UNIVERSITY OF CONNECTICUT

OPIM 5671- DATA MINING AND BUSINESS INTELLIGENCE

Sudip Bhattacharjee

Time Series Forecasting Project Report

Climate Change Data from 1750-2015

By:

OPIM 5671 - B12 Group 3

Sparsha Akula

Kanika Gaikwad

Juhi Gupta

Vasudha Mishra

Shubham Deshmukh

TABLE OF CONTENTS

1. INTRODUCTION

Introducing our research paper that tackles one of the most pressing global challenges of our time - climate change. With a goal of predicting future average temperatures of global lands and oceans, we have developed several forecasting models using cutting-edge strategies taught in class such as Exponential Smoothing, Unobserved Components, and ARIMA. Our team put a strong emphasis on the latter model, testing it rigorously with various autoregressive, differencing, and moving average orders to achieve the most accurate predictions possible.

To ensure the credibility of our results, we gathered our data from a reputable source: Data World (https://data.world/data-society/global-climate-change-data), which was originally shared by Berkeley Earth, a non-profit organization dedicated to analyzing land temperature data for climate science. Our dataset contains a wide range of columns that provide detailed information on ocean and land temperatures dating back to 1750. However, due to discrepancies in some columns, we decided to focus on a smaller but more consistent dataset, covering the years from 1951 to 2015.

2. DATA CLEANING AND PREPARATION

We have received the raw dataset from Berkeley Earth data page and Kaggle platform. This dataset includes five .csv files:

a) Global Land and Ocean Average Temperatures
b) Global average land temperature by country
c) Global average land temperature by state
d) Global Land Temperatures By Major City
e) Global Land Temperatures By City

The dataset contains daily records of average, minimum and maximum land and ocean temperatures of city, country and global data from the 1750s.

Since the dataset has directly been taken from the Berkeley Earth page, it was ensured that the data is accurate and consistent. For our data cleaning section, we considered data from 1900 for city data analysis and data from 1951 for global temperature analysis and forecasting.

We removed all the null values from the data. We use python, excel and sas for the data cleaning process. Please find below attached snapshot of python code used to check overall null values present in global dataset, that we have removed.

```
global_temp.isnull().sum()
```

```
dt                                              0
LandAverageTemperature                         12
LandAverageTemperatureUncertainty              12
LandMaxTemperature                           1200
LandMaxTemperatureUncertainty                1200
LandMinTemperature                           1200
LandMinTemperatureUncertainty                1200
LandAndOceanAverageTemperature               1200
LandAndOceanAverageTemperatureUncertainty    1200
dtype: int64
```

To decide on the level of aggregation to be performed in our dataset, we convert daily records to monthly records and sample the data into regular intervals to perform trend, seasonality, autocorrelation and white noise analysis. Since the dataset was from multiple resources, and for multiple locations, we checked for the time zone differences or time stamp errors.

3. DATA EXPLORATION

Our focus while performing time series exploration in the city and the global temperature is to find if there is any trend, seasonality or any relationship between the temperature present in the historical data over past years.

In Data Exploration we performed the following:

1.Seasonality

2. Trend

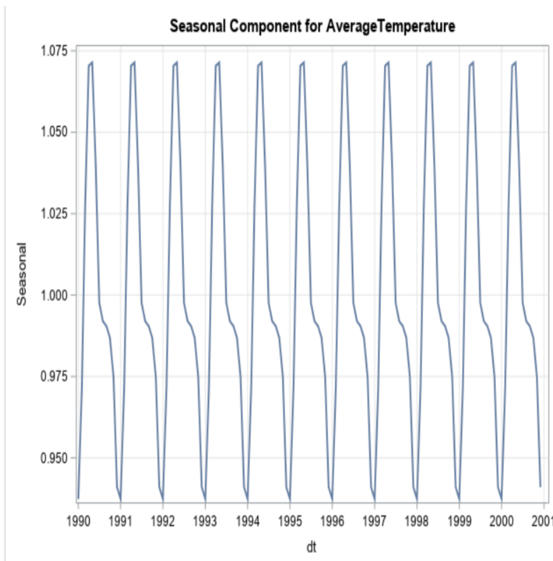
3. Irregular component

4. Autocorrelation

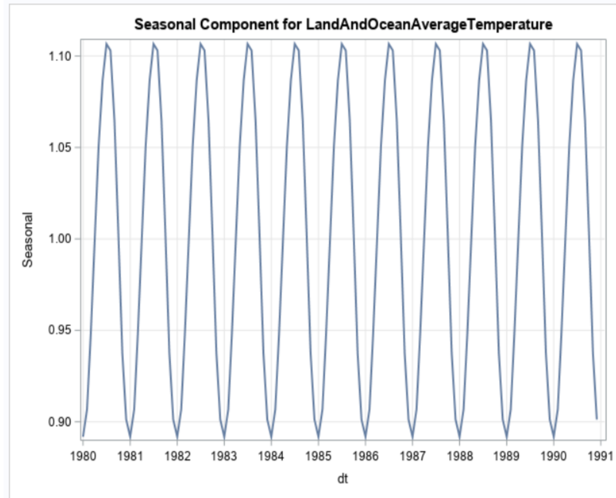
5. White noise

6. Augmented dickey-fuller test

1. **Seasonality:** After examining a time series dataset, we focused on the period between 1980 and 1990 to assess its seasonality. It was not immediately apparent across the entire dataset, but upon closer inspection during that specific timeframe, it became evident that the dataset exhibits a very pronounced and consistent pattern of seasonal fluctuations. This conclusion was drawn based on visual analysis of a graph depicting the data over that specific period of time.

