine learning methods for time series forecasting. For instance, Hill and other colleagues studied the comparison of the neural network model with classical models, including the Box-Jenkins model, single exponential smoothing model, and naive model (Hill, O'Connor, and Remus, 1996). Another study by Alon and their colleagues compared artificial neural networks and classical methods with large economic time series data (Alon, Qi, and Sadowski, 2001). The results of these studies are somehow mixed, but overall, the neural network models have better performance than the classical methods.

Also, Ahmed et al., 2010 performed a comparative study of different machine learning models for time series forecasting. This study claims MLP and GP regression are the best models for time series forecasting with M3 competition data or different types of time series data. However, there are no studies on the comparison of time series forecasting methods with climate time series data. Therefore, this thesis will study four different time series forecasting methods:

- **Classical Time Series Forecasting Methods**
  1. Auto-Regressive Integrated Moving Average
  2. Exponential Smoothing
- **Machine Learning Time Series Forecasting Methods**
  1. Multilayer Perceptron
  2. Gaussian Processes

3.2 Auto-Regressive Integrated Moving Average

3.2.1 Stationarity

A common and conceptually important assumption in time series analysis is the often crucial assumption of *stationarity*. Essentially, stationarity in time series analysis is needed to allow the time series methods to identify the patterns in each interval of time. Intuitively, stationarity captures the recognition that for pattern discovery and extraction to be possible, there has to be something to be extracted and discovered and that pattern must be stable, or constant, or at least less variable with respect to time.

Quintessentially, a time series can be either strictly stationary or weakly stationary.

Definition 3.2.1. A time series is said to be strictly stationary if the probabilistic behavior of sets of random variables,

$$X_{t_1}, X_{t_2}, ..., X_{t_k} \text{ and } X_{t_1+h}, X_{t_2+h}, ...X_{t_k+h}, \tag{3.2}$$

are identical for all integers $t_1, t_2, ...t_k$ and positive integers k and h.

In other words, if the joint probability distributions of random variables are invariant to shifts in time, then the time series is strictly stationary.

Definition 3.2.2. A time series $X_t, t \in \mathbb{T}$ (where $\mathbb{T}$ is the a time index set) is said to be weakly stationary if

1. $E(X_t^2) < \infty$ for all $t \in \mathbb{T}$
2. $E(X_t) = \mu$ for all $t \in \mathbb{T}$
3. $\gamma(X_s, X_t) = Cov(X_s, X_t) = \gamma(X_{s+h}, X_{t+h})$ for all s, t, h $\in \mathbb{T}$. (Hu, 2006).

In other words, a time series is said to be weakly stationary if the time series has a finite variance process (1), a constant first moment (2), and an autocovariance function, $Cov(X_s, X_t)$ depending only on the difference of t and s, $(t - s)$, (3).

Random Walk

A random walk can be expressed as:

$$X_t = X_{t-1} + \epsilon_t \tag{3.3}$$

with $X_t = \sum_{i=0}^{t} \epsilon_i$ and $X_0 = 0$. X_t is independent and identically distributed with mean zero and variance σ^2 (Hu, 2006). Therefore, mean of random walk is $E(X_t) = \mu_X$, and the covariance function is:

$$\gamma(X_t, X_{t+h}) = Cov(X_t, X_{t+h}) = t\sigma^2.$$

Hence, the covariance function is actually dependent on the time t; therefore, the random walk is not stationarity.

White Noise

White noise has no pattern in the time series or just random variation. White noise is a time series with a mean zero and variance σ_ϵ^2, and the covariance function can be written as:

$$Cov(\epsilon_t, \epsilon_{t+h}) = \begin{cases} \sigma_\epsilon^2 & h = 0 \\ 0 & h \neq 0, \end{cases} \tag{3.4}$$

It states that mean and covariance function are independent of time; therefore, the white noise is stationarity. An example of white noise, or known as remainder, is the last panel of Figure 3.1, which is residuals after other components are removed from the time series data.

Unit Root Tests

A time series is considered non-stationary if it exhibits a characteristic known as a "unit root." A unit root is said to exist in a time series when the value of α equals to 1 in the equation below (Prabhakaran, 2022).

$$X_t = \alpha X_{t-1} + \beta X_e + \epsilon_t, \tag{3.5}$$

where X_e is an exogenous variable, ϵ_t is an error term, and α and β are hyperparameters. There are several tests to determine whether the time series data are stationary, such as Augmented Dickey-Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. However, this thesis uses the ADF test to test the null hypothesis that the time series is not stationary. ADF test is an expansion of the Dickey-Fuller test, where the null hypothesis assumes the presence of unit root or $\alpha = 1$ in the equation below (Prabhakaran, 2022).

$$X_t = c + \beta t + \alpha X_{t-1} + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + ... + \phi_p \Delta X_{t-p} + \epsilon_t, \tag{3.6}$$

where X_{t-1} is a lag 1 of time series, $\phi \Delta X_{t-p}$ is a difference term of the time series at (t-p), c is a constant term, β is a coefficient, p is a lag order, and ϵ_t is an error term. If the ADF test's p-value is greater than 0.05, then it fails to reject the null hypothesis and confirms that the data are non-stationary.

3.2.2 Model Components

Autoregressive

Definition 3.2.3. An autoregressive model with an order of p, abbreviated as AR(p), is written as the equation below (Shumway and Stoffer, 2017).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t, \tag{3.7}$$

where X_t is a value of time series that is defined as a function of past p values, $X_{t-1}, X_{t-2}, .. X_{t-p}$, where p determines the number of past values of time series, ϕ_p are constant terms, and ϵ_t is an error term or white Gaussian noise. It can also be written as:

$$X_t = \Sigma_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t. \tag{3.8}$$

If $p = 1$, it is considered AR(1), which is known as the *first-order AR model*. AR(1) is a random walk; therefore, the AR(p) model is an extension of a random walk. Generally, the AR(p) model generates predicted values based on the regression analysis of past time series values.

Moving Average

Definition 3.2.4. A moving average model with an order of q, abbreviated as MA(q), is written as the below equation.

$$X_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ...\theta_q \epsilon_{t-q} + \epsilon_t \tag{3.9}$$

Instead of p, q determines the number of errors of past predictions and θ_q are constant terms as ϕ_p in AR(p) model. It can also be written as:

$$X_t = \Sigma_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{3.10}$$

Unlike the AR (p) model, MA(q) models generate predicted values based on the errors of past forecasts.

Auto-Regressive Integrated Moving Average

ARIMA model is a combination of the AR(p) and MA(q) models, along with the integrated method. As mentioned in Section 3.2.1, if the underlying time series of the data is not stationary according to the unit root tests, the integrated method is applied. This involves differencing the observations to make the time series stationary. Therefore, the ARIMA model requires three parameters, abbreviated as ARIMA(p,d,q). The value of a parameter d is determined by the number of differences applied in the ARIMA model. Additionally, some ARIMA models have a trend where the variance changes over time, so the model has a drift or a constant of c. In terms of X_t, the general ARIMA forecasting equation is:

$$X_t = c + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3.11}$$

3.2.3 Model order selection

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are critical statistical measures for analyzing time series data. The ACF assesses the correlation between time series, considering all the lags or intervals between the time periods, while the PACF only considers certain lags. Lags are a number of intervals between two measurements, or in this case, time series. The order of the MA(q) model is determined by the number of significant lags above the threshold in the ACF plot, while the order of the AR(p) model is determined by the number of significant lags above the threshold in the PACF plot.

To determine the optimal parameter values for the ARIMA model, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are commonly used. These criteria assess the goodness-of-fit of the model to the data while also penalizing models with a larger number of parameters. The AIC measures the information value of the model using maximum likelihood, L, estimates, and the number of parameters in the model k (Equation 3.12). The BIC is similar to the AIC but

has a larger penalty term that considers the number of observations, n, in the data (Equation 3.13).

$$AIC = 2k - 2\ln(L) \tag{3.12}$$

$$BIC = k\ln(n) - 2\ln(L) \tag{3.13}$$

3.2.4 Box-Jenkins Method

The ARIMA model is a form of the Box-Jenkins model, which is a mathematical model that uses regression analysis to fit and forecast time series data. The Box-Jenkins method involves an iterative five-step process to develop a forecasting model for the time series, as outlined by Jenkins, Box, and Reinsel (2011):

1. Ensure the time series are stationary. If the time series is not stationary, then differences can be applied to make it stationary.
2. Identify the best model for the time series. As mentioned, ACF and PACF can be used to select a model that best summarizes the data.
3. Estimate the model parameters. Computer algorithms, such as Julia programming, can be used to specify the coefficients that best fit the selected model.
4. Evaluate the fitted model using the goodness-of-fit tests, such as the AIC and BIC, to measure the model's performance in forecasting.
5. Use the model to forecast future time series values.

3.3 Exponential Smoothing

Exponential Smoothing is one of the traditional time series approaches that forecasts future values of the time series based on the weighted averages of past observations with the assumption that the current observations have more weight than past observations. Since it is more common for time series data to follow an additive model, this thesis only focuses on the additive models of exponential smoothing. There are three main types of exponential smoothing models that can apply to forecast climate factors.

3.3.1 Single Exponential Smoothing

Single Exponential Smoothing (SES) method is the simplest model of exponential smoothing where it doesn't have trend or seasonality components. The SES method only estimates the level component of the time series. A level Component is an average value in the time series. Since this method only computes a single component, it only requires a single parameter, α, that determines the degree of smoothing or the rate at which each observation influences the model.

Definition 3.3.1. The SES formula is given by:

$$L_t = \alpha(X_t) + (1-\alpha)L_{t-1} \qquad 0 < \alpha < 1, \tag{3.14}$$

where L_t is the level component at a time, α is a model parameter of the level component, and X_t is a data point at time t.

The forecasting equation for SES is given by:

$$F_{t+n} = L_t, \qquad n = 1, 2, 3, ... \tag{3.15}$$

where n represents the number of time periods ahead of the current time period t.

3.3.2 Double Exponential Smoothing

Double Exponential Smoothing (DES) method is similar to the SES method, except that this method contains the level and trend components. If the trend displays a constant, this method is commonly known as Holt's Linear method (Dhamodharavadhani and Rathipriya, 2019). Since the DES method involves two components, it involves an additional parameter to the SES method, β, which controls the decay of the influence of the change in trend.

Definition 3.3.2. The DES formula is given by:

$$\begin{aligned} L_t &= \alpha(X_t) + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad 0 < \alpha < 1 \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad 0 < \beta < 1, \end{aligned} \tag{3.16}$$

where T_t is the trend component at time, and β is a model parameter of the trend component.

The forecasting equation of the DES at time t is given by:

$$F_{t+n} = L_t + nT_t, \qquad n = 1, 2, 3, ... \tag{3.17}$$

3.3.3 Triple Exponential Smoothing

Triple Exponential Smoothing (TES) method is an extension of exponential smoothing with an additional component, seasonality. The TES method is also known as Holt-Winters method, which is a developed Holt's Linear method with additional parameters. Therefore, the TES method contains three components, and the additional parameter, γ, controls the influence on seasonality.

Definition 3.3.3. The TES formula of the additive model is given by:

$$\begin{aligned} L_t &= \alpha(X_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad 0 < \alpha < 1 \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad 0 < \beta < 1 \\ S_t &= \gamma(X_t - L_t) + (1-\gamma)S_{t-p} \qquad 0 < \gamma < 1 \end{aligned} \tag{3.18}$$

where S_t is the seasonal component at time, p is a number of periods in a seasonal cycle, such as quarterly, monthly, or weekly, and γ is a model parameter of the seasonal component.

The forecasting equation at time t from the TES method with additive seasonality is given by:

$$F_{t+n} = L_t + nT_t + S_{t-p+n}, \qquad n = 1, 2, 3, ... \tag{3.19}$$

3.3.4 Classification of Exponential Smoothing

Since exponential smoothing does consider three different components within time series data, there is a classification scheme of the exponential smoothing methods. Since this thesis uses Julia programming to develop a model using the exponential smoothing model, the classification of exponential smoothing methods is limited to the availability of the State Space Model Julia package (Hyndman and Athanasopoulos, 2018). Table 3.1 presents a list of all variants of the classification for exponential smoothing methods with a concentration on trend and seasonal components with the additive models only.

		Seasonal	
		N (None)	A (Addictive)
Trend	N (None)	A,N,N	A,N,A
	A (Addictive)	A,A,N	A,A,A
	A_d (Addictive damped)	A, A_d,N	A,A_d,A

TABLE 3.1: Classification of Exponential Smoothing

3.4 Multilayer Perceptron

The Multilayer Perceptron (MLP) is a type of feedforward artificial neural network that is widely used in machine learning applications. Unlike recurrent neural networks (RNNs), which have feedback connections and can process sequential data, MLPs only have forward connections and are primarily used for pattern recognition and classification tasks. The algorithm used to train an MLP is known as backpropagation, which involves adjusting the weights of the network to minimize the difference between the predicted outputs and the actual outputs of the training data. In contrast, RNNs use a variation of backpropagation known as backpropagation through time (BPTT), which allows them to learn from sequences of data over time.

MLPs are a popular model for time series forecasting due to their ability to model complex non-linear relationships between input and output variables. By processing past observations of a time series as input, an MLP can learn to make accurate predictions of future values, making it a powerful tool for tasks such as stock price prediction, weather forecasting, and traffic flow analysis.

Many software packages are available in R, Python, and Julia that simplify the process of implementing and training MLPs for time series forecasting tasks. In R, the "neuralnet" and "nnet" packages provide simple and efficient tools for building and training MLPs. In Python, the "scikit-learn" and "Keras" libraries are popular choices for implementing MLPs, with Keras offering a high-level API for building and training neural networks. Julia has several packages for working with MLPs, including "Flux" and "MLJ" packages. These tools are convenient for experimenting with different network architectures and training strategies, enabling users to develop precise and efficient models for time series forecasting.

3.4.1 Elements of Multilayer Perceptron

As mentioned, MLPs are particularly effective for weather forecasting as they can capture the complex non-linear relationships between different weather variables. The structure typically consists of an input layer, one or more hidden layers of densely connected neurons with non-linear activation functions, and an output layer. The input layer typically consists of lagged observations of weather variables, such as temperature, humidity, and wind speed, while the output layer produces forecasts of future weather conditions. Figure 3.2 depicts an example MLP model with two hidden layers and four hidden neurons in each.

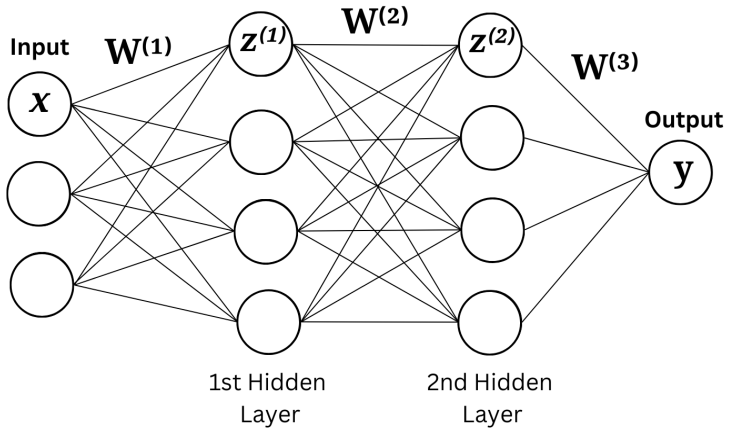


FIGURE 3.2: An example of a simple Multilayer Perceptron model.

Essentially, the computation of this model is conceptually quite simple. Let $\mathbf{x} \in \mathbb{R}^p$ be the input vector, $\mathbf{z}^{(1)} \in \mathbb{R}^{m_1}$ and $\mathbf{z}^{(2)} \in \mathbb{R}^{m_2}$ are the vectors of hidden, latent variables for the first and second hidden layers, respectively, and $y \in \mathbb{R}$ is the output. Suppose we use Figure 3.2 as an example, we have:

$$\begin{aligned} \mathbf{z}^{(1)} &= \sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \\ \mathbf{z}^{(2)} &= \sigma(\mathbf{W}^{(2)}\mathbf{z}^{(1)} + \mathbf{b}^{(2)}) \\ y &= \mathbf{W}^{(3)}\mathbf{z}^{(2)} + \mathbf{b}^{(3)}, \end{aligned} \tag{3.20}$$

The function $\sigma(\cdot)$ is a nonlinear activation function applied elementwise to the output of the linear transformations $\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}$ and $\mathbf{W}^{(2)}\mathbf{z}^{(1)} + \mathbf{b}^{(2)}$, where $\mathbf{W}^{(1)} \in \mathbb{R}^{m_1 \times p}$ and $\mathbf{b}^{(1)} \in \mathbb{R}^{m_1}$ are the weight matrix and bias vector for the first hidden layer, and $\mathbf{W}^{(2)} \in \mathbb{R}^{m_2 \times m_1}$ and $\mathbf{b}^{(2)} \in \mathbb{R}^{m_2}$ are the weight matrix and bias vector for the second hidden layer. Similarly, $\mathbf{W}^{(3)} \in \mathbb{R}^{1 \times m_2}$ and $\mathbf{b}^{(3)} \in \mathbb{R}$ are the weight matrix and bias scalar for the output layer.

