Land Temperature Increase due to Global Warming

Predicting temperature for the next 100 years based on 263 years of global temperature data from 1753 to 2015

Student:

Snehaa Vijayanand Potdar

Term:

Fall 2019

Instructor:

Dr. Jiming Wu

Course:

BAN 693 – Business Analytics Capstone Project

# Table of Contents:

[Table of Contents: 2](#_Toc25267352)

[Motivation for the Project: 3](#_Toc25267353)

[Mankind’s Contribution: 4](#_Toc25267354)

[Impact of Global Warming: 5](#_Toc25267355)

[Change in weather patterns 5](#_Toc25267356)

[Impact on global food supply 5](#_Toc25267357)

[Impact on Flora and Fauna: 6](#_Toc25267358)

[Aim of the Project: 6](#_Toc25267359)

[Data Analysis: 7](#_Toc25267360)

[Predictability of Data: 7](#_Toc25267361)

[Creating a Time Series Object: 8](#_Toc25267362)

[Time Series Component Analysis: 9](#_Toc25267363)

[Splitting data into training and validation partitions: 11](#_Toc25267364)

[Model Development on Training Dataset: 12](#_Toc25267365)

[Model Comparison: 19](#_Toc25267366)

[Model Development on Entire Dataset: 20](#_Toc25267367)

[Conclusion: 22](#_Toc25267368)

[References: 23](#_Toc25267369)

# Motivation for the Project:

On February 14th, 1990, the spacecraft Voyager 1 took an iconic photograph of the Earth from over 4 billion miles away, as it zoomed towards the edge of our Solar System. From this humbling vantage point, our planet appears to be no more than a speck – 0.12 pixels in size – in an enormous arena of darkness. Here is Carl Sagan’s oft-quoted response:

*“Look again at that dot. That’s here. That’s home. That’s us. On it everyone you love, everyone you know, everyone you ever heard of, every human being who ever was, lived out their lives. The aggregate of our joy and suffering, thousands of confident religions, ideologies, and economic doctrines, every hunter and forager, every hero and coward, every creator and destroyer of civilization… every ‘superstar’, every ‘supreme leader’, every saint and sinner in the history of our species lived there – on a mote of dust suspended in a sunbeam.”*



The Earth has always seemed powerful, solid, and reliable to us humans. We are awed by the tempestuous rise and fall of our oceans. Fresh air seems infinite in supply. The sun seems a tranquil yellow circle in the sky, bathing our homes and fields in warmth. As nature wends its cyclical way, we have constructed the grand edifice of human culture – literature, art, music, architecture, philosophies of politics and society, economic systems and governmental regimes. We take it for granted that these things will exist, in all their splendor, for generations to come.

The unfortunate truth is that our human viewpoint is severely limited by our size. Let us take a step back and view humanity for what it is – an incredibly young species, adrift in an unimaginably vast universe. The piece of rock on which we stand is the only thing tethering us to life. Our atmosphere is barely a membrane separating us from the vacuum of space (where we would instantly suffocate and explode), and preventing us from being killed by deadly ultraviolet radiation. Our existence is far more precarious than the solid ground beneath our feet makes it seem.

Despite the billions of galaxies out there, we have not seen any trace of other intelligent civilizations, or even basic life forms. The fact that life exists on our little blue bubble – something we tend to take for granted – is truly remarkable. The universe is an incredibly hostile place. An asteroid impact of sufficient strength could render Earth completely devoid of life. If our planet were twice as close to the center of our galaxy as it happens to be, gamma ray bursts would have prevented the formation or long-term development of life. Notwithstanding external threats, a supervolcanic eruption on Earth could cause mass extinction (which happened 250 million years ago). These are very real possibilities that routinely occur elsewhere in the universe.

Tragically, we humans are currently pushing the boundaries of our luck towards breaking point. We are eroding the conditions of our home planet that have nurtured the beautiful miracle of life, unaware of how tenuous and precious they are.

# Mankind’s Contribution:

Tragically, we humans are currently pushing the boundaries of our luck towards breaking point. We are eroding the conditions of our home planet that have nurtured the beautiful miracle of life, unaware of how tenuous and precious they are.

Since the Industrial Revolution, human activity has been significantly altering a climate that has just the right calibration to support life. Remarkably, Earthly life consists of not just a few strains of microbes, but millions of stunningly diverse species that give this planet color and flair. Climate change has the capacity to be an existential threat. At the very least, it can consign humanity to a terribly frugal existence, while other more helpless forms of life are destroyed.

To climate change skeptics: Species are already going extinct at an alarming rate. The concentration of carbon dioxide and other heat-trapping gases has risen dramatically, at a completely unnatural rate, in the past 50 years, directly attributable to human activity. Several independent studies show that the evidence is overwhelming and scary. Natural causes, like changes in solar energy output, simply cannot account for what we see today. Climate change is real – according to the vast majority of the world’s highly capable scientific community.

To those who acknowledge the existence of climate change, but do not consider it a priority: Climate change demands urgent attention. This is not to say that other issues like poverty and violence do not – they certainly do. But climate change ranks as high as any of them. It has far-reaching implications that go beyond one or two generations. It impacts the future of humanity as a whole, and all other forms of life that are equally entitled to the food, water and air we share.

# Impact of Global Warming:

## Change in weather patterns

It is no surprise that under a Department of Defense Directive, the United States’ military has made responding to climate change a national security priority (5). History has shown, all too often, that humans can be helpless in the face of nature’s wrath. For instance, Hurricane Katrina (2005) destroyed numerous innocent lives, and cost $108 billion in damages (6). Droughts in countries such as India have not only parched crop fields, but also spurred large numbers of farmers to suicide – leaving their children without sustenance or hope. With the climate in human-induced flux, hurricanes like Katrina, droughts, and floods will become much more frequent and intense (7). The human toll is undeniable. For countries with agriculture-based economies, changes in the timeframe and quality of crop growth can be debilitating.

## Impact on global food supply

Consider the fate of something as fundamental as food. Much of the developed world currently takes consumable food and water for granted; we are headed towards a shortage of both these essential resources, due to human profligacy. Our descendants will neither care about, nor be able to contribute to, the development of human culture if they must struggle to eat, drink and breathe every day. Furthermore, we are ravaging our atmospheric shield, directly exposing ourselves to ultraviolet radiation that causes cancer. Imagine how much suffering this would cause to people of all ages, especially those who cannot afford healthcare. We are literally destroying the very things that keep us alive.

## Impact on Flora and Fauna:

Without immediate action, a third of all land plant and animal species on Earth will be extinct by 2050 – a mere 34 years from now (12). The artistry of life is one of the most ineffable features of the cosmos, and it is surely our duty to preserve it for as long as we possibly can.

# Aim of the Project:

The project aims to utilize the data made available publicly by the [Berkeley Earth](https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data), which is affiliated with Lawrence Berkeley National Laboratory.  The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-existing archives. It is nicely packaged and allows for slicing into interesting subsets (for example by country). They publish the source data and the code for the transformations they applied. They also use methods that allow weather observations from shorter time series to be included, meaning fewer observations need to be thrown away.

As part of the project, I have used Time Series Data Analytics techniques to analyze the time series components and then apply forecasting techniques to predict temperature patterns for the next 100 years.

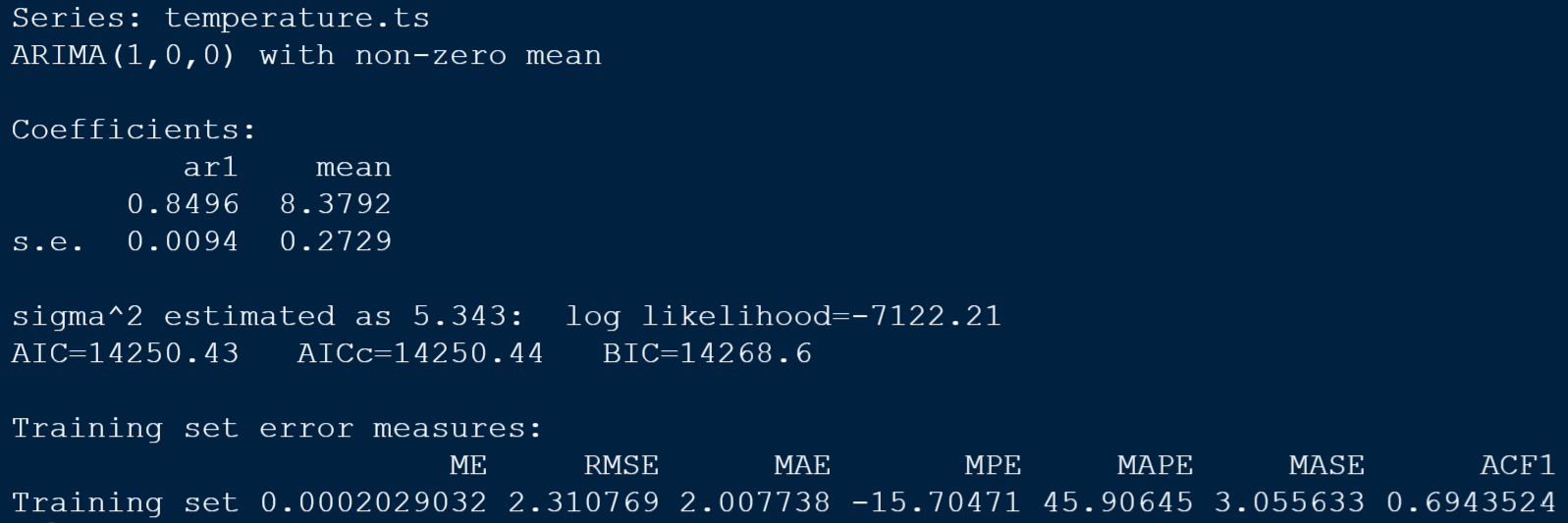
# Data Analysis:

## Predictability of Data:

In order to test the predictability, we will develop an AR(1)[Order = c(1,0,0)] model for the historical data and show the model itself by using the summary function.

