UNIVERSITY OF ICELAND

Magister Scientiarum thesis
in Computational Engineering

Time series methods for improving weather model temperature forecasts

Íris Ósk Hilmarsdóttir

Advisors: Kristján Jónasson and Guðrún Nína Petersen

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science

TIME SERIES METHODS FOR IMPROVING WEATHER MODEL TEMPERATURE FORECASTS

Íris Ósk Hilmarsdóttir

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Computational Engineering

Advisors
Kristján Jónasson
Guðrún Nína Petersen

Faculty Representative
Ólafur Pétur Pálsson

Faculty of Industrial Engineering, Mechanical Engineering
and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavík, May 2023

Time series methods for improving weather model temperature forecasts
60 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Computational Engineering

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
107, Reykjavik, Reykjavik
Iceland

Sími: 525 4000

Bibliographic information:
Íris Ósk Hilmarsdóttir, 2023, Time series methods for improving weather model temperature forecasts, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland.

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavík, Iceland, May 2023

Abstract

The aim of this study is to investigate the current post-processing scheme used to forecast 2-m temperature at the Icelandic Meteorological Office (IMO) and the possibility to improve its performance. The study involves a process of computing and evaluating various time series models in order to find one with minimal prediction error. This study was based on data for the period 2019–2021, which consisted of observations for 2-m temperature and 10-m wind, as well as forecasts for 2-m temperature, relative humidity, 10-m wind, cloud cover, and precipitation used as exogenous variables in the model development. Statistical correction for 24 hour forecast models at noon were computed for 16 selected weather stations in Iceland, both simple correction schemes and ARMA-type models such as AR, MA, ARMA, and ARMAX. The results suggest that it is possible to improve the current post-processing scheme that is used by the IMO for some weather stations, especially those located in fjords and valleys.

Útdráttur

Markmið þessarar rannsóknar er að rannsaka aðferðina sem Veðurstofa Íslands notar nú til að leiðrétta 2-m hitaspá frá veðurlíkani og möguleikann á því að bæta frammistöðu aðferðarinnar. Ýmis tímaraðalíkön eru reiknuð og metin með það að markmiði að lágmarka spáskekkju. Rannsóknin byggir á gögnum frá tímabilinu 2019–2021, sem samanstanda af mælingum af 2-m hita og 10-m vindi, og ennfremur spám fyrir 2-m hita, rakastig, 10-m vind, skýjahulu og úrkomu, notað sem ytri breytur í líkanaþróuninni. Tölfræðileg leiðrétting á 24 tíma spá fyrir 2-m hita kl 12 á hádegi var reiknuð fyrir 16 valdar veðurstöðvar með bæði einföldum leiðréttingarlíkönum og ARMA líkönum eins og t.d. AR, MA, ARMA og ARMAX. Niðurstöður rannsóknar gefa til kynna að hægt sé að bæta núverandi leiðréttingaraðferð sem notuð er á Veðurstofu Íslands fyrir sumar veðurstöðvar, sérstaklega þær sem eru staðsettar í fjörðum og dölum.

Contents

List of Figures

List of Tables

Abbreviations

AR	Auto regressive
ARIMA	Auto regressive integrated moving average
ARMA	Auto regressive moving average
ARMAX	Auto regressive moving average with exogenous variable/s
B	Backshift operator
DMI	Danish meteorological institute
ECMWF	European centre for medium range weather forecasts
F	Wind (m/s)
I	Integrated
IGB	Iceland-Greenland Beta
IMO	Icelandic meteorological of ce
MA	Moving average
MAE	Mean absolute error
N	Cloud cover (%)
NWP	Numerical weather prediction
R	Precipitation (mm/h)

LIST OF TABLES

RH	Relative humidity (%)
RMSE	Root mean squared error
T	Temperature (C)
X	Exogenous

Acknowledgments

The author would like to express her gratitude towards her thesis advisors, Kristján Jónasson and Guðrún Nína Petersen for their continuous guidance and support. Acknowledgements are also due to the Icelandic Meteorological Of ce for sharing data and assistance.

Finally, the author would like to express her sincere gratitude and appreciation to her family for their endless support.

1. Introduction

Operational numerical weather predictions (NWP) have a large impact on society. They are important for a wide range of industries, such as agriculture, energy, transportation and emergency responders as well as the general public. Accurate weather predictions enable individuals and organizations to make informed decisions and take appropriate precautions based on expected weather conditions.

The Icelandic meteorological of ce (IMO) currently uses an operational numerical weather prediction model called HARMONIE-AROME. The model is developed and maintained by many European countries and has been in use in Iceland since the autumn of 2011. The Danish meteorological institute (DMI) runs the Harmonie model for Iceland jointly with the IMO, its surrounding waters and Greenland. The model runs are referred to as IGB.

The IGB model has proven to give good results within the Icelandic forecast domain. Developing accurate numerical weather predictions is approximating the processes in the atmosphere and its boundaries. As most NWP models, IGB has known biases. Post-processing is a method to remove biases [1].

Post-processing numerical weather predictions has been done by the IMO for a long time. The post-processing process started with simple Kalman ltes which later developed into adaptive Kalman Filters. Today, the IMO uses a method developed by Bolli Pálmason which is based on a recursive statistical method, using European centre for medium range weather forecasts (ECMWF) 2-m temperatures and ground station measurements for the post-processing of the IGB model which is referred to as the IMO model in this thesis [2].

The aim of this thesis is to investigate the current post-processing scheme used to forecast 2-m temperature at the IMO and the possibility to improve its performance. The model development process starts with exploring current IMO model and then computing simple correction schemes which indicate what to focus on in further development. Following, more advanced time series models are computed which include different variants of autoregressive moving average models, with and without exogenous variables.

1.1. Overview

After this introduction the thesis is divided into four main chapters.

Background provides a foundation of knowledge on topics that are further discussed and studied in this thesis.

The weather prediction model and data used in this study covers the weather prediction model in this study, its data, the data preprocessing and further model development procedure.

Results presents the results and comparison of post-processing models obtained by model development and study in this thesis.

Discussion and conclusion summarizes the results given by this study, draws main conclusions and suggests possible improvements to the models and future work.

2. Background

The purpose of this chapter is to provide foundation of knowledge on topics that will be further discussed and studied in this thesis. Section 2.1 covers the foundations of time series and Section 2.2 covers time series forecasting. Section 2.3 introduces Box-Jenkins ARIMA models which are studied in this thesis. In Section 2.4, a rolling window method used for evaluation is covered. Forecast bias is introduced in Section 2.5 and the steps of data preprocessing are covered in Section 2.6. Finally, Section 2.7 introduces the statistical methods used in this study.

2.1. Time series

Time series data refers to a sequence of observations in ordered time. These observations are often collected at equally spaced discrete time intervals. Time series are often studied in order to discover a historical pattern that can be exploited in the making of a forecast. To discover this historical pattern, there are several features that are looked closer into, as time series models have several attributes such as trend, seasonal variation, cyclic changes, auto correlation and irregular factors [3].

Trend refers to the long-term change in the mean level of the time series and is essentially an underlying growth or decline component. Seasonal variation or seasonality refers to the periodic uctuation within each year of the time series. Weather time series, for example, usually have a seasonal component. Cyclic changes are similar to seasonal components, as they also have a wavelike pattern but the period is different from a whole year. Such changes can sometimes be explained as a variation across a xed period due to some physical cause that is not seasonal. If the period is not xed, the changes can be explained with so-called autocorrelation.

When all these factors have been accounted for (e.g. with a mathematical model), the remaining variation can be attributed to irregular uctuations and the resulting difference is a series of residuals. If these residuals are not random, this indicates that there are still some patterns that need to be taken into account [3].

It is essential to study all these properties of time series data when using the series for forecasting. When successive time series values are dependent, future values can be predicted using information on patterns.

2.2. Time series forecasting

Forecasting is an important component in various elds of science. This study is concerned with time series forecasting, which has become a popular eld of research in recent years. Time series forecasting consists of predicting future values based on knowledge of the past [4]. There are a variety of different time series forecasting methods available and numerous different methods have been widely studied in recent years. It is important to note that no single procedure is universally applicable [3].

2.3. Box-Jenkins ARIMA models

In 1970, Box and Jenkins described an approach to forecast time series based on linear stochastic processes, called autoregressive integrated moving average or ARIMA models in their fundamental textbook, *Time Series Analysis: Forecasting and Control* [5]. While their method is commonly known as an ARIMA model, many refer to it as the Box-Jenkins methodology. In this study, the method will be referred to as ARIMA models.

In its pure form, ARIMA models consists of three iterative steps of model identi cation, parameter estimation and diagnostic checking. The model identi cation step involves identifying the process that is generating and in uencing the historical pattern. The parameter estimation involves choosing the optimal parameters for the ARIMA model that satisfy certain criteria. Finally, the diagnostic step is an iterative process to guarantee the adequacy of the identi ed model in order to select the optimal tted model [6].