Note that an MLP with two hidden layers is a more complex feedforward neural network that allows for a richer set of nonlinear transformations to be applied to the input. The number of hidden units in each layer m_1 and m_2 are hyperparameters that can be tuned to improve the model's performance on a given task. The activation function $\sigma(\cdot)$ can be chosen from a variety of options, such as the sigmoid, tanh, or ReLU functions. The weights and biases can be learned by minimizing a loss function using an optimization algorithm such as stochastic gradient descent.

Overfitting the MLP model can be convenient; however, it can be avoided by adding a regularization technique. Dropout is a technique that randomly ignores selected neurons during the training, reducing the model's sensitivity to specific neurons' weights. The common value for the dropout rate is $p = 0.2$ in each hidden layer. With the dropout rate as d, the hidden layer from Equation 3.20 can be described as (Srivastava et al., 2014):

$$\begin{aligned} d^{(1)} &\sim \text{Bernoulli}(p), \\ \hat{z}^{(1)} &= d^{(1)} \cdot z^{(1)} \\ z^{(2)} &= \phi^{(2)}(\Sigma W^{(2)} \hat{z}^{(1)} + b^{(2)}). \end{aligned} \tag{3.21}$$

An optimizer is an optimization method that improves the quality of the MLP model and determines optimal hyperparameters with the goal of reducing losses. Adaptive Moment Estimation, or Adam for short, is one of the most commonly stochastic optimization algorithms for neural networks. The Adam optimization method stores exponentially decaying averages of past squared gradients based on the first and second moments of the gradients of the loss function (Bushaev, 2018).

Momentum is an extension to a gradient descent algorithm where it not only accumulates the current steps of the gradient but gathers the historical steps of the gradient to smooth the gradient descent. The first moment is the mean of the gradients m_t and the second moment is the uncentered variance of the gradients v_t. The equations for computing the first and second moments are as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{3.22}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \tag{3.23}$$

where g_t is the gradient at time step t, and β_1 and β_2 are decay rates that control the weight given to past gradients. The hyper-parameters, β_1 and β_2, are typically set to 0.9 and 0.999. Since these first and second moments are biased estimators, the next step of the procedure is to use bias correction to ensure the fact that the moments are initialized to zero. Suppose t is the time step, the bias-corrected estimates are given by:

$$\hat{\mu}_t = \frac{m_t}{1 - \beta_1^t}, \tag{3.24}$$

$$\hat{\Sigma}_t = \frac{v_t}{1 - \beta_2^t}. \tag{3.25}$$

Finally, the Adam optimizer's update equation is:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \tag{3.26}$$

where θ_t is the model parameters at time step t, η is the learning rate, and ϵ is a small constant to prevent division by zero.

In order to estimate the losses of the model, the MLP model requires to include a loss function where it allows the updating parameters to minimize the losses. Since the prediction is real values, this MLP model is considered a regression predictive modeling problem. Therefore, this model uses the default loss function, which is Mean Squared Error (MSE) (Equation 3.27).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.27}$$

3.4.2 Training the Model

One commonly employed approach for training neural networks, particularly multilayer perceptrons (MLPs) with multiple hidden layers, is backpropagation. Its mechanism involves the computation of the gradient of the loss function concerning the network's weight and biases. The algorithm then employs this gradient to update the weights and biases by moving them in the direction opposite to the gradient.

The backpropagation algorithm consists of two stages: forward propagation and backward propagation. In the forward propagation stage, the input data is fed through the network, and each hidden layer applies a nonlinear transformation to the input using the associated weights and biases. Then, the output layer performs a linear transformation of the last hidden layer, producing the final network output.

In the backward propagation stage, the gradient of the loss function with respect to each weight and bias in the network is computed using the chain rule of calculus. This process involves calculating the gradient of the output within each layer with respect to the input of that layer and then combining these gradients using the chain rule. Then, the weights and biases are updated in the opposite direction of the gradient, using a learning rate parameter to control the step size of the update.

The backpropagation algorithm iteratively applies these two stages to the input data, adjusting the weights and biases of the network in order to minimize the loss function. This optimization process enables the network to generate accurate predictions on new data by minimizing the loss function. As mentioned, the MLP model is trained using a backpropagation algorithm to minimize the MSE between the predicted and actual values. During training, the weights and biases of the MLP model are updated using the gradient of the error with respect to these parameters.

Algorithm 1 Backpropagation for MLP with two hidden layers

Require: Training data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, learning rate α, number of hidden units m_1 and m_2

Initialize weights and biases randomly: $\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2, \mathbf{W}_3, \mathbf{b}_3$
for each training example $(\mathbf{x}_i, y_i)$ **do**
    Forward propagation:
    $\mathbf{z}_1 = \sigma(\mathbf{W}_1 \mathbf{x}_i + \mathbf{b}_1)$ ▷ First hidden layer
    $\mathbf{z}_2 = \sigma(\mathbf{W}_2 \mathbf{z}_1 + \mathbf{b}_2)$ ▷ Second hidden layer
    $y\text{pred} = \mathbf{W}_3 \mathbf{z}_2 + \mathbf{b}_3$ ▷ Output layer
    Compute loss $L = \frac{1}{2}(y\text{pred} - y_i)^2$
    Backward propagation:
    $\delta_3 = (y\text{pred} - y_i)$ ▷ Gradient of loss w.r.t. output
    $\delta_2 = \sigma'(\mathbf{z}_2)\mathbf{W}_3^T \delta_3$ ▷ Gradient of loss w.r.t. second hidden layer
    $\delta_1 = \sigma'(\mathbf{z}_1)\mathbf{W}_2^T \delta_2$ ▷ Gradient of loss w.r.t. first hidden layer
    $\mathbf{W}_3 = \mathbf{W}_3 - \alpha \delta_3 \mathbf{z}_2^T$ ▷ Update output layer weights
    $\mathbf{b}_3 = \mathbf{b}_3 - \alpha \delta_3$ ▷ Update output layer biases
    $\mathbf{W}_2 = \mathbf{W}_2 - \alpha \delta_2 \mathbf{z}_1^T$ ▷ Update second hidden layer weights
    $\mathbf{b}_2 = \mathbf{b}_2 - \alpha \delta_2$ ▷ Update second hidden layer biases
    $\mathbf{W}_1 = \mathbf{W}_1 - \alpha \delta_1 \mathbf{x}_i^T$ ▷ Update first hidden layer weights
    $\mathbf{b}_1 = \mathbf{b}_1 - \alpha \delta_1$ ▷ Update first hidden layer biases
end for

This MLP model consists of a batch size where it goes through a specific number of samples before updating the internal parameters. In training the MLP model, the training algorithm must go through a specific number of times to work through the entire dataset, which is called a number of epochs. The number of epochs is typically large, which allows the training algorithm to run until the losses are sufficiently minimized. This MLP model is set to run through 2,000 epochs, which is typically for time series forecasting models.

Overall, the MLP is a powerful tool for time series forecasting, capable of capturing complex nonlinear patterns in the data. However, it can also be prone to overfitting and requires careful tuning of its many hyperparameters. As with any machine learning model, the effectiveness of the MLP depends heavily on the quality and quantity of the input data.

3.5 Gaussian Processes

Gaussian Processes (GPs) are a powerful tool for time series analysis, providing a flexible and non-parametric approach for modeling complex and non-linear relationships in data. In GP regression, a GP is used to model the probability distribution over the function values at each time point in the time series, providing a posterior distribution that captures the uncertainty of the model. By using a GP to model the probability distribution over the function values, GP regression allows for the calculation of prediction intervals that reflect the uncertainty of the model.

3.5.1 Kernels

The fundamental definition of a GP is a collection of random variables in which all the finite-dimensional distributions are joint Gaussian distributions for any finite

number (Quadrianto, Kersting, and Xu, 2010). In the context of time series analysis, a GP can be thought of as a probability distribution over the set of all possible time series functions, with the mean and covariance functions defining the GP's properties. The mean function determines the expected value of the time series at each point, while the covariance function determines the degree of correlation between different points in the time series.

The goal of GP is to learn the underlying distribution from the training data, and in order to do that, GP uses the method of Bayesian inference. Assuming that the underlying function generating the time series values at time t is denoted by f_t, GP regression models the function f_t as a Gaussian process with a mean function μ_t and covariance function Σ_t. Suppose the time series X_t is given the observed values up to time t, i.e., $\mathcal{D}_t = X_1, X_2, \ldots, X_t$, the conditional distribution of X_{t+1} given $\mathcal{D}_t$ is then given by:

$$p(X_{t+1} \mid \mathcal{D}_t) = \mathcal{N}(m_{t+1}, v_{t+1}), \tag{3.28}$$

where the posterior mean and covariance are given by:

$$\mu_{t+1} = k_{t+1}^{\top}(K_t + \sigma^2 I)^{-1} X_t, \tag{3.29}$$

$$\Sigma_{t+1} = k(X_{t+1}, X_{t+1}) - k_{t+1}^{\top}(K_t + \sigma^2 I)^{-1} k_{t+1}, \tag{3.30}$$

with $k_{t+1} = [k(X_1, X_{t+1}), k(X_2, X_{t+1}), \ldots, k(X_t, X_{t+1})]^{\top}$, K_t is the $t \times t$ matrix with entries $K_{i,j} = k(X_i, X_j)$, σ^2 is the noise variance, and I is the identity matrix. The kernel functions $k(X_i, X_j)$ specify the correlation between the predicted and observed time series values. In practice, the mean and covariance functions are specified using a kernel function $k(X_i, X_j)$, which can be optimized by maximizing the log marginal likelihood of the training data.

GP regression for time series analysis commonly utilizes various kernels, among which the linear kernel, the squared exponential kernel, the Matérn kernel, and the periodic kernel are particularly popular. The linear kernel is a simple kernel that simply does Bayesian linear regression (Equation 3.31). The squared exponential kernel, also known as radial basis function (RBF), is a smooth and infinitely differentiable kernel that is commonly used for modeling stationary time series (Equation 3.32). The Matérn kernel is a more flexible kernel that can model non-stationary time series, with the smoothness parameter controlling the degree of differentiability of the function (Equation 3.33). The periodic kernel is a specialized kernel for modeling time series with periodic behavior, such as seasonal fluctuations in economic data or daily temperature patterns (Equation 3.34). Each kernel has its own set of hyperparameters that can be optimized to improve the model's performance on the training data.

$$k_{Linear}(X_i, X_j) = X_i X_j, \tag{3.31}$$

$$k_{RBF}(X_i, X_j) = \exp\left(-\frac{(X_i - X_j)^2}{2\ell^2}\right), \tag{3.32}$$

$$k_{Matern}(X_i, X_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{\ell}|X_i - X_j|\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{\ell}|X_i - X_j|\right), \tag{3.33}$$

$$k_{Periodic}(X_i, X_j) = \exp\left(-\frac{2\sin^2(\pi|X_i - X_j|/p)}{\ell^2}\right), \tag{3.34}$$

The linear kernel is a simple dot product between the two time series. The RBF kernel has a hyperparameter, ℓ, that functions as a characteristic length scale of the kernel. The Matérn kernel has an extra hyperparameter, ν, which functions as a smoothness parameter. K_ν denotes the modified Bessel function of the second kind. The periodic kernel has a hyperparameter of p that represents the periodicity of the kernel. Note that these kernels are stationary and do not depend on time, but only on the distance between the points X_i and X_j.

GPs typically have a kernel composition where several kernels are combined by adding or multiplying the kernels. The kernel composition influences the shape of the resulting distribution. For weather forecasting, the common kernel composition in GPs is:

$$k(X_i, X_j) = k_{Linear}(X_i, X_j) + k_{Periodic}(X_i, X_j) + k_{RBF}(X_i, X_j), \tag{3.35}$$

where the linear kernel provides the trend, the periodic kernel provides the seasonal patterns, and the RBF kernel introduces the non-linear trends.

3.6 Performance Evaluation

An optimal time series forecasting model will be determined by evaluating the performance of forecasting with the testing datasets. To evaluate the performance of four different time series forecasting models, the statistical measures are used in the following: root mean squared error (RMSE) and mean absolute error (MAE). RMSE measures the Euclidean distance between predicted values and observations (Equation 3.36), and MAE measures the magnitude of difference between predicted values and observations (Equation 3.37).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{3.36}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y - \hat{y}| \tag{3.37}$$

Chapter 4

Vector Autoregression

Vector autoregression is an extension of simple autoregressive processes where it also analyzes other variables. VAR models study the relationship of multiple variables within the same time series. Suppose the time series is $X_{i,t}$ where $i = 1, .., m$, and $t = 1, ..., T$, the VAR model with the order of p, or VAR(p), can be written as:

$$\begin{aligned}
X_{1,t} &= \phi_{1,0} + \phi_{1,1}X_{1,t-1} + \phi_{1,2}X_{2,t-1} + \ldots + \phi_{1,m}X_{m,t-1} + \ldots + \phi_{1,1}X_{m,t-p} + \epsilon_{1,t} \\
X_{2,t} &= \phi_{2,0} + \phi_{2,1}X_{1,t-1} + \phi_{2,2}X_{2,t-1} + \ldots + \phi_{2,m}X_{m,t-1} + \ldots + \phi_{2,1}X_{m,t-p} + \epsilon_{2,t} \\
&\vdots \\
X_{m,t} &= \phi_{m,0} + \phi_{m,1}X_{m,t-1} + \phi_{m,2}X_{n,t-1} + \ldots + \phi_{m,m}X_{m,t-1} + \ldots + \phi_{m,1}X_{m,t-p} + \epsilon_{m,t},
\end{aligned} \tag{4.1}$$

where the ϕ_i are coefficients and the error term, ϵ_t, is white noise processes, which is serially uncorrelated. Let $X_t = [X_{1,t}, X_{2,t}, ...X_{m,t}]^\top$ denote an $(m \times 1)$ vector of time series variables. VAR(p) model can be written in matrix form as:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + u_t. \tag{4.2}$$

The ϕ_0 and u_t is $(m \times 1)$ vectors of constant and error terms, and $\phi_1, ..., \phi_p$ are $(m \times m)$ coefficients matrices.

For the study of time series data, VAR(p) might need to add deterministic terms to define the trend and seasonal components of the time series. Additionally, exogenous variables might be required as well. Therefore, the VAR(p) model with the deterministic terms and exogenous variables may be defined as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \xi D_t + \delta A_t u_t, \tag{4.3}$$

where D_t represents an $(m \times 1)$ vector of deterministic components, such as constant, trend, or seasonal variables, A_t represents an $(l \times 1)$ vector of exogenous variables, and ξ and δ are parameter matrices (Zivot and Wang, 2003).

4.1 Type of Models

A VAR model consists of many parameters, which might become difficult to interpret because of the complex interactions between the variables. Therefore, the VAR model breaks down into components with specific focuses and functions.

4.1.1 Structural Model

Definition 4.1.1. A structural VAR with the order of p, abbreviated as SVAR(p), can be described as:

$$BX_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + u_t, \tag{4.4}$$

where B is a main diagonal term or a matrix of B_{ij} that has $i = j$. The matrix B describes the contemporaneous effects of variables.

For instance, a bivariate first-order VAR model is:

$$\begin{aligned} X_{1,t} &= \phi_{1,0} + \phi_{1,1} X_{1,t-1} + \phi_{1,2} X_{2,t-1} - b_{1,2} X_{2,t} + \epsilon_{1,t} \\ X_{2,t} &= \phi_{2,0} + \phi_{2,1} X_{1,t-1} + \phi_{2,2} X_{2,t-1} - b_{2,1} X_{2,t} + \epsilon_{2,t} \end{aligned} \tag{4.5}$$

where $= \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2t} \end{pmatrix} \sim iid \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)$.

It is assumed that the error terms, $\epsilon_{1,t}$ and $\epsilon_{2,t}$, are uncorrelated. Therefore, the coefficient $b_{1,2}$ describes the contemporaneous effect of a change in $X_{2,t}$ on $X_{1,t}$, and similarly, $b_{2,1}$ describes the effect of a change in $X_{1,t}$ on $X_{2,t}$. The next subsection will discuss the analysis of these parameters.

The SVAR model has structural shocks to each variable, which are zero-mean white noise processes that are serially uncorrelated and independent of each other. A general VAR model typically has symmetric covariance matrices, which prevents the analysis of the causal relationship between the variables. Therefore, SVAR models add restrictions on covariance matrices to identify the structural relationships. However, the drawback of this type of model is that it depends on more or less subjective assumptions made by the researchers. Generally, the SVAR model is usually used as an application to economic time series data.

4.1.2 Reduced Form Model

Definition 4.1.2. By multiplying B^{-1} in the structural VAR model

$$X_t = B^{-1}\phi_0 + B^{-1}\phi_1 X_{t-1} + B^{-1}\phi_2 X_{t-2} + \ldots + B^{-1}\phi_p X_{t-p} + B^{-1}u_t, \tag{4.6}$$

the reduced form VAR model is given by

$$X_t = A_0 + A_1 y_{t-1} + A_2 X_{t-2} + \ldots + A_p X_{t-p} + u_t \tag{4.7}$$

where $A_i = B^{-1}\phi_i$ for $i = 0, 1, \ldots, p$ and $u_t = B^{-1}\epsilon_t$ (Kotze, n.d.).