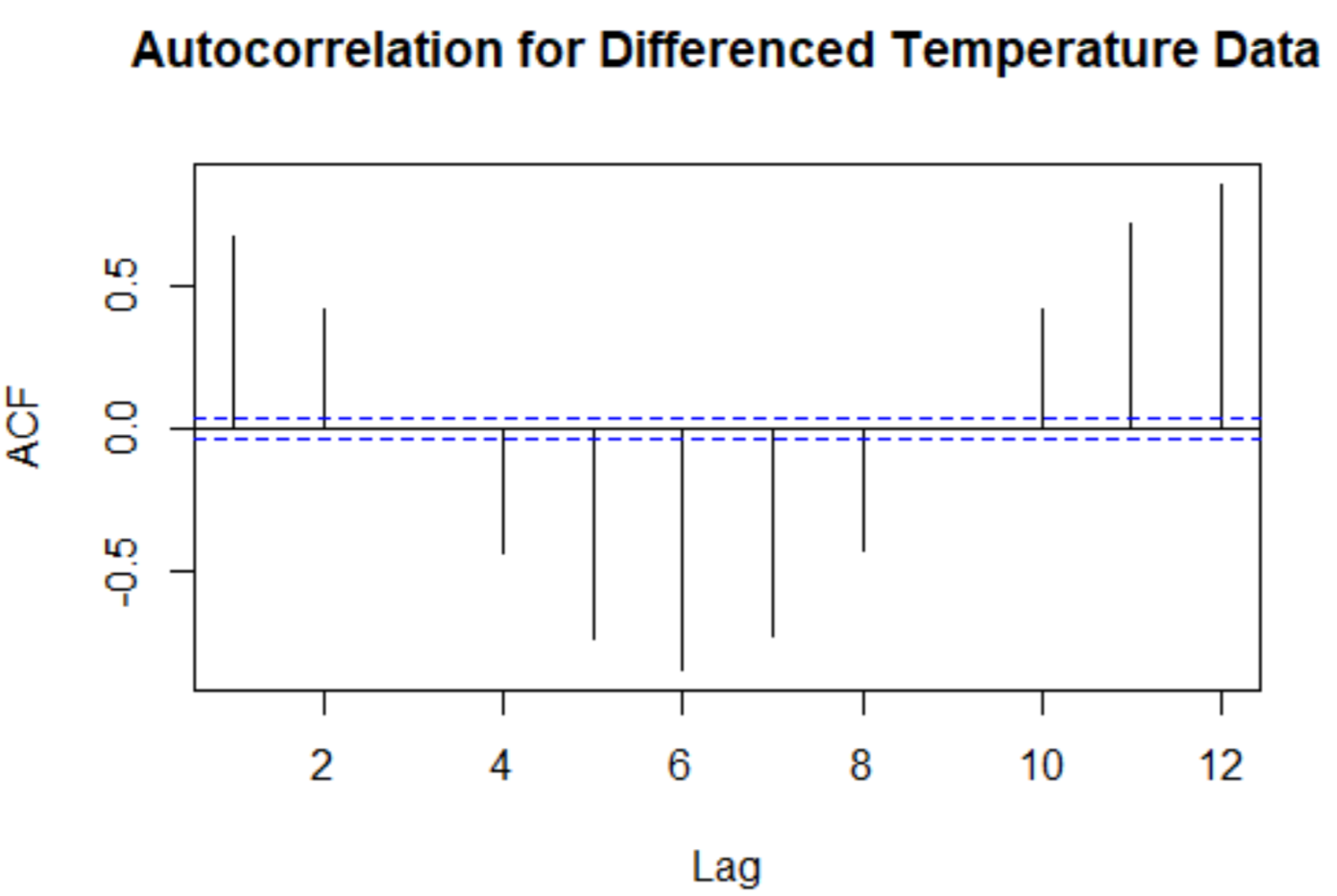
The equation fitted by the AR(1) (autoregressive model of order 1) model is: yt = 8.3792 + 0.8496 yt-1

The correlation coefficient for the autoregressive model is 0.8496 and the intercept is 8.3792. The ar1 co-efficient is not close to 1 which is indicative of the fact that the data is not random walk type data. Further, with a standard error of 0.0094, there is not a very high chance that this co-efficient can be equal to 1. Even if we take 2 standard errors from AR(1), it will still not be equal to 1. This proves that the dataset is not a random walk type dataset and hence can be predicted. We can also check the p-values of the ar1 coefficient by using the coeftest() function which further proves that we are within the 95% confidence interval.



Another way for us to check if the dataset is a random walk or not is to take the differences between dataset with a lag of 1. We are taking the first differences of the data i.e. subtracting the data and the data from the previous historical period to get the differences for the consecutive periods.

In this case, there are 10 autocorrelation coefficients (highest being lag 12) which are statistically significant. Since this is a sizable portion of the dataset, it indicates that the dataset is not a random walk dataset and hence can be predicted.

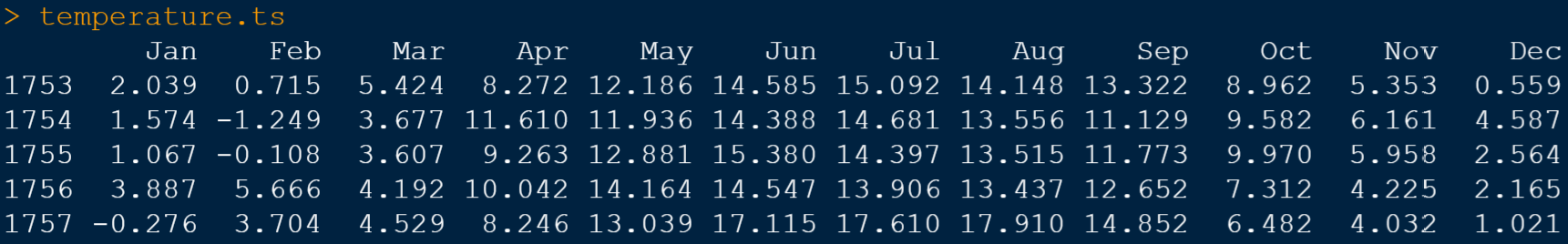


## Creating a Time Series Object:

The ts() function converts a numeric vector into a R time-series object. In this project, we used the following code:

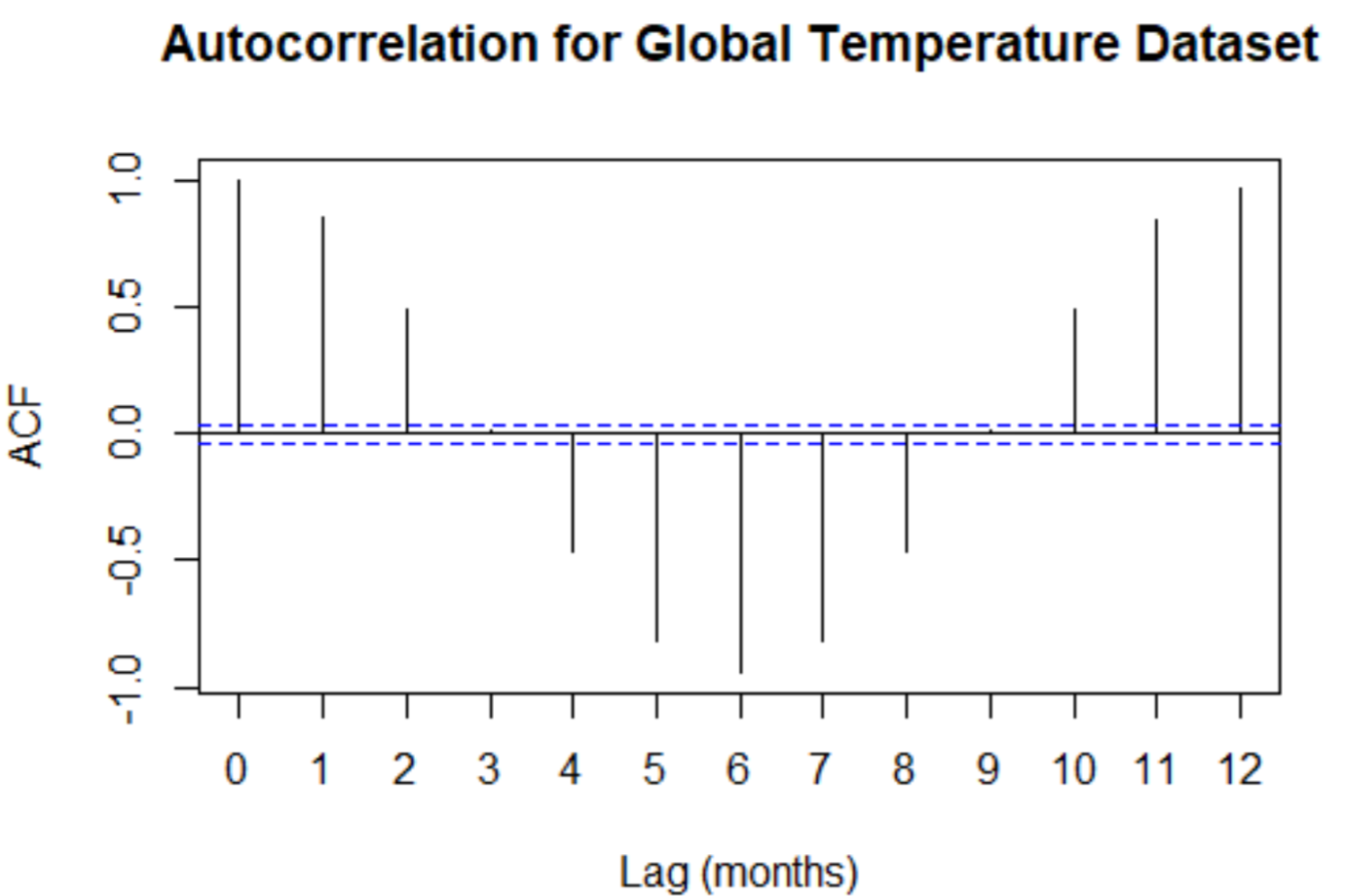
temperature.ts <- ts(temperature.data$LandAverageTemperature, start = c(1753, 1), end = c(2015, 12), freq = 12)

Snippet of the temperature.ts Time Series Object:



## Time Series Component Analysis:

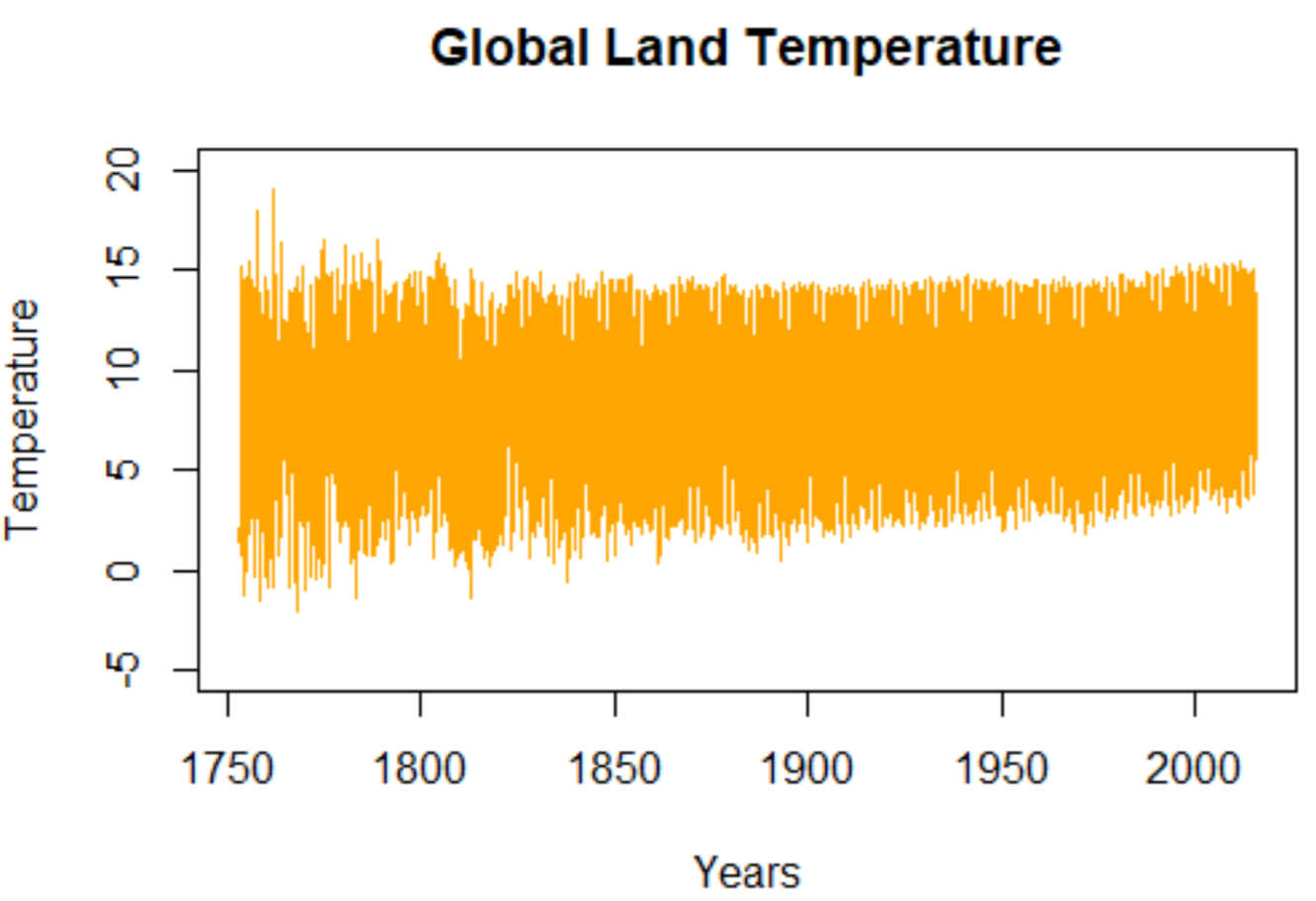
Estimating the autocorrelation function (ACF) at many lags allows us to assess how a time series relates to its past i.e. we can use the graph to analyze autocorrelation co-efficients and answer if the time series has a seasonal/trend component and if the data is random. Since there is strong significant autocorrelation in the first lag, it indicates an upward trend in data. Since we see a strong autocorrelation in lag 12 as well, we can also observe monthly seasonality in the data. The horizontal lines represent the level of significance. For the autocorrelation co-efficients that lie within these dotted lines (@ lags 3, 9), they can be considered to have random data.



