

Modelling and Forecasting Drought events over Punjab using Machine Learning and Time Series Approaches



Shayan Saqib Lodhi
Regn No. 00000365406

A thesis submitted in partial fulfillment of the requirements
for the degree of **Master of Science**
in
Statistics

Supervised by: Dr. Firdos Khan

Department of Mathematics and Statistics

School of Natural Sciences
National University of Sciences and Technology
H-12, Islamabad, Pakistan

2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by **Shayan Saqib Lodhi** (Registration No. **00000365406**), of **School of Natural Sciences** has been vetted by undersigned, found complete in all respects as per NUST statutes/regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/M.Phil degree. It is further certified that necessary amendments as pointed out by GEC members and external examiner of the scholar have also been incorporated in the said thesis.


Signature: 

Name of Supervisor: Dr. Firdos Khan

Date: 08-09-2023

Signature (HoD): 


Date: 01/9/2023

Signature (Dean/Principal): 

Date: 12.9.2023

National University of Sciences & Technology**MS THESIS WORK**

We hereby recommend that the dissertation prepared under our supervision by: "**Shayan Saqib Lodhi**" Regn No. **00000365406** Titled: "**Modelling and Forecasting Drought Events Over the Punjab Using Machine Learning and Time Series Approaches**" accepted in partial fulfillment of the requirements for the award of **MS** degree.

Examination Committee Members1. Name: PROF. TAHIR MAHMOODSignature: 2. Name: DR. SHAKEEL AHMEDSignature: Supervisor's Name: DR. FIRDOS KHANSignature: 


 Head of Department

8/9/23

 Date
COUNTERSIGNEDDate: 12.9.2023


 Dean/Principal

*This thesis is dedicated to my beloved
parents*

Acknowledgements

My deepest thanks to **Allah Almighty**, for His countless blessings and for providing me with the strength to complete this work.

First of all, it is pleasure to express my deep appreciation and sincere gratitude to my supervisor **Dr.Firdos Khan**, for his guidance, encouragement, efforts and invaluable help to complete this work, without him this work would not have been accomplished.

I would also like to thank my loving and caring family especially my parents. Your sacrifice and support mean a lot to me.

Abstract

Droughts are continuous periods of precipitation that are much below average can cause water shortages, decreased soil moisture, and a number of negative impact on ecosystems, agriculture, humans and wildlife. RDI are typically used for monthly and annually data processing to simultaneously calculate precipitation and potential evapotranspiration on an appropriate time scale. Study improves the methods for choosing predictors and create an innovative model to forecast drought in Punjab, Pakistan. This study analyzed precipitation and temperature data from 9 meteorological site using RDI. This study aim is to compare four well-known models: Autoregressive Integrating Moving Average (ARIMA), Support Vector regression (SVR), Neural Network Auto Regressive (NNAR) and Dynamic Linear Model (DLM) in order to assess and compare the effectiveness of various forecasting model in predicting RDI. The training (January 1961 to November 2000) to evaluate its performance and for testing data (December 2000 to December 2016) to estimate their performance. ME, RMSE, MPE and MAPE for performance measure. And for model selection we use AIC and BIC. The selection of best model is based on minimum value of RMSE, ME, MPE, MAPE, AIC and BIC. The evaluation results suggest that NNAR outperforms the other three models. The study found that the NNAR model demonstrated the highest accuracy and prediction for RDI forecasting. The NNAR model performed better than other models capturing RDI changes with unique accuracy. These results demonstrate the effectiveness of the NNAR model in predicting drought conditions and to evaluate the model selection for RDI prediction. These finding help improve planning and management efforts for droughts, enabling efficient use of resources and decision-making. The results of this study provide proof for the decision, particularly droughts, as well as for creating drought mitigation efforts and applying plan into action to lessen the impact of drought in Punjab.

Keywords: Time Series Analysis; Reconnaissance Drought Index; Support Vector Regression; Neural Network Autoregressive; Dynamic Linear Model; Autoregressive Integrated Moving Average.

Contents

List of Abbreviations	v
List of figures	v
List of Tables	vi
1 Introduction	1
1.1 Objectives	3
1.2 Significance of study	3
2 Study Area and Datasets	4
2.1 Study Area	4
2.2 Datasets	6
3 Methodology	7
3.1 RDI	7
3.1.1 Calculation of RDI	7
3.1.2 Calculation of PET	10
3.2 ARIMA	11
3.3 NNAR	13
3.4 SVR	15
3.5 DLM	16

3.6	Model Evaluation	17
3.6.1	Numerical Evaluation	17
3.6.2	Graphical Evaluation	19
3.6.3	Future Projection	19
4	Results and Analysis	21
4.1	Checking the Stationarity	21
4.2	Model structure and learning	21
4.3	Qualitative performance measure	22
4.4	Choice of best model	23
4.5	Forecasting results	24
4.6	Point forecast	24
4.7	Graphical representation of NNAR	27
5	Conclusions	47
	Bibliography	49
	Appendix	56

List of Abbreviations

RDI	Reconnaissance Drought Index
PMD	Pakistan Meteorological Department
DRINC	Drought Indices Calculator
DRINC	Drought Indices Calculator
ARIMA	Autoregressive Integrated Moving Average
SVR	Support Vector Regression
SSM	State Space Models
PACF	Partial Autocorrelation Function
ACF	Autocorrelation Function
ADE	Augmented Dickey-Fuller
NNAR	Neural network autoregression
DLM	Dynamic Linear Model
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
RMSE	Root Mean Square Error
MPE	Mean Prediction Error
ME	Mean Error
MAPE	Mean Absolute Prediction Error
GDP	Gross Domestic Product
PET	Potential Evapotranspiration

List of Figures

2.1	Map of the selected meteorological stations of Punjab, Pakistan	5
3.1	Illustrates the technological road map for this thesis.	8
3.2	The design of NNAR with input, hidden and output layer	14
4.1	Comparison between forecast, observed and predicted for the time period (1962-2019) at Bahawalnagar.	38
4.2	Comparison between forecast, observed and predicted for the time period (1962-2019) at Bahawalpur.	39
4.3	Comparison between forecast, observed and predicted for the time period (1962-2019) at Faisalabad.	40
4.4	Comparison between forecast, observed and predicted for the time period (1962-2019) at Jhelum.	41
4.5	Comparison between forecast, observed and predicted for the time period (1962-2019) at Khanpur	42
4.6	Comparison between forecast, observed and predicted for the time period (1962-2019) at Lahore.	43
4.7	Comparison between forecast, observed and predicted for the time period (1962-2019) at Mianwali.	44
4.8	Comparison between forecast, observed and predicted for the time period (1962-2019) at Multan.	45
4.9	Comparison between forecast, observed and predicted for the time period (1962-2019) at Murree.	46

List of Tables

3.1	Classification of Standardized RDI values.	10
4.1	Model selection among the competing model for Reconnaissance Drought Index (RDI) for all variable at each station using train data.	25
4.2	Model selection among the competing model for Reconnaissance Drought Index (RDI) for all variable at each station using test data.	26
4.3	Goodness-of fit measure for different model of NNAR using train and test data	27
4.4	Forecasted value from NNAR for Bahawalnagar from January, 2018- January, 2019 as a blue line in graph.	28
4.5	Forecasted value from NNAR for Bahawalpur from January, 2018 to January, 2019 as a blue line in graph.	29
4.6	Forecasted value from NNAR for Faisalabad from January, 2018 to January, 2019 as a blue line in graph.	30
4.7	Forecasted value from NNAR for Jhelum from January, 2018 to January, 2019 as a blue line in graph.	31
4.8	Forecasted value from NNAR for Khanpur from January, 2018 to January, 2019 as a blue line in graph.	32
4.9	Forecasted value from NNAR for Lahore from January, 2018 to January, 2019 as a blue line in graph.	33
4.10	Forecasted value from NNAR for Mianwali from January, 2018 to January, 2019 as a blue line in graph.	34

4.11 Forecasted value from NNAR for Multan from January, 2018 to January, 2019 as a blue line in graph.	35
4.12 Forecasted value from NNAR for Murree from January, 2018 to January, 2019 shown in blue line in graph.	36

Chapter 1

Inroduction

Droughts is uncertain phenomena that has been caused by human-caused activities and climate change [1]. Significant changes have been caused at the local level by droughts. When there is much less water available than usual, a region is said to be experiences a drought and for the affected areas, the effects of drought vary. Essentially, a meteorological drought is described as a precipitation deficit across a region for a particular period of time [2]. Due to a lack of precipitation, dry wind and high temperature are highly variables and do vary from region [3].

The financial impact of the droughts in the United States is estimated to be between 6 and 8 billion dollars per years [4]. The IPCC noted in its fifth assessment report the effects of global warming on natural and human systems may be seen across all continents and seas [5]. Alterations to the hydrological cycle are thought to have greater impact on climate change (CC) than droughts, storms and floods [6]. However, the Middle East and North Africa are anticipated to see a 10-25% and 10-40% decline in rainfall and runoff, respectively and 5-20% rise in evaporation [8]. When droughts have negative impacts on agricultural land and cause significant reduction in yearly agricultural production, the related issues become more challenging [7]. The environment here is primarily dry and semi-arid, with an annual rainfall of 250 mm or less in about 60% of Pakistan and 250 to 500 mm in around 24% of the ratios [9]. Pakistan's economy is largely dependent on its agricultural, which is highly exposed to and beneficial to changes caused by the environment. Because agricultural productivity varies according

to change in precipitation, temperature and solar radiations, it has a big impact on the nation's socioeconomic stability. In this area, a trend has led to periodic floods and droughts except for 2004-2005, which was a rainy year. Pakistan has experienced droughts and water stress circumstances recently, which have had an impact on wheat productivity in both rain-fed and irrigated regions. The Punjab province, Pakistan's second-largest and most populated province, was the site of the current study. Punjab contributes the greatest portion of the nation's agriculture output; as a results, it is crucial to sustainable development of the country's economy. The agricultural sector dominates Punjab's landscape in terms of both the overall amount of agricultural land (16.68 million hectares) and its percentage of the nation's total agricultural gross domestic product (GDP) from (PBS), which is 60%(Reviewed on 2 December 2021) [10]. In various climates and for many purposes, a vast variety of weather related droughts indices of different complexity have been used. The RDI, a relatively new index [11], has been proposed at the moment. This index is used widely throughout the world, especially in semi-arid and arid regions and noted for its high sensitivity, adorability and minimal data requirements [12]. Using the Drought Indices Calculator (DriC), an application of specific software designed to offer an accessible interface for the estimation of drought indices [13]. Generally, if temperature rise, water demand increase as well, RDI may be modified to be used as an indication for future drought risk evaluation linked to different water areas.

Data-driven techniques are now the ones that are most frequently used to evaluate and forecast drought [14]. While made more challenging by the fact that defining the beginning and end of drought is highly difficult, numerous drought forecasting techniques have been created to increase the capability of drought forecasting [15]. All these methods are used and compare their performance in drought forecasting in the Punjab, Pakistan. A wide range of topics have seen as increase in use of NNAR and SVR. More specifically several earlier research [16] have demonstrated that these two methods are efficient for drought forecasting in certain study area. Droughts forecasting cannot be predicted with good accuracy using time series stochastic models (ARIMA)[17]. NNAR has some advantage over SVR and ARIMA models for drought forecasting because to

the lower root mean square error and larger regression co-efficient value [18]. A particular type of ANN is the Neural Network Auto-Regressive (NNAR) model, in which the lagged values of time series are used as predictor for the model's input and model's output is predicted values of the series. The NNAR does not include limitations on its parameters to achieve stationarity, which is one of the main difference between it and ARIMA models [19]. The nonlinear autoregressive neural network (NNAR) has been employed in this study [20] since the drought event is a nonlinear issue and a single feature (RDI) has been used as the input parameter. Time series data are viewed by state space models (SSMs) as the output of a dynamic system that is subject to random distribution. A specific instance of SSMs, the Bayesian dynamic model or simply dynamic linear model (DLM) is linear and Gaussian may be used to model, predict time series data and forecast data [21]. The forecasting performance of the ARIMA, NNAR, SVR and DLM were compared to predict the future daily values of the series using the model selection and model evaluation criteria.