Since the error terms in the reduced form VAR model are composites of the structural shocks, it means that a structural shock, ϵ_t, can potentially influence the shocks in all error terms u_t. Going back to the bivariate example of the structural VAR model, the reduced form of the bivariate VAR model is:

$$\begin{aligned} X_{1t} &= A_{01} + A_{11}y_{t-1} + A_{12}X_{t-1} + u_{1t} \\ X_{2t} &= A_{02} + A_{21}y_{t-1} + A_{22}X_{t-1} + u_{2t}, \end{aligned} \tag{4.8}$$

where the covariance matrix is $E(u_t, u_t') = B^{-1}E(\epsilon_t, \epsilon_t')B'^{-1}$. This covariance matrix of the reduced form VAR model can have non-zero off-diagonal terms, which means there is a correlation between error terms. Therefore, this model cannot determine the individual impacts that each structural shock has on the model itself.

There is another form of the reduced form of the VAR model that may be used for impulse response functions in Section 4.4.2:

$$A(L)X_t = A_0 + u_t \tag{4.9}$$

where $A(L) = I_2 - A_iL$ for $i = 1, \ldots, p$. The reduced form VAR model is the simplest model, considering each variable to be a function of its own past values and the past values of other variables. The drawbacks of the reduced form VAR model are that no variable has any direct contemporaneous effect on other variables in the model, and as mentioned, the error terms are correlated.

4.1.3 Recursive Model

The reduced form of the VAR model loses the identification of the relationship in structural shocks, but the recursive VAR model can recover them by applying the Cholesky decomposition. In other words, the recursive VAR model uses a Cholesky decomposition to orthogonalize the disturbances with no direct contemporaneous effects. Since error terms of the reduced form VAR model are correlated, the recursive VAR model includes some contemporaneous values as regressors to construct the error terms in order to not correlate the error terms from each other.

4.2 Development of the Model

The development of the VAR(p) model involves a certain procedure to follow to receive the results of the relationship between multiple variables. Figure 4.1 illustrate the steps involved in constructing the VAR model, which include:

1. **Check for the Stationarity of Time Series Data**

   The stationarity of time series data is critical to the development of the VAR model, as the statistical properties of the model may present inaccurate information. As with the ARIMA model, the ADF test is used in this thesis to determine whether the time series data is stationary.

2. **Combine the Datasets**

   To perform multivariate analysis, the climate dataset is refined to match the time series of either the economic or human health dataset, allowing the datasets to be combined.

3. **Select an Optimal Lag Length**

An optimal lag length for the VAR model is selected using information criteria to determine the performance of the model with a lag length of 1 to 20. More details on this step are presented in the next section of this chapter, Section 4.3.

4. **Train the VAR Model**

   Using R programming, the VAR model with its optimal lag length is fit to the training datasets. The model then presents the summary of VAR Estimation Results, which provides the standard error, t-value, and p-value for each relationship of variables in each number of lags. Additionally, the results provide the covariance and correlation matrix of residuals.

5. **Evaluate the Model**

   The VAR model has many different parameters that may be difficult to interpret due to the complex interaction between variables. Therefore, the dynamic properties of a VAR model may be summarized using three main types of structural analysis: the Granger Causality test, Impulse Response Function, and Forecast Error Variance Decomposition. These details on the three types of analysis are presented in 4.4.

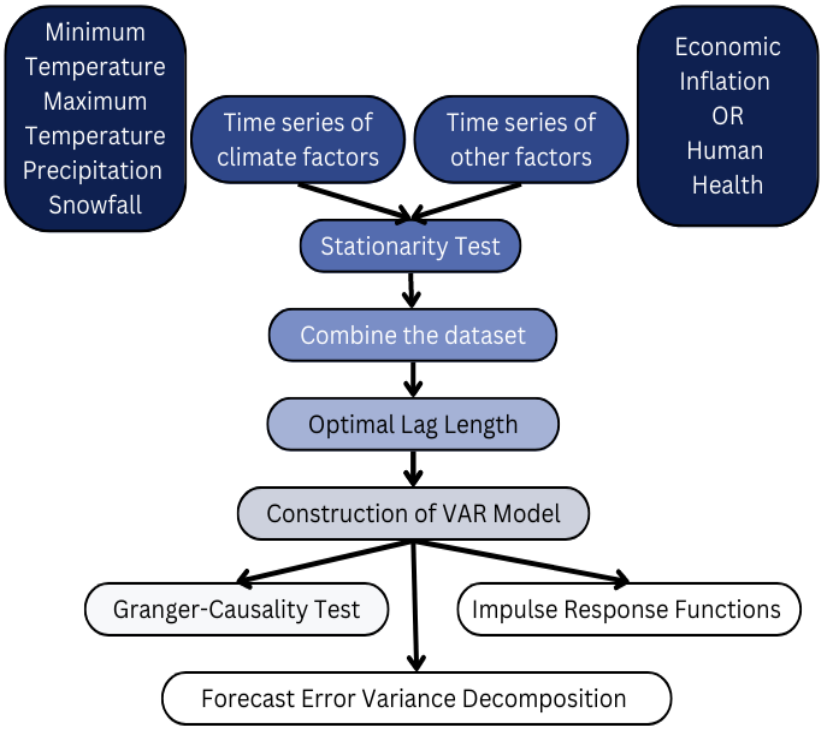


FIGURE 4.1: A flowchart of constructing the VAR model.

4.3 Model Order Selection

Fitting a VAR model involves selecting a lag length, p, and the optimal value for the lag length can be determined using model selection criteria. In this thesis, the VAR(p) model is fitted using R programming. The R function for selecting the lag length in VAR produces four different information criteria that evaluate the performance of

the model with different numbers of lags. The lag length selection R function estimates the VAR model using ordinary least squares (OLS), and the model selection criteria for VAR(p) models take the form:

$$IC(p) = \ln |\tilde{\Sigma}_u(p)| + c_T \cdot \varphi(n, p). \tag{4.10}$$

Here, $\tilde{\Sigma}$ is the estimated covariance matrix of the residual, c_T is a sequence indexed by the sample size T, and $\varphi(n, p)$ is a penalty function (Zivot and Wang, 2003). The information criteria include Akaike information criterion (AIC) (Equation 4.11), Schwarz criterion (SC), or Bayesian information criterion (BIC) (Equation 4.12), Hannan-Quinn Criterion (HQ) (Equation 4.13), and Final Prediction Error (FPE) (Equation 4.14). The formulas for computing these criteria are as follows:

$$AIC(p) = \ln |\tilde{\Sigma}_u(p)| + \frac{2}{T} pn^2, \tag{4.11}$$

$$BIC(p) = \ln |\tilde{\Sigma}_u(p)| + \frac{\ln(T)}{T} pn^2, \tag{4.12}$$

$$HQ(p) = \ln |\tilde{\Sigma}_u(p)| + \frac{2\ln(\ln(T))}{T} pn^2, \tag{4.13}$$

$$FPE(p) = \left(\frac{T + np + 1}{T - np + 1}\right)^n |\tilde{\Sigma}_u(p)|, \tag{4.14}$$

where T is a sample size and n is the total number of equations in the VAR(p), or the number of rows in $X_t \sim (n \times 1)$ in Equation 4.3.

4.4 Model Evaluation

4.4.1 Granger Causality Test

The Granger Causality test provides information about a variable's forecasting ability for other variables. For instance, if a time series X_t can improve the forecast of another variable Y_t, then it said X_t *Granger-cause* Y_t. In other words, Granger Causality may determine the causal influence between variables.

Granger Causality test uses the F-test, which determines the joint significance of the coefficients in all numbers of lags. The null hypothesis of the Granger Causality test is:

$$H_0 : X_t \text{ fails to Granger-cause } Y_t. \tag{4.15}$$

As the p-value is below the significance level, it reveals that a variable is relevant for predicting another variable.

4.4.2 Impulse Response Functions

An impulse response function provides a clearer understanding of the VAR model's dynamic behavior. The impulse response function depends on the moving average representations of coefficients.

Definition 4.4.1. A Wold decomposition theorem states every covariance-stationary time series can be written as the sum of a deterministic component and a component of an infinite moving average representation. Multiplying Equation 4.9 by $A(L)^{-1} = (I_2 - A_1 L)^{-1}$, we get:

$$\begin{aligned} X_t &= \mu + \Psi(L)u_t, \\ \Psi(L) &= \sum_{i=0}^{\infty} \Psi_i L^i, \Psi_0 = I_2, \Psi_i = A_i, \end{aligned} \tag{4.16}$$

where μ is a deterministic component, and $\Psi(L)$ is a component that has an infinite moving average representation.

Definition 4.4.2. Structural Moving Average (SMA) of time series is derived from the infinite moving average representation of the structural shocks, ϵ_t (Zivot and Wang, 2003). Suppose $u_t = B^{-1}\epsilon_t$, Equation 4.16 gives:

$$\begin{aligned} X_t &= \mu + \Psi(L)B^{-1}\epsilon_t, \\ X_t &= \mu + \Theta(L)\epsilon_t, \\ \Theta(L) &= \sum_{i=0}^{\infty} \Theta_i L^i, \Theta_0 = B^{-1}, \Theta_i = A_i B^{-1}. \end{aligned} \tag{4.17}$$

Θ_0 captures the initial impacts of structural shocks and determines the contemporaneous correlation between two variables.

Also, the elements in Θ matrix can give the impulse responses of the variable to changes in the structural errors. Let's go back to an example of the bivariate first order of the VAR model. Suppose the structural VAR model with h steps ahead, may be defined in a matrix form as:

$$\begin{bmatrix} X_{1,t+h} \\ X_{2,t+h} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t+h} \\ \epsilon_{2,t+h} \end{bmatrix} + \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} + \dots \tag{4.18}$$

Therefore, the impulse response of the bivariate VAR model is defined as the following:

$$\begin{aligned} \frac{\partial X_{1,t+h}}{\partial \epsilon_{1,t}} &= \theta_{1,1}, \frac{\partial X_{1,t+h}}{\partial \epsilon_{2,t}} = \theta_{1,2} \\ \frac{\partial X_{2,t+h}}{\partial \epsilon_{1,t}} &= \theta_{2,1}, \frac{\partial X_{2,t+h}}{\partial \epsilon_{2,t}} = \theta_{2,2}, \end{aligned} \tag{4.19}$$

where $\theta_{i,j}$ is $(i,j)$the element of Θ matrix. Thus, the plots of $\theta_{i,j}$ are called the orthogonal impulse response functions.

The impulse response function can't confirm if there is a contemporaneous correlation between the variables in the VAR model. The VAR model's estimated correlated

matrix can confirm if there is a correlation between error terms, but it remains unclear in which direction the causal influences go. There are three different impulse response functions to identify the structural shocks of a VAR model. However, this thesis uses the orthogonal impulse response. This type of response enables the decomposition of the model's variance-covariance matrix. In other words, the impulse response function transforms the VAR model into the recursive model, as mentioned in previous sections.

4.4.3 Forecast Error Variance Decomposition

The forecast error variance decomposition (FEVD) measures the amount of variance of forecast errors in each variable that contributes to other variables. In other words, the FEVD determines the proportion of the variance of the forecasting errors in each variable at time based on the variance in structural shocks.

Definition 4.4.3. Using the Wold decomposition of Equation 4.16, the linear forecast of a variable with h-steps ahead is:

$$X_{t+h|h} = \mu + \Psi_h u_t + \Psi_{h+1} u_{t-1} + \ldots, \tag{4.20}$$

with the forecast error of

$$X_{t+h} - X_{t+h|t} = u_{t+h} + \Psi_1 u_{t+h-1} + \ldots + \Psi_{h-1} u_{t+1}. \tag{4.21}$$

Since $u_t = B^{-1}\epsilon_t$ and $\Theta_i = \Psi_i B^{-1}$, then the forecast error in terms of the structural shocks is:

$$X_{t+h} - X_{t+h|t} = \Theta_0 \epsilon_{t+h} + \Theta_1 \epsilon_{t+h-1} + \ldots + \Theta_{h-1} \epsilon_{t+1}. \tag{4.22}$$

Suppose the VAR model has a bivariate system (Equation 4.18), the forecast errors of the bivariate VAR model is:

$$\begin{bmatrix} X_{1,t+h} - X_{1,t+h|h} \\ X_{2,t+h} - X_{2,t+h|h} \end{bmatrix} = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t+h} \\ \epsilon_{2,t+h} \end{bmatrix} + \cdots + \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{bmatrix}. \tag{4.23}$$

Suppose the first equation of the bivariate VAR model is:

$$X_{1,t+h} - X_{1,t+h|h} = \theta_{1,1}\epsilon_{1,t+h} + \ldots + \theta_{1,1}\epsilon_{1,t+1} + \theta_{1,2}\epsilon_{2,t+h} + \ldots + \theta_{1,2}\epsilon_{2,t+1}, \tag{4.24}$$

the variance of forecast error may be decomposed as:

$$\begin{aligned} VAR(X_{1,t+h} - X_{1,t+h|h}) &= \sigma_1^2(h) \\ &= \sigma_1^2((\theta_{1,1}^{(0)})^2 + \ldots + (\theta_{1,1}^{(h-1)})^2) + \sigma_2^2((\theta_{1,2}^{(0)})^2 + \ldots + (\theta_{1,2}^{(h-1)})^2). \end{aligned} \tag{4.25}$$

Since the FEVD determines the proportion of variances dues to structural shocks, the FEVD for $X_{1,t+h}$ is (Zivot and Wang, 2003):

$$\rho_{1,1}(h) = \frac{\sigma_1^2((\theta_{1,1}^{(0)})^2 + \ldots + (\theta_{1,1}^{(h-1)})^2)}{\sigma_1^2(h)} \tag{4.26}$$

$$\rho_{1,2}(h) = \frac{\sigma_2^2((\theta_{1,2}^{(0)})^2 + \ldots + (\theta_{1,2}^{(h-1)})^2)}{\sigma_2^2(h)} \tag{4.27}$$

Chapter 5

Data Mining

5.1 Data Collection

5.1.1 Data Sources

The climate data is from the National Centers for Environmental Information's (NCEI's) Climate Data Online (CDO), which is under a large scientific and regulatory agency, the National Oceanographic and Atmospheric Administration (NOAA). CDO provides summaries of the historical daily land surface observations around the world.

The economic data is from the U.S. Bureau of Labor Statistics, which compiles the information from the CPI for all urban consumers survey. The CPI for all urban consumers covers approximately 93 percent of the total population, which doesn't include the remaining total population that lives in remote rural areas. CPI considers the prices of all goods and services purchased for consumption by urban households.

The storm events data is from the NCEI's Storm Events Database, which contains documents on the occurrence of storms, unusual weather phenomena, and other significant meteorological events. It also collects the causes of these extreme weather events, such as fatality, injuries, and cost of property damage.

5.1.2 Study Areas

FIGURE 5.1: A map of the United States with five airport sites

This thesis focuses on five populated airport sites in different regions of the United States (Figure 5.1). The United States Census Bureau has four regions: West, Midwest, Northeast, and South (Kirakosyan, 2016).In order to represent the regions, at least one site is selected from each region. The principal characteristics of these sites are summarized in Table 5.1.

Site	Houston	Chicago	Boston	San Francisco	Miami
Airport Code	HOU	ORD	BOS	SFO	MIA
Latitude	29° 38′ 26"	41°58′49"	42°21′56"	37° 37′ 17"	25°47′45"
Longitude	-95° 16′ 26"	87°54′32"	71°0′34"	122°22′44"	80°17′13"
Elevation (feet)	38.43	652.89	10.77	5.73	16.4
County	Harris	Cook	Suffolk	San Mateo	Miami-Dade
Köppen-Geiger Classification	Cfa	Dfa	Dfa	Csb	Am

TABLE 5.1: Characteristics of airport sites

Houston's site is classified as a humid subtropical climate or characterized by hot, long summers with evenly distributed precipitation through the years. Flooding and hurricanes typically occur in Houston during the summer and fall. Chicago's climate is hot summer continental, characterized by hot and humid summers and severe winters. Boston's climate might have classification as Chicago, but Boston experiences hurricane season since the city is on the shore. San Francisco is classified as a dry-summer subtropical climate, known as the Mediterranean climate, which means this city has dry and hot summers with mild winters. Miami has short dry winters and hot and humid summers; therefore, Miami's climate is classified as a tropical monsoon climate. Miami experiences frequent extreme weather events of flooding and hurricanes.

5.1.3 Data Descriptions

The climate dataset comprises 26,538 daily observations with 40 distinct variables, which is an extremely large size to compute efficiently in terms of time. Therefore, the climate dataset is reduced to four different time series measurements in the following: maximum temperatures, minimum temperatures, rainfall, and snow depth. The temperatures are measured in Fahrenheit, while rainfall and snow depth are measured in inches. The climate dataset is recorded at the five airport sites in the United States, and the dataset is collected from January 1st, 1950, to December 31st, 2022.

The economic dataset consists of 258 bi-monthly observations of a consumer price index (CPI) and food-only CPI from January 1980 to December 2022. U.S. Bureau of Labor Statistics combines the three nearest cities within a region, namely Houston-The Woodlands-Sugar Land, Chicago-Naperville-Elgin, Boston-Cambridge-Newton San Francisco-Oakland-Hayward, and Miami-Fort Lauderdale-West Palm Beach, to calculate the CPI. The CPI is a measure of the average change in prices of a basket of consumer goods and services ("Consumer Price Indexes Overview" 2023). The CPI is widely used as a measure of inflation and is closely monitored by businesses,

policymakers, and finance markets. The U.S. Bureau of Labor Statistics recorded different start dates of bi-monthly observations in each city.