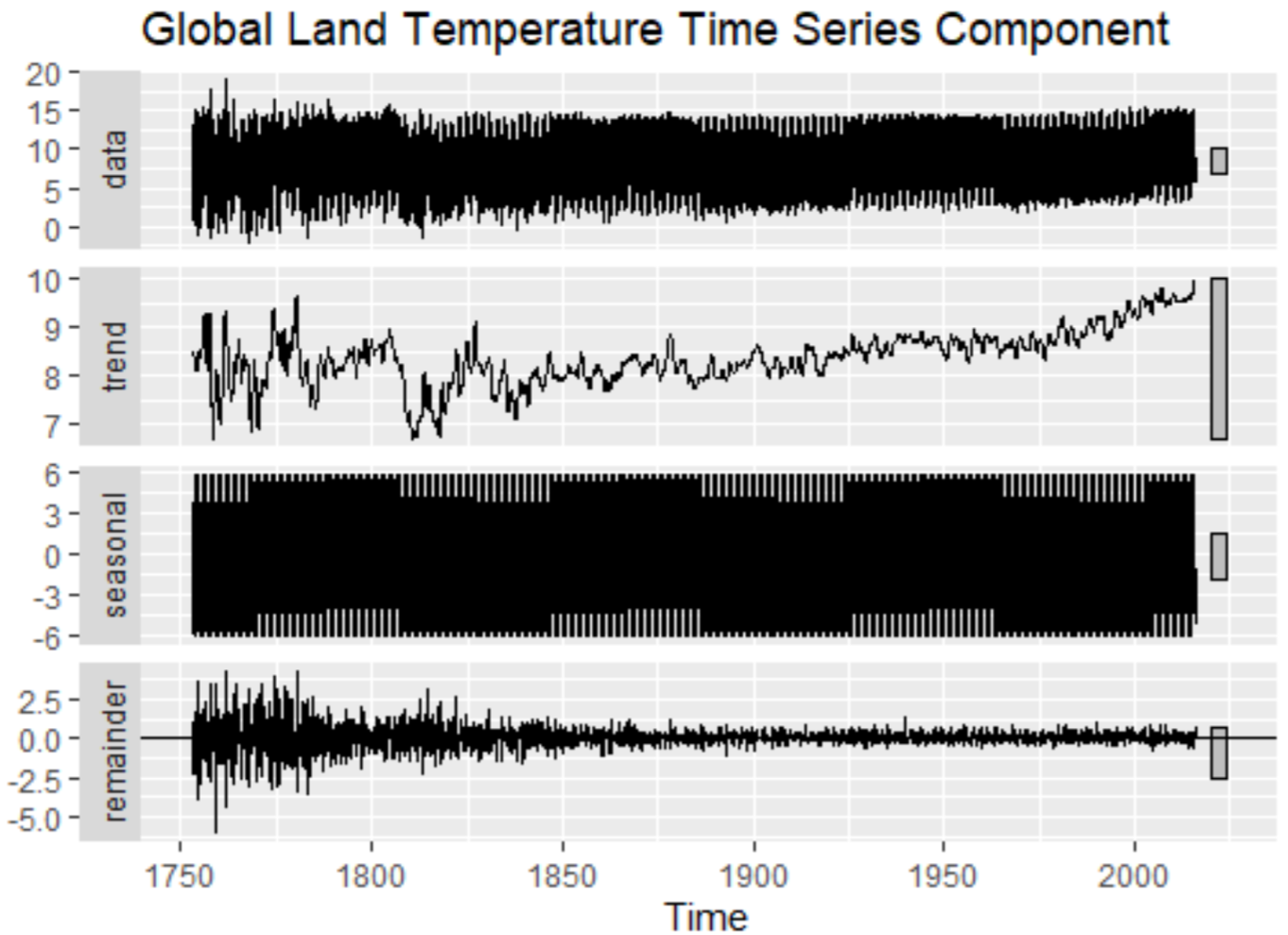
Thus, we can see strong trend relationship and strong seasonality relationship in the Global Temperature data.

The following code was used to generate the plot below:

plot(temperature.ts, xlab = "Years", ylab = "Temperature", ylim = c(-5, 20), main = "Global Land Temperature", col = "orange")



The Time Plot is a line graph that plots each observed value against the time of the observation, with a single line connecting each observation across the entire period. This plot is useful for graphically interpreting time series data.



This Time Plot shows very clearly that the data has a level 🡪 upward trend and seasonal fluctuations. The seasonality results in decreased sales at the beginning of each year and increased sales toward the end of each year. This can be further confirmed by using the stl() plot to understand the components of time-series data.

Also, the plot does not indicate any apparent pattern to the noise (looking at the “remainder” sub-graph) which is preferred.

## Splitting data into training and validation partitions:

We use the last 40 years of data as our validation dataset and the remaining years as our training dataset. This approach was chosen in order to provide as much data as possible in order for the model to be trained with higher accuracy.

The data can be partitioned to the criteria in the question by using the following R code:

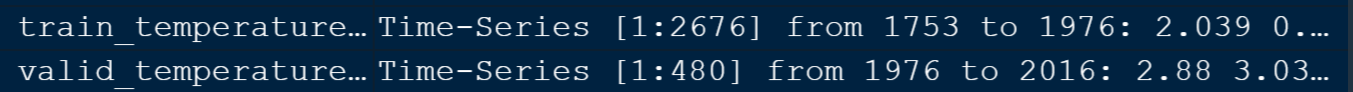
nValid <- 480

nTrain <- length(temperature.ts) – nValid

train\_temperature.ts <- window(temperature.ts, start = c(1753, 1), end = c(1753, nTrain))

valid\_temperature.ts <- window(temperature.ts, start = c(1753, nTrain + 1), end = c(1753, nTrain + nValid))

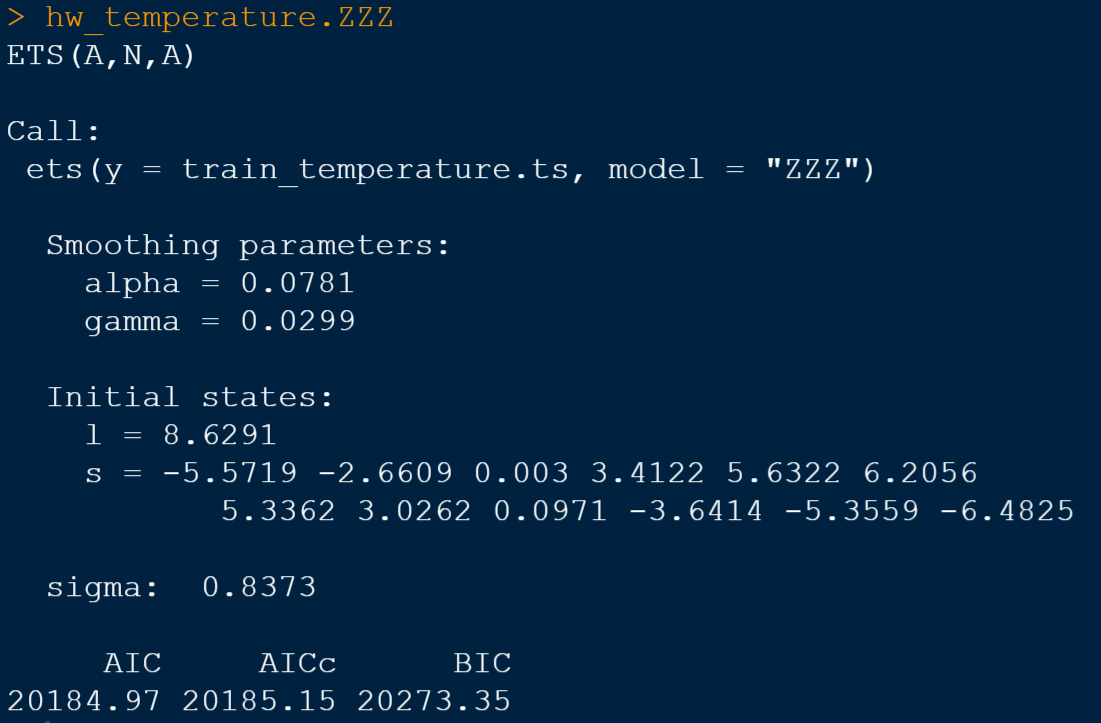
Screenshot of the Values Tab in R Studio:



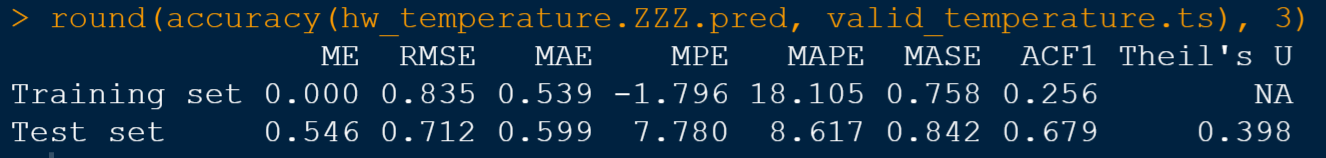
## Model Development on Training Dataset:

Holt – Winter’s Model with Automated Selection of Parameters:

I used the ets() function and since we setup the model such that does not specify the parameters i.e. it does not specify if it is an additive/multiplicative model nor trend/seasonality, then it seeks the optimal combination of the Holt-Winter’s model. The alpha parameter gives us an idea of the smoothing; the beta parameters gives an idea about trend and the gamma parameter about seasonality. They represent weighing the historical data for level, trend and seasonality. The model chosen was ANA and hence we have only the values for alpha and gamma (trend component missing). The seasonal index starts decreasing from the middle of the year and keeps decreasing till end of the year and then increases in the last month.