1.1 Objectives

The objective of the study include the following:

- To determine which forecasting method (ARIMA, SVR, NNAR or DLM) provides the most accurate prediction for drought conditions, the study may compare the performance of different modelling approaches. This goal assist in choosing best approach for this specific application (RDI).

1.2 Significance of study

The following factors make a study on the Punjab droughts in Pakistan significant and extremely valuable. Punjab is an agricultural region, thus droughts can have a serious impact on food productions. In order to help policymakers and farmers create effective methods for droughts resistances.

Chapter 2

Study Area and Datasets

The following section provide the study area and data sets.

2.1 Study Area

Punjab is the province with highest population in the Pakistan, with more than 54% of all people living there, and it is enhanced by the fertile areas of five rivers [22]. Punjab province, the agricultural heartland of Pakistan is located between latitudes 27.70° N and 34.02° E and longitudes 69.32° E and 75.38° N. Punjab has second largest province in term of area $205,344 \text{ km}^2$ ($79,284 \text{ sq mi}$) after Baluchistan[23]. In terms of weather, Punjab has season: the hot season (April to June), the cold or moderate season (October to March) and rainy season (July to September)[24]. Temperature in Punjab range from 2 to 45°C , and average rainfall ranging from 460 to 960 mm . However, in some areas Punjab, temperature can approach 50° . The mean annual precipitation in this province is about 22.18mm . Due to Punjab province's important significance in Pakistani agriculture, it was chosen as the study region. The important cities in Punjab, Pakistan, showed in Figure 3.1 include Bahawalnagar, Bahawalpur, Faisalabad, Khanpur, Lahore, Mianwali, Multan, Murree, Jhelum in figure 3.1. These stations are chosen based on how important the climate is to the population, industries, and agriculture. Due to the threat of drought and the varied weather patterns from region to region, Punjab's agriculture industry is currently facing crisis [25]. Investigating the impacts of drought variability in various Punjab region is important.

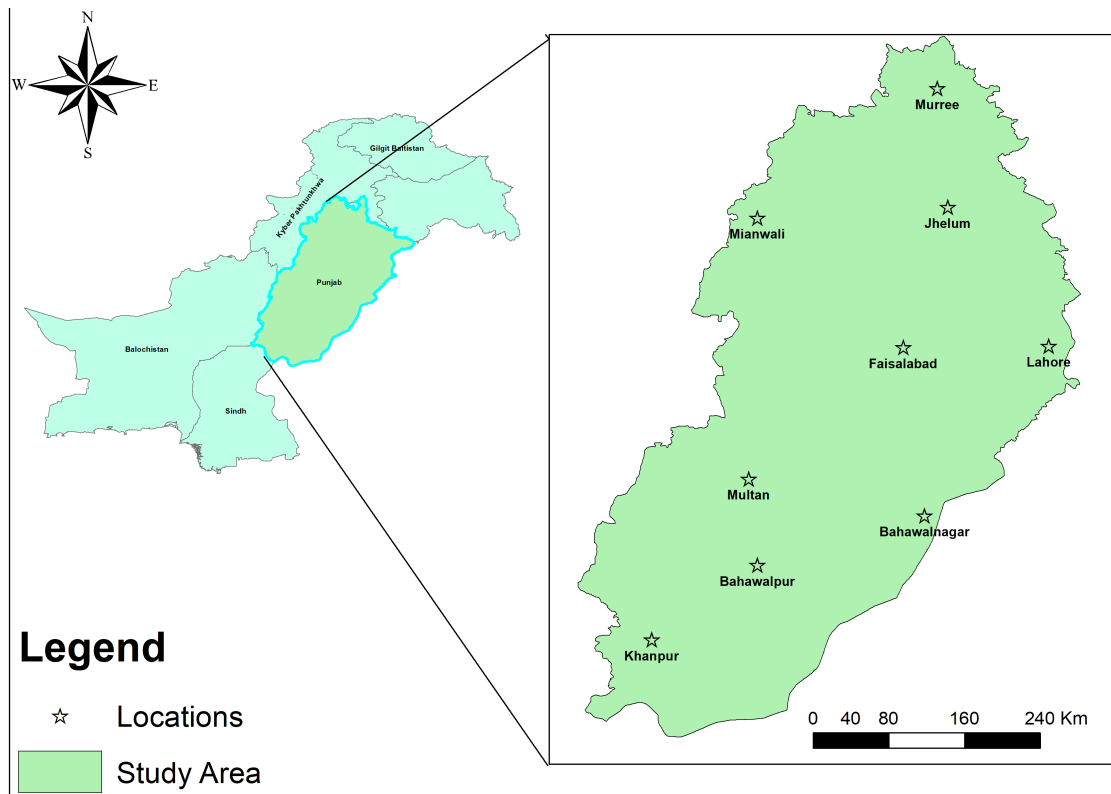


Figure 2.1: Map of the selected meteorological stations of Punjab, Pakistan

2.2 Datasets

The observations used in this study, include the meteorological stations data with 672 observations of Droughts in Punjab between January 1962 to December 2017. The minimum temperature, maximum temperature, precipitation and daily values for the years 1962 to 2017 are converted to monthly data are acquired from Pakistan Meteorological Department (PMD) for Meteorological stations in Punjab. Time series model that was created to compare the ARIMA, NNAR, DLM and SVR models and performance will be thoroughly assessed and the model which performance is best used for future forecasting. I use number of statistical methods together to calculate the forecasts. In particular, ARIMA models, NNAR, SVR and DLM. For phenomena like the droughts patterns, which is characterized by linear and nonlinear dynamic components. The combinations of various time series forecast methods should maximize the chance of capturing both the linear and nonlinear droughts patterns. It is well known from the research work of [26] that combining approaches with special characteristics may improve performance and prediction accuracy.

Chapter 3

Methodology

The following steps form the study's methodology: the RDI, ARIMA, NNAR, SVR and DLM, model evaluation.

3.1 RDI

One of the primary natural factors causing harm to ecosystems, agricultural outputs, water supplies etc is drought. To gauge the severity of the droughts, numerous indexes have been created. The RDI are most recent index used to measure drought.

3.1.1 Calculation of RDI

Climate change has a significant influence on dryland's water supply and demand. Applying the RDI, the dry and wet conditions at the choose station have been evaluated. The RDI, which measure the severity of drought, can be calculated. The initial α and standard form RDI_{st} of RDI are both used to express it. Based on the ratio of prospective precipitation (P) to total precipitation(P), RDI is calculated and on a yearly basis (k=12). The following equation is used to determine the annual values of α_0 for the ith year.

$$\alpha_0^{(i)} = \frac{\sum_{j=1}^{12} P_{ij}}{\sum_{j=1}^{12} PET_{ij}} \quad (3.1)$$

Where $i= 1$ to n and $j=1$ to k . The PET of month j ($j=1$ October, to $j= 12$ September) of years i and n . The P and PET during j th month of i th year are represented by P_{ij}

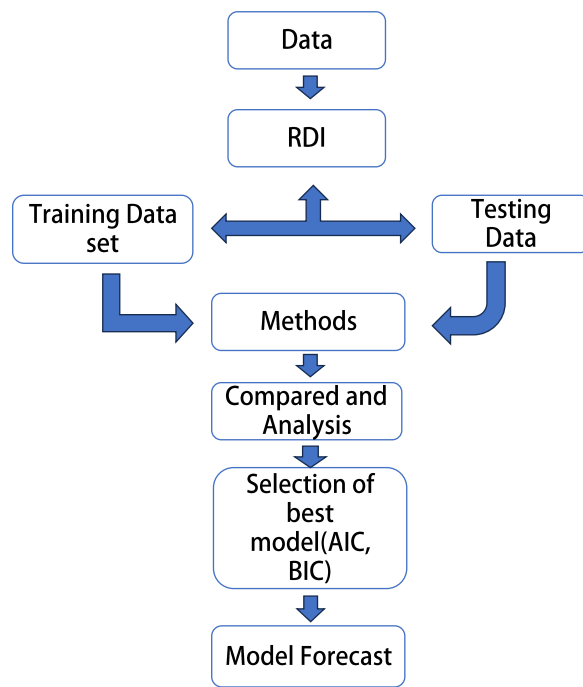


Figure 3.1: Illustrates the techonological road map for this thesis.

and PET_{ij} , respectively in 3.1.

By implementing for each time scale, the Normalized RDI (RDI_n), is the second form is produced.

$$RDI_n^i = \frac{\alpha^i}{\bar{\alpha}_0} - 1 \quad (3.2)$$

The $\bar{\alpha}_0$ represent the average value for I over the n years that are represented in the data. The Standardized RDI (RDI_{st}) is the third form, which is provided in equation 3.2.

$$RDI_k^{(i)} = \frac{y_k^i - \bar{y}_k}{\sigma y_k} \quad (3.3)$$

Where $y^{(i)} = \ln \alpha^{(i)}$, \bar{y}_k arithmetic mean and the σy_k be the standard deviation. The alpha $RDI_{\alpha k}$, normalized RDI_n and standardized RDI_{st} versions of RDI are all possible formulations. While the $RDI_{\alpha k}$ is used as an aridity index, the RDI_{st} is used to measure how severe a drought. The aggregated P and PET theories are the fundamental theoretical underpinnings of the measure [27]. Data analysis revealed that k closely matches the gamma and lognormal distribution across a range of stations and time scale. However, as suggested by [28] and used in this study, the gamma distribution was generally more useful. As a result, by modifying the gamma distribution, the computation of RDI_{st} may be performed under ideal conditions more successfully. The brief overview of the gamma distribution's probability density function [28].

$$g(x) = \frac{1}{\beta \gamma \Gamma(\gamma)} x^{\gamma-1} e^{-\frac{x}{\beta}} \quad (3.4)$$

where x show the rainfall amount, β the scalw and γ is the shape parameter. Now we estimate the values for the both γ and β which is describe by below equation.

$$\gamma = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (3.5)$$

$$\beta = \frac{\bar{x}}{\gamma} \quad (3.6)$$

Where $A = \ln \bar{x} - \frac{\sum \ln(x)}{n}$. Taking the parameters γ and β for specific year, the cumulative probability function of a measured rainfall is calculated [29].

$$G(x) = \frac{1}{\Gamma(\gamma)} \int_0^x \frac{x^{\gamma-1}}{\beta} e^{-\frac{x}{\beta}} \quad (3.7)$$

Drought class	RDI _{st}
Extremely wet	≥ 2.0
Very wet	1.99 to 1.50
Moderately wet	1.49 to 1.00
Normal	0.99 to 0.00
Near normal	0.00 to -0.99
Moderately dry	-1 to -1.49
Severely dry	-1.5 to -1.99
Extremely dry	≤ 2.0

Table 3.1: Classification of Standardized RDI values.

The cumulative probability function changes because the gamma function is undefined at $x=0$ and the distribution of precipitation contains some zeros.

$$H(x) = q + (1 - q)G(x) \quad (3.8)$$

Where $G(x)$ represent the cumulative probability of the incomplete gamma distribution. The values are in Table 3.1.

3.1.2 Calculation of PET

The most often used method for determining PET is the Thornthwaite methodology, which only considers the average temperature [30]. As a result, it can be used in areas where there is a limited supply of data. PET calculation are based on a multiple steps. First, using the equation below, we determine the monthly Thornthwaite heat index i .

$$i = \frac{t^{1.1514}}{5} \quad (3.9)$$

Where temperature (t) in Celsius ($^{\circ}C$) on monthly average. The annual index is then calculated by adding the monthly indexes.

$$I = \sum_{i=1}^n i \quad (3.10)$$

Thirdly, it is assumed that a month consist of 30 days and that a day has 12 hours of sunlight [29].

$$PET_{non-corrected} = 16. \left(\frac{N}{12}\right) \cdot \left(\frac{m}{30}\right) \cdot \left(\frac{10.t}{I}\right)^\alpha \quad (3.11)$$

These figures were then adjusted to reflect the actual amount of sunny days and hours in a month [31], as shown below:

$$PET = PET_{non-corrected} \frac{N}{12} \frac{d}{30} \quad (3.12)$$

For each month, N: is the number of days and d is the number of theoretical sunshine hours.