- **Houston-The Woodlands-Sugar Land** has bi-monthly observations that skip odd numbers recorded from February 1980 to December 2022.
- **Chicago-Naperville-Elgin** has monthly observations, but since most records are bi-monthly, the observations are summarized into bi-monthly, which starts from February 1980 to December 2022.
- **Boston-Cambridge-Newton** has bi-monthly observations that skip even numbers, where it starts from January 1980 to November 2022.
- **San Francisco-Oakland-Hayward** has bi-monthly observations that skip odd numbers recorded from 1980 to 1987 and 1997 to 2022. From 1987 to 1997, the data contains monthly observations. Therefore, these monthly observations are summarized bi-monthly.
- **Miami-Fort Lauderdale-West Palm Beach** has bi-monthly observations that skip even numbers from January 1980 to November 1996 and then switches to bi-monthly observations that skip odd numbers from February 1997 to December 2022.

The storm event dataset comprises various types of storm events, including tornadoes, hailstorms, and floods. However, tornado events were the only ones recorded from 1950 through 1954. From 1955 through 1995, tornado, thunderstorm wind, and hail events were recorded from paper publications, and since 1996, 48 different types of events have been recorded. The data includes information on the number of deaths and injuries for each event occurring between January 1st, 1950, and December 31st, 2022. The storm events dataset covers only counties within the airport sites in the United States. The size of the storm events data varies depending on the number of storm occurrences in each city between 1950 and 2022.

5.2 Data Preprocessing

After collecting the datasets, the next step of the data mining process is to enhance the quality of the datasets and prepare them for algorithms to extract the information and insights. This step involves several strategies to clean and integrate the datasets.

As all climate measurements exhibit a left-skewed distribution in histograms, missing values are replaced with the median value of the variable itself to mitigate the impact of outliers. Dates in all datasets are converted into time series arrays using Julia programming. Since the Bureau of Labor Statistics only collects the CPI values, the change in index values and the inflation rate are computed and added as two new variables to the economic datasets using Excel. The inflation rate, also known as percent change, is calculated using the following equation:

$$\text{Inflation Rate} = \frac{\text{Current CPI}-\text{Prior CPI}}{\text{Prior CPI}} * 100$$

Climate, economic, and storm event datasets have different data resolutions, and each dataset has a unique time series collection. For the VAR model, the climate dataset's daily time resolution is modified to match the time resolution of either

the economic or storm events dataset. Since the economic dataset represents a bi-monthly average, the climate dataset averages the values every two months and then merges them with economic variables.

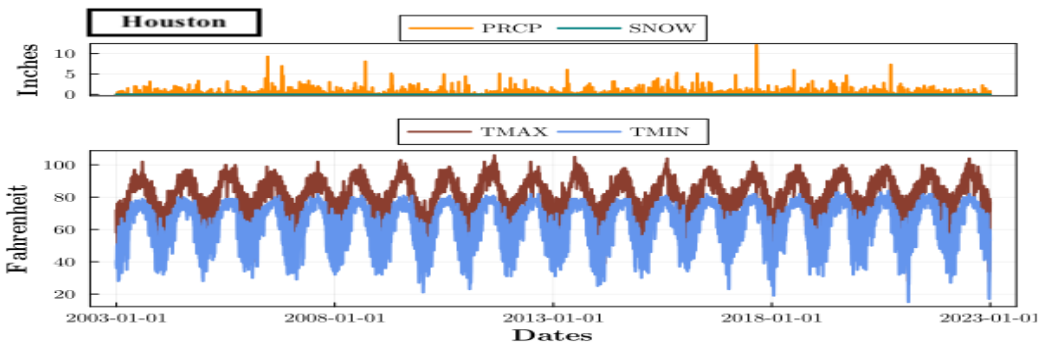
The storm event dataset only consists of the dates of severe weather events occurrences. Therefore, the dates of storm events expand into daily data resolution, where the number of deaths and injuries equals zero if the dates have no occurrence of storm events. However, the storm events database has some dates with several storm events occurring on the same day. To address this, all duplicated dates combine into only one date, which sums up the number of deaths or injuries within the duplicated dates.

The climate dataset is extremely large since it contains almost 75 years of data points in each variable. Therefore, I create a mini-batch of the dataset where I reduce the data size to only last twenty years or 7,305 days. The climate dataset splits into a training set and a testing set in a ratio of 80% to 20% to compare the time series forecasting methods. Since the dataset splits into 80% to 20% of twenty years of observations, the training set has approximately 16 years of time series observations, while the testing set has the most recent four years, or from 2019 to 2022.

5.3 Data Visualization

This thesis focuses on four climate factors, namely maximum temperature, minimum temperature, precipitation, and snowfalls, in five cities in different regions of the United States. The multivariate data analysis in this study considers two variables of the economy, namely the CPI inflation rate and Food only CPI inflation rate, and two variables of human health during the storm events, namely the fatalities and injuries. As different cities have varying degrees of severity in climate factors, the effects of each storm event on the economy and human health may vary.

Figure 5.3 displays the plots of the four climate factors for the five airport sites. While the periodicity of maximum and minimum temperatures may be consistent across all five cities, the maximum or minimum values may differ. For example, Houston is more likely to experience maximum temperatures over 100°F, whereas Miami rarely has maximum temperatures over 100°F. Houston, Miami, and San Francisco have a higher frequency of rainfall over 20 years, while Boston and Chicago have a higher frequency of snowfall. This variation in climate factors may be attributed to the locations of these cities. Cities in the south and west of the United States tend to have warm temperatures with high precipitations, while cities in the north and east are more likely to have cold temperatures with high snowfalls.



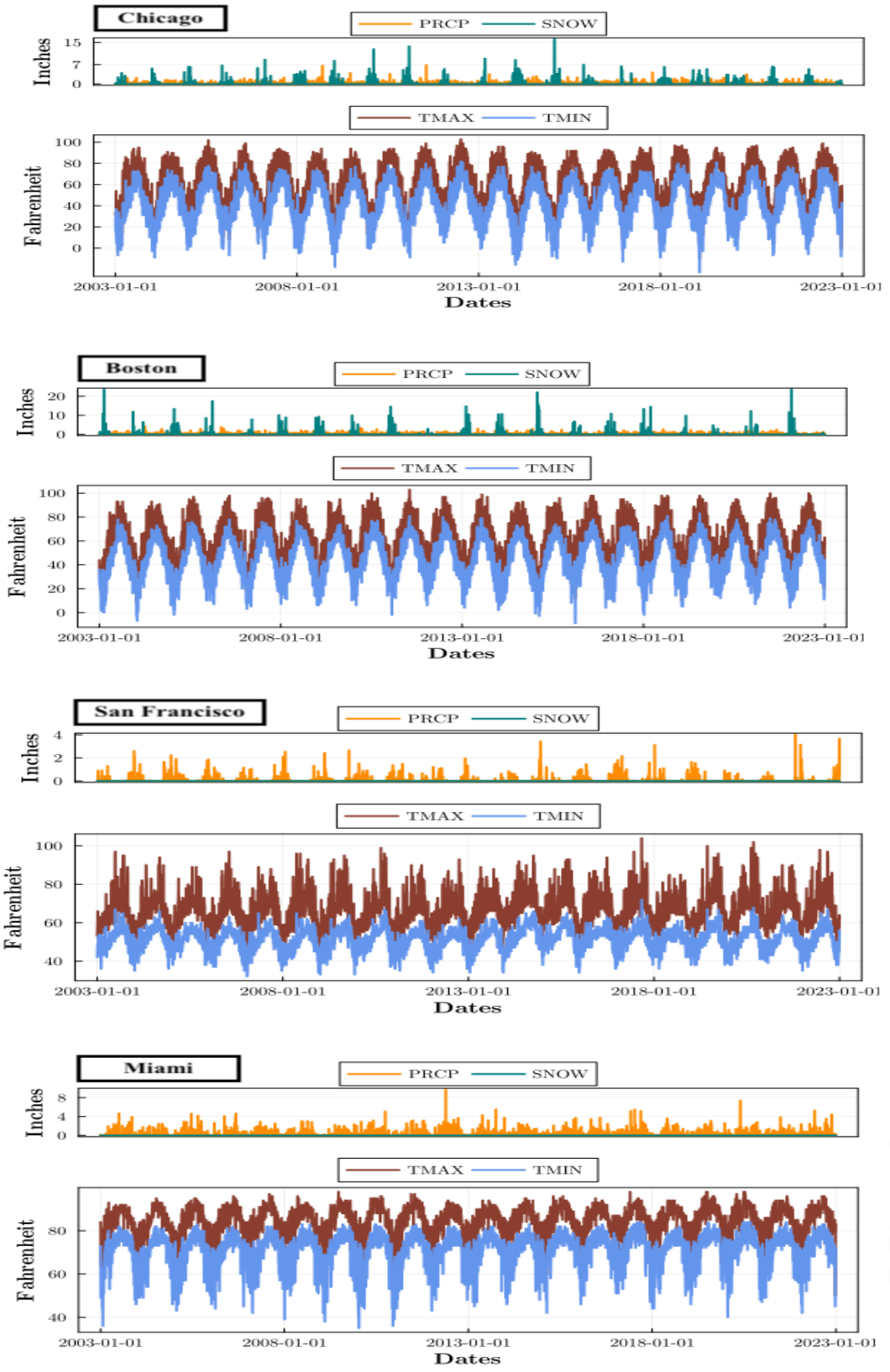


FIGURE 5.3: Plots of climate factors in five different cities from first day of the year 2003 to the last day of the year 2022.

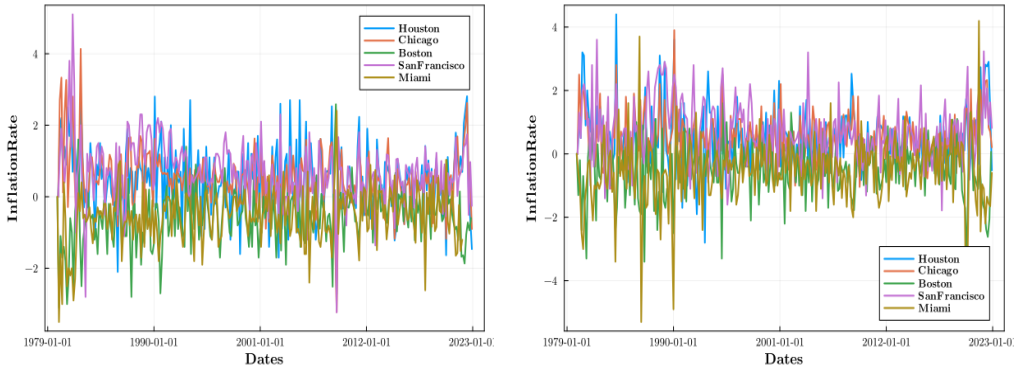


FIGURE 5.4: Plots of the inflation rates of CPI (left) and Food only CPI (right) in five different cities from 1980 to 2022.

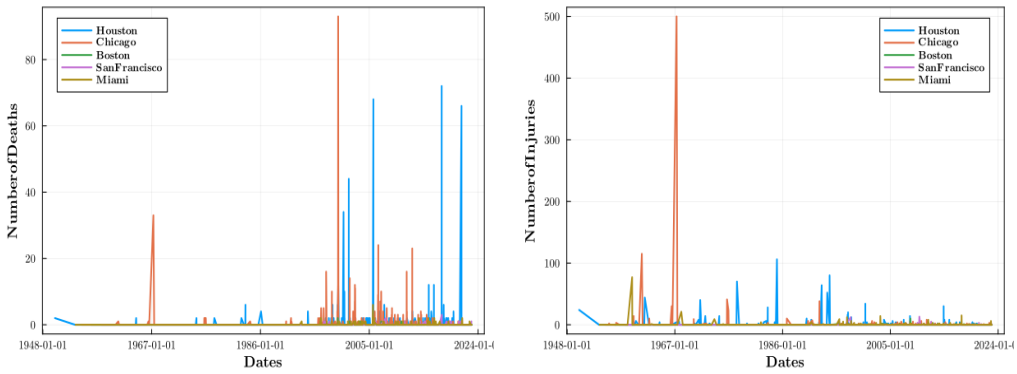


FIGURE 5.5: Plots of the number of fatalities (left) and injuries (right) in five different cities from 1980 to 2022.

The economy plot in Figure 5.4 displays the inflation rates of the CPI and Food only CPI among five cities. Comparing all five cities, San Francisco and Chicago exhibit high inflation rates in either economic factor, while Miami and Boston have low inflation rates. The fluctuations of inflation rates in either factor among the cities appear consistent over the years. On the other hand, the health plot in Figure 5.5 shows the number of deaths and injuries in five cities. Based on the figure, Chicago and Houston seem to have more occurrences of a high number of deaths compared to the other three cities. However, Houston has more recent hits with a high number of deaths. Additionally, Houston has the most recent hits with a high number of injuries, while Chicago experienced a high number of injuries in the 1960s.

Chapter 6

Applications

This chapter presents the findings of this work. The first part of the research focuses on applying four different time series forecasting methods to the climate dataset of five cities and evaluating their forecasting performance using statistical measures. The second part aims to analyze the dynamic relationship between climate aspects and economic factors or human health.

Section 6.1 analysis is carried out in two statistical software tools: Julia and Python. The packages in Julia programming used to develop ARIMA, Exponential Smoothing, and GP models are "Forecast", "StateSpaceModels", and "GaussianProcesses" packages. However, MLP model development was challenging with Julia, so the PyTorch package of Python is used to provide more flexibility in developing the framework for MLP of time series forecasting. Section 6.2 analysis is carried out in the R software environment using the "vars" package.

6.1 Univariate Analysis

The analysis of the time series observations for the five cities starts with Houston as the main city. Before developing any time series forecasting model, the time series needs to be analyzed using the seasonal-trend decomposition (STL). Figure 6.1 shows the components of the time series observations, where the STL divides the time series observations into data (top), trend (first middle), seasonal (second middle), and the remainder (last). This figure clearly indicates that the time series has seasonal components, and the periodicity occurs every year from 2003 to 2022. However, there is no trend in either the maximum or minimum temperatures. Therefore, the time series forecasting models need to consider the seasonal component within the models. The behaviors of seasonal and trend components in Houston's daily observations are similar to those of the other four cities.

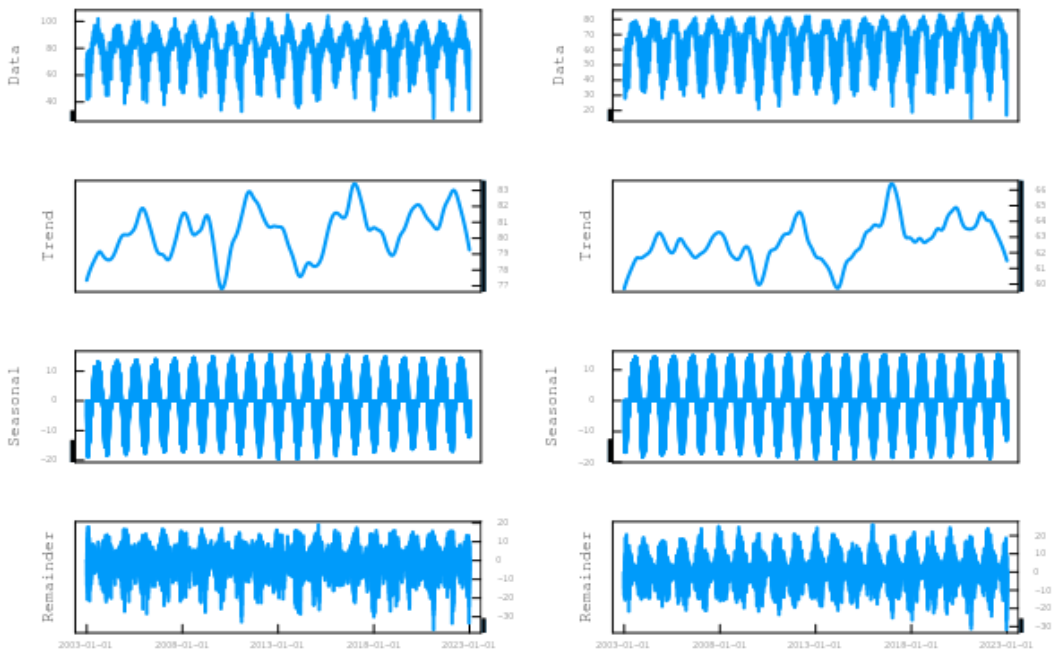


FIGURE 6.1: Plot of time series decomposition of Houston's maximum (right) and minimum (left) temperatures from 2003 to 2022.

6.1.1 ARIMA

Prior to developing the ARIMA model, it is critical to ensure that the temperature variables are stationary to avoid any fabrication of the results, which is a common issue in time series analysis. The ADF test is utilized to determine the stationarity of the time series, as presented in Table 6.1. It is observed that all variables are stationary in their level form, implying that there is no need to add a differencing order into the time series data.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	ADF Values	-10.62	-8.76	-9.22	-15.23	-11.80
	P-Values	<1e-18	<1e-13	<1e-14	<1e-27	<1e-21
TMIN	ADF Values	-10.36	-8.36	-7.37	-11.48	-13.22
	P-Values	<1e-17	<1e-12	<1e-10	<1e-20	<1e-23

TABLE 6.1: Augmented Dickey-Fuller (ADF) Test for Stationarity for Univariate Time Series.