ARIMA models consist of three components, and each of the components helps to model a certain type of pattern. “The “AR“ or autoregressive component attempts to account for the patterns between any one time period and previous periods. The “MA” or moving average component measures the adaptation of new forecasts to prior forecast errors. The “I” or integrated component connotes a trend or other “integrative” process in the data [7].“

All of these components symbolise an ARIMA model which is often denoted as ARIMA(p,d,q) where “p” accounts for the order of the AR term, “q” accounts for the MA term, and “d” accounts for the I term and represents the number of differences given in the series to make it stationary.

There are different variants of ARIMA models, as the models do not always have values for all parameters. When models only consists of autoregressive parameters, the models can be referred to as AR(p) models and models that have only moving average parameters present can be referred to as MA(q) models. Models that are stationary and do not need any differencing, can be referred to as ARMA(p,q) models, as they do not utilize the I term. In this study, the I term is not utilized. When using external variables in the models, the term X is appended (ARIMAX, ARX, MAX, etc.).

A model (p,q) for stationary times series can be defined by the following equation:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (2.1)$$

where the ϕ's represent autoregressive parameters, the θ's represent moving average parameters, the Y's are the measured series and the ϵ's are series of unknown random errors which are assumed to follow a normal distribution.

Models can be described with a backshift operator for easier expression. The backshift operator, B, symbolises changing time period t to time period $t-1$. Using backshift notation, equation 2.1 can be expressed as:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) Y_t = (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t \quad (2.2)$$

Using the backshift operator an ARIMA model can be mathematically expressed as:

$$\phi_p(B) Y_t = \theta_q(B)\epsilon_t \quad (2.3)$$

Which can also be written:

$$(1 - B)^d Y_t = \frac{\theta(B)}{\phi(B)}\epsilon_t \quad (2.4)$$

where the d represents the order of differencing.

External variables can be added to ARMA models and are referred to as exogenous variables. Exogenous variables are used as a weighted input in the model. When using exogenous variables in ARMA models it creates an ARMAX model which simply adds a term on the right hand side:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} + \beta x_t \qquad (2.5)$$

where x_t is a covariate at time t and β is its coefficient. Several exogenous terms can be added to the equation.

2.4. A rolling window forecast

In this thesis, a rolling window forecast will be used for model development. A rolling window model involves fitting a model on a fixed connecting block of previous observations and using that to forecast. To use rolling window forecast, the fixed window size is chosen and the forecast horizon. Then for each rolling window subsample, a model is fitted, forecast, and evaluated. Figure 2.1 demonstrates the process of rolling window forecast used in the thesis. Rolling forecasts are adaptive as the models are modified in each window. They may be more useful on time series problems where recent lag values are more predictive than older lag values. When the forecast horizon is one period, it is called a one-step-ahead forecast and in this study, the one-step-ahead rolling window forecast will be used.

Figure 2.1: A schematic of the process of the rolling window forecast used in the study.

2.5. Forecast bias

When models have a tendency to either systematically over-predict or under-predict, it can be described as forecast bias. The bias in forecasting models can be caused by a range of different factors. In order to overcome biases in models, there are many bias correction methods that can be utilized. In this thesis, different bias correction methods will be analyzed.

2.6. Data preprocessing

For the model development in this master's thesis, the steps of data preprocessing covered below will be followed. Data preprocessing is an essential step in the model development to enhance data ef ciency. It is one of the most important steps which deals with data preparation and transformation of the dataset. The preprocessing process includes several techniques and steps which will be covered below.

Data reduction: This step includes selecting a part of the data that the researcher wants to explore further. This step is useful as it gives the researcher a better overview of the data and it becomes more manageable to work with.

Data integration: This step is used to merge data from different and multiple sources into a single larger dataset that will be used further in the study. Data is often derived from multiple sources and therefore this step is necessary in many studies.

Data transformation: For this step, the data is transformed to meet the researcher's goal. That could include things like structuring the data, changing the format, attribute selection, and aggregation.

Data cleansing: In order to reach the best results, the dataset needs to be clean to ensure suf cient data quality. If data cleansing is not performed, results from studies could be inaccurate, biased or irrelevant. Data cleansing techniques include identifying and possibly lling in missing data, reducing noisy data, and identifying and removing duplicates and identi ed erroneous values.

2.7. Statistical measures

The performance of the statistical learning method on a certain dataset is usually evaluated by measuring how well its predictions match the observed data. There are various statistical measures that can be used to evaluate different forecasting methods and they can all be useful as they evaluate different parts of the method results. In the following sections, the measures used in this study will be described.

RMSE: The root mean squared error (RMSE) is frequently used in measuring model performance. It is essentially the standard deviation of the prediction errors. RMSE is commonly used in forecasting, climatology and regression analysis to evaluate models [8]. RMSE is given by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \tag{2.6}$$

where $\hat{y}_i$ denotes the predicted value for the i_{th} observation, the y_i denotes the observed value for the i_{th} observation and n denotes the total number of observations.

MAE: The mean absolute error (MAE) measures the mean absolute value of the difference between the forecast value and the observed value. The formula is:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2.7}$$

where $\hat{y}_i$ denotes the predicted value for the i_{th} observation, y_i denotes the observed value for the i_{th} observation and n denotes the total number of observations.

Correlation: Correlation is a statistical measure that expresses the connection existing between independent variables and how strongly they are linearly related.

Multicollinearity: Multicollinearity is a statistical expression that exists when an independent variable is strongly correlated to one or more of the other variables in the dataset. This is to be avoided as multicollinearity can weaken the statistical signi cance of an independent variable and will therefore result in less valid statistical speculation.

To evaluate the performance of the forecasting models in this master thesis, MAE will be the primary measure. MAE is a standard performance metric that is resistant to outliers as it gives linear scores that averages all individual differences with equal weight. MAE is also less sensitive to outliers than RMSE. However, all of the measures mentioned above are used in the study in some way.

3. The weather prediction model and data used in this study

This chapter covers the numerical weather prediction model (NWP) used in this study, its data, the data preprocessing and further model development procedure. In Section 3.1, the history, development and description of the Harmonie IGB model is covered. In Section 3.2, the data used in this study is studied and presented. Section 3.3 covers the important steps of data preprocessing and all of its process. Finally, Section 3.4 covers data exploration and visualization in order to give readers knowledge on the data in an easily presentable way.

3.1. The Harmonie IGB model

The operational numerical weather prediction model currently used by the Icelandic Meterological Office (IMO). It has been in use in Iceland since the autumn of 2011. The model is developed and maintained jointly by many European countries. It is originally based of the AROME model developed by Meteo France.

The Harmonie domain run for Iceland has a horizontal grid-point spacing of 2.5 km in both directions with 65 vertical fields and it is run at 6 hour intervals for the full 66 hour forecast. The Danish meteorological institute (DMI) runs the Harmonie model for Iceland jointly with the IMO, its surrounding waters and Greenland which is referred to as Harmonie-Iceland Greenland Beta, hereafter referred to as IGB. In this thesis, the IGB model is studied.

The Harmonie model has proven to give good results within the Icelandic forecast domain. However, there are studies that show that there exists bias, that is removed by post-processing the forecasts [1]. Post-processing for numerical weather prediction output is an ongoing active research field. The IMO has been a participant in that field for many years as they have conducted research on various dynamic weather forecasting models. The challenge with NWP outputs is that they are calculated in a grid, some coarse and therefore there is the tendency for bias, which can occur due to factors such as height errors or other errors due to an incomplete description in the models. Height errors occur when the model's grid points have different elevation than weather stations, leading to a discrepancy in the predicted values. Additionally, models most often have an incomplete description of the atmosphere, making reliable forecasts difficult to obtain.

To overcome these challenges, the IMO has used post-processing techniques to eliminate bias. At the IMO, the process of post-processing started in 1997 with Knútur Árnason using Kalman filtering, which is a popular post-processing method. Later in 2006, Philippe Crochet improved upon this method by using adaptive Kalman filtering [9]. In the years 2014-2016, Bolli Pálmason developed a method based on a recursive statistical method using ECMWF 2-m temperatures forecasts and station measurements [2]. The method will be covered again in Section 4.1.

3.2. The data used in this study

The data for model development in this master thesis, was obtained from the IMO; it consists of two different datasets, one containing weather forecasts and the other containing weather observations, both for about 300 different weather stations across Iceland.

The dataset IGB, consists of weather forecasts made by the IMO for 320 different weather stations. The dataset used here has a date range from August 2018 to March 2022 and it has nine different variables. An overview of the dataset can be seen in Table 3.1.

The observation dataset, consists of historical weather observations from the IMO database for 289 different weather stations. The dataset consists of four different variables which can be seen in Table 3.2, along with a description.

In Table 3.3, the comparison of the two different datasets used in this master thesis can be seen. The most noticable difference between the two is the size of them, as the IGB forecast dataset has 104,217,811 data points while the observation dataset only has 9,927,346. That can simply be explained as forecasts are initialized four times a day, with output every hour and forecasting range of 66 hours while the observation dataset only contains a single observation for each hour for each station.

Table 3.1: IGB forecast dataset variable information.

Variable	Description
Weather station	Weather station numerical identi er
Time	Date and time of forecast
D_{kl}	Forecast period
T	2-m temperature (C)
RH	2-m relative humidity (%)
F	10-m wind (m/s)
N	Cloud cover (%)
R	Precipitation (mm/h)

Table 3.2: Observation dataset variable information.