3.2 ARIMA

The ARIMA model originated in 1970 by Box and Jenkins [32] widely used in the social sciences, economics and finance. The ARIMA models is expressed as ARIMA (p, d, q) where p denotes the number of parameters in the autoregressive (AR) models, d represents the differencing degree, q indicates the number of parameters in moving average(MA) models. By describing autocorrelation in the data, ARIMA models provide an alternative method for time series forecasting. In recently years, ARIMA linear model has been often applied to predict droughts [33]. The general non-seasonal ARIMA model's equation is expressed as [10].

$$Z_t = \frac{\theta(\beta)\alpha_t}{\phi(\beta)\nabla^d} \quad (3.13)$$

$$\theta(\beta) = (1 - \theta_1\beta - \theta_2\beta^2 - \dots - \theta_q\beta^q) \quad (3.14)$$

$$\phi(\beta) = (1 - \phi_1\beta - \phi_2\beta^2 - \dots - \phi_p\beta^p) \quad (3.15)$$

In equation 3.13 Z_t represent observed time series and β is Back shift operator. Polynomial of order p and q are $\theta(\beta)$ and $\phi(\beta)$. The orders p and q respectively, represent the non-seasonal auto-regression and moving average orders. ∇^d specifies the differencing process, d regular differencing. Three model building processes and a forecasting step

are including in the Box and Jenkins methodology.

- Identification: The time series is thought to be stationary under the Box-Jenkins models. The statistical properties of a stationary time series, such as the mean is constant through time. The technique of transformation and differencing were applied to non-stationary series in this research.
- Diagnostic test: Evaluate a preliminary model's goodness of fit, a variety of diagnostic approach can be applied. The models in this study were examined using a variety of diagnostic statistic and graph. To examine the white noise in residuals, the Ljung-Box (LB) [34] test was applied. ACF and PACF plots of the residuals were also used to evaluate the absence of correlation between the residuals. With the use of Durbin-Watson (DW) statistics [35] it was also determined whether there was any correlation between the residuals.
- Forecasting: For predicting, the model with lowest RMSE was selected.

To determine whether or not our data are steady, we performed unit root test DF and ADF. Conduct were fixed to satisfy the model's basic elements. Identifying a time series stationary was the initial step in developing ARIMA models. After non-stationarity has been eliminated, the data's correlation structure was determined using ACF and PACF. After the ACF and PACF were to identify the significant lags. Based on the model's accuracy and precision, a set of ARIMA models was chosen. Time series consists of linear and nonlinear components and different model approaches that forecast future values using previous observed data and separate dataset into two sets with ratio of 70% for training and 30% for testing set. The results of the RDI with monthly data and AIC and BIC used.

3.3 NNAR

Artificial neural networks (ANNs) are a well-liked method for modelling nonlinear time series. ANNs are machine learning model that may identify nonlinear, complex correlation between variables since they are based on the structures of the neurons. ANNs solve various weather patterns, stock market trends and traffic flow and many other forecasting issues. The "nnetar ()" function included in Hyndman's package "forecast" (in the R environment) was used to identify NNAR models. The primary focus was on NNAR model and using training and test data sets. The objective was to determine the "prediction performance" of the produced NNAR models. However, in this study NNAR, ARIMA were created under the R software environment and the "forecasting performance" of these model approach was compared. First data were transformed using Box-Cox method before model was estimated. The hidden layer chosen automatically throughout the modelling process, we concentrated on the NNAR model. Just as lagged values in a linear autoregressive model, lagged values of the time series can actually be used as input data to neural network with time-series data. In NNAR models, complicated nonlinear interactions and functional form are represented by a network of neuron or nodes. The neurons of a simple neural framework are arranged in two layers: the original time series is identified in the bottom layer, which the forecasts are identified in the top layer. When an intermediary layer containing "hidden neuron" is added, the final model transform into a nonlinear one, although it is still equal to simple linear regression 3.2. The NNAR models called NNAR (p, k) models that are used to forecast time series models. Where p is the non-seasonal delays as input, k is the number of nodes in hidden layer. The neural network is feed forward neural network which involves a linear combination function and activation function.

$$net_j = \sum w_{ij}y_{ij} \quad (3.16)$$

The sigmoidal function, which was proposed by Master [1993] [36], is the activation function that is used most commonly.

$$f(y) = \frac{1}{1 + e^{-y}} \quad (3.17)$$

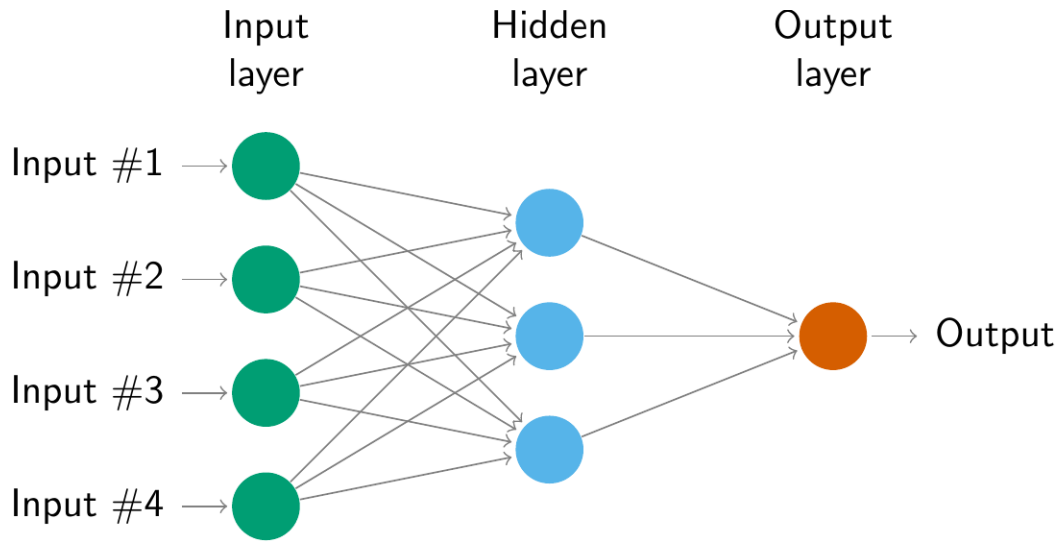


Figure 3.2: The design of NNAR with input, hidden and output layer
<https://otexts.com/fpp3/figs/nnet2-1.png>

After combining the inputs using a linear function, the output is then sent through a non-linear sigmoid activation function. In R Core Team (2014), the back propagation algorithm is used to update the neural network's weights. It is possible to understand the weights as in the case of normal regression models, which use a closed form for the minimization of sum of squared errors, because the weights are derived through computational techniques 3.2. It is true that [37] advise against interpreting the fitted weights. The MAE, RMSE, MAPE, MASE is used to check out the goodness of fit. The equation for NNAR model as follow [38].

$$y_t = f(y_{t-1} + \epsilon_t) \quad (3.18)$$

where in 3.18 $y_{(t-1)} = (y_{(t-1)}, y_{(t-2)}, \dots, y_{(t-n)})$, is vector that contains number of lagged values for observed data, f is the neural network with n number of neurons in the single layer and ϵ_t is the error term. We can simulate future sample pathways of this model iteratively, by randomly generating a value for, ϵ_t either by resampling from historical values.

$$y_{T+1}^* = f(y_t + \epsilon_{T+1}^*) \quad (3.19)$$

3.19 is used to forecast for the time series.

3.4 SVR

Support vector machine (SVMs) were developed by Vapnik in 1997 [39] to describe the characteristics of learning machine and their ability to simplify unknown substance [40]. Support Vector Regression (RDI) is a machine learning technique that is particularly relevant for RDI forecasting due to its ability. For the linear separable data for training set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in \{x \times y\}$ with a collection of sample points $X = \{x_1, x_2, \dots, x_n\}$ derived from R^n with target value $Y = \{y_1, y_2, \dots, y_n\}$ with the input and target value $\{x'_i, y'_i\}$ is the monthly record of RDI where $y_i \in R$. The optimization problem is solved by:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (3.20)$$

$$y_i[(w \cdot x_i) + b] \geq 1, i = 1, 2, 3, \dots, l \quad (3.21)$$

where in equation 3.21, the best solution determines the classification surface w^* and b^* is:

$$(w^* \cdot b + b^*) + b^* = 0 \quad (3.22)$$

The important qualities must be preserved when using the SVM approach, namely SVR to tackle the regression problem. The "loss model" is expressed as follows using "loss function" based on the Vapnik standard as a frame work.

$$\min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^m (\epsilon_i + \xi_i^*) \quad (3.23)$$

The decision boundary of the support vector expands as C in equation 3.23 σ is the nuclear density, ξ_i and ξ_i^* are slack variables that specify the model's error requirements and c is the penalty parameter. The relationship between sigma and gamma in the radial basis function (RBF) formula. In the RBF, sigma and gamma are determined as follow:

$$\text{gamma} = \frac{1}{2\sigma^2} \quad (3.24)$$

Therefore, a decrease in σ or an increase in gamma result in reduction in the RBF's width in equation 3.24. According to conventional neural network, which conform to

the empirical risk minimization concept, SVR models support the structural risk minimization principle [41]. All SVR models were developed using online SVR program [42] which may be used to develop support vector machine for regression. The training set was made up of 70% of the data, while the training set was made up of the remaining 30%. The data unlike neural networks, can only be divided into two sets, with the calibration set serving as a substitute for the training and testing sets in neural networks. The range of all input and output was set to 0 to 1. A constant called the c parameter minimizes the model space and determines how complicated the solution. Where ϵ is the loss function that represents the regression vector without all of the input data and C is a positive constant that is a C is a positive constant that is a capacity control parameter [43]. These three parameters were chosen through a process of trial and error. These set of variables lowest RMSE value from testing data is selected.

3.5 DLM

A category of non-stationary time series models is defined by DLMS. Trends, seasonality and autoregressive components can all be modelled using DLMS. The dynamic Bayesian model is founded on the idea that uncertainties must be represented by a probability distribution as formalized by West and Harrison [44]. The fundamental method of DLM modelling is to develop a structural model for the time series, representing all parameters and observable data. A common method for smoothing and predicting time series is DLMS. The dynamic linear model with an observation equation and a model/evolution equation is:

$$y_t = F_t' \theta_t + \epsilon_t \quad (3.25)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (3.26)$$

The equation 3.25 is observational equation and 3.26 is evolution equation.

$t=1,2,\dots,n$.

Observation made at time t is y_t . although clarity is assumed, it might actually be a vector. The vector at time t with size is a row vector of covariate at time t

$\theta_t = (\theta_{t,1}, \theta_{t,2}, \dots, \theta_{t,p})$ with dimension $p \times 1$. F'_t is a row vector (dimension $(p \times 1)$) of covariate at time t . G_t is a matrix of dimension $p \times p$ known as evolution or transition matrix. We will mainly talk about the Bayesian analysis of these models, whose Kalman filter are equivalent [44] Bayesian Forecasting and Dynamic Models, [45]. The expression (P_t/θ_{t-1}) represents the probability of transition state. This probability captures the Markov property, which states that the current state θ_t depends only on the previous state θ_{t-1} and not on any earlier states.

$$\theta_t/\theta_{t-1} \sim N(G_t\theta_{t-1}, w_t) \quad (3.27)$$

$$y_t/\theta_t \sim N(F'_t, \epsilon_t) \quad (3.28)$$

The time series aim is future forecasting observation, with the length of the forecast depending on the nature and objective of the study. To achieve this, the first step is to estimate the next value of the state vector θ_{t+1} based on the available data $(y_{1:t})$. Subsequently, the next observation y_{t+1} can be forecasted using estimate state and same step is used for k step forecasting [46]. The selection of model is best we use AIC and BIC and for performance we use ME, RMSE, MAPE, MPE. For each method and each modelling time forecasting is obtained.