Time series observations are correlated through time; it's important to assess the nature of correlation through ACF and PACF plots. Figure 6.2 shows the lag order plot with a maximum of 25 lags and the strength of correlation between data points. The ACF plots indicate a slow decline with lags, while PACF plots have few significant lags. Therefore, Houston's temperatures are well-suited for a pure auto-regressive (AR) model. Furthermore, since the time series data is stationary, the plots do not exhibit a significant seasonal component within the data. The ACF and PACF plots of the other four cities exhibit similar behaviors to Houston's maximum and minimum temperatures, leading this thesis to focus on the pure AR model.

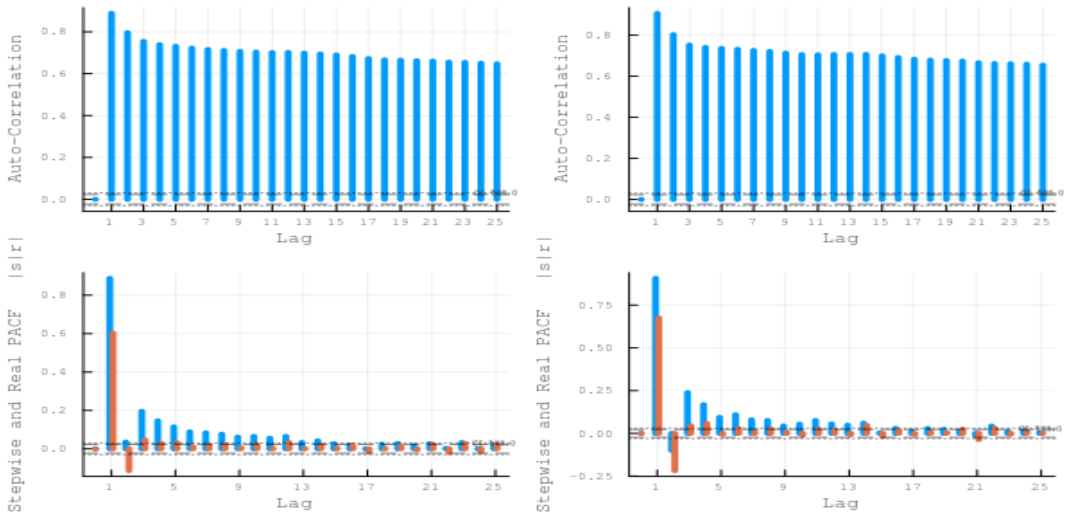


FIGURE 6.2: ACF (top) and PACF (bottom) of Houston's maximum (right) and minimum (left) temperatures from 2003 to 2022.

Once the time series observations are split into a training and testing set of 80% to 20%, as mentioned in Section 5.2, an optimization method is developed to determine the optimal lag length for the AR model of the time series observations by measuring with AIC (Equation 3.12) and BIC (Equation 3.13). Figure 6.3 illustrates the AIC and BIC scores where the optimization method runs from lag 1 to lag 110 of the pure AR model in two different variables of Houston temperatures. The lower the scores of AIC or BIC, the better the model fits the time series observations.

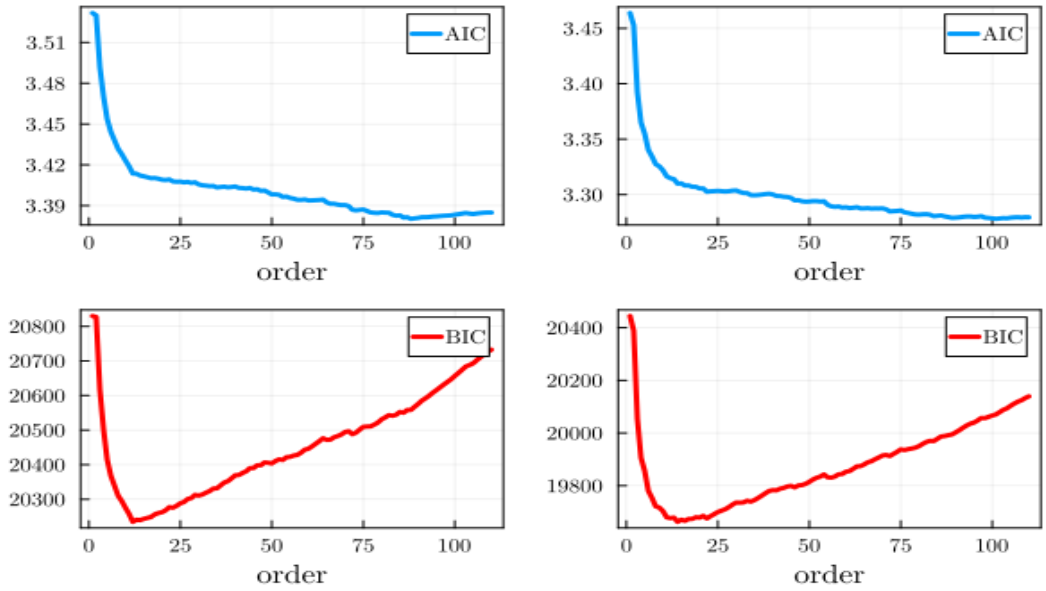


FIGURE 6.3: Plot of AIC (top) and BIC (bottom) scores with lag order of pure AR model for Houston's maximum (right) and minimum (left) temperatures from 2003 to 2022.

Based on Figure 6.3 and Table 6.2, it is evident that the optimal lag order of the pure AR model ranges from lag 80 to 110 based on the AIC score, for all five cities, except for San Francisco's maximum temperatures, which have an optimal lag order of 19.

Although it may seem unusual for San Francisco to have a lower lag order than other cities, this is likely due to the unique shape of the maximum temperature data for San Francisco, as shown in Figure 5.3. This study uses the AIC score as the primary information criterion for selecting the lag order of the pure AR model.

Variables	City	p of AIC	AIC score	p of BIC	BIC score
TMAX	Houston	88	3.38	12	20234.66
	Chicago	109	4.04	13	24019.23
	Boston	108	4.06	15	24151.47
	San Francisco	19	3.08	17	18400.89
	Miami	100	2.30	18	14255.13
TMIN	Houston	101	3.28	14	19662.43
	Chicago	106	3.66	12	21777.16
	Boston	108	3.23	10	19492.65
	San Francisco	97	1.95	11	12293.25
	Miami	106	2.55	12	15539.55

TABLE 6.2: Model Order Selection for pure AR model of Maximum and Minimum Temperatures among five cities.

The optimal AIC score for Houston's maximum temperature is at lag 88, while for Houston's minimum temperature, it is at lag 101. The AR model with the optimal lag is used to forecast the maximum temperatures for the next 1,461 days, which is the length of the testing dataset. Figure 6.4 illustrates the forecasting of the testing set using the AR model with its optimal lag order. In the figure, the red line represents the actual values of the testing set, while the blue line represents the forecast values of the AR model. As discussed in Section 3.6, to evaluate the model's performance, two different statistical methods, RMSE and MAE, are used to measure the difference between actual values and forecast values of the AR model to determine the model's performance.

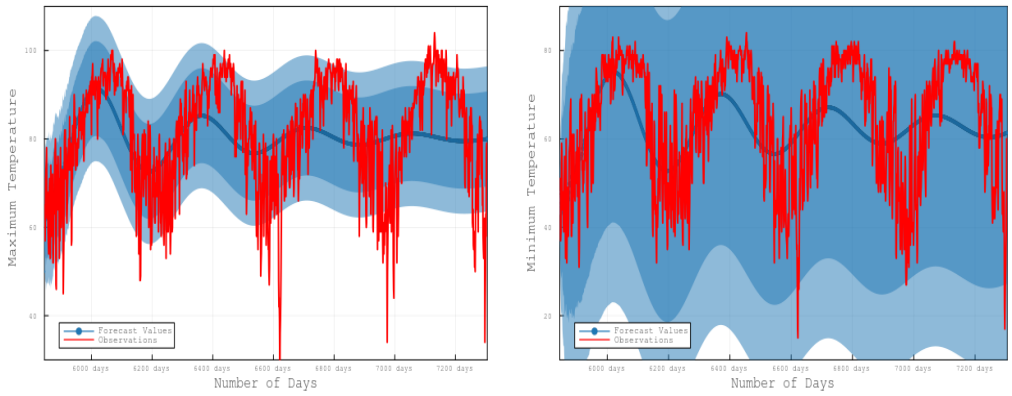


FIGURE 6.4: Plots of forecasting the maximum (left) and minimum (right) temperatures of Houston using AR model with its optimal lag order of 88 and 101.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	RMSE	11.24	16.75	13.74	8.62	5.45
	MAE	8.90	14.10	11.33	6.69	4.29
TMIN	RMSE	11.60	16.75	11.59	5.50	6.82
	MAE	9.65	12.90	9.43	4.53	5.48

TABLE 6.3: Model Performance Measures for AR Model's Forecasting on Maximum and Minimum Temperatures among five cities.

Table 6.3 presents the evaluation results of the pure AR model's performance with the optimal lag order using two different statistical approaches. The model's performance across all cities is comparable, with lower RMSE or MAE scores indicating better performance. Notably, the AR model demonstrated good forecasting performance for Miami's maximum and San Francisco's minimum temperatures.

6.1.2 Exponential Smoothing

The time series decomposition confirmed the presence of seasonal components in the maximum and minimum temperatures. Therefore, this thesis focuses on only one exponential smoothing model: Triple Exponential Smoothing (TES) with no trend. Since the training set covers 16 years, it implies that there is a period in every single year. Figure 6.5 illustrates the forecasting of the time series observations of the testing set, utilizing the TES model with no trend and a periodicity of 16 seasons. Like Figure 6.4, the red line represents the actual observations in the testing set. However, it might be difficult to see that the blue line within the blue area indicates the forecast values based on the TES model, which is not closely aligned with the actual observation data.

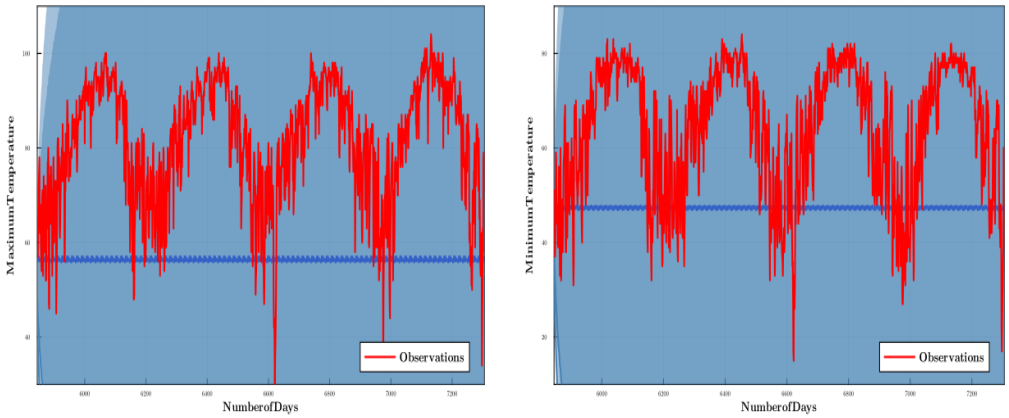


FIGURE 6.5: Plots of forecasting the maximum (left) and minimum (right) temperatures of Houston using TES the model.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	RMSE	28.14	30.28	25.54	14.62	7.11
	MAE	25.63	25.09	20.87	12.07	5.92
TMIN	RMSE	21.50	23.67	22.52	9.41	7.32
	MAE	18.65	19.29	18.48	8.00	5.56

TABLE 6.4: Model Performance Measures for TES Model's Forecasting on Maximum and Minimum Temperatures among five cities.

While the TES model may not perform as well as the AR model in some cases, it has shown comparable results to the AR model for San Francisco and Miami. Both models performed well in forecasting the maximum and minimum temperatures for these two cities, as indicated by Table 6.4. However, the confidence intervals for the TES model, shown in Figure 6.5, are considerably larger than those of the AR model, to the extent that they cover much of the plot.

6.1.3 MLP

In this thesis, the structure of the MLP model consists of three layers as follows: input layer, hidden layer, and output layer. The input layer includes three input neurons representing three past observations in a sequence, and the output neuron is the forecast observation. The hidden layer is established with four hidden neurons and with an activation function of a rectified linear unit (ReLU). The ReLU activation function sets any negative values to 0 (Equation 6.1). This function is easy and fast to compute and provides simple nonlinear transformation.

$$\sigma(u) = max(0, u) \tag{6.1}$$

Prior to training the model, the training set splits into a sequence of three observations, representing three input neurons for the model. The training and testing datasets are divided into a batch of 12 samples to update the optimal parameters within the model. The model is trained for 2,000 epochs to update the optimal parameters and determine the most minimized losses. Figure 6.6 presents the decrease in losses through the numbers of epochs, and the lowest loss is recorded in epoch 279 for Houston's maximum temperatures and epoch 1,441 for Houston's minimum temperatures

Figure 6.7 illustrates the MLP model's forecasting time series observations of the testing set. In the figure, the actual values (red line) are closely aligned to the forecast values of the MLP model (blue line). According to Table 6.5, the MLP model's performance is better than either AR or TES models. Like the other two models, San Francisco and Miami have better model performance with their time series observations compared to the other three cities. Additionally, the minimum temperatures show better model performance than the maximum temperatures.

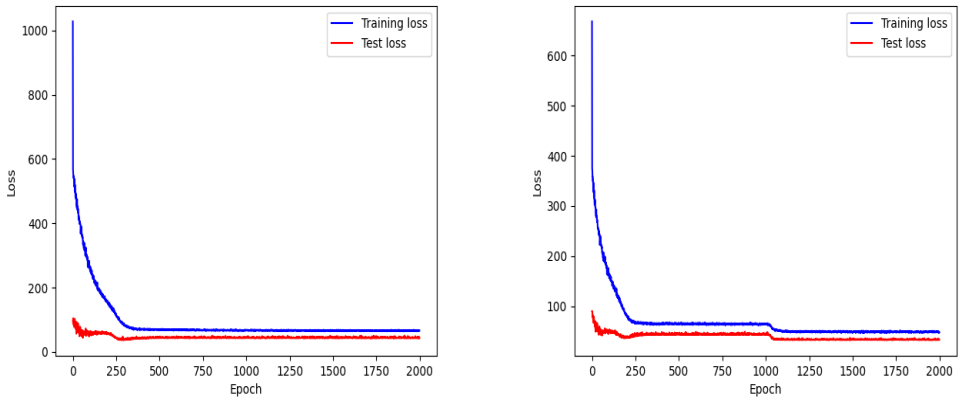


FIGURE 6.6: Line plots of losses in training and testing set over the number of epochs for Houston's maximum (left) and minimum (right) temperatures.

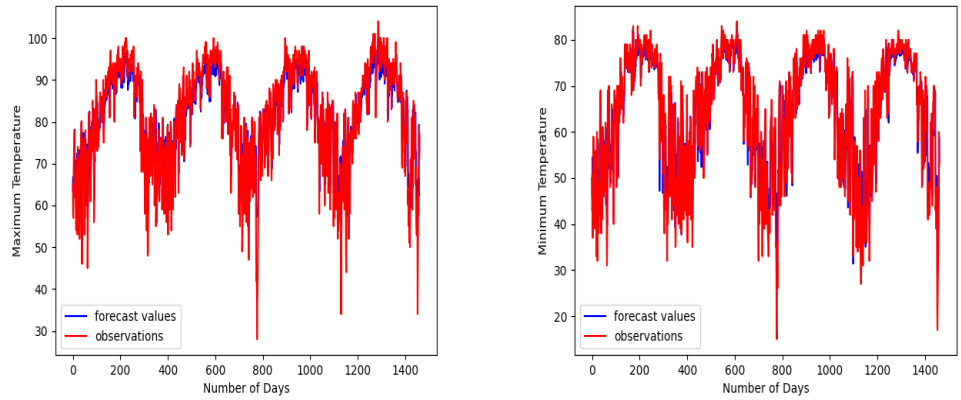


FIGURE 6.7: Plots of forecasting the maximum (left) and minimum (right) temperatures of Houston using MLP.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	RMSE	6.09	8.24	8.25	4.99	3.54
	MAE	4.42	6.33	6.53	3.81	2.29
TMIN	RMSE	5.71	6.56	5.42	2.81	4.00
	MAE	4.02	5.01	4.07	2.13	2.80

TABLE 6.5: Model Performance Measures for MLP Model's Forecasting on Maximum and Minimum Temperatures among five cities.

6.1.4 Gaussian Processes

Section 3.5 suggests that the kernel composition of

$$k(X_i, X_j) = k_{Linear}(X_i, X_j) + k_{Periodic}(X_i, X_j) + k_{RBF}(X_i, X_j) \tag{6.2}$$

is the best for weather forecasting. However, in this thesis, the kernel composition is slightly modified because the time series decomposition reveals that there are no trend components in the maximum or minimum temperatures. Therefore, the linear kernel is replaced with a white noise kernel since there are some noises in the time series decomposition. The kernel composition for this thesis includes:

- a noise kernel with a standard deviation of 1.0
- a periodic kernel with a length scale of 0.0, a standard deviation of 1.0, and a periodicity of 1.0
- and an RBF kernel with a length scale of 4.0 and standard deviation of 0.0.

The parameter values for each kernel are selected based on several trials and errors using RMSE as the criterion, as shown in Table 6.6. With this fixed kernel composition, the GP model runs with a zero mean function since it is the standard mean function for time series observation data, indicating that the mean is always at zero.