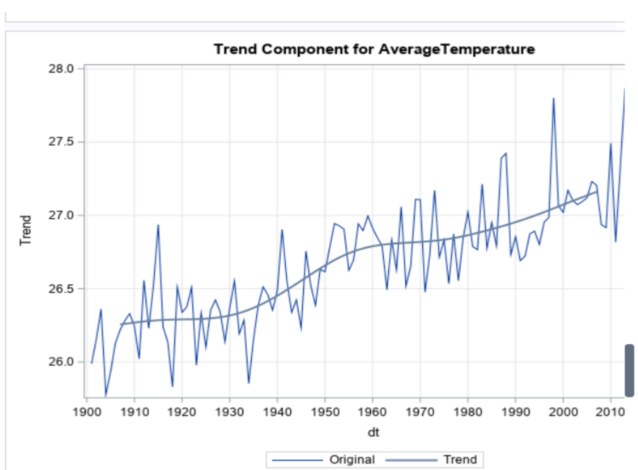


City temperature Seasonality

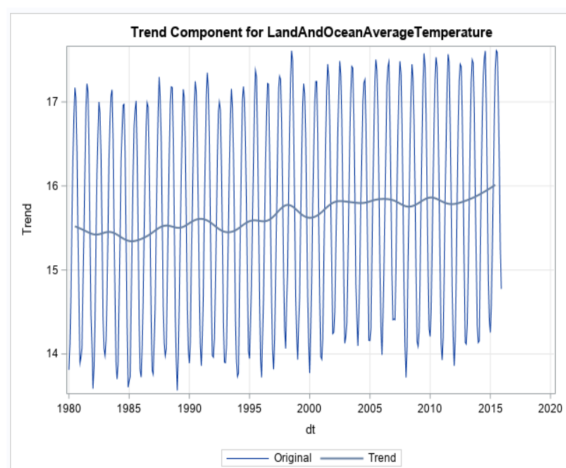


LandandOceanAveragetemperature Seasonality

2. Trend: This time series dataset exhibits a slight positive trend, it indicates that the overall values are gradually increasing over time.

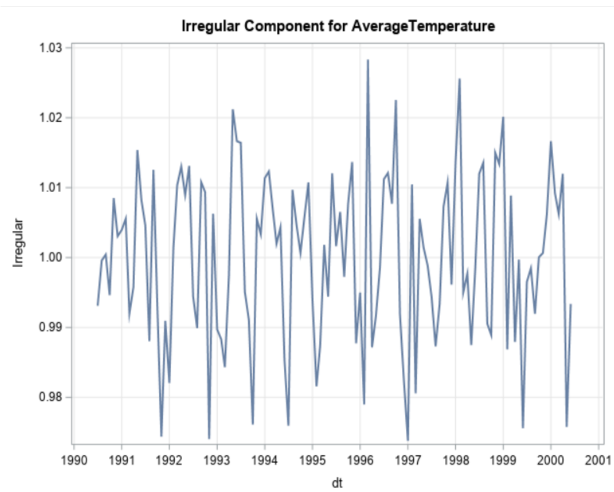
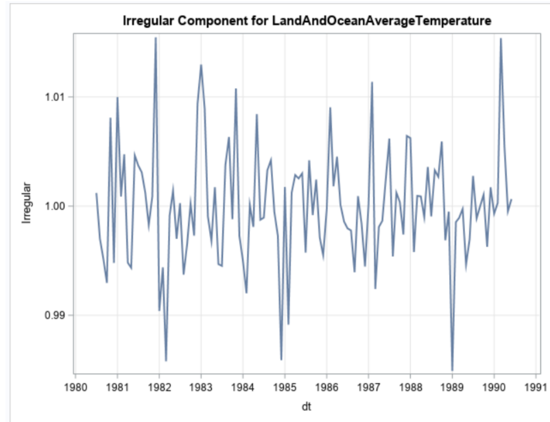


City temperature Trend



LandandOceanAveragetemperature Trend

3. Irregular component: This is the irregular component of the time series forecasting data set as it helps us to better understand the underlying patterns and characteristics of the data, and to develop more accurate models.

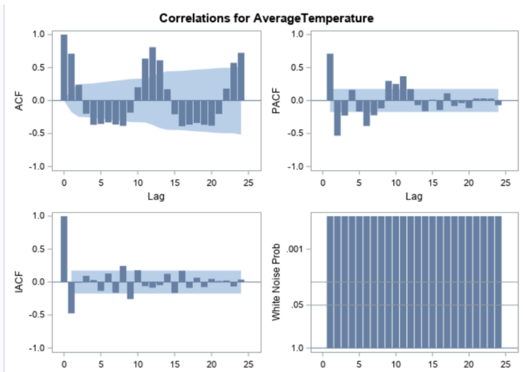
 City: Irregular Component	 LandandOceanAverageTemperature Irregular component

4. **Autocorrelation:** Below graph represents the Autocorrelation of the data. Autocorrelation measures the correlation between a variable and itself at different time lags. Positive autocorrelation at a particular lag indicates that values of the time series tend to be similar at lags of that length, while negative autocorrelation indicates that values tend to be different. The autocorrelation function (ACF) is a tool used to visualize autocorrelation at different lags, and can be used to identify patterns and dependencies in the data. Understanding autocorrelation can help in developing more accurate forecasting models for time series data.

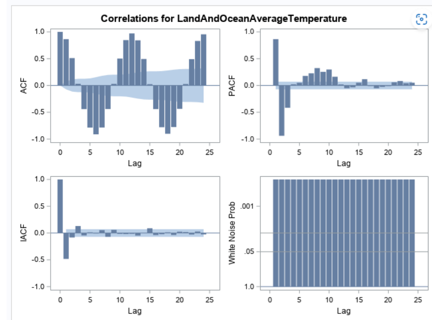
5. **Partial autocorrelation:** We also perform partial autocorrelation analysis to measure the linear relationship between a time series variable and its lagged values, while controlling for the influence of all other intermediate lags. We have also used the PACF plot in the time series exploration section to identify the order of AR (auto regressive) order in our time series analysis.

6. **White Noise Test:** Based on the White Noise Probability graph, our dataset has passed the white noise test. This means that there is no evidence of randomness or uncorrelated variation in our data. As a result, we can confidently say that our time series has some underlying patterns or trends, and we can use these patterns to forecast future values.