3.6 Model Evaluation

Evaluating the performance in statistical modelling, analysis is an important stage. Effectiveness of calibrated models is assessed numerically by various statistical measures as well as graphically by comparing the observed and predicted value and forecast these values which are listed in the subsections that follow.

3.6.1 Numerical Evaluation

How closely the test data predictions and observed values match determines the performance of models. The performance obtained for the ARIMA, NNAR, SVR and DLM are compared using four distinct prediction consistency measures: ME 3.29, RMSE, MPE 3.31 and MAPE 3.32. And for model selection we use AIC 3.33 and BIC 3.34.

This is an estimate of the random component standard deviation in data. Model performance was assessed using RMSE. The optimal model has a minimal RMSE, This is an estimate of the random component's standard deviation in data.[47]. Four distinct metrics were used to numerically evaluate the model's performance:

$$ME = \frac{1}{n} \sum_{t=1}^n [y_t - \hat{y}_t] \quad (3.29)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3.30)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (3.31)$$

$$MPE = \sum_{t=1}^n |y_t - \hat{y}_t| \quad (3.32)$$

n represent points, y_t (observed), \hat{y}_t (predicted) stand for the observed and predicted data from the model, respectively in equation. A model is said to have higher accuracy when its value for RMSE is lower when compared to the performance of other models. RMSE quantifies the difference between observed data and model presented data. The MAPE is one of the most often used error metrics in time series forecasting 3.32. It is calculated by dividing the actual values by the absolute difference between the values that were observed and those that were anticipated. The normal percentage error of the forecast is measured by term MPE. ME determine the average of the forecast and actual value differences.

Despite the fact that both AIC and BIC are driven by Maximum Likelihood estimate and penalize free parameter to prevent overfitting, they do so in ways that produce noticeably different behavior. Assuming regularly distributed errors and other well-behaving assumptions, the following approach is one that is frequently used.

$$AIC = -2 \ln(L) + 2k \quad (3.33)$$

$$BIC = -2 \ln(L) + k \ln(N) \quad (3.34)$$

The number of parameters is denoted by k 3.33, captures the complexity of a model. $\ln(L)$, the log-likelihood of the model on the data, captures the goodness of fit. And n is number of data points. In this context of maximum likelihood estimation, AIC and BIC reward the goodness of fit, the likelihood function and penalize the complexity of a model and the number of parameters.

3.6.2 Graphical Evaluation

Using a yearly graph, the predicted data are contrasted with the data that was observed during 52-years period (1961-2017) for the purpose of graphic evaluation. The graph for comparing the observed and predicted values to evaluate the performance of various models, such as ARIMA, NNAR, SVR and DLM. For this purpose, a graph can make that contrasts each model's projected values with the observed values. The time period or data point can be presented on x-axis and the values can be represented on the y-axis in a line graph. The observed data first convert to time series data for further predictions. The lines represent the predicted value and observed values on the graph and on the basis of that forecasting the values. The time period or data point can be presented on x-axis and the values can be represented on the y-axis in a line graph. The observed data first convert to time series data for further predictions. The lines represent the predicted value and observed values on the graph and on the basis of that forecasting the values. The find out about each model's performance and identify the one that offers the best correspond for the information provided by comparing observed and predicted values and looking at the error metrics.

3.6.3 Future Projection

The future projection is to find out more about each model's performance and identify the one that which provide the best correspond for the information provided by comparing the observed and predicted values and looking at the error metrics. First fitting the model (ARIMA, SVR, NNAR and DLM) and then training each model with the ratio of 70:30. The 70% for the training set and 30% for the testing set. Determine

the forecast for the number of years in the future for which years we determine the forecast accuracy. The observed and predicted values and from that forecast the values for next two years.

Chapter 4

Results and Analysis

This chapter is divided into two parts. In first part the result about the performance of the model and the forecast result are presented while second part is about the discussion.

4.1 Checking the Stationarity

In this research, the series variances were normalized and stationarity transformed using the Box and Cox [48] power transformation. This unit root test may be used to distinguish between a stationary and non-stationary time series [49]. Augmented Dicky-Fuller employed to evaluate the stationary behavior of time series processing in this study. Regression models use nonstationary time series may produce false results [50]. Additionally, any non-stationary series first order differencing seems to be series is steady (p-value < 0.01). As a result, it was assumed that each series integrated model term or order of differencing (d) had a single unit root.

4.2 Model structure and learning

It is difficult to give the results for all seven stations in detail, split data which consists of training data testing data is chosen. The observed and predicted values for RDI produced by each of the nine station in Punjab, Pakistan using the various suggested models. The testing set is used to check the performance of these four models. The

following is a summary of the details. According to [51], the autocorrelation function is statistically practical method for obtaining a time series description [52]. ARIMA models were developed using various combinations that resulted in the model with best fit was chosen as the one with smallest AIC. The outcome of the different performance indicator of the ARIMA model at each of nine stations. A time series model's lack of fit might be investigate using the LB test [50]. Additionally, it has been demonstrated that neural networks with a single hidden layer may accurately approximate any function when the model include enough weight [54]. The forecast stream flow for droughts, model including support vector regression (SVR) was applied. The C , γ and ϵ are parameters values were determined by trial and error. The objective of this trial and error was to reduce the RMSE for selecting the model [43]. We establish the prior distribution for and this quadruple as well as walk through the DLM specification [53]. The ability to combine regression, trend, seasonal components to produce a single DLM is a significant benefit of dynamic linear model (DLM). The model's performance and selection can be detected and check which model is best on the basis of that model we forecast our results.

4.3 Qualitative performance measure

In this study, various performance metrics are taken into account while comparing the models. There is no single measure that can be used for all data-driven techniques, which is the main reason for the use of many measurements. Various metrics can also account for possible consequences that may be hidden in data set. The model fitting and prediction using the ARIMA, NNAR, SVR and DLM models for the prediction period of January 1962 to December 2017. Although all models work, it is not clear which model is the most effective. The ME, MPE, RMSE and MAPE from testing data are selected and forecasting accuracy measure to evaluate the provided model's quality. Table 3.1 display the performance and results of ARIMA, NNAR, SVR and DLM. The other four criteria are model performance criteria, whereas the AIC and BIC are used to pick models. The NNAR model's performed for RDI provide satisfactory and better

accuracy result than other models. We listed the primary forecast accuracy metrics and the top-chosen parameter for the models in the table.

4.4 Choice of best model

The choice of best model depends on RMSE, ME, MPE and MAPE and we compare the performance metrics across the models and identify the model with high accuracy and lowest error. Depending on the particular requirements and feature of RDI, the most appropriate model will be chosen. It is recommended to test our various models, evaluate their efficacy and choose the one that provides a perfect combination of precision. In the table we see that the highlighted AIC, BIC, ME, RMSE, MAPE, MPE perform best on training set in all of our research. The value for these criteria vary depending on the station and model. The values for the various models can be compared to find the model that performs the best for each station in terms of accuracy and goodness of fit. Better performance models are those with lower RMSE, MPE and MAPE values as well as lower AIC, BIC for all models [4.3](#).

For the station "Bahawalnagar," the "NNAR" model has the lowest AIC (268.12), BIC (271.37), RMSE (0.288), MPE (0.008) and MAPE (5.612) values among all the models, making it the best model for this station. NNAR is more effectively for the station "Bahawalpur," with the lowest AIC (795.4), BIC (808.7), RMSE (0.097) and MPE (0.002) values. NNAR is more effectively and has the lowest MPE (0.002) and second lowest RMSE (0.010) values for the station "Faisalabad," as well as the lowest AIC (375.4) and BIC (388.7) values. Therefore, for this station, NNAR is more effectively. NNAR is more effectively for the station "Jhelum," having the lowest AIC (785.4) and BIC (798.6) values, as well as the lowest RMSE (0.102) and MPE (0.009) values. NNAR is more effectively for the station "Khanpur" having lowest AIC (706.0) and BIC (719.2) values. NNAR is more effectively for the station "Lahore," having the lowest AIC (375.5) and (388.7) values. NNAR is more effectively for the station "Mainwali" having the lowest AIC (295.7) and BIC (308.9) values. NNAR is more effectively for the station "Multan," having the lowest AIC (545.0) and BIC (558.2)

values. NNAR is more effectively for the station "Murree," having the lowest AIC (545.027) and BIC (558.272) values.

4.5 Forecasting results

It is shown in the table that NNAR models has higher forecast quality and assess that the data better than the rest of the models, making it more appropriate model for predicting the RDI. We therefore, used the NNAR model to predict the forecasted RDI for time series from January 1, 2018, to December 31,2020. The best model for the RDI with more accurate value are NNAR. Lewis'view [55] state that all prediction models are extremely accurate since MAPE is less than 10. The prediction specially display a forecast accuracy range between 95.68% and 97.84%. The NNAR estimation (testing) methodology has been used. There is no set approach for determining the best network design in neural network. As a result, Predictions of the RDI in the A variety of parameter settings have been used to explore the Punjab region while taking into account hidden layer. In order to increase the number of observation for comparison in the prediction, having the number of observation for training from January 1961 to November 2000 and for testing data from December 2000 to December 2016. Due to its demonstrated efficiency accuracy, and dependability, the Levenberg- Marquardt (LM) algorithm was employed the train the NNAR model [56]. The forecast for the next 2 years. Numerous simulations were run when the NNAR was begin developed in an effort to determine the optimal model. We compare four models and it can be seen that there are no significant difference between them.

4.6 Point forecast

The predicted annual values for the year 2018-2020. It is important to note that, unlike other forecasting models implemented in R, the "nnetar" function only provides point forecast at this moments, not that 80%and 95% prediction intervals 4.4. The results suggest that the droughts for all the stations will be increased and decreased which mean drought condition change and not same and the trends during the next two years

Station	Models	Model selection criteria		Model Performance Evaluation criteria			
		AIC	BIC	ME	RMSE	MPE	MAPE
Bahawalnagar	ARIMA	465.16	519.06	0.512	0.829	0.658	3.69
	NNAR	259.52	268.5	0.005	0.291	0.225	3.807
	SVR	566.52	625.13	0.008	0.349	0.221	7.661
	DLM	614.4	715.6	0.749	1.110	0.937	8.112
Bahawalpur	ARIMA	515.9	528.9	0.002	0.318	2.743	8.512
	NNAR	640.54	654.01	0.002	0.390	0.023	1.009
	SVR	886.45	987.22	0.015	0.3684	0.2424	8.789
	DLM	3076.4	3089.8	0.331	0.362	0.231	9.324
Faisalabad	ARIMA	560.8	580.8	0.224	1.153	0.798	8.432
	NNAR	409.31	422.79	0.004	0.326	0.199	1.196
	SVR	713.7	842.3	0.009	0.264	0.992	5.663
	DLM	3108.7	3122.2	0.001	0.408	0.242	7.621
Jhelum	ARIMA	672.4	713.9	0.557	1.167	0.904	4.198
	NNAR	596.6	619.1	0.000	0.376	0.245	1.0265
	SVR	564.25	680.83	0.003	0.379	0.248	2.326
	DLM	3041.1	3054.6	0.012	0.387	0.241	9.876
Khanpur	ARIMA	896.9	907.6	0.287	0.904	1.189	7.944
	NNAR	575.34	597.8	-0.002	0.371	0.2038	1.234
	SVR	413.1	523.9	-0.231	0.361	0.209	2.039
	DLM	3006.8	3020.2	0.005	0.378	0.1908	9.987
Lahore	ARIMA	629.1	633.5	0.599	1.0334	0.852	5.234
	NNAR	214.27	223.25	0.012	0.281	0.168	3.661
	SVR	1143.9	1234.7	0.001	0.333	0.218	7.124
	DLM	884.7	897.5	0.004	0.344	0.203	8.427
Mainwali	ARIMA	144.5	149	-0.144	1.033	0.645	6.933
	NNAR	-54.08	-36.12	0.001	0.229	0.146	1.258
	SVR	998.6	1011.8	0.004	0.261	0.156	6.896
	DLM	559.6	573.1	-0.054	0.269	0.170	6.978
Multan	ARIMA	469.8	472.9	-0.144	0.813	0.645	9.723
	NNAR	370.94	388.9	0.001	0.318	0.234	3.897
	SVR	672.9	722.3	0.005	0.312	0.264	2.663
	DLM	790.67	804.14	0.007	0.320	0.998	3.992
Murree	ARIMA	234.5	345.9	0.945	1.276	0.789	1.745
	NNAR	136.59	172.51	0.001	0.318	0.234	2.998
	SVR	554.0	887.9	0.009	0.368	0.332	4.723
	DLM	790.6	804.1	0.0012	0.320	0.235	3.399

Table 4.1: Model selection among the competing model for Reconnaissance Drought Index (RDI) for all variable at each station using train data.