Kernel Composition	RMSE Values
N(1) + P(0,1,1) + RBF(4,0)	12.51
N(1) + P(0,1,1)	12.81
P(0,1,1) + RBF(4,0)	12.81
N(1) + P(0,1,2) + RBF(4,0)	Failed due to infinite sin values
N(1) + P(0,1,0) + RBF(4,0)	Failed due to infinite sin values
N(1) + P(0,1,1) + RBF(2,0)	12.80
N(1) + P(0,1,1) + RBF(5,0)	12.83
N(1) + P(0,1,1) + RBF(4,4)	76
N(1) + P(0,1,0) + RBF(4,4)	76
N(1) + P(0,1,1) + RBF(3,0)	12.86
N(1) + P(1,1,1) + RBF(4,0)	Failed due to infinite sin values

TABLE 6.6: Trials and Errors with Values for Kernel in Gaussian Process, using Houston's Maximum Temperatures

Figure 6.8 illustrates the GP model's forecasting of the time series observations of the testing set. The model did not perform as well as the MLP model in Figure 6.7, as the actual values (red line) are not closely aligned with the forecast values (blue line). The model performance of GP is worse than that of the MLP model, as confirmed by 6.7. Additionally, the fixed kernel composition of GP has failed for three cities' maximum and minimum temperatures due to infinite values in the sine function. GP models are known for their instability and ill-conditioning, which means that the fixed kernel composition might be ineffective for these cities due to errors in inverse matrices. Based on the model performance of only two cities, GP appears to have better performance with the data of maximum temperatures than minimum temperatures, which is contrary to the results of the other three models.

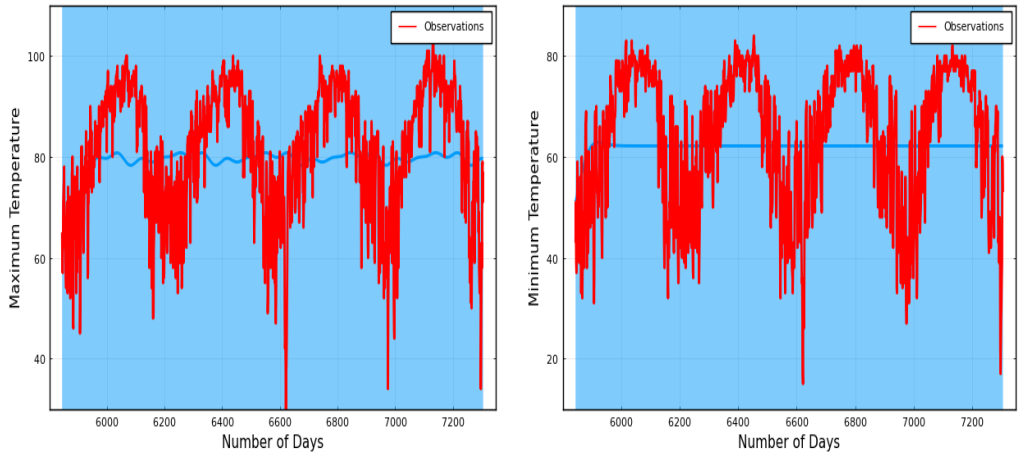


FIGURE 6.8: Plots of forecasting the maximum (left) and minimum (right) temperatures of Houston using GP.

Variables	Measures	Houston	Chicago	Boston	San Francisco	Miami
TMAX	RMSE	12.51	–	17.94	–	–
	MAE	10.36	–	15.24	–	–
TMIN	RMSE	13.77	–	16.36	–	–
	MAE	11.96	–	13.91	–	–

TABLE 6.7: Model Performance Measures for Gaussian Processes Forecasting on Maximum and Minimum Temperatures among five cities.

6.1.5 Temperature Projection

According to the figures and the model performance evaluation, the MLP model outperformed the other three statistical models. Figure 6.9 illustrates the maximum and minimum temperature projections for Chicago and Boston from 2023 to 2030, using the MLP model. The linear trend in Chicago indicates a 2.81°F (1.56°C) increase in maximum temperatures and a 4.30°F (2.39°C) increase in minimum temperatures within 28 years. In other hands, Boston's trend line indicates a 3.22°F (1.79°C) increase in maximum temperatures and a 3.56°F (1.98°C) increase in minimum temperatures within 28 years. The other three cities have smaller increases in either maximum or minimum temperatures compared to Chicago and Boston.

Although the MLP model performed well, Figure 6.9 indicates that temperature projections flatten over time, indicating that the model cannot detect extreme temperatures or peaks. Since the maximum and minimum temperatures are increasing over time, it is more likely that severe storms will occur in the coming years. For instance, the state of Illinois experienced 82 extreme weather events between 2003 and 2022, causing overall damages/costs that reached or exceeded $1 billion (Smith, 2023). Similarly, Massachusetts experienced 20 such events in the same period. Therefore, the number of extreme weather events is likely to rise in the future as the maximum and minimum temperatures continue to increase.

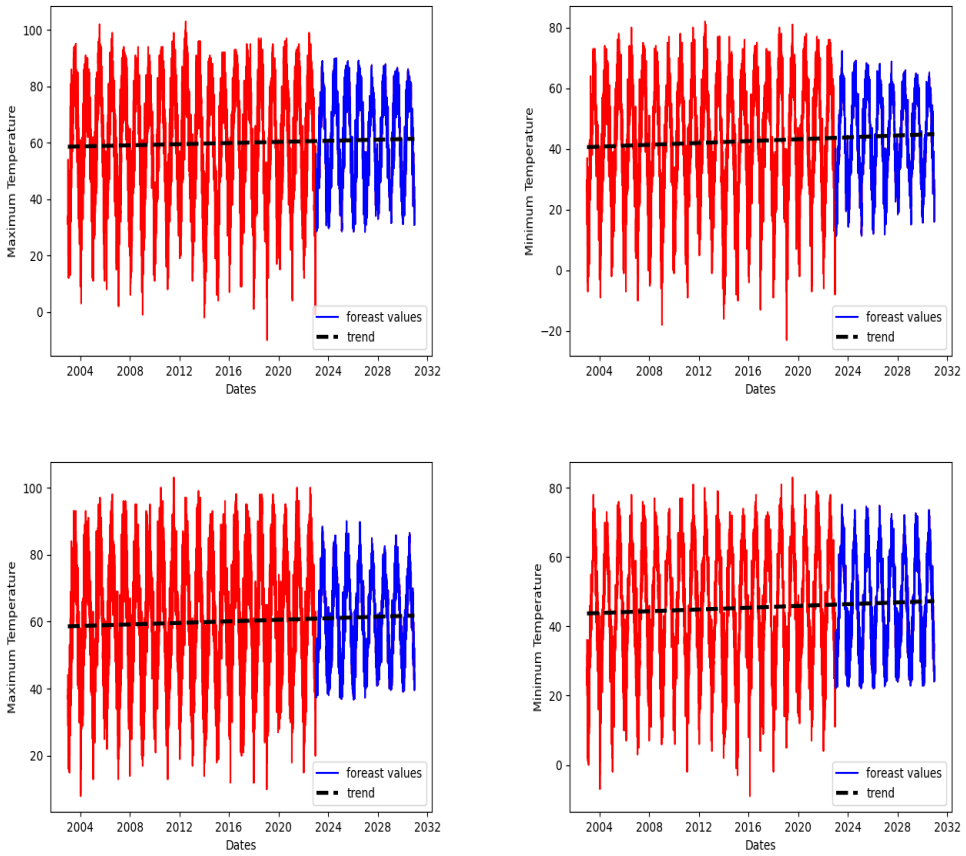


FIGURE 6.9: Projection plots of the maximum (left) and minimum (right) temperatures of Chicago (top) and Boston (bottom) using MLP from 2003 to 2030.

6.2 Multivariate Analysis

6.2.1 Stationarity Tests

The stationarity of time series is essential for studying the relationship between climate and economy or human health. The ADF test is performed to ensure the stationary of the time series. Since the stationarity test already ran with maximum and minimum temperatures, Table 6.8 presents the results of the ADF tests with remaining climate variables and the indicators of the economy and human health. It was found that the snowfall data of San Francisco and Miami had errors when running the ADF tests due to linear algebra. Nonetheless, the snowfall data is not included in the VAR analysis since snowfall rarely occurs in either city.

Variables	City	Dickey-Fuller Values	P-Values	Dickey-Fuller Values	P-Values
		PRCP	PRCP	SNOW	SNOW
Climate	Houston	-47.6963	<1e-99	-51.0996	<1e-99
	Chicago	-48.8002	<1e-99	-41.1644	<1e-99
	Boston	-49.9515	<1e-99	-41.2793	<1e-99
	San Francisco	-38.4492	<1e-99	–	–
	Miami	-5.92199	<1e-06	–	–
		CPI	CPI	Food-only CPI	Food-only CPI
Economy	Houston	-4.61125	0.0001	-4.97277	0.0001
	Chicago	-5.59242	<1e-05	-4.95713	<1e-04
	Boston	-4.88938	<1e-04	-5.31986	<1e-05
	San Francisco	-3.64425	0.0050	-4.02376	0.0013
	Miami	-4.49194	0.0002	-6.37931	<1e-07
		Fatalities	Fatalities	Injuries	Injuries
Human Health	Houston	-65.0115	<1e-99	-66.6538	<1e-99
	Chicago	-65.7422	<1e-99	-66.6735	<1e-99
	Boston	-66.6852	<1e-99	-66.7065	<1e-99
	San Francisco	-65.8107	<1e-99	-66.0953	<1e-99
	Miami	-64.8741	<1e-99	-66.6683	<1e-99

TABLE 6.8: Augmented Dickey-Fuller (ADF) Test for Stationarity of Multivariate Time Series.

6.2.2 VAR Order Selection and Estimation

Several information criteria, namely AIC, HQ, BIC, and FPE, are used to determine the optimal lag length of the VAR model, as presented in Table 6.9. Following the same example of ARIMA models, this thesis uses AIC as the primary information criterion to select the optimal lag length of the VAR model. Table 6.9 only shows the results for the VAR model of climate factors with the CPI. As there are four endogenous variables, besides the climate factors, to analyze each city, the model order selection is performed for each city's each endogenous variable. Therefore, a total of 20 VAR models are computed for all five cities and their economic and human health variables. Table 6.10 provides the optimal lag length of VAR models for each endogenous variable of each city.

p	AIC	HQ	BIC	FPE
1	-18.37	-18.20	-17.94	1.06e-08
2	-20.24	-19.92	-19.46	1.63e-09
3	-20.31	-19.85	-19.17	1.52e-09
4	-20.23	-19.63	-18.74	1.64e-09
5	-20.41	-19.67	-18.57	1.37e-09
6	-20.61	-19.72	-18.41	1.13e-09
7	-20.48	-19.45	-17.93	1.29e-09
8	-20.38	-19.21	-17.48	1.43e-09
9	-20.30	-18.99	-17.04	1.56e-09
10	-20.20	-18.75	-16.59	1.74e-09

TABLE 6.9: Order Selection Statistics for VAR Model of Houston's climate and CPI dataset.

Variables	Houston	Chicago	Boston	San Francisco	Miami
CPI	6	6	6	7	6
Food-only CPI	6	9	6	6	6
Fatalities	13	18	15	17	19
Injuries	13	18	15	19	19

TABLE 6.10: Model Order Selection for VAR Models among five cities

6.2.3 Granger Causality

Using the optimal lag length of VAR models, the Granger Causality test determines the causality between climate variables and the four endogenous variables in each city. This study uses a significance level of 0.5% to test the null hypothesis, which states that the combination of climate factors does not Granger-cause an endogenous variable. The statistical scores presented in Table 6.11 indicate that the null hypothesis is rejected only in a few cases, including Chicago, Boston, and San Francisco, where there is evidence that the combination of climate factors *Granger-cause*CPI. Therefore, there is a causal relationship between climate factors and CPI in these three cities.

Variable	Measures	Houston	Chicago	Boston	San Francisco	Miami
CPI	F-statistics	1.3116	2.5107	2.032	3.401	1.5945
	P-values	0.1441	8.143e-05	0.002388	4.348e-07	0.05485
Food-only CPI	F-statistics	1.0919	1.5269	1.4473	1.6966	1.7507
	P-values	0.3453	0.02509	0.07552	0.03463	0.02693
Fatalities	F-statistics	0.8298	0.96042	0.50976	1.1803	1.0286
	P-values	0.804	0.5733	0.9994	0.1773	0.4155
Injuries	F-statistics	1.3124	0.92835	0.44112	1.0229	1.3848
	P-values	0.06481	0.6497	0.9999	0.4271	0.02887

TABLE 6.11: Results of Granger Causality Test for Climate Variables.

6.2.4 Impulse Response Functions

The impulse response function analyzes the effects of climate factors on an endogenous variable over time. With the optimal lag length of the VAR model, the impulse response traces the impacts of an exogenous shock to an endogenous variable over a 12-bimonthly period. Figure 6.10 presents the impulse responses of CPI to the climate factors for the city of Houston. All five subplots within the figure illustrate the fluctuations of the climate shocks on the CPI. The dynamic relationships of Food-only CPI and climate factors have similar behavior to the CPI, where climate factors have fluctuating positive and negative effects on the CPI and Food-only CPI over time (Appendix A.1). The CPI and Food-only CPI in the other four cities have similar responses to the effects of climate factors.

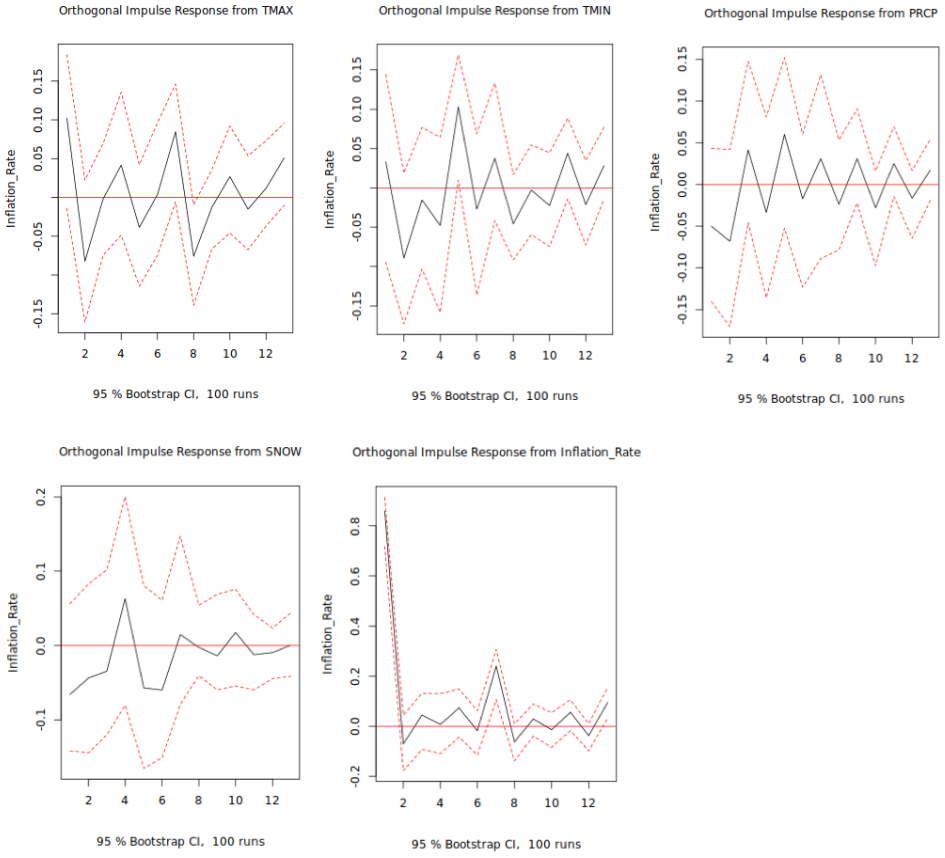


FIGURE 6.10: Impulse Response Function of CPI on Climate Factors of Houston.

The impulse response tracks the effects of a weather shock on fatalities or injuries over a 30-day period, and the responses to maximum and minimum temperatures in 6.11 exhibit similar behavior to that observed in Figure 6.10. However, the effects of precipitation and snowfall depth on fatalities are more complex, with precipitation having a positive effect within the first two to three days, which then decays to zero over time, while snowfall depth has an opposite effect, where it has a negative effect within the first two to three days and then decay to zero. The response of fatalities to its own instantaneous one standard deviation has an immediate positive shock and then decays to zero, just like the effect of precipitation and snow. Injuries show similar dynamic responses to climate factors as fatalities, except for the effect of snow, which exhibits an immediate positive impact on the sixth day, followed by a decay to zero (Appendix A.2).