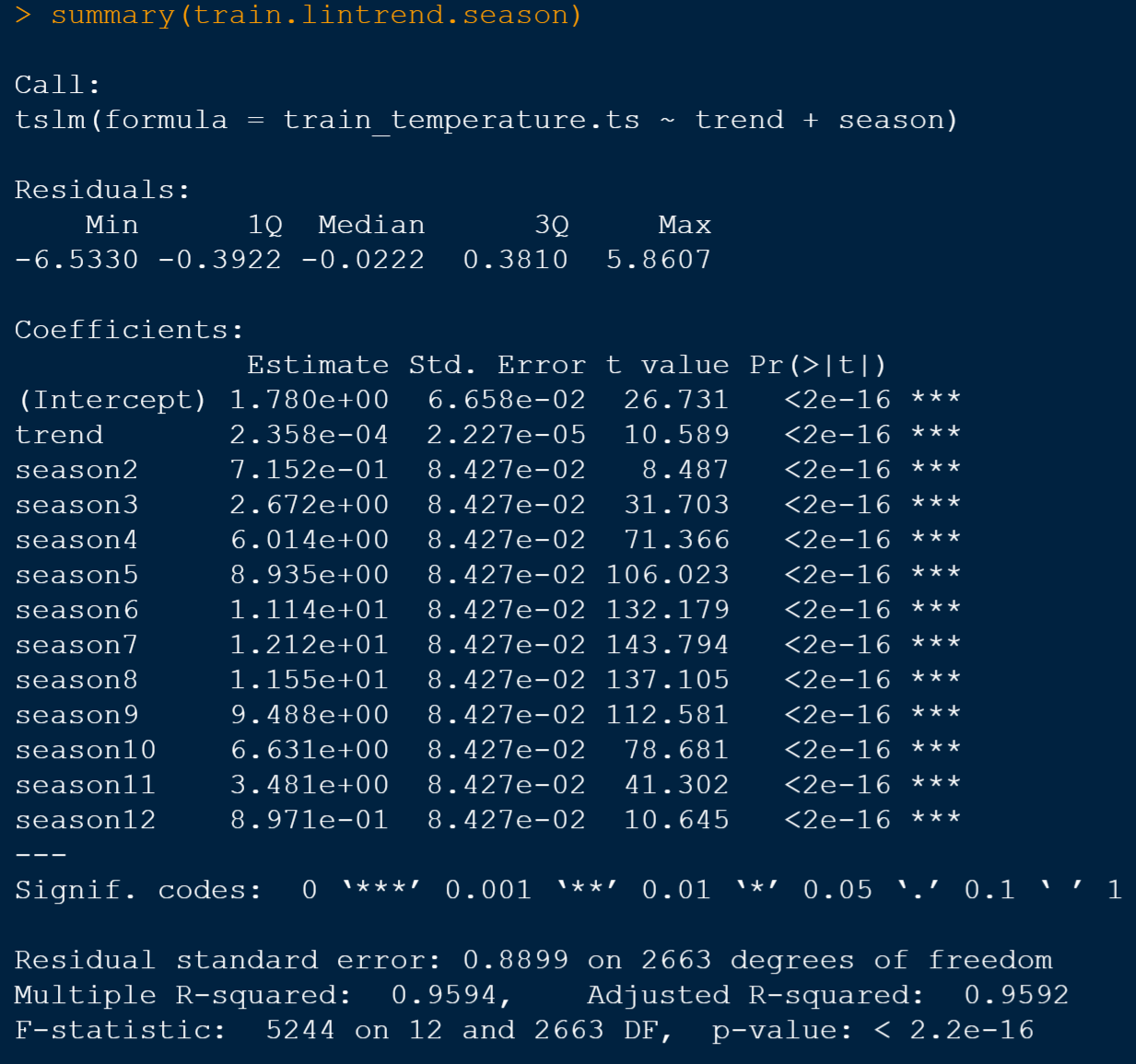
Accuracy Metrics:



To begin with, the HW forecast is over-predicting (since the ME is positive). On average, the forecast is producing 8.617% error when using the Holt-Winter’s model based on the MAPE metric. The RMSE for the Holt-Winter’s model forecast developed is 0.712.

Regression Model with Linear Trend and Seasonality:

The adjusted R2 value for the model is 0.9592 which is high. This indicates that 95.92% of the variation in the forecast can be explained by the variation in the seasonal parameters and period number. All the p-values are below the alpha threshold of 5%. The F-statistic looks fine as well (below the alpha threshold). This model is valid/statistically significant.



Model Equation:

**yt = 1.78 + 2.35E-4 t + 7.15E-1 D2 + 2.67 D3 + 6.01 D4 + 8.93 D5 + 1.11E+1 D6 + 1.21E+1 D7 + 1.15E+1 D8 + 9.48 D9 + 6.63 D10 + 3.48 D11 + 8.97E-1 D12**

t = Period number

D2 = 1 if Month 2 else equal to 0 (Binary dummy variable for Month 2)

D3 = 1 if Month 3 else equal to 0 (Binary dummy variable for Month 3)

D4 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 4)

D5 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 5)

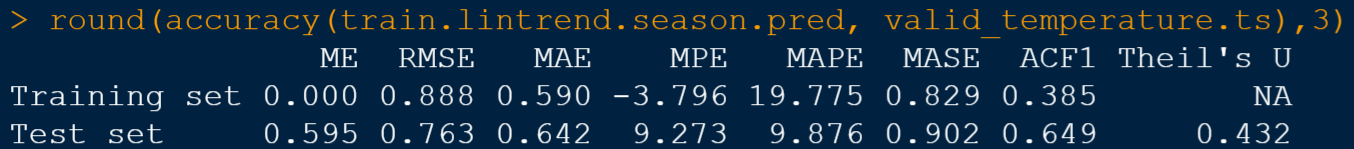
..

………..

D12 = 1 if Month 12 else equal to 0 (Binary dummy variable for Month 12)

If all D­I = 0 (i = 2, 3, 4, …., 12), then it is Month 1

Accuracy Metrics:



To begin with, the linear regression with seasonality model forecast is over-predicting (since the ME is positive). On average, the forecast is producing 9.876% error when using the Linear Regression with Seasonality’s model based on the MAPE metric. The RMSE for the Linear Regression with Seasonality’s model forecast developed is 0.763.

Regression Model with Quadratic Trend and Seasonality:

The adjusted R2 value for the model is 0.9604 which is high. This indicates that 96.04% of the variation in the forecast can be explained by the variation in the seasonal parameters and period number. All the p-values are below the alpha threshold of 5%. The F-statistic looks fine as well (below the alpha threshold). This model is valid/statistically significant.

Model Equation:

**yt = 2.127 - 5.41E-4 t + 2.9E-7 t2 + 7.15E-1 D2 + 2.67 D3 + 6.01 D4 + 8.93 D5 + 1.11E+1 D6 + 1.21E+1 D7 + 1.15E+1 D8 + 9.48 D9 + 6.63 D10 + 3.48 D11 + 8.97E-1 D12**

t = Period number

D2 = 1 if Month 2 else equal to 0 (Binary dummy variable for Month 2)

D3 = 1 if Month 3 else equal to 0 (Binary dummy variable for Month 3)

D4 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 4)

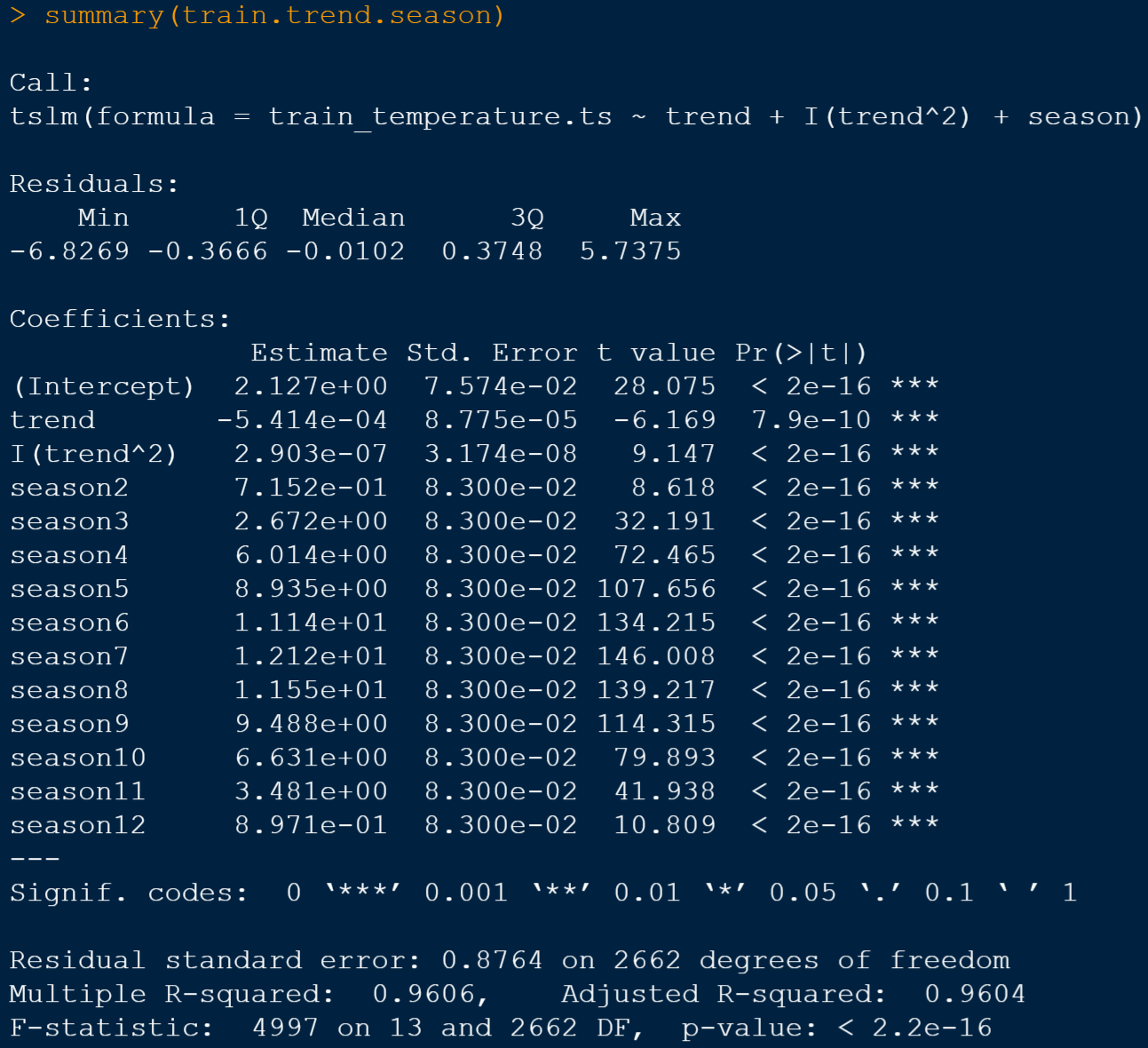
D5 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 5)

..