Therefore, we can proceed with the development of a time series forecasting model to make predictions about future values with a reasonable degree of accuracy.



City Temperature: Correlations



Global Temperature: Correlations

7. **Augmented dickey-fuller test:** Our ADF test yielded a p-value (Pr<Tau) of less than 0.0001 for both the trend and single mean components. This indicates that the null hypothesis of non-stationarity can be rejected at a statistically significant level. Therefore, we can conclude that the dataset is stationary. In other words, the ADF test results suggest that the time series data does not have a unit root and is stationary, which is a desirable property for many time series analysis techniques.

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.8006	0.5090	-0.62	0.4502		
	1	-1.4869	0.3970	-0.82	0.3613		
	2	-1.1239	0.4522	-0.76	0.3903		
Single Mean	0	-370.770	0.0001	-14.65	<.0001	107.38	0.0010
	1	-1220.49	0.0001	-24.84	<.0001	308.41	0.0010
	2	-2078.74	0.0001	-24.23	<.0001	293.65	0.0010
Trend	0	-395.649	0.0001	-15.21	<.0001	115.63	0.0010
	1	-1362.06	0.0001	-26.19	<.0001	342.91	0.0010
	2	-2839.08	0.0001	-26.31	<.0001	346.08	0.0010

City Temperature

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.5761	0.5538	-0.50	0.4984		
	1	-6.9742	0.0683	-1.88	0.0579		
	2	-1.5722	0.3851	-0.89	0.3330		
Single Mean	0	-105.063	0.0001	-7.55	<.0001	28.49	0.0010
	1	-4260.73	0.0001	-46.81	<.0001	1095.60	0.0010
	2	3412.986	0.9999	-37.86	<.0001	716.59	0.0010
Trend	0	-107.500	0.0001	-7.61	<.0001	29.00	0.0010
	1	-4866.94	0.0001	-49.60	<.0001	1230.02	0.0010
	2	2466.904	0.9999	-45.79	<.0001	1048.48	0.0010

LandandOceanAverageTemperature

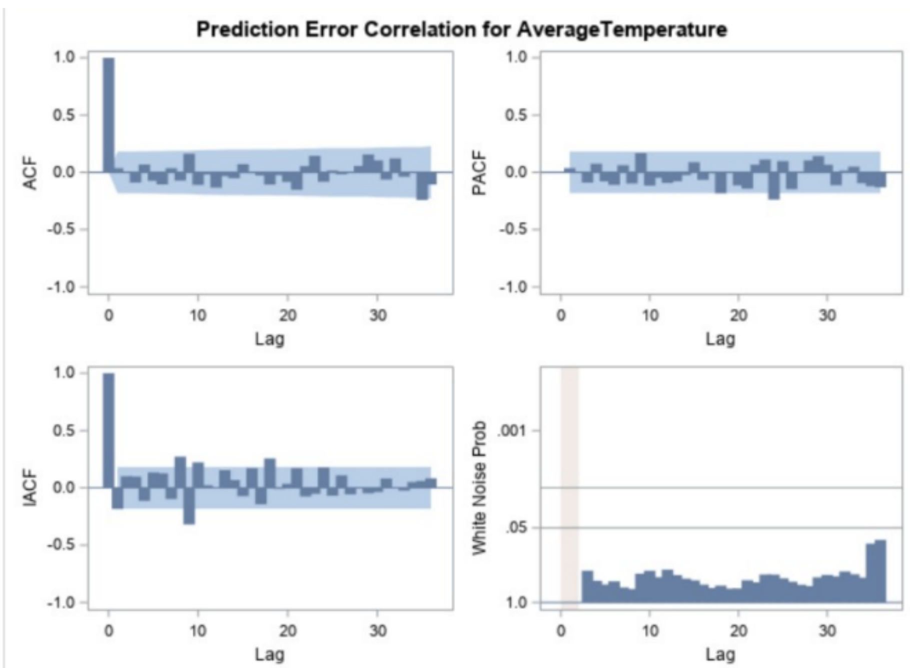
4. Modeling

We first perform temperature analysis and forecast in the city temperature dataset to generalize how temperature will vary in the upcoming 100 years.

After performing time series exploration, we figured out that our dataset contains both trend and seasonality for both city and global temperature analysis. Following this, we start modeling our dataset with an exponential smoothing model with a forecasting model as winter additive model.

4.1 City Temperature analysis:

Exponential smoothing:

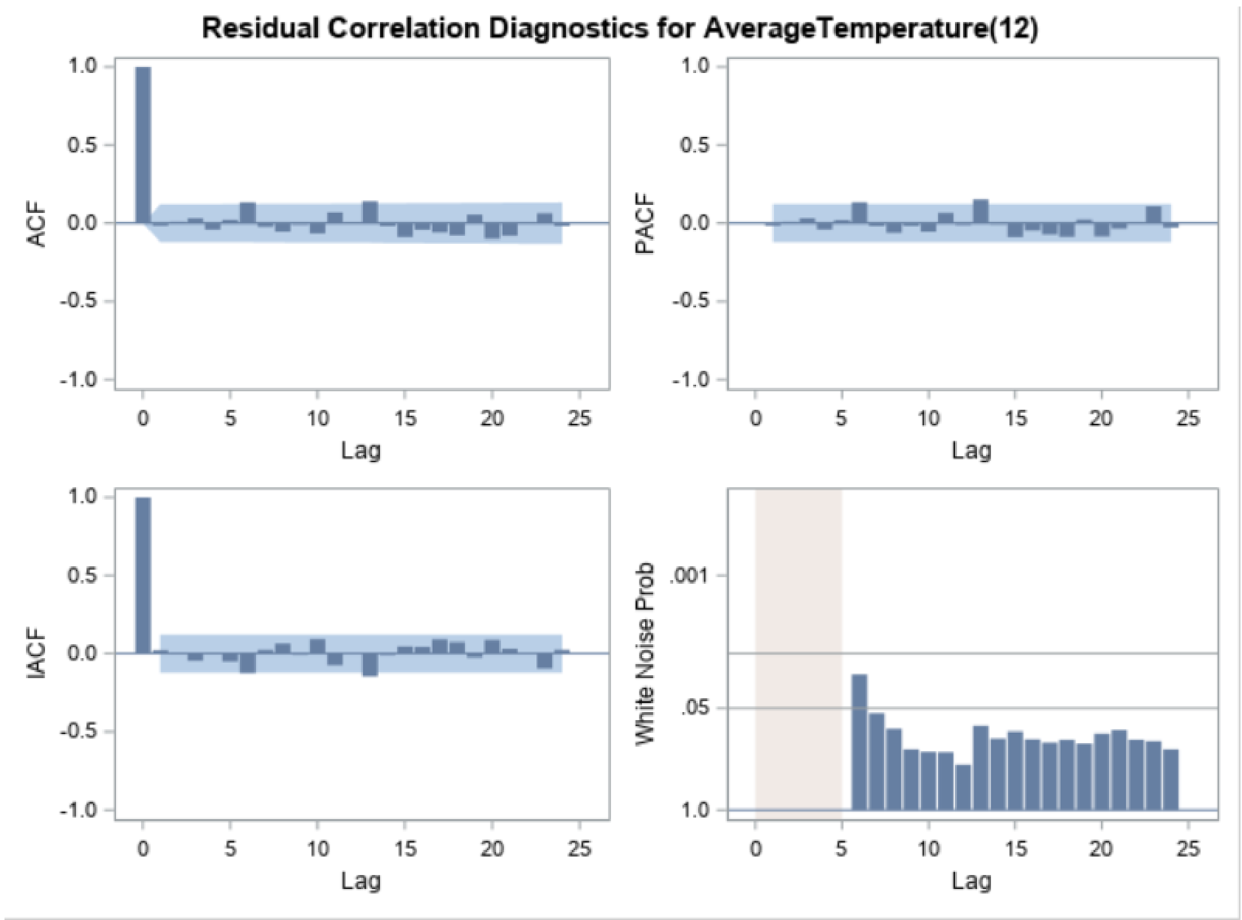


Since, there were white noise after we fit winter additive exponential smoothing model, this can be one of the good model, which indicates that random values in the forecast will not be included. So, we calculated the RMSE, AIC, and SBC values to do model comparison later.

RMSE	MAPE	MAE	AIC	SBC
0.3733290194	1.1026075202	0.2972280417	-232.470838	-226.8958546
0.4263071101	1.2250661606	0.3291073214	-20.46228664	-20.46228664

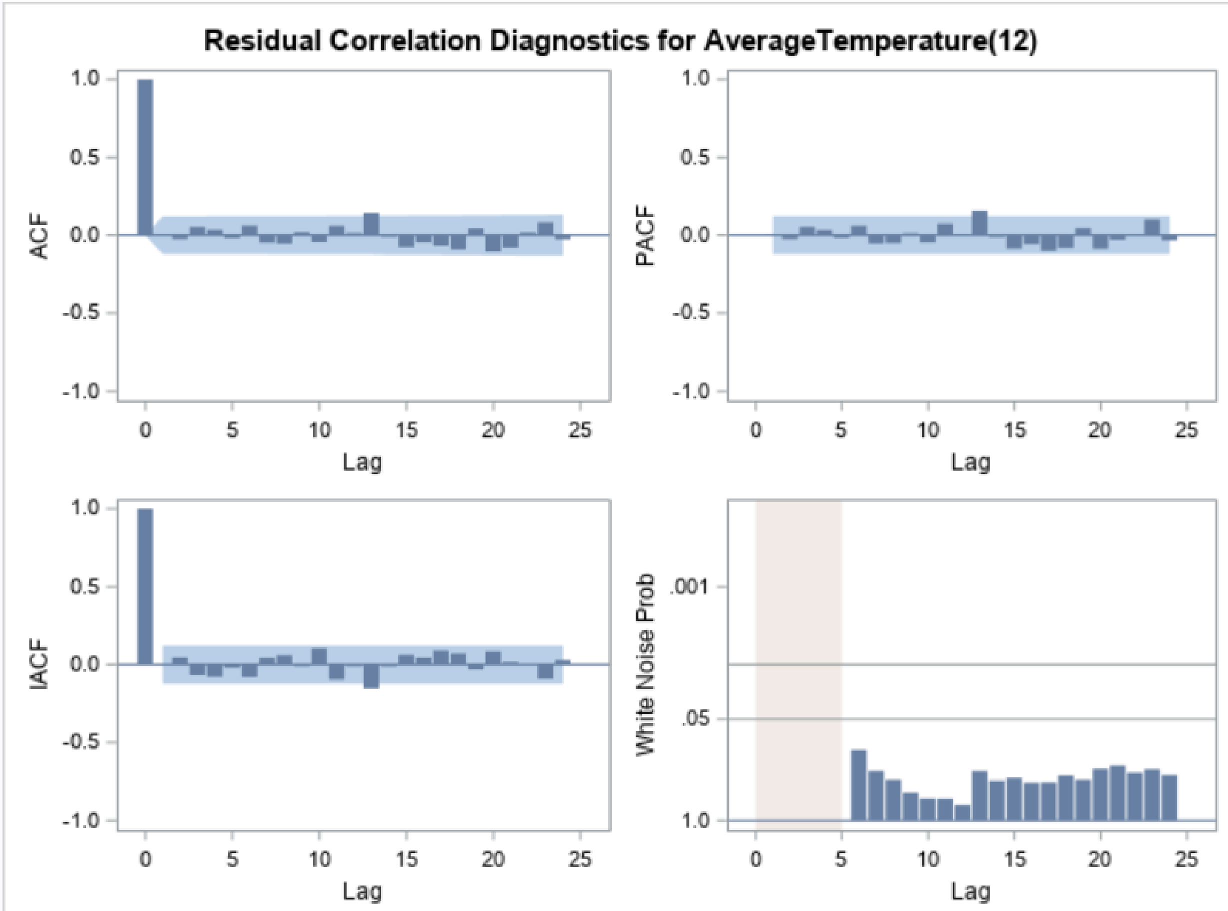
ARIMA: While performing time series exploration and analyzing ACF and PACF plots, we got to know that we could fit in both AR and MA models. Also, since our dataset has seasonality. We would be performing Seasonal ARIMA. Based on the relations between different lags of ACF and PACF values, we have tried various combinations, and were able to get combinations as **A(p,d,q)(P,D,Q)s = (3,0,1)(0,1,1).**

We save the model output as CityModel1 for our model comparison.