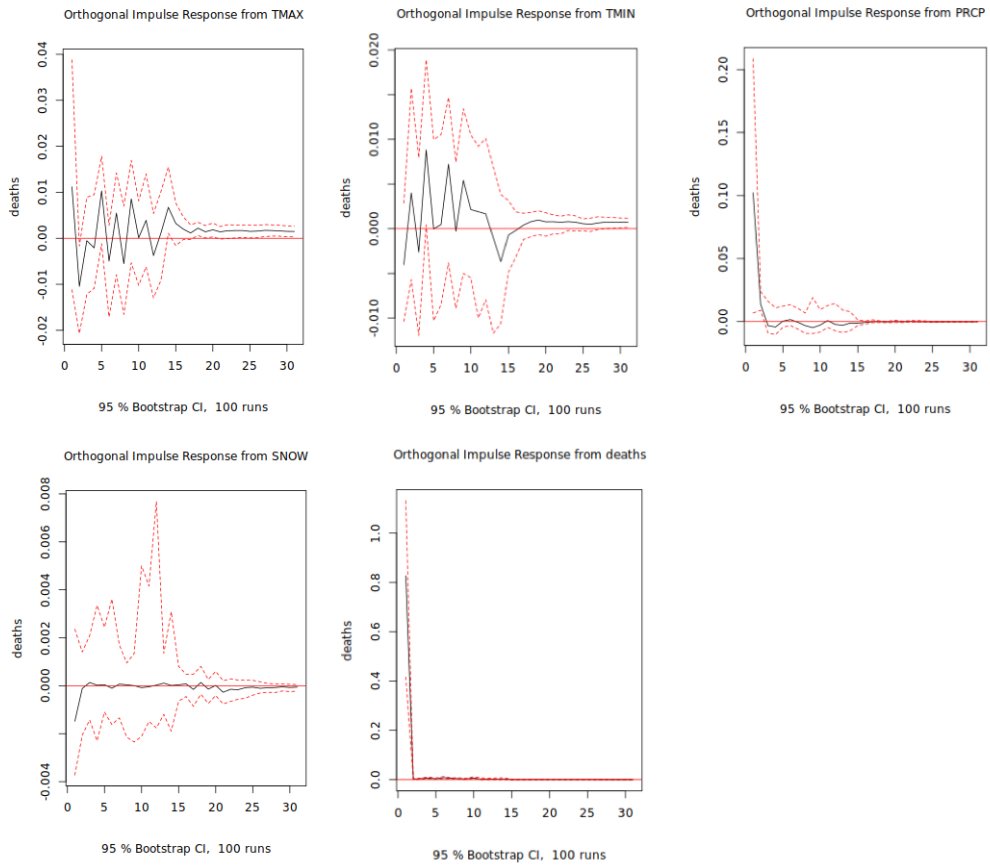
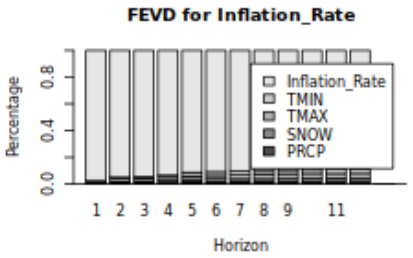


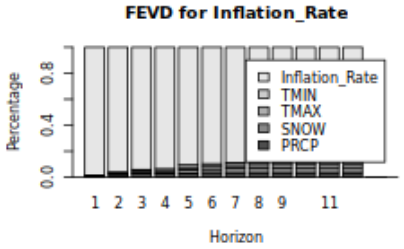
FIGURE 6.11: Impulse Response Function of Fatalities on Climate Factors of Houston.

6.2.5 Forecast Error Variance Decomposition

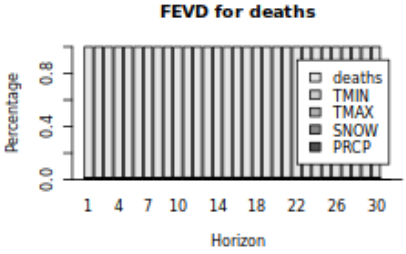
FEVD demonstrates the significance of an exogenous shock in explaining the endogenous variable in the VAR model. With the optimal lag length of the VAR model, FEVD quantifies the amount of contribution of each exogenous shock to the endogenous variable over a specific period. In the case of economic indicators, FEVD calculates their contributions to the endogenous variable over a period of 12 bi-month future values, while for human health indicators, it determines the contributions over a period of 30 days future values. Figure 6.12 illustrates that each endogenous variable is mainly affected by shocks to itself, contributing approximately 90% to 100% of the variation. The remaining percentage is from the exogenous shocks. For economic factors, the contribution of its variable decreases over time as other exogenous shocks have more contribution to the endogenous variable. The contribution of exogenous variables in the other four cities is similar to the pattern presented in Figure 6.12.



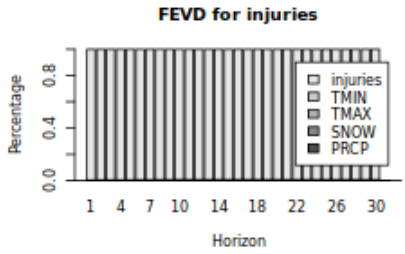
(A) CPI



(B) Food-only CPI



(C) Fatalities



(D) Injuries

FIGURE 6.12: Forecast Error Variance Decomposition of Houston Dataset

Chapter 7

Conclusion

7.1 Summary of Findings

In this thesis, I developed four different statistical models for time series forecasting of temperatures in five cities. Each city has a different range of minimum and maximum temperatures depending on its location, such as coastal areas or mountain regions. The aspect of temperatures within all cities presents a significance of seasonal components with no trend. The foundation of these developed time series methods is based on recent advancements in machine learning and statistical methods, which provide the optimal framework for each model for weather forecasting.

Based on the statistical measures of RMSE and MAE, the forecasting performance of the MLP model has the highest accuracy in predicting the maximum and minimum temperatures of all cities. ARIMA comes as the second-best model, followed by Exponential Smoothing and GP. However, the fixed kernel composition of GP has failed to compute the maximum and minimum temperatures of three cities. If the model performance only measures the results of these two cities, the GP model performed better than the exponential smoothing. Among the three methods, excluding the GP, San Francisco and Miami have the best model performance compared to the other three cities. One of the potential reasons is that these two cities have a smaller amplitude of temperature fluctuation compared to Houston, Chicago, and Boston.

The time series analysis of multivariate datasets in climate, economy, and human health claims that there is a relationship between the CPI and climate factors in several cities. For instance, the Granger Causality test confirms the causality between CPI and climate factors in Chicago, San Francisco, and Miami. The impulse response function presents the fluctuation in the effects of climate factors on the CPI. However, the Granger Causality test affirms that the combination of climate factors fails to Granger-Cause the Food-only CPI, fatalities, and injuries of storm events. The impulse response function also presents the fluctuation in the effects of climate factors on these variables.

7.2 Limitations

This thesis might state that the MLP model is the best model for time series forecasting of maximum and minimum temperatures. However, there are limitations within the development of models that need to be considered in the following:

- This thesis limits to one structure of the MLP model, which is one hidden layer of four hidden neurons with a ReLU activation function. Also, the MLP model

has the optimization method of Adam, where there are many options of optimization methods to select from. There are many possibilities for selecting the number of hidden layers, hidden neurons, activation functions, and optimization methods to improve the MLP model's performance.

- This thesis also has a limit of one fixed kernel composition of GP. There are over 50 potential kernel combinations for forecasting the time series; however, Julia programming doesn't have an optimization method to determine the best kernel composition based on the dataset.
- With the fixed kernel composition, each kernel has a fixed value of standard deviation and length scale, where each dataset has various standard deviation and length scales. As mentioned, Julia programming doesn't contain an optimization method to determine the optimal value of standard deviation, and the length scale of each kernel depends on the dataset.
- Last but not least, the dataset is extremely large, and it can crash the typical computer if these methods run on the entire dataset. Therefore, this research only concentrates on a mini-batch of a 20-year dataset instead of a 75-year. If high-performance computing technology is accessible, these machine learning techniques might give a better model performance.

Also, this thesis's multivariate analysis might state that there is only one correlation between climate factors and CPI. There are limitations within the VAR analysis that needs to be considered in the following:

- The Storm Events Database only contains the dates of storm events occurrence, which means the data resolution of the storm event database is mismatched to the daily resolution of climate factors. In order to match the data resolution of climate factors, the data resolution of storm events becomes daily observations where the dates of no occurrence have a value of zero in fatalities and injuries and the median value of climate factors. Adding zeros to the dataset of storm events creates biases that might affect the VAR analysis.
- The Storm Events Database only focuses on the variables among the selected counties. However, some extreme weather events hit several counties at once. For instance, a tornado usually hits several counties and causes a number of fatalities and injuries in total. In other words, selecting a specific county might not provide the full information about the effect of extreme weather on a number of fatalities and injuries.

7.3 Future Work

Time series analysis contains many more methods that this research has not explored, and the statistical techniques continue advancing. Also, there are areas to which research could be extended. For instance, the empirical comparison study can expand to more machine learning methods for weather time series forecasting, such as LSTM and CNN.

Several directions that this research can explore in the foreseeable future:

- Deepen the understanding of the machine learning techniques of Multilayer Perceptron and Gaussian Processes with climate time series forecasting.
- Develop optimization methods of Gaussian Processes' kernel compositions for time series datasets.
- Perform machine learning approaches, such as Support Vector Machines (SVM) or Decision Trees, to analyze the impact of climate aspects on health conditions, including cardiovascular and respiratory illness.
- Explore the impact of climate change on the factors that play a role in the quality of life, such as environment, psychological status, and socioeconomic interactions.

As part of this thesis project, I have created a GitHub repository where all of the code related to my research is stored. The repository is publicly accessible and intended to serve as a resource for anyone interested in the work I have done. By making my code available to the public, I hope to encourage others to build upon my research and to contribute to the wider community of scholars in this field. You can find the repository at `https://github.com/kamkinast24/Time-Series-Approach`.

Appendix A

Additional Graphs

A.1 Impulse Response Functions

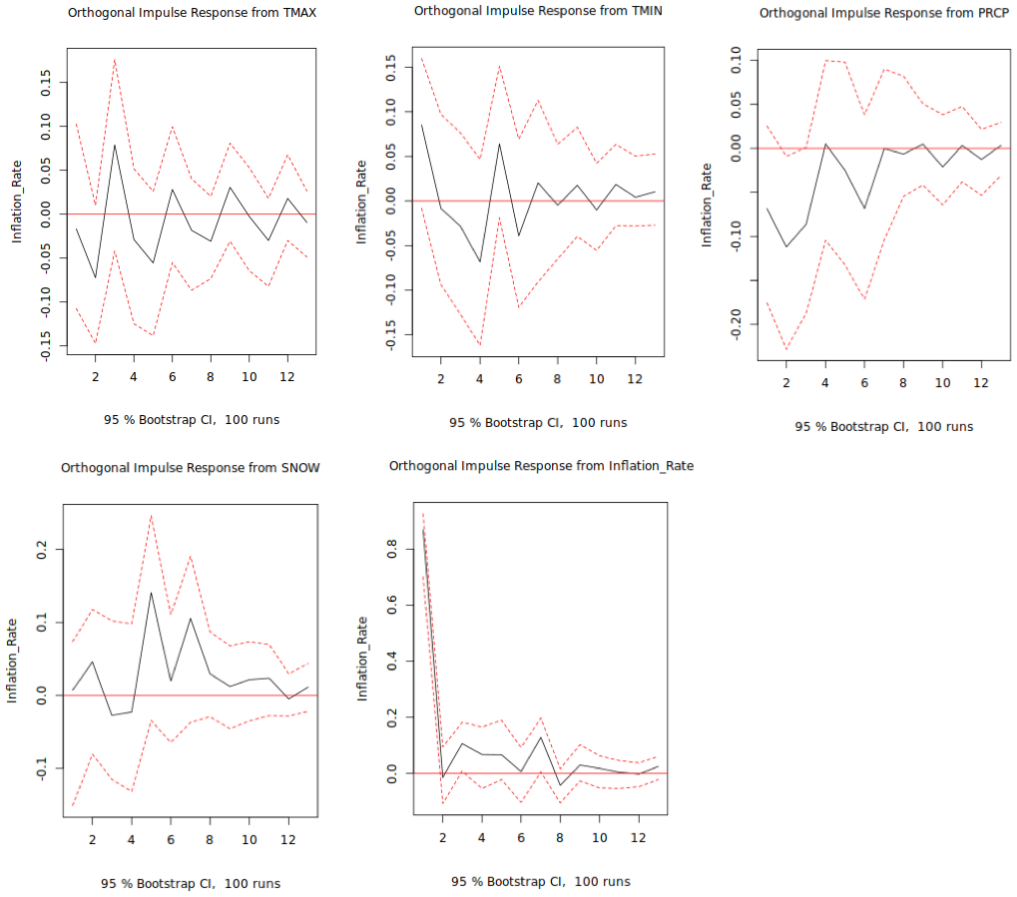


FIGURE A.1: Impulse Response Function of Food-only CPI on Climate Factors of Houston.

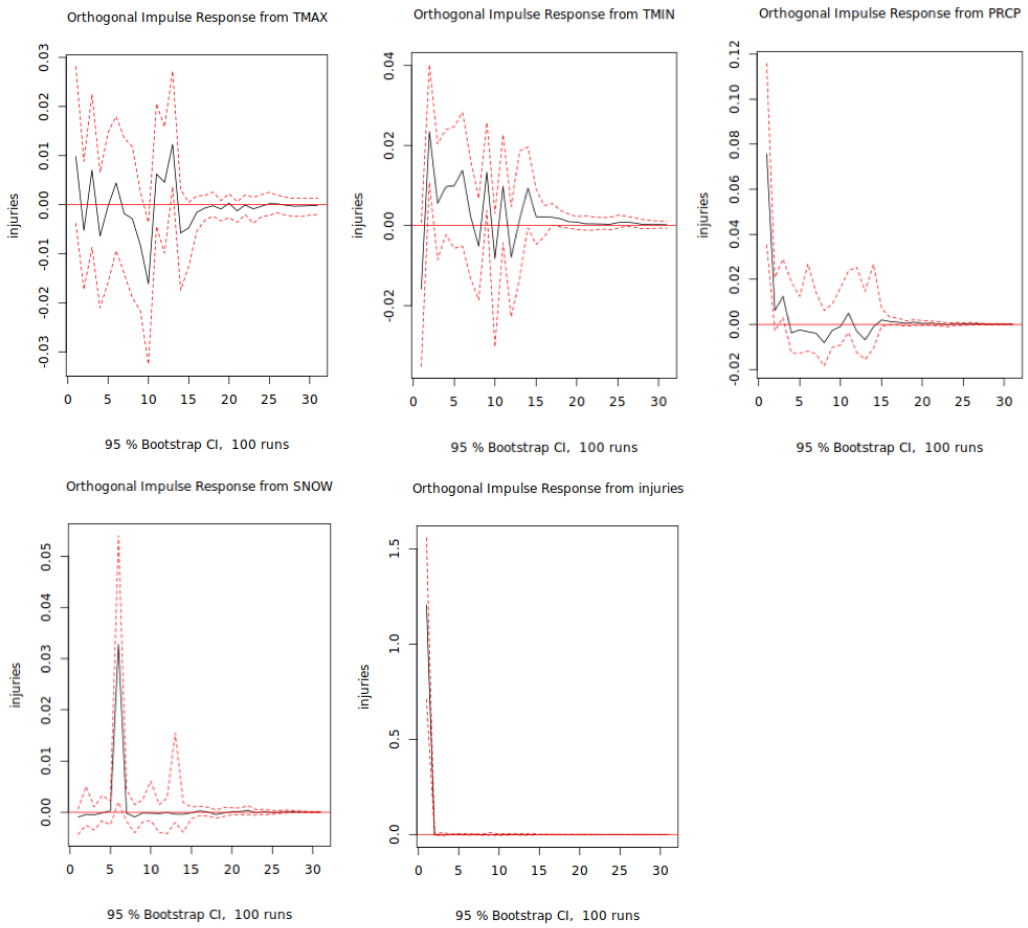
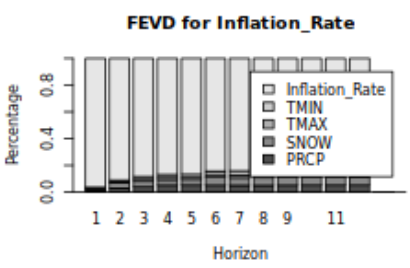
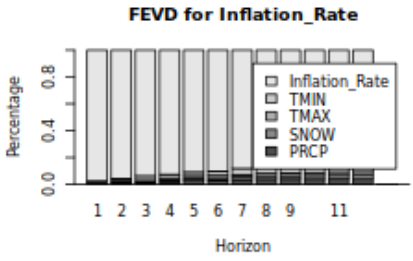


FIGURE A.2: Impulse Response Function of Injuries on Climate Factors of Houston.