Variable	Description
Weather station	Weather station numerical identi er
Time	Date and time of observation
T_{obs}	2-m temperature (C)
F_{obs}	10-m wind (m/s)

Table 3.3: Datasets used for model development.

	IGB forecast dataset	Observation dataset
Description	Weather forecasts	Weather observation
Date range	23.08.2018 - 20.03.2022	01.01.2018 - 20.03.2022
Weather stations	320	289
Data points	104,217,811	9,927,346

3.3. Data preprocessing

For the important procedure of data preprocessing in this master thesis, the steps covered in Section 2.6 will be followed. As stated there, data preprocessing is an essential process in every study to ensure data quality. Without the steps of data preprocessing, the data might be unreliable, meaning the results are not suf cient and can not be utilized and concluded from.

Data reduction

Data ltering was performed on two variables. D_{kl} was ltered at 24 hour forecasts, as that is the forecast period the study will be based on.

$$D_{kl} = 24 \tag{3.1}$$

The time variable was ltered at noon or at 12:00.

$$Time = 12{:}00 \tag{3.2}$$

Therefore, the study will be based on models that have a 24 hour forecast period and are creating a forecast at noon for noon the following day. The forecast period and time were chosen in consultation with the IMO.

Data integration

For the development of the time series model in this study, the IGB and the observation datasets previously mentioned, were combined into a larger dataset which was used as the ltered dataset for the model development process. The datasets were joined by weather stations and time. After joining the two datasets, there were 275 common weather stations, which was explored further.

Data transformation

A new variable was added to the filtered dataset,

$$x_t = T_{observation} - T_{forecast} \tag{3.3}$$

measuring the difference between the observation 2-m temperature and the forecast 2-m temperature for each time t, i.e. the forecast temperature error.

This variable is used as the dependent or the endogenous variable in the model development as the study is interested in developing post-processing models that correct single forecasts based on past bias of the uncorrected forecasts, thus decreasing the forecast error.

Data cleansing

Following the previously discussed steps, data exploration and cleansing were performed on the filtered dataset. The data exploration was performed with the purpose of assessing the quality of the data, and then data cleansing followed, with the purpose of preparing the dataset for further model development.

First, there were a few unrealistic or extreme values present in the dataset which would have affected the results if they were not taken care of. These values were replaced with missing values.

A program in R was written to gather important information for each weather station such as the time range and periods with missing values. Using this program, a table was constructed which gave a good overview of the data quality present at each station.

The information in the table for each weather station consists of:

- Start date, i.e. the first day of observation and forecast
- End date, i.e. the last day of observation and forecast
- Number of days, i.e. the number of days between start and end date
- Fraction missing, i.e. missing days divided by number of days

- Number of days in the longest missing streak, i.e. the largest data gap
- Fraction missing besides the longest streak, i.e. how much is missing that is not part of the longest missing streak

The fraction missing besides the longest streak (M) is calculated as:

$$M = \frac{(\text{Number of days} \cdot \text{Fraction Missing}) - \text{Number of days in the longest missing streak}}{\text{Number of days} - \text{Number of days in the longest missing streak}} \quad (3.4)$$

Table A.1 in the appendix demonstrates the full information for each weather station in the dataset.

In order to maintain data quality, weather stations which did not meet the criteria:

$$M < 2\% \quad (3.5)$$

were excluded.

Overall, the data quality was good as the stations did not have a lot of missing values and most of them had none. However, a total of 8 stations did not meet this criteria, leaving 267 stations of sufficient data quality.

For the stations that still had some missing values but passed the criteria and are therefore a part of the filtered dataset, linear interpolation was used to fill in missing values. Linear interpolation is a technique used to fill in missing values by utilizing the known values of adjacent data points.

Finishing this procedure, data cleansing had been completed and data quality guaranteed, meaning that the dataset could be used further in the model process.

Filtered dataset for model development procedure

An overview of the filtered dataset is shown in Table 3.4, and the variable information is in Table 3.5.

Table 3.4: Overview of the filtered dataset.

Description	Dataset used for model development containing both observations and forecasts
Date range	23.08.2018 - 20.03.2022
Weather stations	267
Data points	322,838 (number of days times number of stations)
Forecast length	24
Forecast hour	12 UTC

Table 3.5: Filtered dataset variable information.

Variable	Description
Weather station	Weather station numerical identifier
Time	Date and time
D_{kl}	Forecast period
T	Forecast 2-m temperature (°C)
RH	Forecast 2-m relative humidity (%)
F	Forecast 10-m wind (m/s)
N	Forecast cloud cover (%)
R	Forecast precipitation (mm/h)
T_{obs}	Observed 2-m temperature (°C)
F_{obs}	Observed 10-m wind (m/s)
x_t	Error (T_{obs} - T) (°C)

3.4. Data exploration and visualization

There are various visualization techniques used in modern systems, that provide readers with information to get familiar with data. These techniques can identify numerous things such as patterns, correlations, causalities, and outliers.

Below are several figures and tables for exploring the dataset used in model development, which will provide key insights and information.

In Table 3.6, the correlation matrix of the residuals of an AR(1) model and exogenous variables is computed. This is done to see which variables have the largest correlation to the residuals in order to choose which variables to use as exogenous variables in future model development. As the table shows, there is not an obvious correlation to

Table 3.6: Correlation matrix of residuals and exogenous variables.

	Residuals	T (C)	RH (%)	F (m/s)	N (%)	R (mm/h)
1. Residuals	1					
2. Temperature - T (C)	–0.141	1				
3. Relative humidity - RH (%)	0.127	–0.204	1			
4. Wind - F (m/s)	–0.018	–0.145	0.084	1		
5. Cloud cover - N (%)	–0.018	0.034	0.372	0.095	1	
6. Precipitation - R (mm/h)	–0.040	–0.020	0.303	0.202	0.144	1

any variable. It is also apparent that none of the variables are noticeably correlated to each other. This suggests that there is no multicollinearity present in the dataset which is preferable.

Figure 3.1 shows the distribution of average 2-m temperature for weather stations in the dataset. The lowest average temperature is observed for mountain station 4275, Gagnheiði, while the highest average temperature is observed for an island station 6012, Surtsey, south of Iceland.

Figure 3.1: The distribution of average observed 2-m temperature (°C) for the weather stations in the dataset during the time range 23.08.2018 – 20.03.2022.

(a) The distribution of highest 2-m temperature.

(b) The distribution of lowest 2-m temperature.

Figure 3.2: The distribution of highest and lowest observed 2-m temperature (°C) for weather stations in the dataset during the time range 23.08.2018 – 20.03.2022.

Figure 3.2 shows the distribution of the highest and lowest 2-m temperatures recorded across different stations. The highest 2-m temperature of 27.1C was observed at station 5965, Þórudalur (elevation: 300 m) located in east Iceland, while the lowest 2-m temperature of -26.2C was recorded at station 4300, Mývatn (elevation: 282 m), situated in east Iceland as well.

Next in Figure 3.3, we consider the distribution of mean absolute error (MAE) of 2-m temperature for all stations in the dataset, essentially the difference between the observed temperature and the temperature forecast by the IMO (see Equation 3.3). This figure is presented in order to see the range of MAE in the IGB model for different stations. The figure shows that the majority of stations have a MAE of around 1.5 C while there are some stations that are above that. The average MAE for all the stations is 1.53 C.

Figure 3.3: The distribution of mean absolute error of 2-m temperature (°C) for weather stations in the dataset during the time range 23.08.2018 – 20.03.2022.

Table 3.7 shows the weather stations with the highest mean absolute error of 2-m temperature computed by the IGB model. Most of these stations are located in fjords. This suggests that the model has a harder time accurately forecasting weather predictions for stations located in these areas, i.e. complicated orography.

Table 3.7: Weather stations in the ltered dataset with the highest mean absolute error of IGB 2-m temperature (°C) forecast.

Station	Name	MAE
4182	Seyðisfjörður	4.69
32635	Botn í Súganda rði	4.35
3463	Möðruvellir	4.34
4180	Seyðisfjörður - Vestdalur	4.20
33661	Ólafsfjarðarvegur	3.82

On the other hand, Table 3.8 shows the weather stations with the lowest MAE of 2-m temperature. Most of these stations are on peninsulas or on islands, which suggests that the model has a good ability to accurately forecast temperature in these areas.

Table 3.8: Weather stations in the ltered dataset with the lowest mean absolute error of IGB 2-m temperature (°C) forecast.

Station	Name	MAE
6012	Surtsey	0.65
6015	Vestmannaeyjabær	0.73
4867	Fontur	0.80
31640	Reykjanesviti	0.83
1350	Ke avíkur ugvöllur	0.85

The mean absolute error of the IGB forecast 2-m temperature (°C) was also computed for different months to see if certain months were harder to accurately predict. In Figure 3.4, March–April have the highest MAE while November–December have the lowest. However the MAE for other months are similar.

Figure 3.4: Mean absolute error of 2-m temperature (°C) in the IGB model for different months in the dataset.

Table 3.9: Overview of all 16 weather stations studied in the thesis and their attributes.