AIC	486.7712
SBC	508.3839

ARIMA(2,0,2)(0,1,1)s



The output of the above model is stored in CityModel2 sas file for model comparison.

Using macros code, we have calculated MAPE and RMSE values to verify the model accuracy. We have also analyzed AIC and SBC values for the model comparison later.

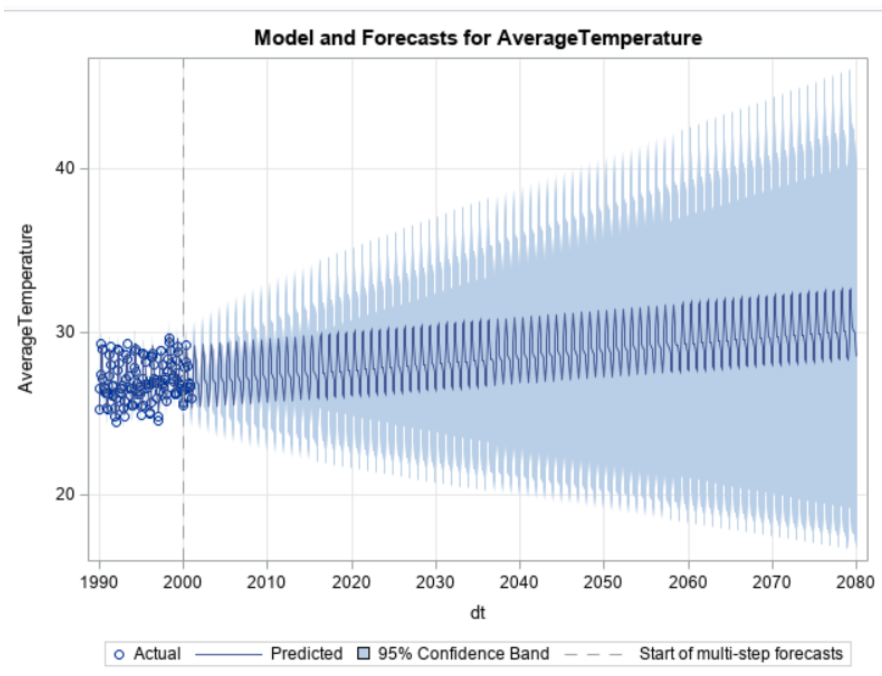
Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
'AverageTemperature'n	work.CityModel1	284	1.56%	0.42178	0.33269	0.57679

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
'AverageTemperature'n	work.CityModel2	284	1.55%	0.41867	0.32652	0.57142

After performing model comparison for both exponential smoothing and SARIMA model for accuracy and model fit, we got below results:

Parameter	Exponential Smoothing	ARIMA (2,0,2)(0,1,1)s
AIC	-20.46	486.77
SBC	-20.46	508.38
MAPE	1.22	1.55
RMSE	0.42	0.57

From above parameter analysis for model comparison and model accuracy, since the values for MAPE, AIC and SBC in exponential smoothing model are smaller than ARIMA model. We can say that the Exponential smoothing model fits better than the ARIMA model. So, we have referred to the time forecast for the winters additive model to see temperature variation in upcoming years.



4.2 Global temperature analysis:

After observing an increasing trend in city-level temperature data, we have expanded our analysis to include global data for average land and ocean temperatures. This approach allows us to examine the historical behavior of temperature trends on a larger scale and make predictions about future temperature increases. Our time series project will involve analyzing this global data to gain insights into the patterns and trends of temperature rise and the potential implications for the future.

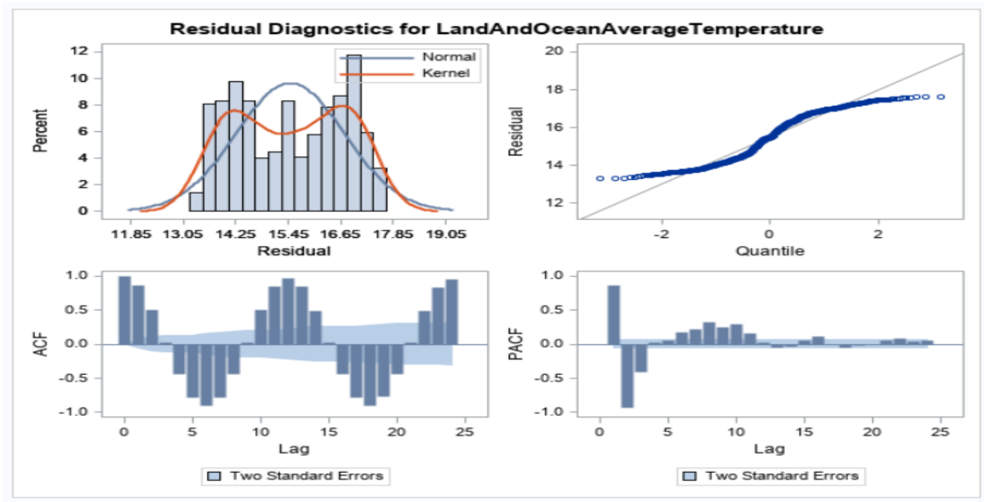
Exponential Smoothing:

Having both trend and seasonality in the global dataset relations, we fitted exponential smoothing winter's additive modeling algorithm to examine the target variable, LandandOceanAverageTemperature.