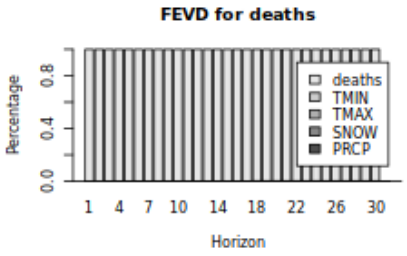
A.2 Forecast Error Variance Decomposition



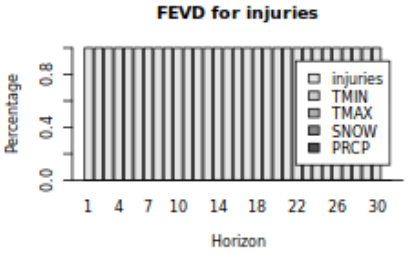
(A) CPI



(B) Food-only CPI

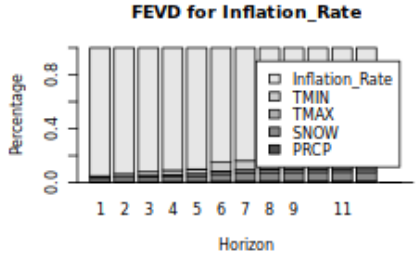


(C) Fatalities

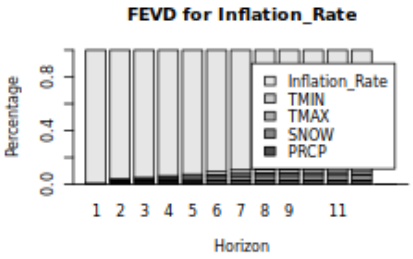


(D) Injuries

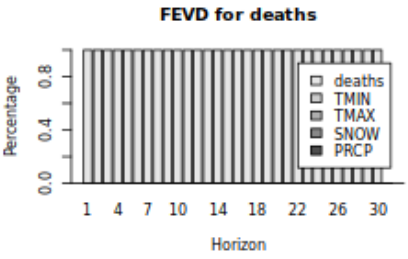
FIGURE A.3: Forecast Error Variance Decomposition of Chicago Dataset



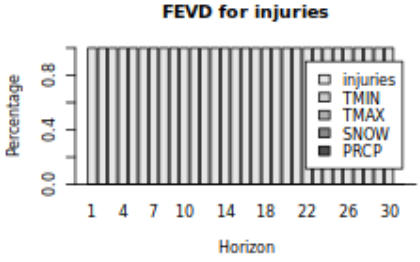
(A) CPI



(B) Food-only CPI

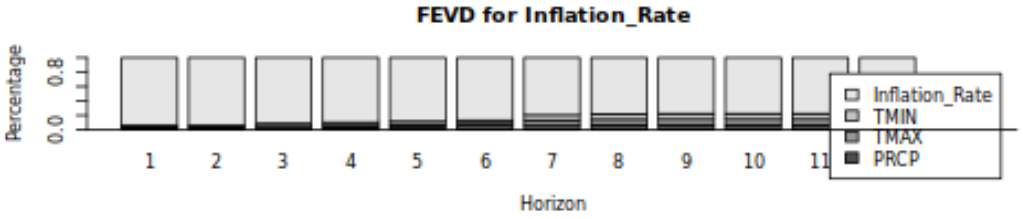


(C) Fatalities

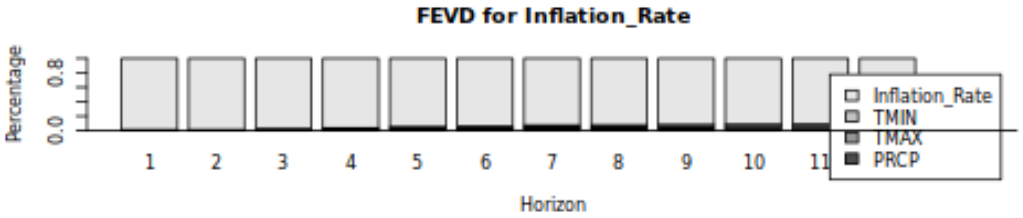


(D) Injuries

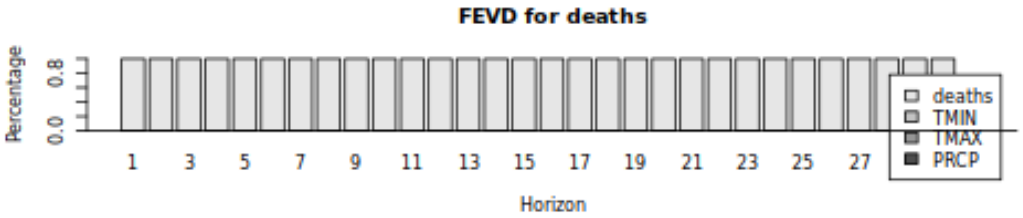
FIGURE A.4: Forecast Error Variance Decomposition of Boston Dataset



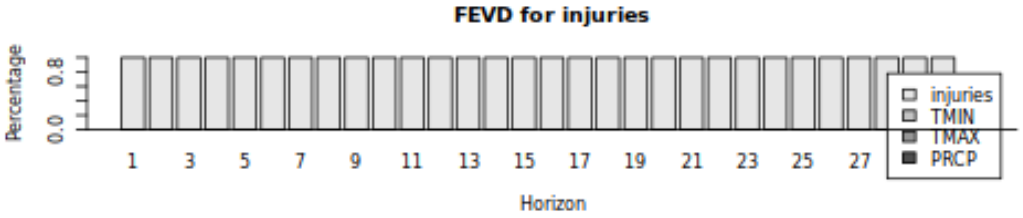
(A) CPI



(B) Food-only CPI

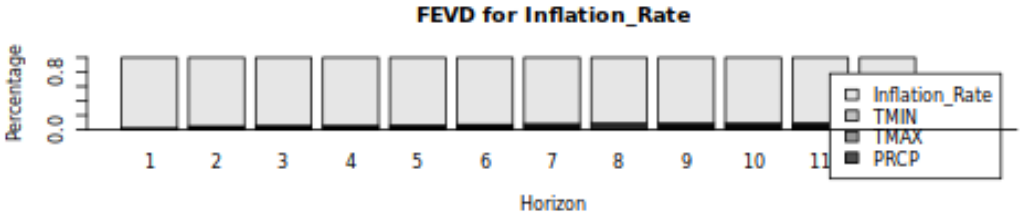


(C) Fatalities

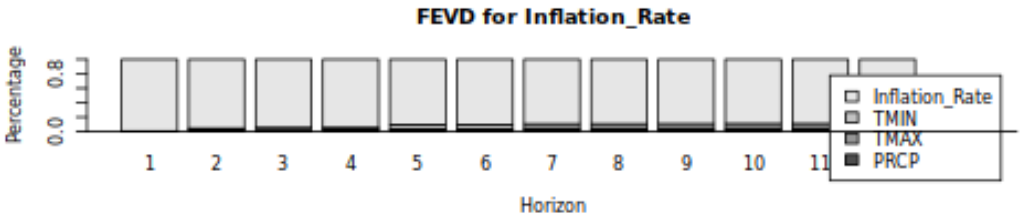


(D) Injuries

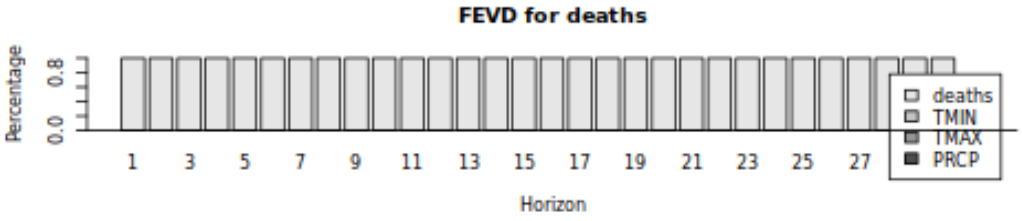
FIGURE A.5: Forecast Error Variance Decomposition of San Francisco Dataset



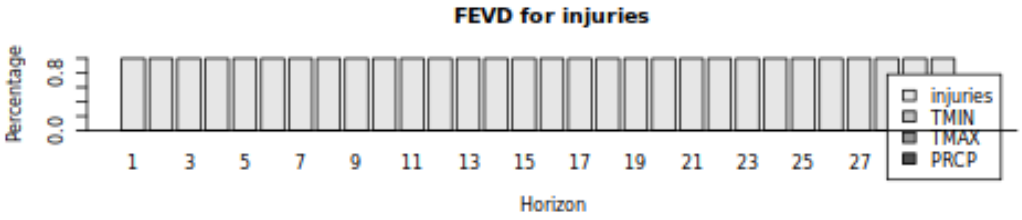
(A) CPI



(B) Food-only CPI



(C) Fatalities



(D) Injuries

FIGURE A.6: Forecast Error Variance Decomposition of Miami Dataset

Appendix B

Selected Codes

B.1 Julia

B.1.1 Optimization Function of AR (p) model

```
using Forecast

aic_ar_pred = []
bic_ar_pred =[]

for i in 1:110
    AR_model = ar(Vector{Float64}(y_train), i)
    d = AR_model.ic::Dict
    aic_ar = get(d,"AIC",4)
    bic_ar = get(d,"BIC",4)
    push!(aic_ar_pred, aic_ar)
    push!(bic_ar_pred, bic_ar)
end

println(findmin(aic_ar_pred))
println(findmin(bic_ar_pred))
```

B.1.2 Forecasting using AR and TES model

```
using Forecast; using StateSpaceModels

ar_temp = ar(Vector{Float64}(y_train),88)

fc_ar = Forecast.forecast(ar_temp,length(y_test));
plot(fc_ar)

model_ets = ExponentialSmoothing(Vector{Float64}(y_train);
                    trend = false, seasonal = 16)
StateSpaceModels.fit!(model_ets)

fc_ets = StateSpaceModels.forecast(model_ets, length(y_test))
plot(model_ets, fc_ets)
```

B.2 Python

B.2.1 Defining MLP's architecture

```
import torch
import torch.optim as optim
import torch.nn as nn
import torch.nn.functional as F

n_steps = 3
device = 'cuda' if torch.cuda.is_available() else 'cpu'

class MLP(nn.Module):

    def inital(self, input_size):
        super(MLP, self).inital()
        self.linear1 = nn.Linear(input_size, 4)
        self.output = nn.Linear(4, 1)
        self.dropout = nn.Dropout(p=0.2)

    def forward(self, x):
        x = self.linear1(x)
        x = F.relu(x)
        x = self.dropout(x)
        x = self.output(x)
        return x

torch.manual_seed(42)

model = MLP(n_steps).to(device)
loss_fn = nn.MSELoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)
```

B.3 R

B.3.1 VAR Model Analysis

```
library('vars')
library(astsa)

VARselect(CPI_df, lag.max = 15, type ='const')
var_model_df <- VAR(CPI_df, p = 6, type = 'const')

causality(var_model_df, cause=c("PRCP","SNOW","TMAX","TMIN"))

irf_var <- irf(var_model_df,n.ahead = 30, response = "injuries",
                    ortho = TRUE)
plot(irf_var)

fevd <- fevd(var_model_df,n.ahead = 30)
plot(fevd)
```

Bibliography

Ahmed, Nesreen K. et al. (2010). "An Empirical Comparison of Machine Learning Models for Time Series Forecasting". In: *Econometric Reviews* 29.5-6, pp. 594–621.

Ahuja, Sahil and Abhimanyu Kumar (Apr. 2022). "Expectation-based probabilistic naive approach for forecasting involving optimized parameter estimation". In: *Arabian Journal for Science and Engineering* 48.2, 1363–1370.

Albrecht, Frauke (Feb. 2023). *A Brief History of Neural Nets*.

Alon, Ilan, Min Qi, and Robert J. Sadowski (2001). "Forecasting aggregate retail sales:: a comparison of artificial neural networks and traditional methods". In: *Journal of Retailing and Consumer Services* 8.3, pp. 147–156. ISSN: 0969-6989.

Aubury, M. and W. Luk (Jan. 1996). "Binomial filters". In: *VLSI Signal Processing* 12.1.

Beutler, T. et al. (2022). "Deep Learning Approach Towards Precipitation Nowcasting: Evaluating Regional Extrapolation Capabilities". In: *EGUsphere* 2022, pp. 1–17.

Bushaev, Vitaly (Oct. 2018). "Adam — latest trends in deep learning optimization." In: *Towards Data Science*.

Chatwaranon, Yutti and Satean Tunyasrirut (2022). "Weather Forecast Comparison Between ARIMA and Exponential Smoothing in the Cloud". In: *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1–4.

Chen, Buo-Fu, Yu-Te Kuo, and Treng-Shi Huang (2023). "A deep learning ensemble approach for predicting tropical cyclone rapid intensification". In: *Atmospheric Science Letters*, e1151.

Cleveland, Robert B et al. (1990). "STL: A seasonal-trend decomposition". In: *J. Off. Stat* 6.1, pp. 3–73.

Climate at a Glance: National Time Series (Mar. 2023).

"Consumer Price Indexes Overview" (Jan. 2023). In: *Consumer Price Indexes*.

Dhamodharavadhani, S. and R. Rathipriya (2019). "Region-Wise Rainfall Prediction Using MapReduce-Based Exponential Smoothing Techniques". In: *Advances in Big Data and Cloud Computing*. Ed. by J. Dinesh Peter, Amir H. Alavi, and Bahman Javadi. Singapore: Springer Singapore, pp. 229–239. ISBN: 978-981-13-1882-5.

Ebi, Kristie L. et al. (2021). "Extreme Weather and Climate Change: Population Health and Health System Implications". In: *Annual Review of Public Health* 42.1, pp. 293–315.

Hajat, S. et al. (2016). "Public health vulnerability to wintertime weather: time-series regression and episode analyses of national mortality and morbidity databases to inform the Cold Weather Plan for England". In: *Public Health* 137, pp. 26–34. ISSN: 0033-3506.

Hill, Timothy, Marcus O'Connor, and William Remus (Nov. 1996). "Neural Network Models for Time Series Forecasts". In: *Management Science* 42, pp. 1082–1092.

Hu, Ling (2006). *Lecture 1: Stationary time series*.

Hyndman, Robin John and George Athanasopoulos (2018). *Forecasting: Principles and Practice*. English. 2nd. Australia: OTexts.

Jenkins, Gwilym M, George EP Box, and Gregory C Reinsel (2011). *Time series analysis: forecasting and control*. Vol. 734. John Wiley & Sons.

Keeler, Paul (Jan. 2022). *Wiener or brownian (motion) process*.

Kim, Hee Soo (test record), Christian Matthes, and Toan Phan (Aug. 2011). *Extreme Weather and the Macroeconomy*. Working Paper 21-14. Federal Reserve Bank of Richmond.

Kirakosyan, Arman (July 2016). "Extreme Weather Impact on Public Health and Economy across the United States". In.

Kotze, Kevin (n.d.). *Structural vector autoregression models - GitHub Pages*.

Kumar, Ajitesh (Aug. 2022). "Different types of time-series forecasting models". In: *Data Analytics*.

Kumar, N. and A. Middey (2023). "Extreme climate index estimation and projection in association with enviro-meteorological parameters using random forest-ARIMA hybrid model over the Vidarbha region, India". In: *Environmental Monitoring and Assessment* 195.380.

Kurniawan, Isman, Lusi Sofiana Silaban, and Devi Munandar (2020). "Implementation of Convolutional Neural Network and Multilayer Perceptron in Predicting Air Temperature in Padang". In: *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*.

Lai, Yuchuan and David A. Dzombak (2020). "Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation". In: *Weather and Forecasting* 35.3, pp. 959 –976.

Muriuki, Samuel Waiguru, Joseph Mung'atu, and Antony Gichuhi Waititu (2018). "Structural Vector Autoregressive (SVAR) Analysis of Maize Prices and Extreme Weather Shocks". In: *Journal of data science* 4, p. 79.

Newbold, P. and C. W. J. Granger (1974). "Experience with Forecasting Univariate Time Series and the Combination of Forecasts". In: *Journal of the Royal Statistical Society. Series A (General)* 137.2, pp. 131–165. ISSN: 00359238.

Nishimura, Taku et al. (2021). "Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City, Japan". In: *Sustainable Cities and Society* 74, p. 103203. ISSN: 2210-6707.

Park, Inyoung et al. (2019). "Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network". In: *Atmosphere* 10.11. ISSN: 2073-4433.

Paudel, Dilli et al. (2022). "Machine learning for regional crop yield forecasting in Europe". In: *Field Crops Research* 276, p. 108377. ISSN: 0378-4290.

Prabhakaran, Selva (Apr. 2022). "Augmented jdickey-fuller (ADF) test - must read guide". In: *Machine Learning Plus*.

Quadrianto, Novi, Kristian Kersting, and Zhao Xu (2010). "Gaussian Process". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, pp. 428–439. ISBN: 978-0-387-30164-8.

Roberts, S. et al. (2013). "Gaussian processes for time-series modelling". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1984, p. 20110550.

Seneviratne, S.I. et al. (2021). "Weather and Climate Extreme Events in a Changing Climate". In: *In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY: Cambridge University Press.

Shumway, Robert H. and David S. Stoffer (2017). "ARIMA Models". In: *Time Series Analysis and Its Applications: With R Examples*. Cham: Springer International Publishing, pp. 75–163. ISBN: 978-3-319-52452-8.

Smith, Adam (Jan. 2023). "2022 U.S. Billion-dollar Weather and Climate Disasters in Historical Context". In: *U.S. Billion-Dollar Weather and Climate Disasters (2023)*. NOAA National Centers for Environmental Information (NCEI).

Spadon, Gabriel et al. (2020). "Pay Attention to Evolution: Time Series Forecasting with Deep Graph-Evolution Learning". In: *CoRR* abs/2008.12833.

Srivastava, Nitish et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958.

Subrahmanyam, Kandula V et al. (Nov. 2021). "Prediction of heavy rainfall days over a peninsular Indian station using the Machine Learning Algorithms". In: *Journal of Earth System Science* 130.4.

Wang, Xin, Wenke Wang, and Bing Yan (2020). "Tropical Cyclone Intensity Change Prediction Based on Surrounding Environmental Conditions with Deep Learning". In: *Water* 12.10. ISSN: 2073-4441.

Zivot, Eric and Jiahui Wang (2003). "Vector Autoregressive Models for Multivariate Time Series". In: *Modeling Financial Time Series with S-Plus®*. New York, NY: Springer New York, pp. 369–413. ISBN: 978-0-387-21763-5.