Station	Models	Model selection criteria		Model Performance Evaluation criteria			
		AIC	BIC	ME	RMSE	MPE	MAPE
Bahawalnagar	ARIMA	453.81	493.54	0.003	0.345	4.331	1.045
	NNAR	268.12	271.37	-0.001	0.228	0.008	5.612
	SVR	912.37	921.39	-0.613	0.532	0.001	8.101
	DLM	715.92	724.74	0.006	0.816	-5.81	8.112
Bahawalpur	ARIMA	855.2	877.3	-0.003	0.318	2.743	8.512
	NNAR	795.4	808.7	0.000	0.097	2.358	3.595
	SVR	831.9	840.8	-1.468	1.926	-4.35	4.256
	DLM	911.4	920.4	0.003	0.362	1.247	7.483
Faisalabad	ARIMA	629.0	633.4	0.002	0.404	0.998	8.432
	NNAR	375.4	388.7	-0.002	0.010	0.002	6.177
	SVR	818.2	831.4	-0.021	0.165	-6.877	1.226
	DLM	912.3	921.3	0.003	0.404	0.237	26.177
Jhelum	ARIMA	869.4	891.4	0.002	0.382	1.030	21.891
	NNAR	785.4	798.6	0.000	0.102	0.009	1.766
	SVR	884.2	893.2	0.609	0.707	0.609	5.271
	DLM	909.9	919.0	0.002	0.362	1.247	7.482
Khanpur	ARIMA	793.6	711.2	-0.012	0.359	0.204	93.860
	NNAR	706.0	719.2	0.000	0.233	0.138	3.756
	SVR	864.5	873.5	-0.3607	0.761	0.610	4.821
	DLM	1698.7	1706.6	0.005	0.327	0.196	9.838
Lahore	ARIMA	629.1	633.5	0.005	0.325	0.197	4.012
	NNAR	375.5	388.7	-0.0001	0.103	0.067	3.193
	SVR	818.2	831.5	-1.398	1.875	1.424	3.258
	DLM	912.3	921.3	0.005	0.327	0.196	9.838
Mainwali	ARIMA	306.8	311.2	-0.004	0.263	0.169	6.586
	NNAR	295.7	308.9	0.094	0.095	0.067	2.751
	SVR	889.7	898.7	0.186	0.411	0.294	3.864
	DLM	698.1	706.9	0.005	0.327	0.196	9.838
Multan	ARIMA	653.23	670.9	-0.003	0.321	0.236	4.369
	NNAR	545.0	558.2	-0.0006	0.095	0.069	2.261
	SVR	881.9	890.9	-0.758	0.939	0.760	6.262
	DLM	698.1	706.9	0.005	0.327	0.196	9.838
Murree	ARIMA	653.2	670.8	-0.0003	0.321	0.236	9.738
	NNAR	545.027	558.272	-0.0002	0.089	0.066	2.264
	SVR	1308.6	1316.9	-0.758	0.939	0.760	6.262
	DLM	1698.4	1706.7	0.005	0.327	0.196	9.838

Table 4.2: Model selection among the competing model for Reconnaissance Drought Index (RDI) for all variable at each station using test data.

Station	Models	Model selection criteria		Model Performance Evaluation criteria			
		AIC	BIC	ME	RMSE	MPE	MAPE
Bahawalnagar(Training)	NNAR(23,9)	456.3	472.8	0.009	0.567	0.378	3.987
	NNAR(16,6)	542.88	672.9	2.998	0.736	0.019	10.98
	NNAR(13,8)	779.8	812.8	1.914	0.456	0.459	8.194
	NNAR(24,14)	968.9	1090.6	0.184	0.993	2.561	4.790
Bahawalnagar (Testing)	NNAR(23,9)	2098.9	2109.6	0.987	0.8234	2.743	8.512
	NNAR(16,6)	1098.7	1116.9	1.789	1.876	3.987	9.845
	NNAR(13,8)	918.0	920.9	2.789	1.205	3.427	7.093
	NNAR(24,14)	2918.6	3032.1	1.372	0.236	2.743	8.256

Table 4.3: Goodness-of fit measure for different model of NNAR using train and test data

in Table 4.5. As a result, the model projects that would rise and fall in these stations, the drought predicted to be equivalent to -0.634 to 1.365 for Bahawalnagar and same forecasting procedure for all other stations. Positive forecasting data correspond to data when the model forecasts increased precipitation or an excess of water. The region is predicted to have above-average precipitation and high amount of water availability than the historical norm, according to favorable data. Negative forecasting values represent time when the model anticipate less precipitation or a shortage of water. These low level indicates that the region will likely see below-average precipitation, which would cause water scarcity and the possibility of droughts. The point forecast values for all the stations are in Table 4.4-4.12.

4.7 Graphical representation of NNAR

In the graph black line represent the observed, red line represent the predicted and blue for forecasted values for all stations. We take one station Bahawalnagar with NNAR (26,14) model with last 26 dataset $(y_{(t-1)}, y_{(t-2)}, , y_{(t-26)})$ used as input anticipate output y_t with 14 neuron in hidden layers. The forecast value in the graph Bahawalnagar the minimum value is -0.968 and maximum value is 1.365 respectively. The term "Nonlinear Neural Autoregressive" or "NNAR," designates an autoregressive model with nonlinear functions, which is what the forecasting model is. The forecasted variable's past values are used by autoregressive models to project its future values.

Months	Point Forecast
Jan-18	-0.634
Feb-18	-0.381
Mar-18	-0.606
Apr-18	-0.660
May18	-0.946
Jun-18	-0.860
Jul-18	-0.968
Aug-18	-0.753
Sep-18	-0.686
Oct-18	-0.745
Nov-18	-0.598
Dec-18	-0.406
Jan-19	-0.206
Feb-19	-0.186
Mar-19	0.115
Apr-19	0.310
May19	0.465
Jun-19	0.572
Jul-19	0.684
Aug-19	0.820
Sep-19	0.971
Oct-19	1.013
Nov-19	1.264
Dec-19	1.365

Table 4.4: Forecasted value from NNAR for Bahawalnagar from January, 2018-January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	0.282
Feb-18	0.507
Mar-18	0.275
Apr-18	-0.004
May18	0.061
Jun-18	0.589
Jul-18	0.409
Aug-18	0.688
Sep-18	0.907
Oct-18	-0.036
Nov-18	0.451
Dec-18	0.406
Jan-19	0.031
Feb-19	-0.0967
Mar-19	0.121
Apr-19	0.23
May19	0.298
Jun-19	-0.054
Jul-19	0.194
Aug-19	-0.056
Sep-19	-0.216
Oct-19	0.087
Nov-19	-0.369
Dec-19	-0.038

Table 4.5: Forecasted value from NNAR for Bahawalpur from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	0.283
Feb-18	0.507
Mar-18	0.277
Apr-18	-0.004
May18	0.062
Jun-18	0.589
Jul-18	0.409
Aug-18	0.688
Sep-18	0.907
Oct-18	-0.036
Nov-18	0.451
Dec-18	0.406
Jan-19	0.031
Feb-19	-0.0967
Mar-19	0.121
Apr-19	0.235
May19	0.298
Jun-19	-0.054
Jul-19	0.194
Aug-19	-0.059
Sep-19	-0.216
Oct-19	0.087
Nov-19	-0.369
Dec-19	-0.0383

Table 4.6: Forecasted value from NNAR for Faisalabad from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	1.496
Feb-18	1.433
Mar-18	1.146
Apr-18	1.115
May18	1.097
Jun-18	1.043
Jul-18	1.132
Aug-18	0.756
Sep-18	1.178
Oct-18	0.205
Nov-18	0.351
Dec-18	0.371
Jan-19	0.443
Feb-19	0.509
Mar-19	0.519
Apr-19	0.517
May19	0.509
Jun-19	0.595
Jul-19	0.431
Aug-19	0.680
Sep-19	0.250
Oct-19	0.537
Nov-19	0.368
Dec-19	0.289

Table 4.7: Forecasted value from NNAR for Jhelum from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	1.249
Feb-18	1.354
Mar-18	1.373
Apr-18	1.313
May-18	1.277
Jun-18	1.185
Jul-18	0.863
Aug-18	1.001
Sep-18	0.845
Oct-18	0.618
Nov-18	0.765
Dec-18	0.8984
Jan-19	1.267
Feb-19	1.567
Mar-19	1.768
Apr-19	2.031
May-19	2.003
Jun-19	2.083
Jul-19	1.916
Aug-19	1.935
Sep-19	1.627
Oct-19	1.6234
Nov-19	1.463
Dec-19	1.439

Table 4.8: Forecasted value from NNAR for Khanpur from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	0.342
Feb-18	0.332
Mar-18	0.342
Apr-18	0.177
May-18	0.083
Jun-18	0.007
Jul-18	0.279
Aug-18	-0.030
Sep-18	0.419
Oct-18	0.772
Nov-18	0.990
Dec-18	0.839
Jan-19	0.702
Feb-19	0.773
Mar-19	0.637
Apr-19	0.758
May-19	0.714
Jun-19	0.710
Jul-19	0.672
Aug-19	0.826
Sep-19	0.609
Oct-19	0.373
Nov-19	0.286
Dec-19	0.492

Table 4.9: Forecasted value from NNAR for Lahore from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	0.342
Feb-18	0.332
Mar-18	0.342
Apr-18	0.177
May-18	0.083
Jun-18	0.007
Jul-18	0.279
Aug-18	-0.030
Sep-18	0.419
Oct-18	0.772
Nov-18	0.990
Dec-18	0.839
Jan-19	0.701
Feb-19	0.773
Mar-19	0.637
Apr-19	0.758
May-19	0.711
Jun-19	0.717
Jul-19	0.672
Aug-19	0.826
Sep-19	0.601
Oct-19	0.377
Nov-19	0.286
Dec-19	0.492

Table 4.10: Forecasted value from NNAR for Mianwali from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	-0.619
Feb-18	-0.427
Mar-18	-0.642
Apr-18	-0.679
May-18	-0.931
Jun-18	-1.018
Jul-18	-1.109
Aug-18	-0.930
Sep-18	-0.889
Oct-18	-1.072
Nov-18	-1.043
Dec-18	-0.944
Jan-19	-0.830
Feb-19	-0.880
Mar-19	-0.607
Apr-19	-0.623
May-19	-0.609
Jun-19	-0.495
Jul-19	-0.361
Aug-19	-0.185
Sep-19	-0.229
Oct-19	-0.359
Nov-19	-0.340
Dec-19	-0.500

Table 4.11: Forecasted value from NNAR for Multan from January, 2018 to January, 2019 as a blue line in graph.

Months	Point Forecast
Jan-18	-0.617
Feb-18	-0.441
Mar-18	-0.599
Apr-18	-0.659
May-18	-0.990
Jun-18	-1.039
Jul-18	-1.180
Aug-18	-0.922
Sep-18	-0.860
Oct-18	-0.916
Nov-18	-0.872
Dec-18	-0.682
Jan-19	-0.565
Feb-19	-0.509
Mar-19	-0.215
Apr-19	-0.033
May-19	0.023
Jun-19	0.048
Jul-19	0.122585
Aug-19	0.155
Sep-19	0.136
Oct-19	0.159
Nov-19	0.349
Dec-19	0.299

Table 4.12: Forecasted value from NNAR for Murree from January, 2018 to January, 2019 shown in blue line in graph.