	AIC	RMSE	AICC	SBC
1	-694.8979815	0.6397169441	-694.8928401	-690.2386876

During our analysis, we conducted a White Noise Probability test on the model residuals and found that they were not distributed as white noise. This indicates that our model had systematic bias and may not be suitable for accurate predictions. Additionally, we noted a high root mean square error (RMSE) of 64%, which suggested that our model was not very accurate. Despite these challenges, our project aimed to identify temperature trends and provide insights that could inform climate change policy and scientific research. To achieve more accurate predictions, we will continue to explore new modeling techniques to advance our understanding of this critical issue.

Unobserved Component Analysis: However, our initial ESM model was not suitable, and we subsequently utilized an unobserved component model for our analysis.



Likelihood Based Fit Statistics	
Statistic	Value
Diffuse Log Likelihood	-3247
Diffuse Part of Log Likelihood	0
Non-Missing Observations Used	780
Estimated Parameters	1
Initialized Diffuse State Elements	0
Normalized Residual Sum of Squares	780
AIC (smaller is better)	6495.6
BIC (smaller is better)	6500.3
AICC (smaller is better)	6495.6
HQIC (smaller is better)	6497.4
CAIC (smaller is better)	6501.3

Likelihood Optimization Algorithm Converged in 0 Iterations.

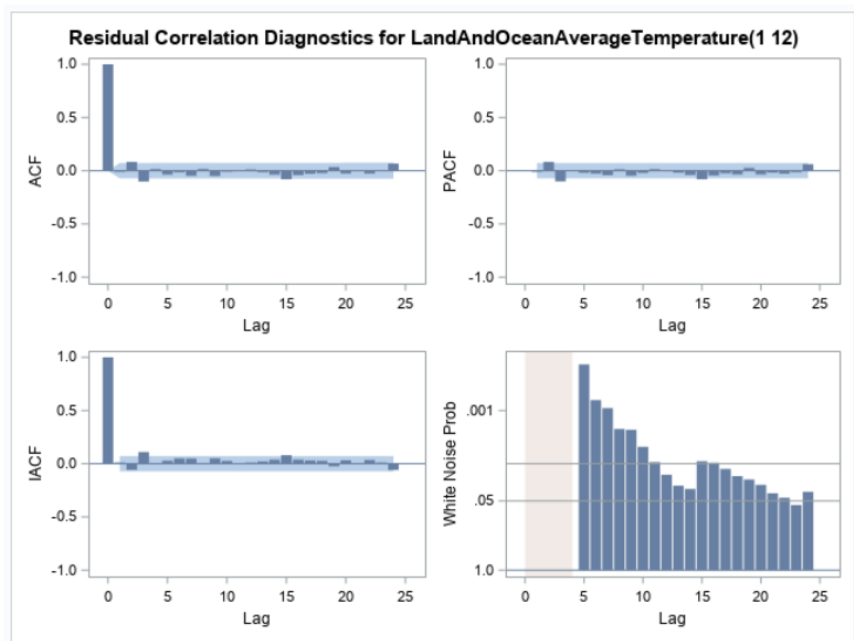
Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	241.59077	12.23342	19.75	<.0001

Fit Statistics Based on Residuals	
Mean Squared Error	241.67242
Root Mean Squared Error	15.54582
Mean Absolute Percentage Error	100.00000
Maximum Percent Error	100.00000
R-Square	-156.86015
Adjusted R-Square	-156.86015
Random Walk R-Square	-589.76306
Amemiya's Adjusted R-Square	-157.26596
Number of non-missing residuals used for computing the fit statistics = 779	

Unfortunately, we discovered that this model was not parsimonious, which was reflected in higher Akaike and Bayesian Information Criterion (AIC and BIC) values than other models we examined. Additionally, the residuals from this model were not independently and identically distributed, further indicating its unsuitability for our dataset. These findings emphasize the importance of careful model selection and the need for robust analysis techniques to identify the most appropriate models for a given dataset. Despite these challenges, our project is committed to advancing our understanding of temperature trends and providing valuable insights that can inform policy and scientific research. We will continue to refine our approach and explore new modeling techniques to achieve more accurate predictions and insights.

ARIMA: We have decided to utilize the ARIMA (AutoRegressive Integrated Moving Average) model to analyze our temperature data. ARIMA is a powerful and widely used time series modeling technique that can capture complex patterns and trends in our dataset.

ARIMA(1,1,1)(1,1,1)s



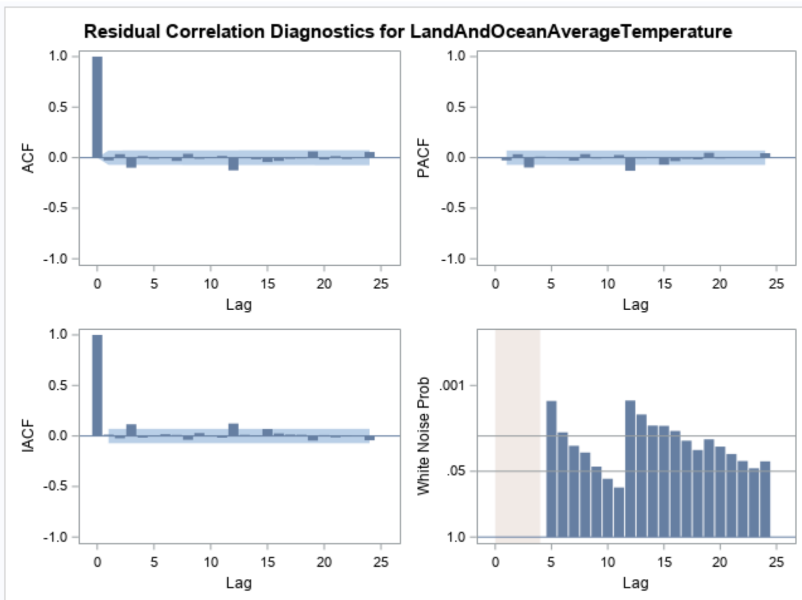
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.00004018	0.0001146	0.35	0.7259	0
MA1,1	0.58683	0.05809	10.10	<.0001	1
MA2,1	0.94733	0.01770	53.51	<.0001	12
AR1,1	0.12711	0.07089	1.79	0.0730	1
AR2,1	-0.12159	0.03770	-3.23	0.0013	12