Station	Name	Acronym	Latitude (°N)	Longitude (°W)	Height (m)	Category
1475	Reykjavík	REY	64.13	21.90	52	A
1590	Skálafell	SKÁ	64.24	21.46	771	C
2323	Tálknafjörður	TÁL	65.63	23.83	9	B
2642	Ísafjörður	ÍSA	66.06	23.17	2	B
3471	Akureyri	AKU	65.70	18.11	31	A
3752	Siglufjörður	SIG	66.13	18.92	6	B
4275	Gagnheiði	GAG	65.22	14.26	949	B
6935	Hveravellir	HVE	64.87	19.56	641	A
2050	Stykkishólmur	STY	65.07	22.73	12	A
3317	Blönduós	BLÖ	65.66	20.29	8	A/B
4867	Fontur	FON	66.38	14.53	43	B
5309	Fagurhólsmýri	FAG	63.87	16.64	9	B
5872	Teigarhorn	TEI	64.68	14.34	21	A
6017	Stórhöfði	STÓ	63.40	20.29	118	A
6515	Hjarðarland	HJA	64.25	20.33	88	A
6802	Húsafell	HÚS	64.70	20.87	133	B

3.4.1. Data selection

The weather stations studied in Chapter 4 are chosen in two steps.

The subset consists of 16 different stations which were chosen in consultation with the IMO. Initially, there were eight station chosen and later in the model development process eight more stations were added. The weather stations are distributed across Iceland and have a range of different attributes. The IMO categorizes the stations, based on level of service which reflects in the data quality of each station. Therefore, most stations were chosen in category A and B. The stations are listed in Table 3.9. The filtered dataset used for model development consists of years 2019–2021.

4. Results

The purpose of this chapter is to present the results given by the model development. Comparison of different model performances is presented, both simple correction schemes and more advanced ARMA models. In Section 4.1, the rst steps of the model development process are presented. Three simple correction schemes models initially studied and computed are described along with their statistical results. Following, Section 4.2 covers further study of the time series models, where different variants of ARMA models are studied and computed. Section 4.3, presents results from all the models studied and computed for the rst eight weather stations. In Section 4.4, additional weather stations are added in order to further validate previous results. Finally, Section 4.5 covers overall comparison of MAE and RMSE for all models and all stations studied in this thesis.

4.1. Simple correction models

This section covers the rst steps of the model development and two simple models that were studied, the method of Bolli Pálmason mentioned in Section 3.1 and a method that uses average past bias to correct the temperature prediction.

Raw error

First, the raw error of the IGB 2-m temperature forecast was computed in order to have a baseline for further model development.

$$x_t = T_{observation} \quad T_{forecast} \tag{4.1}$$

with $T_{forecast}$ being the numerical weather prediction output (NWP output).

IMO correction model

In order to correct for the 2-m temperature bias in the IGB model, the IMO has been using the model,

$$L_t = 0.9L_{t-1} + 0.1x_{t-1} \tag{4.2}$$

where L_t is used for model correction

$$\hat{S}_t = S_t + L_t \tag{4.3}$$

This can also be written as an ARMA(1,1) model with coefficients θ = (1.0, -0.9)

$$x_t = 1.0x_{t-1} - 0.9\epsilon_{t-1} + \epsilon_t \tag{4.4}$$

This bias correction method is developed by Bolli Pálmason at the IMO [10]. It is based on Boi's statistical method for forecasting extreme daily temperatures using 2-m temperature forecasts and ground station measurements from the European centre for medium range weather forecasts (ECMWF) [2].

60 day bias correction model

One of the simplest ways to implement a bias correction model is to take the average bias for the last n days,

$$\bar{x}_t = \frac{x_{t-n} + x_{t-n+1} + \dots + x_{t-1}}{n} \tag{4.5}$$

and using that bias, $\bar{x}_t$, for model correction,

$$\hat{S}_t = S_t + \bar{x}_t \tag{4.6}$$

Here, n was chosen to be 60, given promising results. It is possible that other values might improve the results. In some ways this choice of n is arbitary. The model will also be referred to as 60 BC.

Model comparison

Mean absolute error (MAE) was used to compare the different simple correction schemes model performance for eight weather stations. MAE was chosen as the performance measure as it is more robust and less sensitive to outliers compared to root mean squared error. MAE gives equal weight to each error, resulting in a more even measure of the model t.

Table 4.1 shows the MAE of 2-m temperature forecasts (24 hours, valid at 12 UTC): raw NWP output, the current IMO model and the 60 day bias correction model. The NWP error varies a lot from station to station. Figures 4.1 and 4.2 show the monthly MAE for for the eight weather stations.

Table 4.1: Mean absolute error of 2-m temperature (°C) for simple models in the time range 2019–2021.

Station	NWP output	IMO model	60 day bias correction
1475 Reykjavík	0.96	0.95	0.96
1590 Skálafell	2.10	1.20	1.17
2323 Tálknafjörður	2.27	1.21	1.13
2642 Ísafjörður	3.10	1.52	1.38
3471 Akureyri	1.62	1.51	1.50
3752 Siglufjörður	3.56	1.88	1.81
4275 Gagnheiði	1.93	1.29	1.27
6935 Hveravellir	1.67	1.21	1.20

The plots seem to have a pattern of minima of MAE the most often in Sep–Oct. This is not necessarily surprising because as shown in Figure 3.4, the time period Sep–Oct has the second lowest MAE overall.

The results in Figures 4.1 and 4.2, show that each station has its own attributes. The Reykjavík station has a low bias and height of 52 meters above sea level. Its landscape is relatively well explained by IGB due to its simplicity. As the NWP output MAE is already low, it is hard to improve with statistical post-processing.

The Skálafell station is on a delimited mountain, 771 meters above sea level. The NWP output is volatile while the other two models show more constant results. In Figure 4.1, it is noticeable that there is a seasonal effect in the NWP output. The seasonal effect in the NWP output is also, but less noticeably, present in other stations.

The Tálknafjörður station is 8 meters above sea level and in a narrow fjord. The peaks in May–Jun 2019 for the two computed models are interesting and were explored further in Figure 4.3. The gure shows each day in that period. As can be seen in the gure, there are some days with higher MAE than in general impacting the average for the period.

The Ísafjörður station is located in a narrow fjord as well. The two computed models, are able to give a much lower MAE than the raw NWP output MAE.

The Akureyri station is another fjord station, and is located 31 meters above sea level. The results indicate similarities with the Reykjavík station, not having a complicated orography and therefore, having overall low MAE for all models.

The Siglufjörður station is 6 meters above sea level and in a fjord. The results for Siglufjörður show a very high raw NWP output MAE, suggesting that the model does not have a good predictability of the station due to complex landscape. The other two models have much lower MAE but still relatively high compared to other stations.

The Gagnheiði station is 949 meters above sea level and is a highland station in the interior of Iceland. The results give a lower MAE for the IMO and 60 day bias correction model.

The last station, Hveravellir, is located 641 meters above sea level and is located between large ice cops and has a harsh environment. The results are quite volatile.

The raw NWP output, from the IGB model has a high MAE for stations which are located in fjords and have complex orography. The post-processing models are able to improve the results and even though there are sudden changes in the raw bias, they are still able to catch the changes.

Overall, these results demonstrate that for most stations, the 60 day bias correction model gives the lowest MAE of 2-m temperature, hence the best results. It is interesting that the results have lower MAE than the IMO model, suggesting that the IMO model places too much weight on newer observations while not weighing the earlier observation enough, which the 60 day bias correction model essentially does. The parameters in the IMO model could be revised and this will be considered in the next section.

Figure 4.1: Mean absolute error of 2-m temperature (°C) for simple models in the time range 2019–2021 for Reykjavík, Skálafell, Tálknafjörður and Ísafjörður.

Figure 4.2: Mean absolute error of 2-m temperature (°C) for simple models in the time range 2019–2021 for Akureyri, Siglufjörður, Gagnheiði and Hveravellir.

Figure 4.3: Mean absolute error of 2-m temperature (°C) for two simple models for the Tálknafjörður station in the time range 1.5.2019–30.6.2019.

4.2. Autoregressive moving average models

This section covers further study of time series models. Autoregressive moving average (ARMA) methodology is studied further, with different variants of ARMA models.

Following the computation of the simple correction models in Chapter 4.1, the next step of the process was to experiment with more advanced ARMA models. The models are computed with the Python class statsmodels.tsa.arima.ARIMA which is also used to forecast. This class chooses the optimal parameters for each model based on minimizing root mean squared error. The one-step-ahead rolling window forecast discussed in Section 2.4 was used to test the models with 60 training days and one test day. Numerous models were computed in order to find the one with minimum prediction error. All combinations of models with p,q = [(1,0), (2,0), (0,1) (1,1)] were tested. Other values p,q = (0,2) and (2,2) were also tested, but did not deliver promising results. The models computed are discussed below.

Models with no exogenous variables

Four different models with no exogenous variables were computed, ARMA(1,1), AR(1), AR(2) and MA(1). These models were created with optimal parameters chosen by the Python class mentioned above.

Models with exogenous variables

An interesting aspect of the model development was to see if exogenous variables that had not been used in previous models, would improve the forecasting accuracy. Therefore, ARMAX(1;1), ARX(1), ARX(2) and MAX(1) models were computed for different combinations of exogenous variables, using the Python class. The exogenous variables are 2-m temperature, relative humidity, 10-m wind, cloud cover and precipitation, see Table 3.5.