The (13,7): (p, k) is the parameters of NNAR model are indicating by numerical in brackets: p: The total number of inputs that are non-seasonal lags. Since p in this instance is 13, the model will use 13 previous observations as input to forecast the following values in the time series 4.1. k: The quality of nodes in the neural network's hidden layer. In this instance, k is 7, indicating that there are 7 neurons in the hidden layers. NNAR (13,7) denotes that the forecasting model for Bahawalnagar is an autoregressive model with a neural network topology with 7 neurons in the hidden layer and use 13 previous data as input. It takes advantages of these features to forecast future values in time series with nonlinear functions, making the model well suited to identifying intricate correlations and patterns in the data. NNAR (25,13) denotes that the forecasting model for Bahawalpur is an autoregressive model with a NNAR topology with 13 neurons in the hidden layer and use 25 previous data as input 4.2. NNAR (25,13) denotes that the forecasting model for Faisalabad is an autoregressive model with a neural network topology with 13 neurons in the hidden layer and use 25 previous data as input 4.3. NNAR (26,14) denotes that the forecasting model for Jhelum is an autoregressive model with a neural network topology with 14 neurons in the hidden layer and use 26 previous data as input 4.4. NNAR (13,7) denotes that the forecasting model for Khanpur is an autoregressive model with a neural network topology with 7 neurons in the hidden layer and use 13 previous data as input 4.5. NNAR (25,13) denotes that the forecasting model for Lahore is an autoregressive model with a neural network topology with 13 neurons in the hidden layer and use 25 previous data as input 4.6. NNAR (25,13) denotes that the forecasting model for Mainwali is an autoregressive model with a neural network topology with 13 neurons in the hidden layer and use 25 previous data as input 4.7. NNAR (26,14) denotes that the forecasting model for Multan is an autoregressive model with a neural network topology with 13 neurons in the hidden layer and use 25 previous data as input 4.8. NNAR (26,14) denotes that the forecasting model for Murree is an autoregressive model with a neural network topology with 13 neurons in the hidden layer and use 25 previous data as input 4.7.

Forecasts from NNAR(13,7)

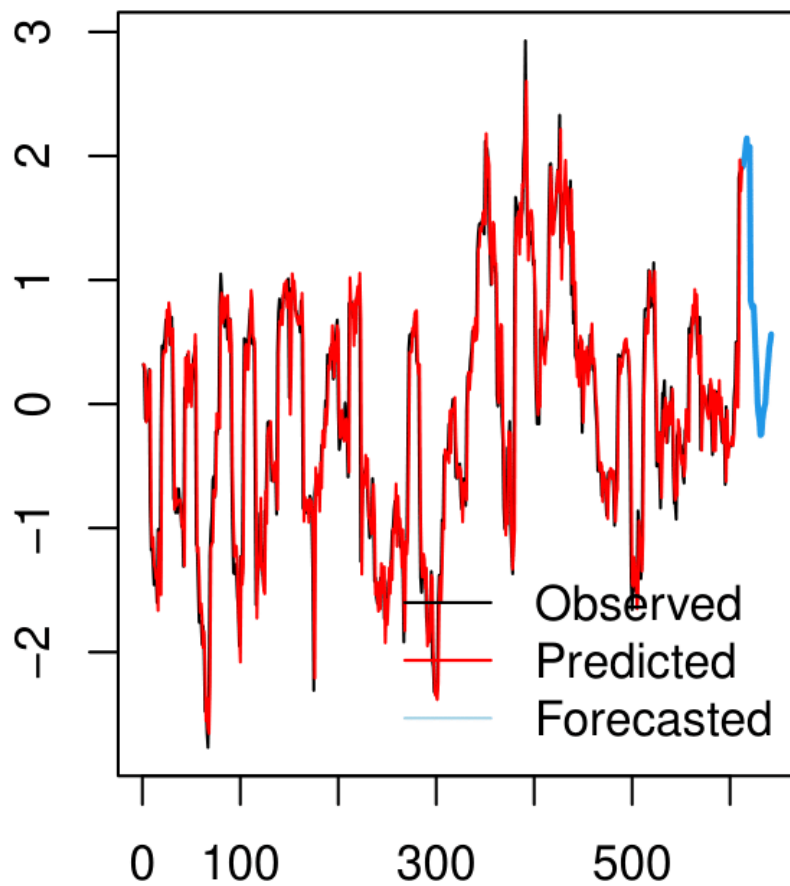


Figure 4.1: Comparison between forecast, observed and predicted for the time period (1962-2019) at Bahawalnagar.

Forecasts from NNAR(25,13)

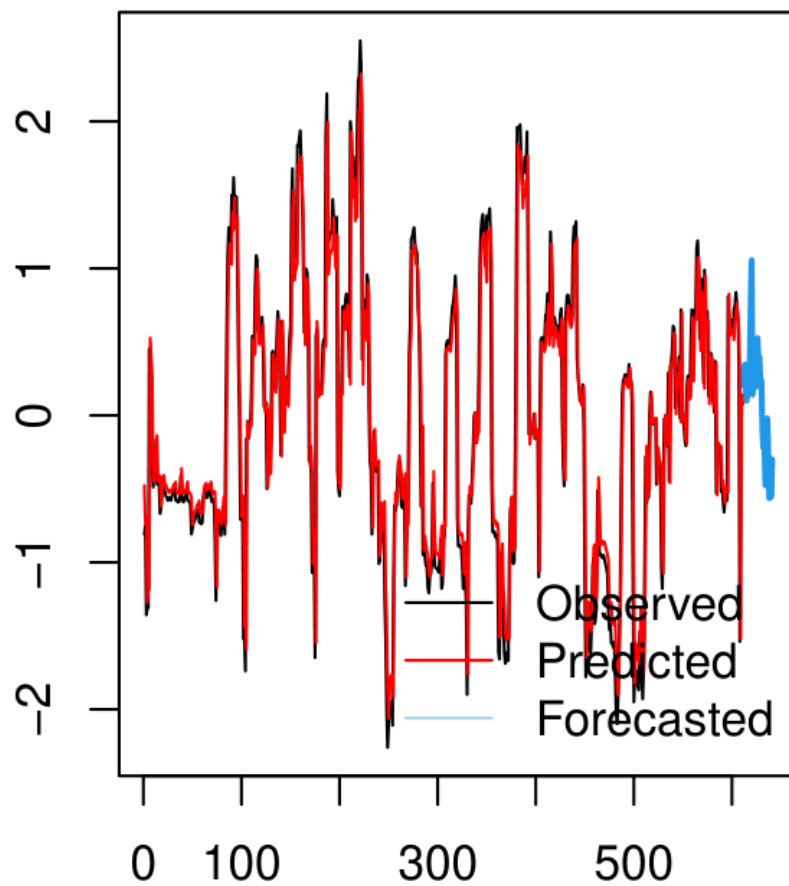


Figure 4.2: Comparison between forecast, observed and predicted for the time period (1962-2019) at Bahawalpur.

Forecasts from NNAR(25,13)

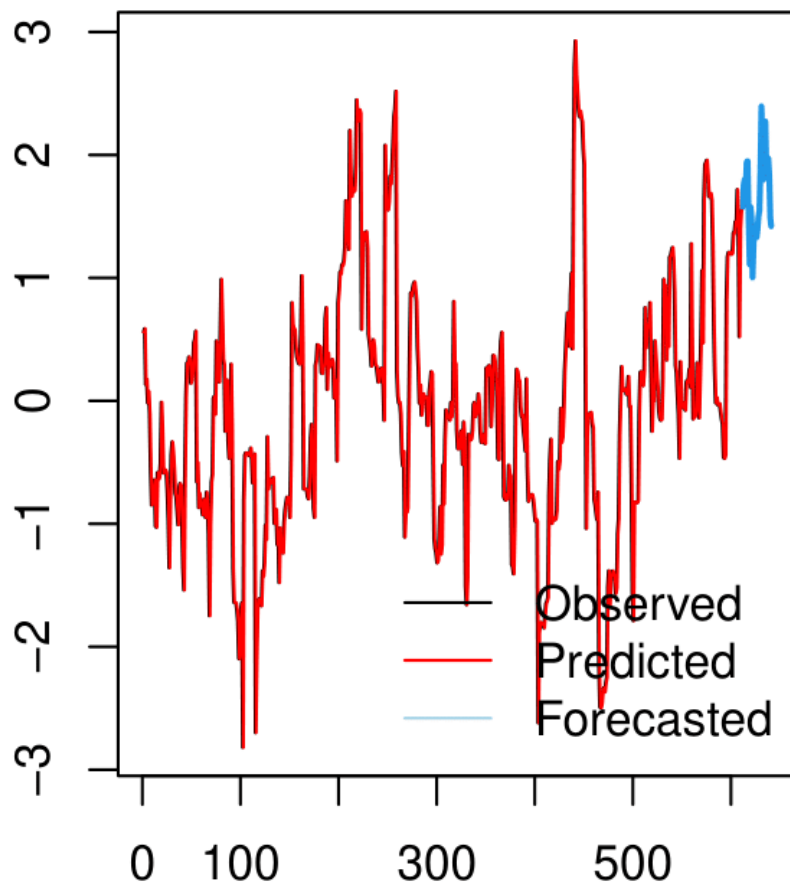


Figure 4.3: Comparison between forecast, observed and predicted for the time period (1962-2019) at Faisalabad.

Forecasts from NNAR(26,14)

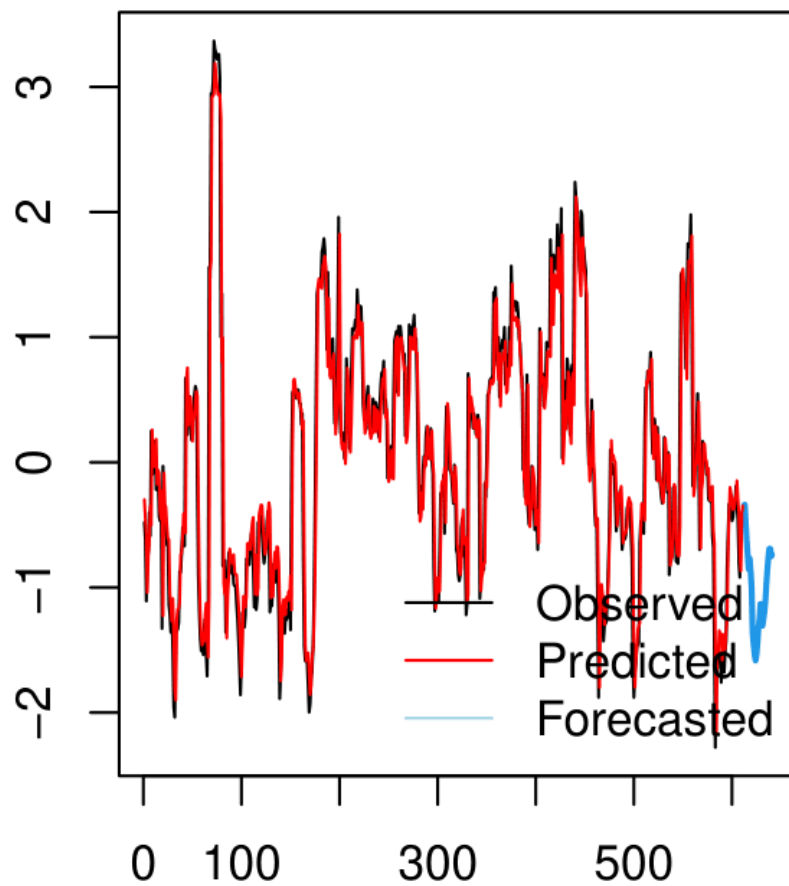


Figure 4.4: Comparison between forecast, observed and predicted for the time period (1962-2019) at Jhelum.