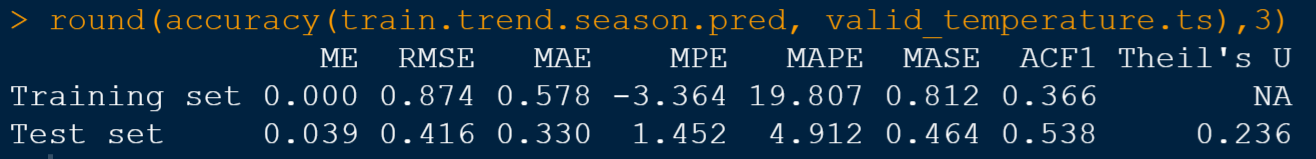
………..

D12 = 1 if Month 12 else equal to 0 (Binary dummy variable for Month 12)

If all D­I = 0 (i = 2, 3, 4, …., 12), then it is Month 1



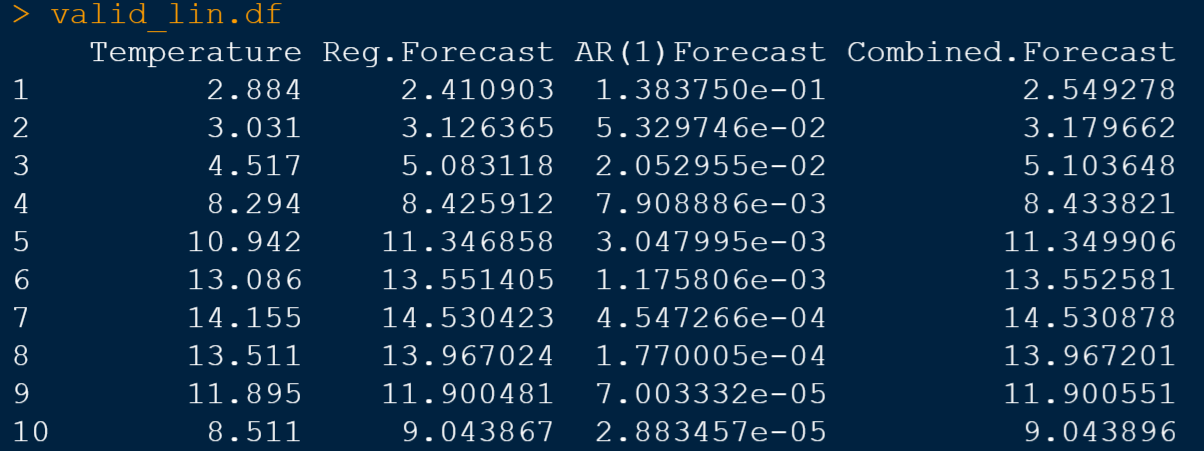
Accuracy Metrics:



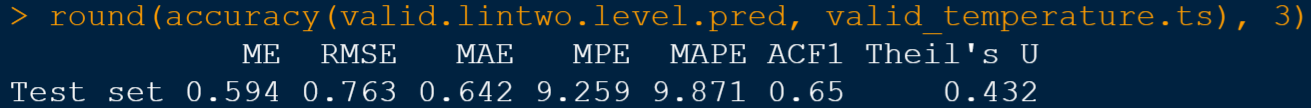
To begin with, the quadratic regression with seasonality model forecast is over-predicting (since the ME is positive). On average, the forecast is producing 4.912% error when using the Quadratic Regression with Seasonality’s model based on the MAPE metric. The RMSE for the Quadratic Regression with Seasonality’s model forecast developed is 0.416.

Two – Level Model with Linear Trend + Seasonality and AR(1):

I created a two-level forecast for the validation dataset by combining the regression with linear trend and seasonality model with the autoregressive model of order 1. The validation data is in the first column, the regression forecast in the second column, the AR(1) forecast based on autoregressive model for residuals is in the third column and the last column is the combined forecast.



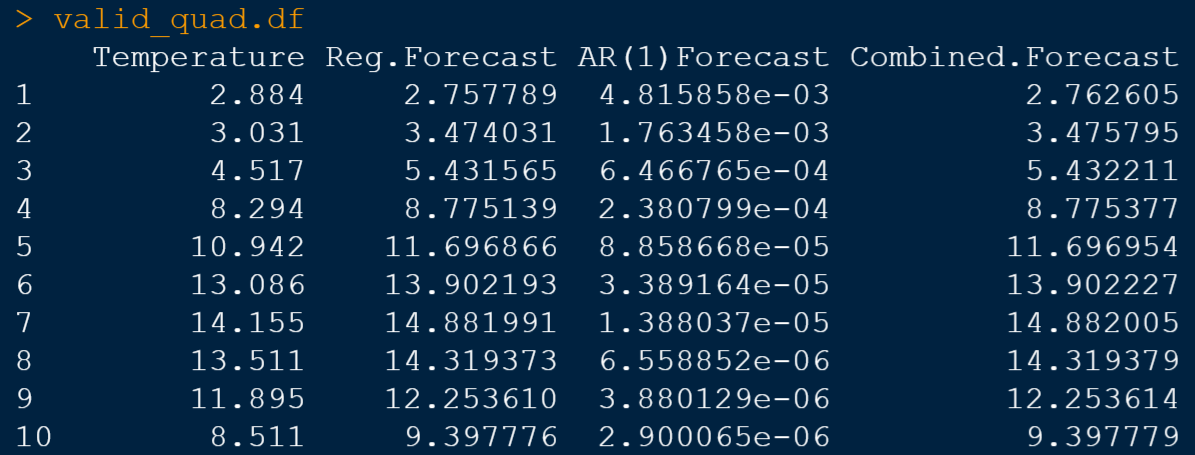
Accuracy Metrics:



To begin with, the two level model forecast (using linear trend + seasonality and AR(1)) is over-predicting (since the ME is positive). On average, the forecast is producing 9.871% error when using the two-level model based on the MAPE metric. The RMSE for the two level model forecast developed is 0.763.

Two – Level Model with Quadratic Trend + Seasonality and AR(1):

I created a two-level forecast for the validation dataset by combining the regression with quadratic trend and seasonality model with the autoregressive model of order 1. The validation data is in the first column, the regression forecast in the second column, the AR(1) forecast based on autoregressive model for residuals is in the third column and the last column is the combined forecast.



Accuracy Metrics:



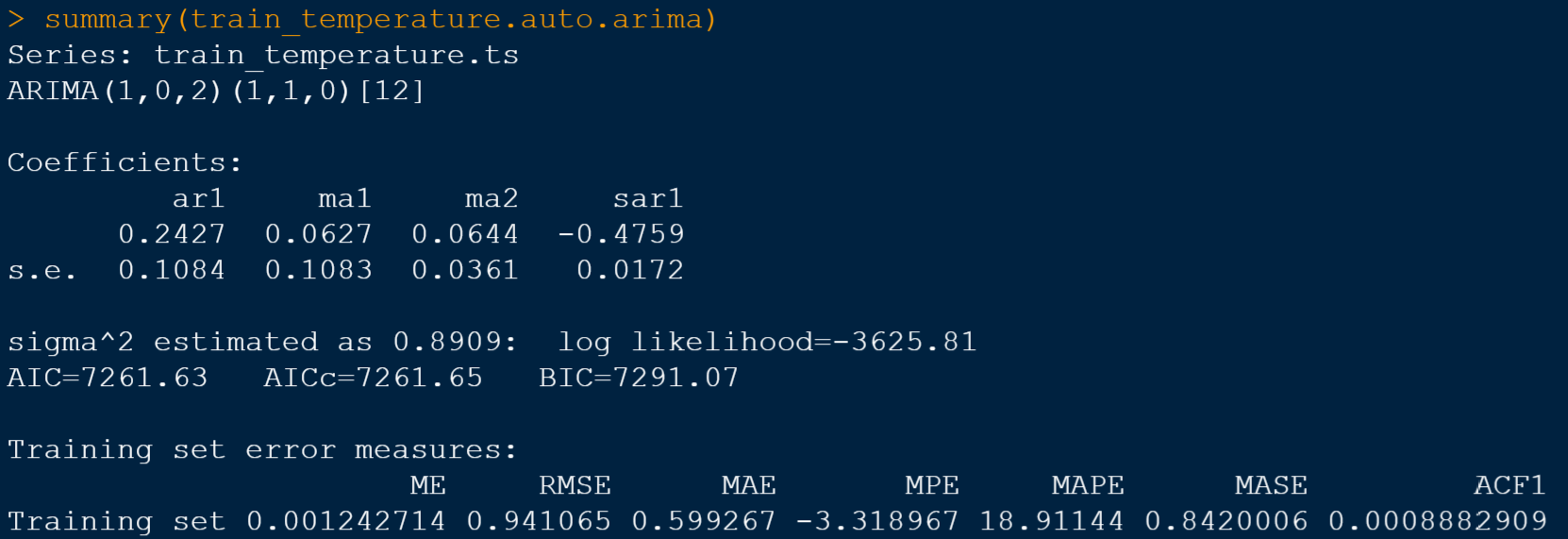
To begin with, the two level model forecast (using quadratic trend + seasonality and AR(1)) is over-predicting (since the ME is positive). On average, the forecast is producing 4.912% error when using the two-level model based on the MAPE metric. The RMSE for the two- level model forecast developed is 0.416.

ARIMA Model with automated selection of options:

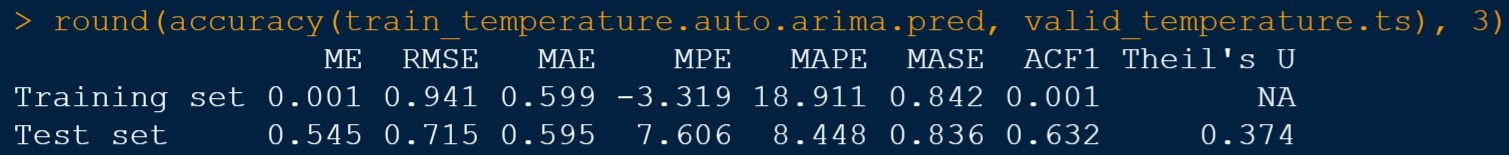
The auto.arima() function is used to identify optimal ARIMA model and its (p, d, q)(P, D, Q) parameters. Based on the auto.arima(), the model that we get for this dataset is ARIMA(1,0,2)(1,1,0)[12]. To elaborate further, we have autoregressive model of order 1 (ar1), moving average of order 2 for error lags and for the seasonal part, we do have an autoregressive portion or order 1 and a differencing of order 1 with monthly seasonality.

The model can defined as follows:

yt = 0.2427yt-1 + 0.0627Ԑt-1 + 0.0644Ԑt-2 – 0.4759 (yt-1 – yt-13)

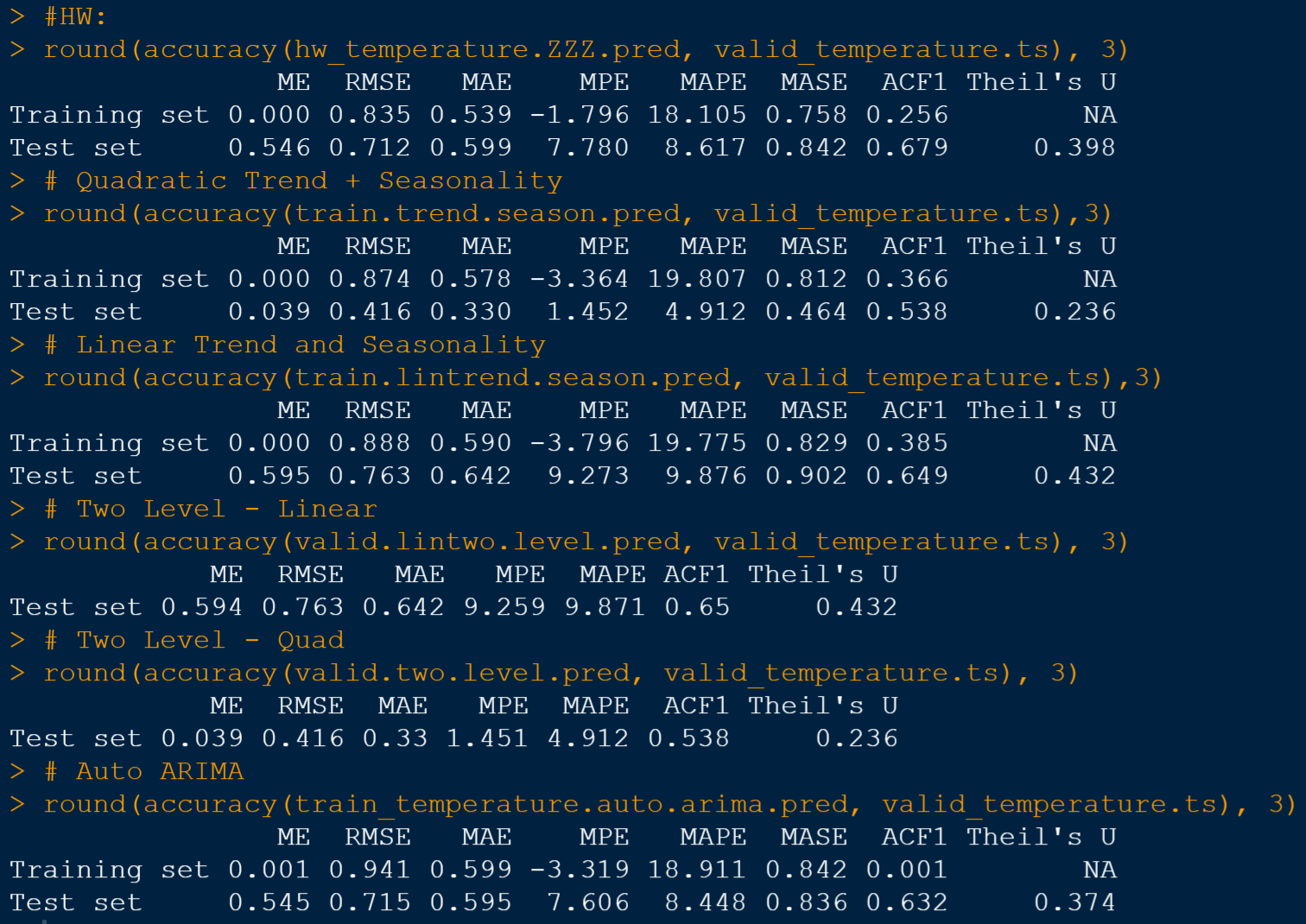


Accuracy Metrics:



To begin with, the ARIMA model is over-predicting (since the ME is positive). On average, the forecast is producing 8.448% error when using the two-level model based on the MAPE metric. The RMSE for the two- level model forecast developed is 0.715.