4.3. Model comparison

This section compares the models of Section 4.1 and 4.2. Mean absolute error is used to evaluate the models.

Table 4.2 presents the MAE of 2-m temperature for all models computed in the study. The results are promising as the MAE is somewhat lower for many stations. Overall, the models which use exogenous variables give better results, suggesting the benefits of using exogenous variables in this weather post-processing model. The process started with computing different combinations of ARMAX models with two different combinations of exogenous variables, but as the results were promising, models with more combinations of exogenous variables were computed.

The exogenous variable temperature T has the highest correlation to the residuals in Table 3.6, so the first ARMAX model was created with only exogenous variable T. Following, a model with all the exogenous variables T, RH, F, N, R was also computed, returning encouraging results. Following these two promising ARMAX models, two additional ARMAX models were computed. The first of these had exogenous variables T and RH, as they are the two variables that have the highest correlation to the residuals in Table 3.6. This gave results that were overall better than just using T as an exogenous variable. However, the results were not as good as the ARMAX model using all of the exogenous variables. The final ARMAX model computed was with exogenous variables T, RH, R as it has the three variables with the highest correlation as shown in Table 3.6. This model delivered results which were not quite as promising as previous results.

In Table 4.2 the MAE for the ARMA$(1;1)$, is lower than than the MAE for the IMO model. The difference between the two is that the IMO model has fixed parameters for ARMA$(1;1)$ as (1.0,–0.9) whereas the ARMA$(1;1)$ model generated by Python has optimal parameters for each station and training period.

(a) AR parameters

(b) MA parameters

Figure 4.4: Parameters for ARMA(1,1) model computed by Python class statsmodels.tsa.arima.ARIMA.

It is interesting to take a closer look at which parameters are being chosen by the Python class. Figure 4.4 shows the parameters for the ARMA(1,1) models computed for the full time range and all stations. For the AR term, the parameter range –0.4 to 0.7 represents the middle 50%. The IMO model uses parameter 1.0 for the AR term which is outside this 50% interval. The 50% central interval for the MA term ranges from –0.3 to 0.2. The IMO model uses -0.9, which is also outside the 50% interval. Overall, the comparison of parameters between the two post-processing models is interesting as it suggests the possibility of improving the IMO model by updating its parameters.

Table 4.2: Mean absolute error of 2-m temperature (°C) in the time range 2019–2021 for all models computed in the study for the eight selected weather stations. See section 4.1 for de nition of the simple models, Section 4.2 for de nition of autoregressive integrated moving average models and Table 3.9 for weather station abbreviations.

		1475 REY	1590 SKÁ	2323 TÁL	2642 ÍSA	3471 AKU	3752 SIG	4275 GAG	6935 HVE	Average
Simple models	NWP	0.96	2.10	2.27	3.10	1.62	3.58	1.93	1.66	2.15
	IMO	0.95	1.20	1.21	1.52	1.51	1.88	1.29	1.21	1.35
	60 BC	0.96	1.17	1.13	1.38	1.50	1.81	1.27	1.20	1.30
ARMA	(1,1)	0.98	1.15	1.13	1.35	1.48	1.74	1.24	1.20	1.28
	(1,0)	0.97	1.14	1.11	1.34	1.48	1.72	1.24	1.19	1.27
	(0,1)	0.97	1.15	1.12	1.35	1.48	1.74	1.24	1.19	1.28
	(2,0)	0.98	1.15	1.12	1.35	1.49	1.73	1.24	1.19	1.28
ARMAX - T	(1,1)	0.99	1.13	1.09	1.32	1.49	1.73	1.24	1.22	1.28
	(1,0)	0.98	1.12	1.09	1.32	1.49	1.71	1.23	1.21	1.27
	(0,1)	0.98	1.12	1.10	1.32	1.49	1.72	1.24	1.22	1.27
	(2,0)	0.99	1.11	1.09	1.32	1.50	1.72	1.23	1.22	1.27
ARMAX - T,RH	(1,1)	0.96	1.06	1.09	1.29	1.50	1.70	1.18	1.21	1.25
	(1,0)	0.95	1.05	1.07	1.29	1.49	1.70	1.17	1.20	1.24
	(0,1)	0.95	1.05	1.08	1.29	1.49	1.71	1.18	1.21	1.24
	(2,0)	0.96	1.05	1.08	1.30	1.50	1.70	1.18	1.21	1.25
ARMAX - T,RH,R	(1,1)	0.97	1.07	1.10	1.31	1.54	1.71	1.19	1.22	1.26
	(1,0)	0.96	1.05	1.08	1.30	1.51	1.70	1.19	1.21	1.25
	(0,1)	0.96	1.05	1.09	1.29	1.52	1.71	1.19	1.21	1.25
	(2,0)	0.97	1.06	1.10	1.31	1.53	1.70	1.19	1.22	1.26
ARMAX - T,RH,F,N,R	(1,1)	0.99	1.07	1.10	1.29	1.53	1.63	1.18	1.24	1.25
	(1,0)	0.97	1.06	1.10	1.28	1.50	1.63	1.16	1.22	1.24
	(0,1)	0.97	1.05	1.10	1.28	1.51	1.63	1.16	1.22	1.24
	(2,0)	0.99	1.05	1.10	1.30	1.53	1.63	1.17	1.24	1.25

4.4. Adding weather stations

In order to further validate previous results, all models were tested on eight additional stations, thus 16 stations in total. This was done to see if these results gave similar results as earlier, confirming the previous outcomes. Table 4.3 shows MAE of 2-m temperature for all models computed for the eight new weather stations.

Overall the new stations that were added in this section, have a relatively low MAE for the IMO model, making it more challenging for other models to improve the forecasting accuracy. However there are some models which were computed that improve the overall performance of the IMO model for these eight stations. The 60 day bias correction model has a marginally lower MAE than the IMO model in this section as well, confirming again that the parameters could be looked closer into for the IMO model. The AR(1) model has the same MAE as the IMO model or lower for all stations. It is the only model that does not have a higher MAE than the IMO model for any station.

It is also interesting to study the results for different stations. As mentioned earlier they all have their different attributes and therefore their forecasting accuracy is different. Throughout this study, it is clear that the NWP model does not have a good forecasting accuracy for stations located in fjords. However, the post-processing models are able to improve the forecasting accuracy for these stations notably. Many of the first eight stations were located in such areas that the NWP model does not predict very well while the second eight stations studied, are located in better locations for the NWP model. For that reason, the second eight stations do not show as much improvement from other models as before because their forecasting accuracy was already good.

Table 4.3: Mean absolute error of 2-m temperature (°C) in the time range 2019–2021 for all models computed in the study for the second eight weather stations selected. See section 4.1 for de nition of the simple models, Section 4.2 for de nition of autoregressive integrated moving average models and Table 3.9 for weather station abbreviations.

		2050 STY	3317 BLÖ	4867 FON	5309 FAG	5872 TEI	6017 STÓ	6515 HJA	6802 HÚS	Average
Simple models	NWP	0.95	1.30	0.80	1.18	1.31	1.10	1.31	1.96	1.24
	IMO	0.93	1.30	0.70	1.07	1.22	0.76	1.04	1.51	1.07
	60 BC	0.93	1.31	0.71	1.05	1.22	0.78	1.05	1.46	1.06
ARMA	(1,1)	0.94	1.32	0.70	1.06	1.22	0.76	1.04	1.49	1.07
	(1,0)	0.93	1.30	0.69	1.05	1.22	0.76	1.03	1.47	1.06
	(0,1)	0.94	1.30	0.70	1.05	1.22	0.77	1.03	1.47	1.06
	(2,0)	0.95	1.32	0.71	1.06	1.23	0.77	1.05	1.49	1.07
ARMAX - T	(1,1)	0.95	1.32	0.71	1.06	1.24	0.77	1.06	1.50	1.08
	(1,0)	0.94	1.29	0.71	1.04	1.23	0.77	1.05	1.46	1.06
	(0,1)	0.94	1.30	0.72	1.04	1.23	0.78	1.05	1.46	1.06
	(2,0)	0.95	1.31	0.72	1.05	1.24	0.78	1.07	1.48	1.08
ARMAX - T,RH	(1,1)	0.94	1.30	0.72	1.05	1.26	0.78	1.06	1.50	1.08
	(1,0)	0.93	1.29	0.72	1.03	1.24	0.78	1.04	1.48	1.06
	(0,1)	0.93	1.29	0.72	1.04	1.24	0.78	1.04	1.49	1.07
	(2,0)	0.94	1.31	0.73	1.05	1.26	0.79	1.06	1.49	1.08
ARMAX - T,RH,R	(1,1)	0.95	1.32	0.73	1.06	1.28	0.78	1.06	1.52	1.09
	(1,0)	0.94	1.30	0.73	1.04	1.25	0.78	1.05	1.49	1.07
	(0,1)	0.94	1.31	0.73	1.05	1.25	0.78	1.05	1.51	1.08
	(2,0)	0.96	1.33	0.74	1.06	1.27	0.79	1.07	1.51	1.09
ARMAX - T,RH,F,N,R	(1,1)	0.97	1.35	0.74	1.07	1.24	0.77	1.08	1.55	1.09
	(1,0)	0.95	1.33	0.74	1.05	1.20	0.75	1.06	1.52	1.08
	(0,1)	0.96	1.34	0.74	1.05	1.21	0.76	1.06	1.53	1.08
	(2,0)	0.97	1.35	0.75	1.06	1.23	0.77	1.08	1.53	1.09

4.5. Average MAE and RMSE for all 16 weather stations

This section covers the overall comparison of the models. It compares all models and all stations with mean absolute error (MAE) and root mean squared error (RMSE) of 2-m temperature (°C).