Forecasts from NNAR(13,7)

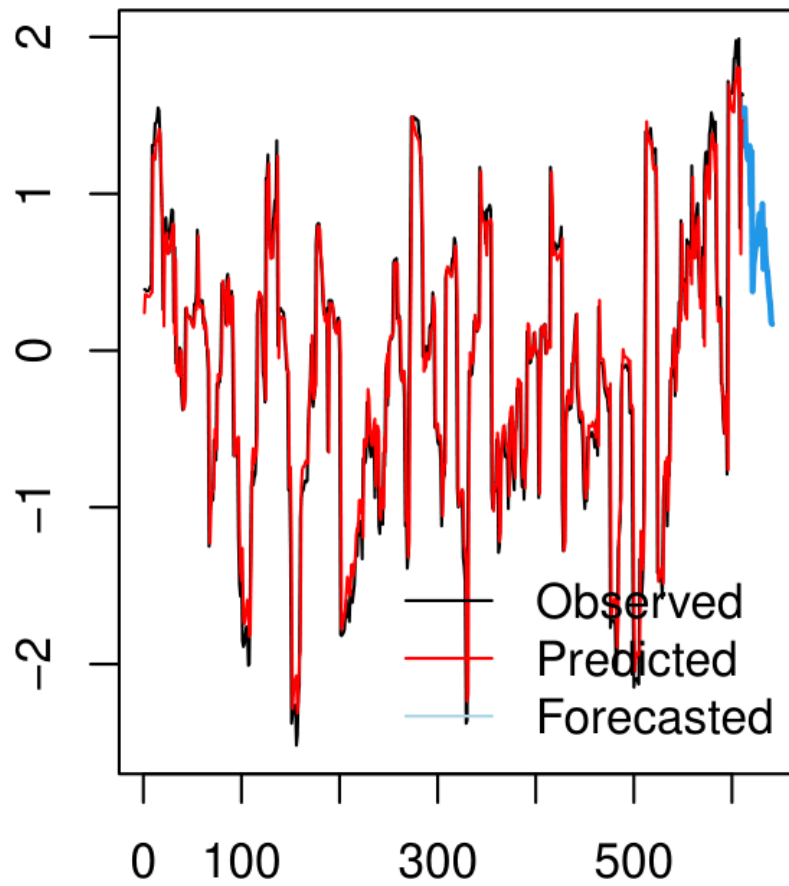


Figure 4.5: Comparison between forecast, observed and predicted for the time period (1962-2019) at Khanpur

Forecasts from NNAR(25,13)

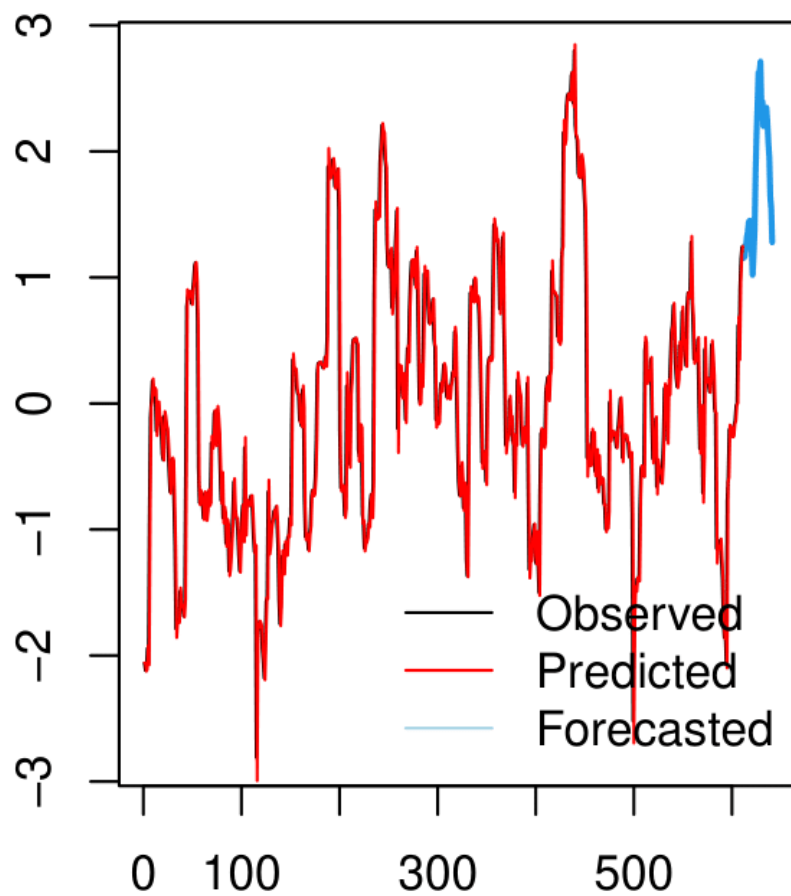


Figure 4.6: Comparison between forecast, observed and predicted for the time period (1962-2019) at Lahore.

Forecasts from NNAR(25,13)

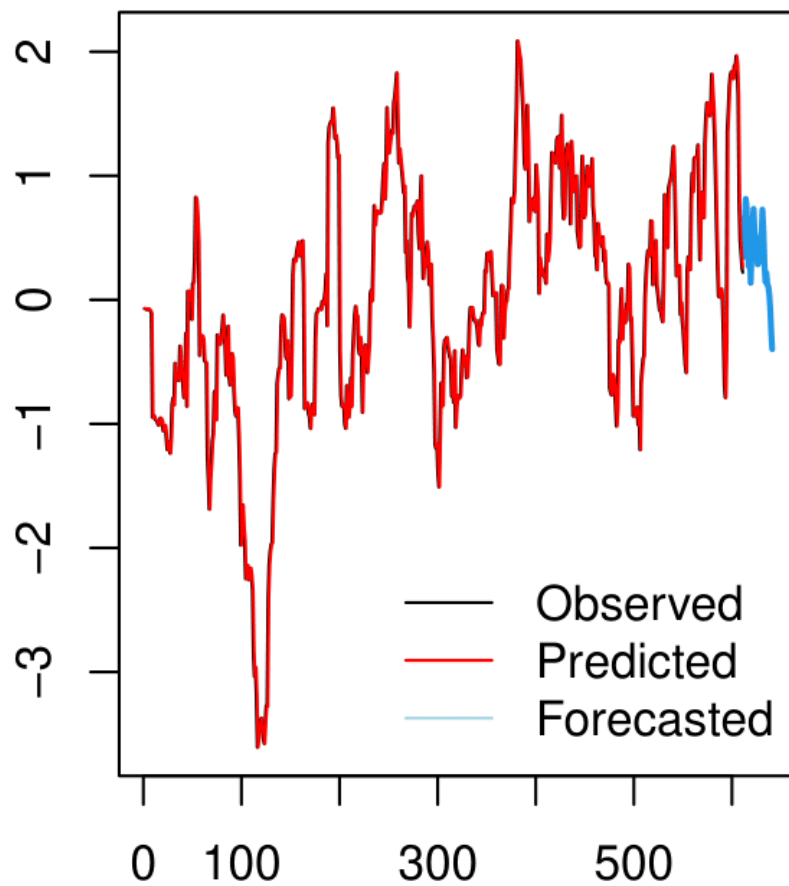


Figure 4.7: Comparison between forecast, observed and predicted for the time period (1962-2019) at Mianwali.

Forecasts from NNAR(26,14)

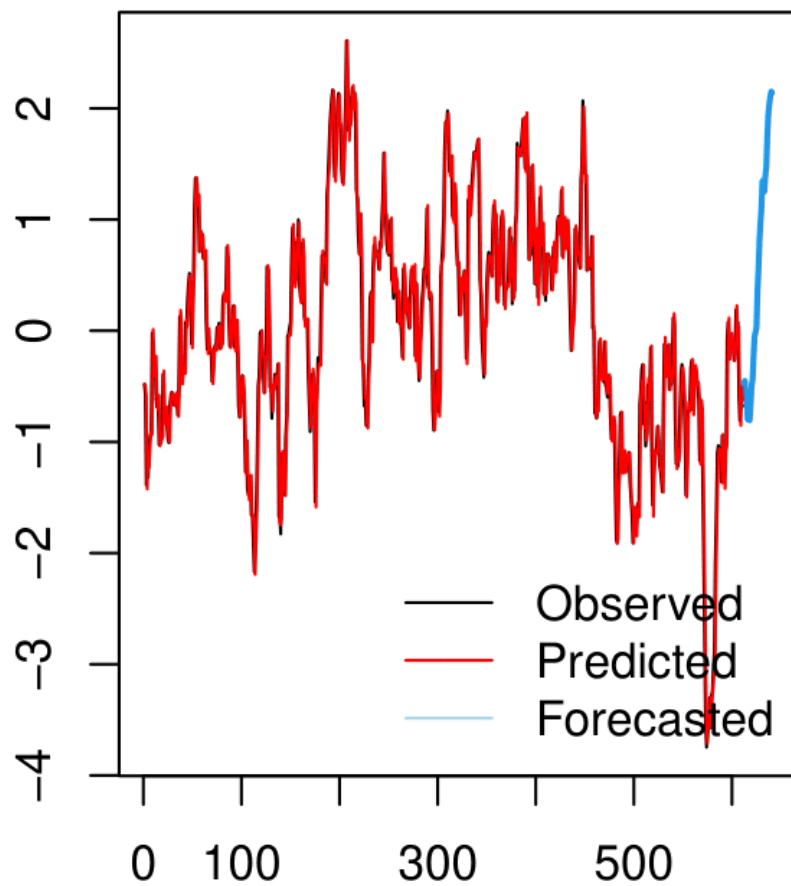


Figure 4.8: Comparison between forecast, observed and predicted for the time period (1962-2019) at Multan.

Forecasts from NNAR(26,14)

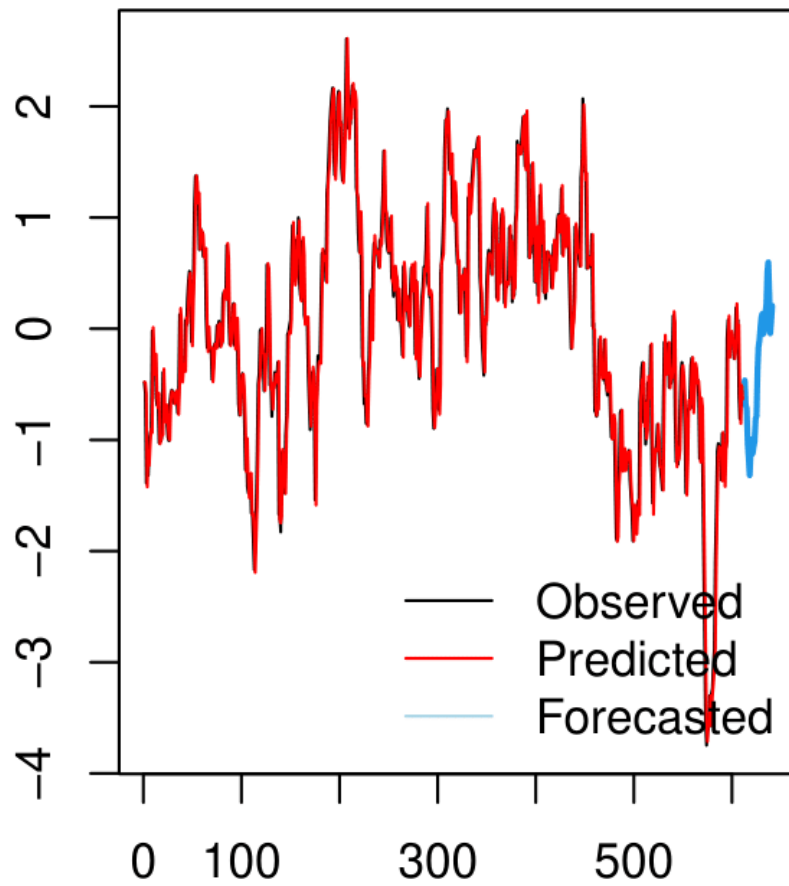


Figure 4.9: Comparison between forecast, observed and predicted for the time period (1962-2019) at Murree.

Chapter 5

Conclusions

In fact, more accurate estimates are essential to improve the multi-step forecast required for the most immediate best management practices. The (12-24 month) drought index is proven to be easier to predict with shorter lead time using ARIMA model. Although this linear technique falls short of perfectly capturing the non-linearity component of time series, machinelearning models are rapidly taking its place. In recent studies on drought forecasting machine learning models such as NNAR and SVR have certainly been utilized extensively since they can predict drought circumstances for which there isn't a strong mathematical solution. It has been demonstrated that machine learning models are capable of detecting non-stationarity and the data that are not linear in time series. According to research, modeling with machine learning models can accurately predict drought indices (RDI).