## Model Comparison:



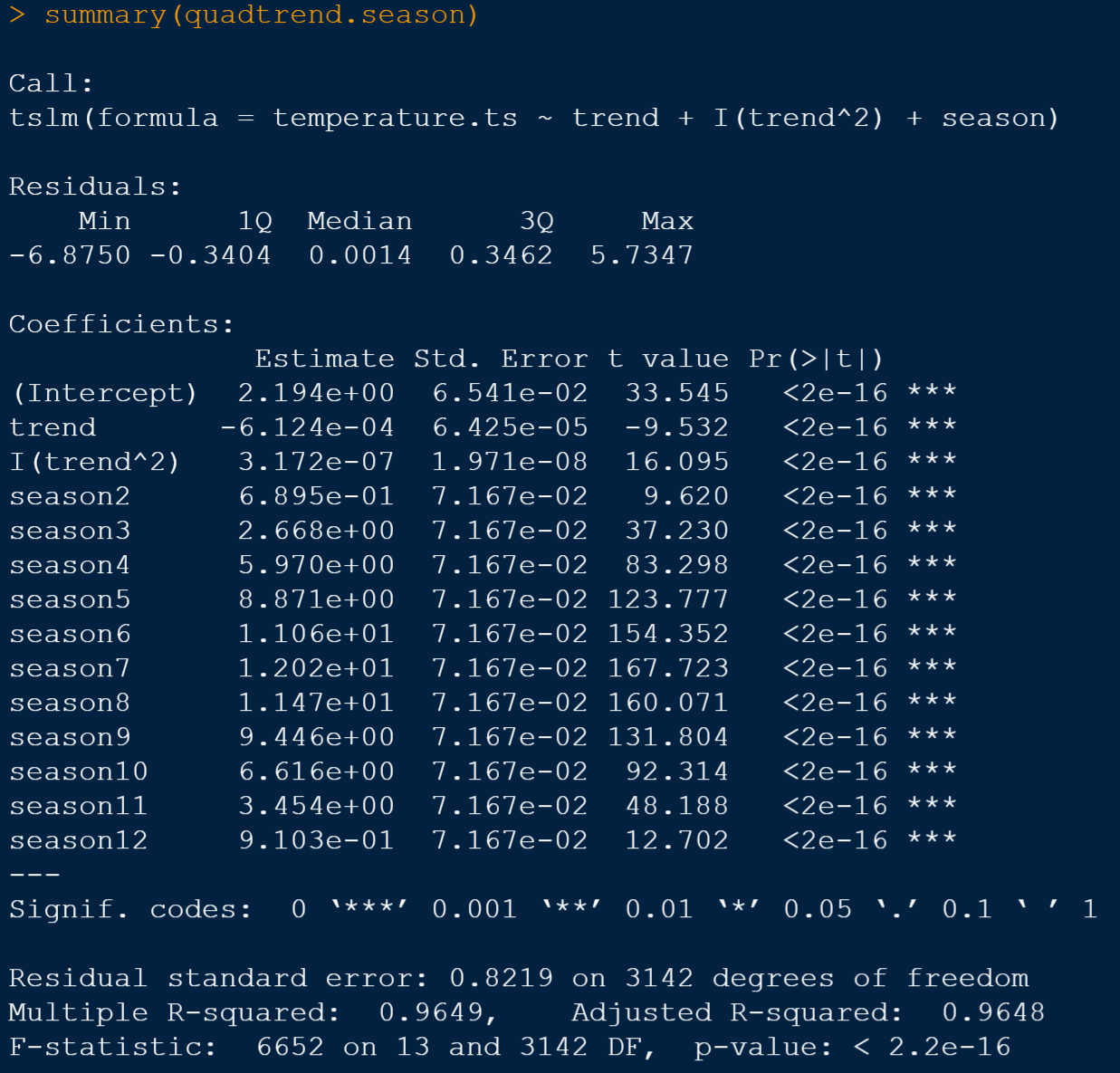
All the forecasts are over-predicting (since the ME is positive). Based on the MAPE metric, the quadratic trend and seasonality and the two level model based on the quadratic trend and seasonality + AR(1) give us the lower RMSE. Since we want to choose the most parsimonious model, we will use the quadratic trend and seasonality.

Hence, I would conclude that the best forecast is provided by the regression model with quadratic trend and seasonality.

The next step would be to fit the quadratic trend + seasonality model on the entire dataset.

## Model Development on Entire Dataset:

After fitting the model on the entire dataset, the details of the model are as follows:



Model Equation:

**yt = 2.194 – 6.12E-4 t + 3.17E-7 t2 + 6.89E-1 D2 + 2.66 D3 + 5.97 D4 + 8.871 D5 + 1.106E+1 D6 + 1.202E+1 D7 + 1.147E+1 D8 + 9.446 D9 + 6.616 D10 + 3.45 D11 + 9.103E-1 D12**

t = Period number

D2 = 1 if Month 2 else equal to 0 (Binary dummy variable for Month 2)

D3 = 1 if Month 3 else equal to 0 (Binary dummy variable for Month 3)

D4 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 4)

D5 = 1 if Month 4 else equal to 0 (Binary dummy variable for Month 5)

..

………..

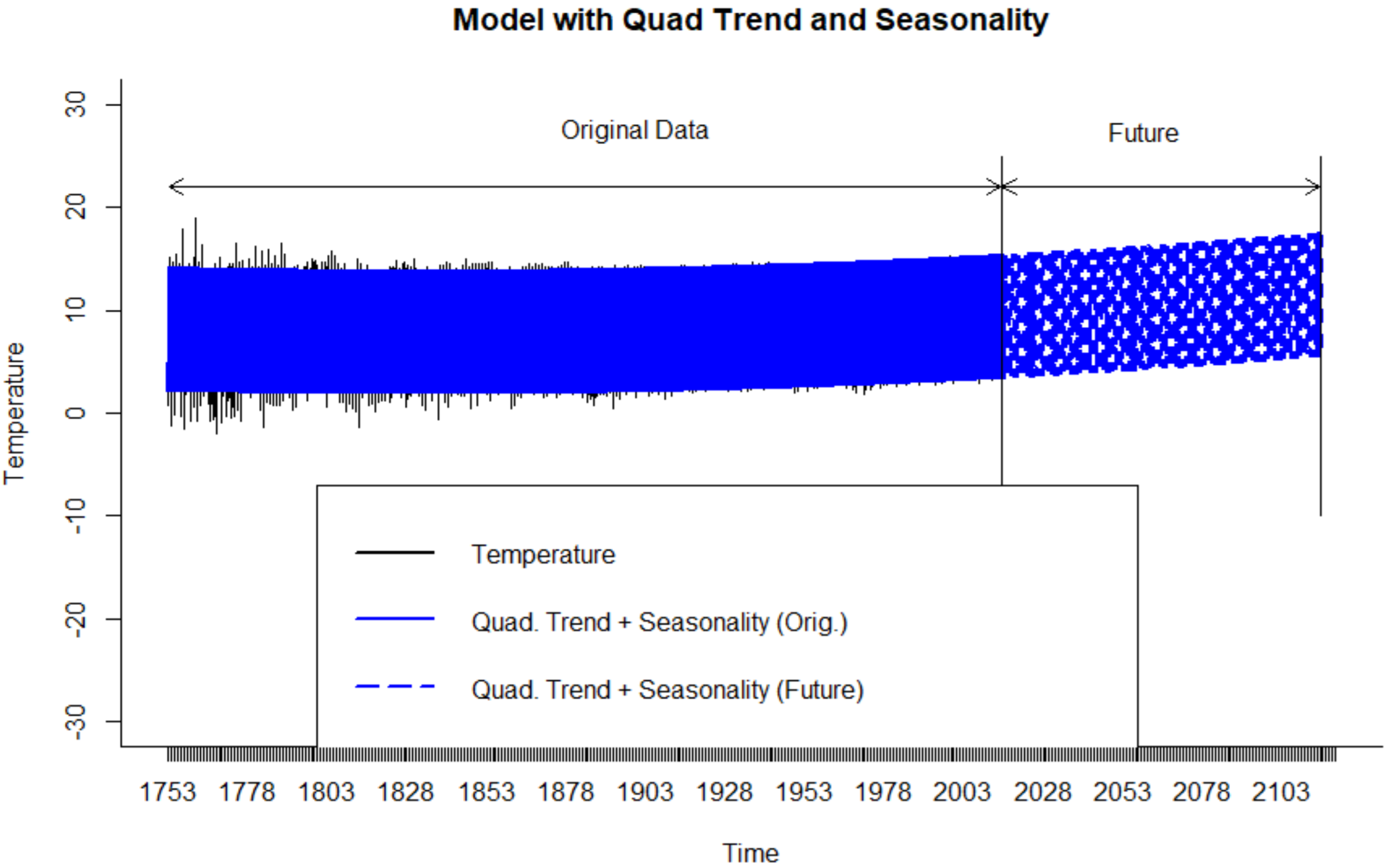
D12 = 1 if Month 12 else equal to 0 (Binary dummy variable for Month 12)

If all D­I = 0 (i = 2, 3, 4, …., 12), then it is Month 1

We will now use this model to predict land temperature data for the next 100 years.

## Conclusion:

The model predicts that there is a going to be an approximate increase of 2 degrees Celsius in the global temperature by the year 2115.



This is even more evident when we plot the average yearly temperature for every 10th year from 1900 to 2110. The temperature increase per decade is in excess of 0.8 degree Celsius every 50 years since 2000 as compared to 0.6 years per degree Celsius previous to 2000.

This is very concerning as this study assumes that the factors influencing the temperature increase will continue to change at the same rate that they have been in the past. In reality, factors influencing the temperature increase could get worse and as a result, we could experience the temperature increase in an accelerated timeframe.



An average increase of 2 degrees Celsius in global temperature would lead to the following effects within the next 100 years:

An increase of 2 degrees Celsius would almost double the water deficit and would lead to a drop-in wheat and maize harvests (food deficit).

Reduced snowpack and shifting rainfall patterns may reduce hydropower in the long run leading to energy deficits

Increase in temperature would lead to melting of the ice-caps causing sea levels to rise leading to coastal cities getting submerged and destroying human property and lives

It is worthwhile noting that all the above analysis has been carried out using historical data without any consideration to factors that could contribute to an temperature increase in the present or the future.

## References:

1. <https://harvardsciencereview.com/2016/12/20/why-should-we-care-about-climate-change/>
2. <https://www.livescience.com/58891-why-2-degrees-celsius-increase-matters.html>
3. <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>
4. BAN 673 Time Series Analytics Lecture Notes (Prof. Z. Radovilsky)(California State University – East Bay)