For even more validation, an addition to the MAE measurement previously used, RMSE of 2-m temperature was calculated for all models and all stations for further model comparison. The bene ts of RMSE include the penalization of large errors in models, which is often desirable in weather forecasting models. The Python class used to compute the models also minimizes RMSE.

Table 4.4 shows the nal comparison of all the models studied and computed in this thesis. The table shows the average MAE and average RMSE of 2-m temperature over all 16 stations for all models. The results show that the 60 day bias correction model has both lower average MAE and RMSE of 2-m temperature than the IMO model does, once again con rming the importance of choosing the optimal parameters.

The results also show that all models computed have a marginally lower average MAE and RMSE of 2-m temperature than the IMO model does. The results in Table 4.4 also show that the AR(1) model or when using exogenous variables, the ARX(1) model have the lowest average MAE and RMSE compared to other models in the same block. The model with the lowest MAE of 2-m temperature is the ARMAX(1; 0) or ARX(1) model with exogenous variables T, RH. The model has an average MAE of 2-m temperature of 1.151 °C while the IMO model has an average MAE of 2-m temperature of 1.207 °C. This model also has a smaller RMSE of 2-m temperature, at 1.526 °C while the IMO model has a RMSE of 2-m temperature of 1.590 °C. The ARMAX(1; 0) model using all exogenous variables has a similar RMSE but the MAE is a little bit higher.

Table 4.4: Average mean absolute error and root mean squared error of 2-m temperature (°C) in the time range 2019–2021 for all models computed in the study for the selected 16 weather stations. See section 4.1 for de nition of simple models and 4.2 for de nition of autoregressive integrated moving average models.

		MAE	RMSE
Simple models	NWP	1.695	2.051
	IMO	1.207	1.590
	60 BC	1.182	1.572
ARMA	(1,1)	1.176	1.565
	(1,0)	1.166	1.552
	(0,1)	1.170	1.557
	(2,0)	1.177	1.565
ARMAX - T	(1,1)	1.177	1.560
	(1,0)	1.164	1.543
	(0,1)	1.168	1.547
	(2,0)	1.174	1.556
ARMAX - T,RH	(1,1)	1.162	1.540
	(1,0)	1.151	1.526
	(0,1)	1.155	1.530
	(2,0)	1.163	1.540
ARMAX - T,RH,R	(1,1)	1.175	1.553
	(1,0)	1.161	1.536
	(0,1)	1.165	1.542
	(2,0)	1.175	1.552
ARMAX - T,RH,F,N,R	(1,1)	1.174	1.547
	(1,0)	1.156	1.527
	(0,1)	1.162	1.534
	(2,0)	1.172	1.544

5. Discussion and conclusion

The model development process in this study included many steps performed, such as data preprocessing, data exploration and visualization, data selection, model assembly and model evaluation which are all covered in the thesis. The goal of this process was to study and improve the performance and forecast accuracy of the post-processing scheme for 2-m temperature forecasts which is currently being used by the IMO.

In the study, both new models were computed as well as the same model being used by the IMO but with different parameters. The process started with computing different simple models which suggested the weaknesses of the current IMO scheme. Following these suggestions, more advanced variations of autoregressive moving average models were computed and studied. Eight more stations were also added for further validation.

5.1. Discussion of the primary results of this study

The results from this thesis suggest there are some ways to improve the performance of the post-processing scheme being used by the IMO in order to get more accurate temperature forecast.

As noted in the study, it is possible to revise the parameters of the IMO model. In the study, the 60 day bias correction model had a lower MAE than the IMO model, suggesting that the parameters that the IMO uses are placing too much weight on newer observations while not weighing the earlier observations enough. The IMO model currently uses an ARMA$(1;1)$ model with 1.0 for the AR term and -0.9 for the MA term. The ARMA$(1;1)$ model computed in this study by the Python class statsmodels.tsa.arima.ARIMA chooses parameters –0.4 to 0.7 as the middle 50% for the AR term and the range –0.3 to 0.2 represents the 50% central interval for the MA term. Both parameters used by the IMO are outside the 50% interval. The study concluded that the IMO model is sensitive to its parameter selection and it would be possible to improve the forecast accuracy by weighing the past observations differently by updating the parameters of this model alone.

The other possibility to improve the model would be using ARMAX models which use exogenous variables as weighted input in the model. An interesting part of the study was to see if using exogenous variables temperature, relative humidity, wind, cloud cover, and precipitation would improve the forecasting accuracy. The computation of the first few ARMAX models gave promising results and therefore, more ARMAX models with different combinations of exogenous variables were computed. All of the ARMAX models computed outperformed the IMO model, however some only slightly. The model with the lowest average MAE was an ARMAX(1,0) or ARX(1) model with exogenous variables T and RH, suggesting post-processing 2-m temperature forecasts can be improved with exogenous variables temperature and relative humidity.

5.2. Future Work

This study focused on 24 hour forecasts of 2-m temperature, valid at noon, using 60 days for the bias correction model. The study suggests that the current scheme can be improved. However, there are a few avenues that need to be studied further.

Future work should include increasing the number of weather stations to the comparison. In this study, there were 16 stations used. It is possible that this limited number of selected stations may bias the results. It would also be possible to add a wider range of forecast periods and forecast times, as this study only focuses on a forecast period of 24 hours and time at noon. It would be interesting to see if the results were similar with a wider range of data selection. Data selection was performed in this study in order to identify trend and pattern more easily in a larger dataset. Future work could include expanding the data selection in order to get wider results.

Future work could also include adding newer data to the dataset. This study used data for 2019–2021. In order to validate the models, it would be beneficial to apply the method on unseen data.

Another direction could be to research the 60 day bias correction model further. This model gave promising results throughout this thesis, and almost always outperformed the IMO model. However, the 60 day selection length is artificial and different time length might improve the results.

As models using exogenous variables gave promising results in this study, it would be logical to study them further and try other exogenous variables which were not used in this study, such as wind direction. Neural networks used to post-process output from atmospheric models is also an obvious choice to investigate.

This study hopefully encourages meteorologists at the IMO to reevaluate their post-processing scheme. It suggests many different ways of how the forecasting accuracy of the current scheme could be improved with different parameters and other time series models, possibly including exogenous variables.

Bibliography

[1] N. Nawri, “Evaluation of HARMONIE reanalyses of surface air temperature and wind speed over Iceland,” Report no.: VÍ 2014-005, The Icelandic Meterological Of ce, 2014.

[2] P. Boi, “A statistical method for forecasting extreme daily temperatures using ECMWF 2-m temperatures and ground station measurements,” *Meteorological Applications*, vol. 11, no. 3, pp. 245–251, 2004. DOI: https://doi.org/10.1017/S1350482704001318. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1017/S1350482704001318. [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1017/S1350482704001318.

[3] S. Green, “Time series analysis of stock prices using the Box-Jenkins approach,” *Electronic Theses and Dissertations*, 2011. [Online]. Available: https://digitalcommons.georgiasouthern.edu/etd/668.

[4] G. Bontempi, S. Ben Taieb, and Y.-A. Le Borgne, “Machine learning strategies for time series forecasting,” in *Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15-21, 2012, Tutorial Lectures*, M.-A. Aufaure and E. Zimányi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 62–77, ISBN: 978-3-642-36318-4. DOI: 10.1007/978-3-642-36318-4_3. [Online]. Available: https://doi.org/10.1007/978-3-642-36318-4_3.

[5] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day series in time series analysis and digital processing). Holden-Day, 1970, ISBN: 9780816210947. [Online]. Available: https://books.google.is/books?id=5BVfnXaq03oC.

[6] K. Taneja, S. Ahmad, K. Ahmad, and S. Attri, “Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach,” *Atmospheric Pollution Research*, vol. 7, no. 4, pp. 585–596, 2016, ISSN: 1309-1042. DOI: https://doi.org/10.1016/j.apr.2016.02.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1309104215301082.

[7] E. Stellwagen and L. Tashman, “Arima: The models of Box and Jenkins,” *Foresight: Int. J. Appl. Forecast.*, pp. 28–33, Jan. 2013.

[8] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature,” Geoscientific model development, vol. 7, no. 3, pp. 1247–1250, 2014.

[9] P. Crochet, “Adaptive kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in iceland,” Meteorological Applications, vol. 11, no. 2, pp. 173–187, 2004. DOI: https://doi.org/10.1017/S1350482704001252. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1017/S1350482704001252 [Online]. Available: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1017/S1350482704001252.