Constant Estimate	0.000039
Variance Estimate	0.009833
Std Error Estimate	0.099164
AIC	-1333.02
SBC	-1309.8
Number of Residuals	767

Correlations of Parameter Estimates					
Parameter	MU	MA1,1	MA2,1	AR1,1	AR2,1
MU	1.000	0.003	-0.052	0.005	-0.017
MA1,1	0.003	1.000	0.026	0.866	-0.032
MA2,1	-0.052	0.026	1.000	0.001	0.311
AR1,1	0.005	0.866	0.001	1.000	-0.034
AR2,1	-0.017	-0.032	0.311	-0.034	1.000

Based on the AIC and SBC values, the current ARIMA model appears to be the best fit for the dataset thus far, with a low probability of white noise. However, it may still be worthwhile to continue searching for a better model that can provide a more optimal fit.

ARIMA(1,0,1)(1,0,1)s



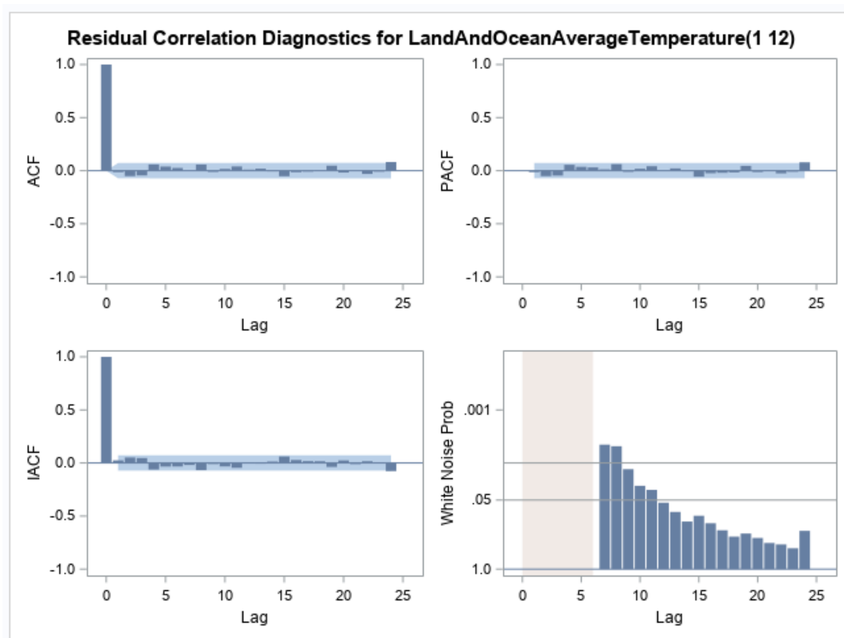
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	15.57411	2.32759	6.69	<.0001	0
MA1,1	0.39989	0.03806	10.51	<.0001	1
MA2,1	0.90917	0.01941	46.83	<.0001	12
AR1,1	0.93540	0.01642	56.96	<.0001	1
AR2,1	0.99995	0.00003291	30380.1	<.0001	12

Constant Estimate	0.000052
Variance Estimate	0.009945
Std Error Estimate	0.099723
AIC	-1302.75
SBC	-1279.45
Number of Residuals	780

Correlations of Parameter Estimates					
Parameter	MU	MA1,1	MA2,1	AR1,1	AR2,1
MU	1.000	0.089	0.095	0.199	0.096
MA1,1	0.089	1.000	0.158	0.544	0.207
MA2,1	0.095	0.158	1.000	0.400	0.663
AR1,1	0.199	0.544	0.400	1.000	0.328
AR2,1	0.096	0.207	0.663	0.328	1.000

The probability of white noise is very low,and the current model did not yield better AIC and SBC values compared to the ARIMA(1,1,1)(1,1,1)s model. Therefore, it would be prudent to reject this model and continue the search for a better fitting model that can accurately capture the underlying patterns in the dataset.

ARIMA(2,1,2)(1,1,1)s



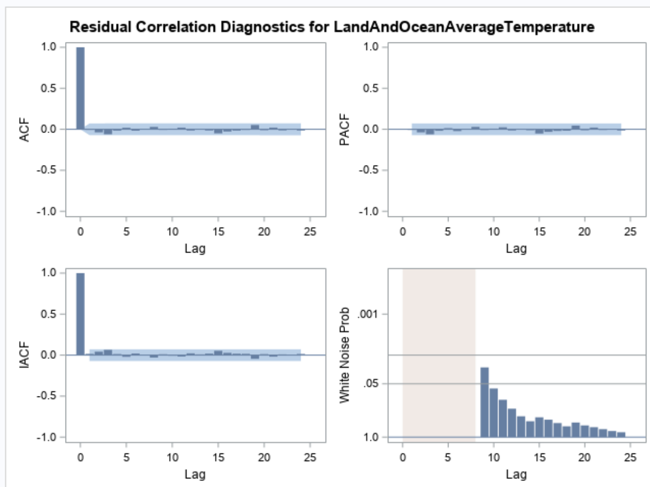
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.00003696	8.36421E-6	4.42	<.0001	0
MA1,1	0.89840	3.06872	0.29	0.7697	1
MA1,2	0.10158	0.32302	0.31	0.7532	2
MA2,1	0.95231	0.01795	53.04	<.0001	12
AR1,1	0.41050	0.12117	3.39	0.0007	1
AR1,2	0.34225	0.08469	4.04	<.0001	2
AR2,1	-0.08967	0.03798	-2.36	0.0182	12