The study attempted to predict a short-term meteorological drought situation for Punjab, which is located in central-east Pakistan. Reconnaissance Drought Index (RDI), a drought measure has been predicted using ARIMA, NNAR, SVR and DLM. These methods are developed for nine meteorological station across Punjab, including Bahawalnagar, Bahawalpur, Lahore, Jhelum, Faisalabad, Mainwali, Khanpur, Murree and Multan for future forecasting.

This study compared various time-series models using drought data collected in Pakistan from January, 1962 to December, 2017 as training and testing data set. Training set, we calculated the models and on the training set, we evaluated them. For the

ARIMA, NNAR, SVR and DLM models, we computed the RMSE, ME, MPE and MAPE. The NNAR models outperforms the other competing models in term of predicting since its out-of-sample RMSE, MPE were the lowest among all other models. Currently, the NNAR models outperforms the other models in terms of forecast, evaluation and quality. According to the comparison criteria, the NNAR model provides the most appropriate outcome. The NNAR model is the best model for estimating the drought condition in Punjab region since it will give management insight into how to prevent the rise in droughts.

Bibliography

- [1] Anne F Van Loon and Henny AJ Van Lanen. Making the distinction between water scarcity and drought using an observation-modeling framework. *Water Resources Research*, 49(3):1483–1502, 2013.
- [2] Ashok K Mishra and Vijay P Singh. A review of drought concepts. *Journal of hydrology*, 391(1-2):202–216, 2010.
- [3] PG Gore and KC Sinha Ray. Variability in drought incidence over districts of maharashtra. *Mausam*, 53(4):533–538, 2002.
- [4] Stanley A Changnon. Shifting economic impacts from weather extremes in the united states: A result of societal changes, not global warming. *Natural Hazards*, 29(2):273–290, 2003.
- [5] Oleg Smirnov, Minghua Zhang, Tingyin Xiao, John Orbell, Amy Lobben, and Josef Gordon. The relative importance of climate change and population growth for exposure to future extreme droughts. *Climatic Change*, 138:41–53, 2016.
- [6] MG Rabbani, A Atiq Rahman, Nazria Islam, D Michel, and A Pandya. Climate change and sea level rise: issues and challenges for coastal communities in the indian ocean region. *Coastal zones and climate change*, pages 17–29, 2010.
- [7] Mia Reinhard, M Rebetez, and R Schlaepfer. Recent climate change: rethinking drought in the context of forest fire research in ticino, south of switzerland. *Theoretical and Applied Climatology*, 82:17–25, 2005.

- [8] Dorte Verner, David Treguer, John Redwood, Jens Christensen, Rachael McDonnell, Christine Elbert, Yasuo Konishi, and Saad Belghazi. Climate variability, drought, and drought management in morocco’s agricultural sector. 2018.
- [9] David R Archer, Nathan Forsythe, Hayley J Fowler, and Syed Muhammed Shah. Sustainability of water resources management in the indus basin under changing climatic and socio economic conditions. *Hydrology and Earth System Sciences*, 14(8):1669–1680, 2010.
- [10] Caily Schwartz, W Lee Ellenburg, Vikalp Mishra, Timothy Mayer, Robert Griffin, Faisal Qamer, Mir Matin, and Tsegaye Tadesse. A statistical evaluation of earth-observation-based composite drought indices for a localized assessment of agricultural drought in pakistan. *International Journal of Applied Earth Observation and Geoinformation*, 106:102646, 2022.
- [11] G Tsakiris, I Nalbantis, D Pangalou, D Tigkas, and H Vangelis. Drought meteorological monitoring network design for the reconnaissance drought index (rdi). In *Proceedings of the 1st International Conference on Drought management: scientific and technological innovations*. Zaragoza, Spain: option Méditerranéennes, series A, volume 80, pages 57–62, 2008.
- [12] Arash Asadi and F Vahdat. The efficiency of meteorological drought indices for drought monitoring and evaluating in kohgilouye and boyerahmad province, iran. *Int. J. Mod. Eng. Res*, 3(4):2407–2411, 2013.
- [13] Dimitris Tigkas, Harris Vangelis, and George Tsakiris. Drinc: a software for drought analysis based on drought indices. *Earth Science Informatics*, 8:697–709, 2015.
- [14] Dehe Xu, Qi Zhang, Yan Ding, and Huiping Huang. Application of a hybrid arima–svr model based on the spi for the forecast of drought—a case study in henan province, china. *Journal of applied meteorology and climatology*, 59(7):1239–1259, 2020.

- [15] KF Fung, YF Huang, CH Koo, and YW Soh. Drought forecasting: A review of modelling approaches 2007–2017. *Journal of Water and Climate Change*, 11(3):771–799, 2020.
- [16] Mukesh K Tiwari, Ki-Young Song, Chandranath Chatterjee, and Madan M Gupta. River-flow forecasting using higher-order neural networks. *Journal of Hydrologic Engineering*, 17(5):655–666, 2012.
- [17] AK Mishra and VR Desai. Drought forecasting using stochastic models. *Stochastic environmental research and risk assessment*, 19:326–339, 2005.
- [18] A Belayneh, J Adamowski, B Khalil, and BJJOH Ozga-Zielinski. Long-term spi drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508:418–429, 2014.
- [19] Ruben Thoplan. Simple v/s sophisticated methods of forecasting for mauritius monthly tourist arrival data. *International Journal of Statistics and Applications*, 4(5):217–223, 2014.
- [20] Bahram Choubin, Shahram Khalighi-Sigaroodi, Arash Malekian, and Özgür Kişi. Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological sciences journal*, 61(6):1001–1009, 2016.
- [21] Soohyeon Kim, Jihyo Kim, and Eunnyeong Heo. Speculative incentives to hoard aluminum: Relationship between capital gains and inventories. *Resources Policy*, 70:101901, 2021.
- [22] Mohamad Kassem, Amjad Ali, and Marc Audi. Unemployment rate, population density and crime rate in punjab (pakistan): an empirical analysis. *Bulletin of Business and Economics (BBE)*, 8(2):92–104, 2019.

- [23] Syeda Maria Ali, Bushra Khalid, Asma Akhter, Aneeza Islam, and Shahzada Adnan. Analyzing the occurrence of floods and droughts in connection with climate change in punjab province, pakistan. *Natural Hazards*, 103:2533–2559, 2020.
- [24] Robert J Hijmans, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15):1965–1978, 2005.
- [25] Nausheen Mazhar, Safdar Ali Shirazi, Lindsay C Stringer, Rachael H Carrie, and Martin Dallimer. Spatial patterns in the adaptive capacity of dryland agricultural households in south punjab, pakistan. *Journal of Arid Environments*, 194:104610, 2021.
- [26] John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the operational research society*, 20(4):451–468, 1969.
- [27] H Vangelis, D Tigkas, and G Tsakiris. The effect of pet method on reconnaissance drought index (rdi) calculation. *Journal of Arid Environments*, 88:130–140, 2013.
- [28] G Tsakiris, Dialecti Pangalou, and H Vangelis. Regional drought assessment based on the reconnaissance drought index (rdi). *Water resources management*, 21:821–833, 2007.
- [29] Mohammed Sanusi Shiru, Shamsuddin Shahid, Noraliani Alias, and Eun-Sung Chung. Trend analysis of droughts during crop growing seasons of nigeria. *Sustainability*, 10(3):871, 2018.
- [30] Charles Warren Thornthwaite. An approach toward a rational classification of climate. *Geographical review*, 38(1):55–94, 1948.
- [31] Kamal Ahmed, Shamsuddin Shahid, and Nadeem Nawaz. Impacts of climate variability and change on seasonal drought characteristics of pakistan. *Atmospheric research*, 214:364–374, 2018.

- [32] George Box. Box and jenkins: time series analysis, forecasting and control. In *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*, pages 161–215. Springer, 2013.
- [33] Bahram Choubin and Arash Malekian. Combined gamma and m-test-based ann and arima models for groundwater fluctuation forecasting in semiarid regions. *Environmental Earth Sciences*, 76:1–10, 2017.
- [34] Taewook Lee. Wild bootstrap ljung–box test for residuals of arma models robust to variance change. *Journal of the Korean Statistical Society*, 51(4):1005–1020, 2022.
- [35] Kenneth J White. The durbin-watson test for autocorrelation in nonlinear models. *The Review of Economics and Statistics*, pages 370–373, 1992.
- [36] Timothy Masters. *Practical neural network recipes in C++*. Morgan Kaufmann, 1993.
- [37] Julian Faraway and Chris Chatfield. Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 47(2):231–250, 1998.
- [38] Gaetano Perone. Comparison of arima, ets, nnar and hybrid models to forecast the second wave of covid-19 hospitalizations in italy. *arXiv preprint arXiv:2010.11617*, 2020.
- [39] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- [40] Jian Zhou, Yingui Qiu, Shuangli Zhu, Danial Jahed Armaghani, Chuanqi Li, Hoang Nguyen, and Saffet Yagiz. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting tbm advance rate. *Engineering Applications of Artificial Intelligence*, 97:104015, 2021.

- [41] Vladimir Cherkassky and Yunqian Ma. Practical selection of svm parameters and noise estimation for svm regression. *Neural networks*, 17(1):113–126, 2004.
- [42] Francesco Parrella. Online support vector regression. *Master’s Thesis, Department of Information Science, University of Genoa, Italy*, 69, 2007.
- [43] Ozgur Kisi and Mesut Cimen. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*, 399(1-2):132–140, 2011.
- [44] Mike West and Jeff Harrison. Introduction to the dlm: The first-order polynomial model. *Bayesian Forecasting and Dynamic Models*, pages 32–67, 1997.
- [45] Eric Zivot and Jiahui Wang. *Modeling financial time series with S-PLUS*, volume 2. Springer, 2006.
- [46] Firdos Khan, Shaukat Ali, Alia Saeed, Ramesh Kumar, and Abdul Wali Khan. Forecasting daily new infections, deaths and recovery cases due to covid-19 in pakistan by using bayesian dynamic linear models. *Plos one*, 16(6):e0253367, 2021.
- [47] Shawn P Serbin, Aditya Singh, Brenden E McNeil, Clayton C Kingdon, and Philip A Townsend. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications*, 24(7):1651–1669, 2014.
- [48] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.
- [49] Joseph Plasmans, William Verkooijen, and Hennie Daniels. Estimating structural exchange rate models by artificial neural networks. *Applied Financial Economics*, 8(5):541–551, 1998.

- [50] Afshin Maleki, Simin Nasseri, Mehri Solaimany Aminabad, and Mahdi Hadi. Comparison of arima and nnar models for forecasting water treatment plant's influent characteristics. *KSCE Journal of Civil Engineering*, 22:3233–3245, 2018.
- [51] Le Jian, Yun Zhao, Yi-Ping Zhu, Mei-Bian Zhang, and Dean Bertolatti. An application of arima model to predict submicron particle concentrations from meteorological factors at a busy roadside in hangzhou, china. *Science of the Total Environment*, 426:336–345, 2012.
- [52] Chuanjin Yu, Yongle Li, Huoyue Xiang, and Mingjin Zhang. Data mining-assisted short-term wind speed forecasting by wavelet packet decomposition and elman neural network. *Journal of Wind Engineering and Industrial Aerodynamics*, 175:136–143, 2018.
- [53] Howard Demuth, Mark Beale, and Martin Hagan. Neural network toolbox. *For Use with MATLAB. The MathWorks Inc*, 2000, 1992.
- [54] Roberto Battiti, Nhat Thang Le, and Alessandro Villani. Location-aware computing: a neural network model for determining location in wireless lans. 2002.
- [55] Marta Fraile and Michael S Lewis-Beck. Economic and elections in spain (1982–2008): Cross-measures, cross-time. *Electoral Studies*, 31(3):485–490, 2012.
- [56] Jan Adamowski and Christina Karapataki. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ann learning algorithms. *Journal of Hydrologic Engineering*, 15(10):729–743, 2010.

Appendix