[10] B. Pálmason, “Leiðrétting á hitaspám,” The Icelandic Meterological Office, 2016.

A. Appendix

Table A.1: The full table of data exploration performed in Chapter 3. An overview of the principal information for each weather station.

Station	Start date	End date	No. days	Missing (%)	Longest streak	M (%)
1350	24.8.2018	20.3.2022	1304	0	0	0
1361	24.8.2018	20.3.2022	1304	0	0	0
1370	10.10.2020	20.3.2022	526	0	0	0
1395	24.8.2018	20.3.2022	1304	0	0	0
1453	24.8.2018	20.3.2022	1304	0	0	0
1470	26.5.2021	20.3.2022	298	0	0	0
1471	11.12.2019	20.3.2022	830	0	0	0
1473	24.8.2018	20.3.2022	1304	0	0	0
1474	21.6.2019	20.3.2022	1003	0.41	1	0.41
1475	24.8.2018	20.3.2022	1304	0	0	0
1477	24.8.2018	20.3.2022	1304	0.16	2	0.16
1478	3.10.2019	20.3.2022	899	0	0	0
1479	28.8.2018	20.3.2022	1300	0.09	1	0.08
1480	24.8.2018	20.3.2022	1304	0.08	1	0.08
1481	24.8.2018	20.3.2022	1304	0	0	0
1482	1.12.2018	20.3.2022	1205	0	0	0
1485	24.8.2018	20.3.2022	1304	0.72	9	0.72
1486	24.8.2018	20.3.2022	1304	0	0	0
1487	24.8.2018	20.3.2022	1304	0	0	0
1490	24.8.2018	20.3.2022	1304	0	0	0
1493	24.8.2018	20.3.2022	1304	0	0	0
1496	24.8.2018	20.3.2022	1304	0	0	0
1578	24.8.2018	20.3.2022	1304	0	0	0
1590	24.8.2018	20.3.2022	1304	1.44	9	1.44
1596	24.8.2018	20.3.2022	1304	0	0	0
1673	24.8.2018	20.3.2022	1304	0	0	0
1679	24.8.2018	20.3.2022	1304	1.81	16	1.82

1685	24.8.2018	20.3.2022	1304	0	0	0
1689	24.8.2018	20.3.2022	1304	0	0	0
1779	24.8.2018	20.3.2022	1304	0	0	0
1781	24.8.2018	20.3.2022	1304	0	0	0
1868	24.8.2018	20.3.2022	1304	0	0	0
1881	24.8.2018	20.3.2022	1304	0	0	0
1919	24.8.2018	20.3.2022	1304	0	0	0
1924	24.8.2018	20.3.2022	1304	0	0	0
1936	24.8.2018	20.3.2022	1304	0	0	0
1938	24.8.2018	20.3.2022	1304	0	0	0
2050	24.8.2018	20.3.2022	1304	0	0	0
2175	24.8.2018	20.3.2022	1304	0	0	0
2197	24.8.2018	20.3.2022	1304	0	0	0
2266	24.8.2018	20.3.2022	1304	0	0	0
2304	24.8.2018	20.3.2022	1304	0	0	0
2315	24.8.2018	20.3.2022	1304	0.08	1	0.08
2319	24.8.2018	20.3.2022	1304	0	0	0
2320	24.8.2018	20.3.2022	1304	0	0	0
2323	24.8.2018	20.3.2022	1304	0.32	4	0.31
2370	24.8.2018	20.3.2022	1304	0	0	0
2390	24.8.2018	20.3.2022	1304	0	0	0
2428	24.8.2018	20.3.2022	1304	0	0	0
2480	24.8.2018	20.3.2022	1304	0	0	0
2481	24.8.2018	20.3.2022	1304	0	0	0
2530	15.12.2020	20.3.2022	460	0	0	0
2631	24.8.2018	20.3.2022	1304	0	0	0
2636	24.8.2018	20.3.2022	1304	1.03	13	1.03
2640	25.8.2018	20.3.2022	1303	0	0	0
2641	24.8.2018	20.3.2022	1304	0	0	0
2642	24.8.2018	20.3.2022	1304	0	0	0
2643	24.8.2018	20.3.2022	1304	0	0	0
2646	24.8.2018	20.3.2022	1304	0	0	0
2655	24.8.2018	20.3.2022	1304	0	0	0
2692	24.8.2018	20.3.2022	1304	0	0	0
2693	10.10.2020	20.3.2022	526	0	0	0
2738	24.8.2018	20.3.2022	1304	0	0	0
2862	24.8.2018	20.3.2022	1304	0.40	5	0.39
2941	24.8.2018	20.3.2022	1304	11.44	92	12.23
3007	24.8.2018	20.3.2022	1304	0	0	0
3054	24.8.2018	20.3.2022	1304	0	0	0

3103	24.8.2018	20.3.2022	1304	0	0	0
3223	24.8.2018	20.3.2022	1304	0	0	0
3225	24.8.2018	20.3.2022	1304	0	0	0
3242	24.8.2018	20.3.2022	1304	0	0	0
3292	24.8.2018	20.3.2022	1304	0	0	0
3317	24.8.2018	20.3.2022	1304	0	0	0
3371	24.8.2018	20.3.2022	1304	0	0	0
3380	24.8.2018	20.3.2022	1304	0	0	0
3433	24.8.2018	20.3.2022	1304	0	0	0
3463	24.8.2018	20.3.2022	1304	0	0	0
3471	24.8.2018	20.3.2022	1304	0	0	0
3474	24.8.2018	20.3.2022	1304	0	0	0
3477	24.8.2018	20.3.2022	1304	0	0	0
3490	24.8.2018	12.10.2021	1145	0	0	0
3591	24.8.2018	20.3.2022	1304	0	0	0
3596	24.8.2018	20.3.2022	1304	0	0	0
3658	24.8.2018	20.3.2022	1304	0	0	0
3696	24.8.2018	20.3.2022	1304	0	0	0
3720	24.8.2018	20.3.2022	1304	0	0	0
3751	14.11.2020	20.3.2022	491	0	0	0
3752	24.8.2018	20.3.2022	1304	0	0	0
3754	24.8.2018	26.6.2021	1037	0	0	0
3779	24.8.2018	20.3.2022	1304	0.16	2	0.16
3797	24.8.2018	20.3.2022	1304	0	0	0
3976	3.9.2018	20.3.2022	1294	0	0	0
4019	24.8.2018	20.3.2022	1304	0	0	0
4060	24.8.2018	20.3.2022	1304	0.08	1	0.08
4180	24.8.2018	20.3.2022	1304	0.19	2	0.18
4181	24.8.2018	20.3.2022	1304	0.43	2	0.43
4182	24.8.2018	20.3.2022	1304	0	0	0
4193	24.8.2018	20.3.2022	1304	0	0	0
4271	24.8.2018	20.3.2022	1304	0	0	0
4275	24.8.2018	20.3.2022	1304	0	0	0
4300	24.8.2018	20.3.2022	1304	0	0	0
4323	24.8.2018	20.3.2022	1304	0	0	0
4380	24.8.2018	20.3.2022	1304	0	0	0
4406	24.8.2018	20.3.2022	1304	0	0	0
4455	24.8.2018	20.3.2022	1304	0	0	0
4472	24.8.2018	20.3.2022	1304	0	0	0
4500	24.8.2018	20.3.2022	1304	0	0	0

4614	24.8.2018	20.3.2022	1304	0.08	1	0.08
4652	24.8.2018	20.3.2022	1304	0	0	0
4828	24.8.2018	20.3.2022	1304	0	0	0
4830	24.8.2018	20.3.2022	1304	0	0	0
4867	24.8.2018	20.3.2022	1304	0	0	0
4912	24.8.2018	20.3.2022	1304	0	0	0
4921	4.10.2018	20.3.2022	1263	0	0	0
5210	24.8.2018	20.3.2022	1304	0	0	0
5309	24.8.2018	20.3.2022	1304	0	0	0
5316	24.8.2018	20.3.2022	1304	0	0	0
5544	24.8.2018	20.3.2022	1304	0	0	0
5777	24.8.2018	20.3.2022	1304	0	0	0
5825	24.8.2018	12.9.2021	1115	0	0	0
5847	24.8.2018	5.9.2021	1108	1.94	20	1.95
5872	24.8.2018	20.3.2022	1304	0	0	0
5885	24.8.2018	20.3.2022	1304	0	0	0
5932	24.8.2018	7.3.2019	195	5.82	11	6.11
5933	24.8.2018	20.3.2022	1304	0	0	0
5940	24.8.2018	20.3.2022	1304	3.08	39	3.14
5943	24.8.2018	25.8.2021	1097	0	0	0
5960	24.8.2018	20.3.2022	1304	0	0	0
5965	24.8.2018	20.3.2022	1304	0	0	0
5969	24.8.2018	7.9.2021	1110	0	0	0
5970	24.8.2018	20.3.2022	1304	0.08	1	0.08
5975	24.8.2018	20.3.2022	1304	0	0	0
5981	24.8.2018	20.3.2022	1304	0.08	1	0.08
5982	24.8.2018	20.3.2022	1304	0	0	0
5988	24.8.2018	20.3.2022	1304	0	0	0
5990	24.8.2018	20.3.2022	1304	0	0	0
5992	24.8.2018	20.3.2022	1304	0	0	0
5993	24.8.2018	20.3.2022	1304	0	0	0
6012	24.8.2018	20.3.2022	1304	0	0	0
6015	24.8.2018	20.3.2022	1304	0	0	0
6017	24.8.2018	20.3.2022	1304	0	0	0
6045	24.8.2018	20.3.2022	1304	0	0	0
6134	24.8.2018	20.3.2022	1304	0.08	1	0.08
6176	24.8.2018	18.3.2022	1302	0.49	6	0.48
6208	24.8.2018	20.3.2022	1304	0	0	0
6222	24.8.2018	20.3.2022	1304	0	0	0
6235	24.8.2018	20.3.2022	1304	0.08	1	0.08

6237	24.8.2018	20.3.2022	1304	0	0	0
6272	24.8.2018	20.3.2022	1304	0	0	0
6300	21.6.2019	20.3.2022	1003	0	0	0
6310	24.8.2018	20.3.2022	1304	0	0	0
6315	24.8.2018	20.3.2022	1304	0	0	0
6420	24.8.2018	20.3.2022	1304	0	0	0
6424	24.8.2018	20.3.2022	1304	0	0	0
6430	24.8.2018	20.3.2022	1304	0	0	0
6459	24.8.2018	6.10.2021	1139	0	0	0
6472	24.8.2018	20.3.2022	1304	0	0	0
6499	24.8.2018	20.3.2022	1304	0	0	0
6515	24.8.2018	20.3.2022	1304	0	0	0
6546	24.8.2018	27.9.2021	1130	0	0	0
6657	24.8.2018	20.3.2022	1304	0	0	0
6670	24.8.2018	20.3.2022	1304	0	0	0
6745	24.8.2018	20.3.2022	1304	3.62	29	3.68
6748	24.8.2018	20.3.2022	1304	0	0	0
6760	24.8.2018	20.3.2022	1304	0	0	0
6776	24.8.2018	29.7.2021	1070	0	0	0
6802	24.8.2018	20.3.2022	1304	0	0	0
6935	24.8.2018	20.3.2022	1304	0	0	0
6975	24.8.2018	20.3.2022	1304	0	0	0
7365	26.3.2021	17.2.2022	328	0	0	0
7475	24.8.2018	20.3.2022	1304	0	0	0
7659	24.8.2018	20.3.2022	1304	7.68	35	7.86
7736	24.8.2018	20.3.2022	1304	0	0	0
7738	24.8.2018	20.3.2022	1304	0	0	0
7753	24.8.2018	20.3.2022	1304	0.08	1	0.08
7790	24.8.2018	7.12.2021	1201	0	0	0
31109	24.8.2018	17.3.2022	1301	0	0	0
31122	24.8.2018	17.3.2022	1301	0	0	0
31363	24.8.2018	17.3.2022	1301	0	0	0
31364	24.8.2018	17.3.2022	1301	0	0	0
31365	24.8.2018	17.3.2022	1301	0	0	0
31380	24.8.2018	17.3.2022	1301	0	0	0
31387	24.8.2018	17.3.2022	1301	0	0	0
31392	24.8.2018	17.3.2022	1301	0.16	2	0.16
31399	24.8.2018	17.3.2022	1301	0.32	4	0.32
31475	24.8.2018	17.3.2022	1301	0	0	0
31488	24.8.2018	17.3.2022	1301	0	0	0

31572	24.8.2018	17.3.2022	1301	0	0	0
31577	24.8.2018	17.3.2022	1301	0	0	0
31578	9.1.2021	17.3.2022	432	0	0	0
31579	24.8.2018	17.3.2022	1301	0	0	0
31591	24.8.2018	17.3.2022	1301	0	0	0
31599	24.8.2018	17.3.2022	1301	0	0	0
31640	9.4.2021	17.3.2022	342	0	0	0
31674	24.8.2018	17.3.2022	1301	0	0	0
31840	24.8.2018	17.3.2022	1301	0	0	0
31882	24.8.2018	17.3.2022	1301	0	0	0
31931	24.8.2018	17.3.2022	1301	0	0	0
31932	24.8.2018	17.3.2022	1301	0	0	0
31942	24.8.2018	17.3.2022	1301	0	0	0
31948	24.8.2018	17.3.2022	1301	0	0	0
31950	24.8.2018	17.3.2022	1301	0	0	0
31958	24.8.2018	17.3.2022	1301	0	0	0
31985	24.8.2018	21.2.2022	1277	0	0	0
32097	24.8.2018	17.3.2022	1301	0	0	0
32179	24.8.2018	17.3.2022	1301	0	0	0
32190	24.8.2018	17.3.2022	1301	0	0	0
32224	24.8.2018	17.3.2022	1301	0	0	0
32322	24.8.2018	17.3.2022	1301	4.73	37	4.84
32355	24.8.2018	17.3.2022	1301	0	0	0
32365	24.8.2018	17.3.2022	1301	0	0	0
32372	24.8.2018	17.3.2022	1301	0	0	0
32377	24.8.2018	17.3.2022	1301	1.74	22	1.75
32390	24.8.2018	17.3.2022	1301	0	0	0
32474	24.8.2018	17.3.2022	1301	0	0	0
32533	24.8.2018	17.3.2022	1301	0	0	0
32635	24.8.2018	17.3.2022	1301	0.71	9	0.71
32654	24.8.2018	17.3.2022	1301	0	0	0
33204	24.8.2018	17.3.2022	1301	0	0	0
33357	12.9.2018	17.3.2022	1282	0.48	6	0.48
33394	24.8.2018	17.3.2022	1301	0	0	0
33419	24.8.2018	17.3.2022	1301	0	0	0
33424	24.8.2018	17.3.2022	1301	0	0	0
33431	24.8.2018	17.3.2022	1301	0	0	0
33451	6.12.2018	17.3.2022	1197	0	0	0
33480	24.8.2018	17.3.2022	1301	0	0	0
33487	24.8.2018	17.3.2022	1301	0	0	0

33495	24.8.2018	17.3.2022	1301	0	0	0
33563	24.8.2018	17.3.2022	1301	0	0	0
33576	24.8.2018	17.3.2022	1301	0	0	0
33643	24.8.2018	17.3.2022	1301	0	0	0
33652	9.1.2021	17.3.2022	432	0	0	0
33654	24.8.2018	17.3.2022	1301	3.80	48	3.90
33661	24.8.2018	17.3.2022	1301	0	0	0
33750	24.8.2018	17.3.2022	1301	0	0	0
33751	24.8.2018	17.3.2022	1301	0	0	0
34073	24.8.2018	17.3.2022	1301	0.08	1	0.08
34081	6.12.2018	17.3.2022	1197	0	0	0
34087	24.8.2018	17.3.2022	1301	0	0	0
34148	24.8.2018	17.3.2022	1301	0	0	0
34175	24.8.2018	17.3.2022	1301	0	0	0
34238	24.8.2018	17.3.2022	1301	0	0	0
34326	24.8.2018	17.3.2022	1301	0	0	0
34348	24.8.2018	17.3.2022	1301	0	0	0
34382	24.8.2018	17.3.2022	1301	0	0	0
34413	18.9.2018	17.3.2022	1276	0	0	0
34450	24.8.2018	17.3.2022	1301	0	0	0
34559	24.8.2018	17.3.2022	1301	0	0	0
34700	24.8.2018	17.3.2022	1301	2.93	25	2.96
34732	24.8.2018	17.3.2022	1301	0	0	0
34733	24.8.2018	17.3.2022	1301	0	0	0
35107	6.12.2018	17.3.2022	1197	0	0	0
35116	14.5.2019	17.3.2022	1038	0	0	0
35305	24.8.2018	17.3.2022	1301	0	0	0
35315	24.8.2018	17.3.2022	1301	0	0	0
35666	24.8.2018	17.3.2022	1301	0	0	0
35769	24.8.2018	17.3.2022	1301	0	0	0
35880	24.8.2018	17.3.2022	1301	0	0	0
35884	24.8.2018	17.3.2022	1301	0	0	0
35963	24.8.2018	17.3.2022	1301	0	0	0
35965	24.8.2018	17.3.2022	1301	0.87	7	0.87
35985	24.8.2018	17.3.2022	1301	0	0	0
36049	24.8.2018	17.3.2022	1301	0	0	0
36122	24.8.2018	17.3.2022	1301	0	0	0
36127	24.8.2018	17.3.2022	1301	0.32	2	0.32
36132	24.8.2018	17.3.2022	1301	0	0	0
36156	24.8.2018	17.3.2022	1301	0	0	0

36270	24.8.2018	17.3.2022	1301	0	0	0
36308	24.8.2018	17.3.2022	1301	0	0	0
36386	24.8.2018	17.3.2022	1301	0.57	6	0.56
36391	24.8.2018	17.3.2022	1301	0	0	0
36411	24.8.2018	17.3.2022	1301	0	0	0
36415	24.8.2018	17.3.2022	1301	0	0	0
36504	24.8.2018	17.3.2022	1301	0	0	0
36519	24.8.2018	17.3.2022	1301	0	0	0