Constant Estimate	9.957E-6
Variance Estimate	0.009344
Std Error Estimate	0.096662
AIC	-1360.41
SBC	-1327.91
Number of Residuals	767

Correlations of Parameter Estimates							
Parameter	MU	MA1,1	MA1,2	MA2,1	AR1,1	AR1,2	AR2,1
MU	1.000	0.372	0.341	-0.006	0.059	-0.014	0.027
MA1,1	0.372	1.000	0.903	-0.086	0.186	-0.066	0.041
MA1,2	0.341	0.903	1.000	-0.103	-0.238	0.335	0.035
MA2,1	-0.006	-0.086	-0.103	1.000	0.026	-0.059	0.311
AR1,1	0.059	0.186	-0.238	0.026	1.000	-0.940	-0.006
AR1,2	-0.014	-0.066	0.335	-0.059	-0.940	1.000	-0.029
AR2,1	0.027	0.041	0.035	0.311	-0.006	-0.029	1.000

The current model has generated superior results with improved AIC and SBC values when compared to all previously tested models. Despite the high probability of white noise, this model is considered the best fit for the dataset. Overall, these findings suggest that the current model should be favored and can be relied upon to accurately capture the underlying patterns present in the data.

ARIMA(2,0,2)(2,0,2)s



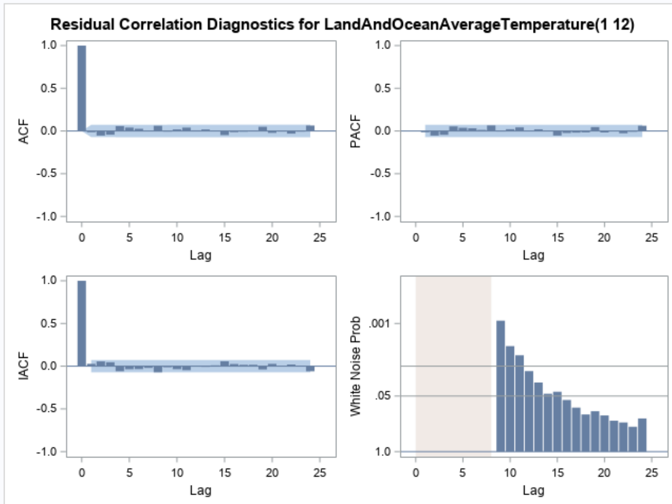
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	15.57972	1.96884	7.91	<.0001	0
MA1,1	-0.39660	0.14340	-2.77	0.0057	1
MA1,2	0.23819	0.08121	2.93	0.0034	2
MA2,1	0.59692	0.22360	2.67	0.0076	12
MA2,2	0.23904	0.20409	1.17	0.2415	24
AR1,1	0.12005	0.13461	0.89	0.3725	1
AR1,2	0.73502	0.12925	5.69	<.0001	2
AR2,1	0.58274	0.21303	2.74	0.0062	12
AR2,2	0.41714	0.21301	1.96	0.0502	24

Constant Estimate	0.000272
Variance Estimate	0.009745
Std Error Estimate	0.098716
AIC	-1316.31
SBC	-1274.37
Number of Residuals	780

Correlations of Parameter Estimates									
Parameter	MU	MA1,1	MA1,2	MA2,1	MA2,2	AR1,1	AR1,2	AR2,1	AR2,2
MU	1.000	0.000	0.064	0.006	0.004	0.001	0.049	-0.003	0.003
MA1,1	0.000	1.000	-0.753	-0.060	0.055	0.972	-0.938	-0.029	0.029
MA1,2	0.064	-0.753	1.000	0.067	-0.051	-0.833	0.897	0.031	-0.031
MA2,1	0.006	-0.060	0.067	1.000	-0.992	-0.068	0.076	0.987	-0.987
MA2,2	0.004	0.055	-0.051	-0.992	1.000	0.063	-0.062	-0.985	0.985
AR1,1	0.001	0.972	-0.833	-0.068	0.063	1.000	-0.970	-0.037	0.037
AR1,2	0.049	-0.938	0.897	0.076	-0.062	-0.970	1.000	0.037	-0.037
AR2,1	-0.003	-0.029	0.031	0.987	-0.985	-0.037	0.037	1.000	-1.000
AR2,2	0.003	0.029	-0.031	-0.987	0.985	0.037	-0.037	-1.000	1.000

Although the current model has a relatively high white noise probability and low AIC and SBC values, it may still provide a good fit for our data. However, it should be noted that the previous model had better AIC and SBC values, indicating that it may have been a stronger candidate for accurately capturing the underlying patterns in the data.

ARIMA(2,1,2)(2,1,2)s



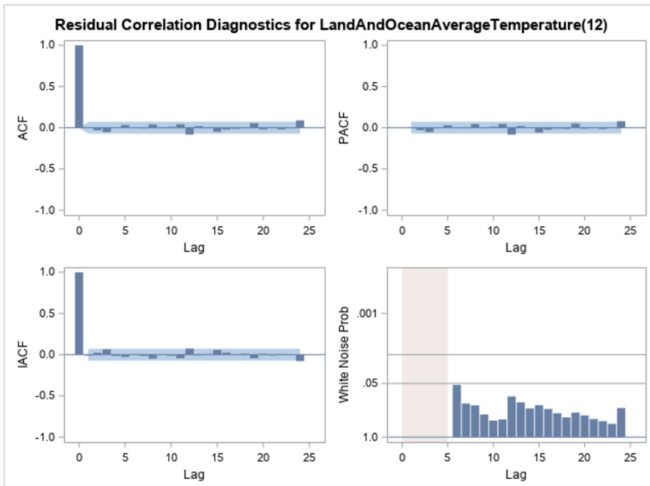
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.00003529	8.78616E-6	4.02	<.0001	0
MA1,1	0.87911	13.08268	0.07	0.9464	1
MA1,2	0.12089	1.57197	0.08	0.9387	2
MA2,1	-0.01418	0.05210	-0.27	0.7855	12
MA2,2	0.92623	0.04880	18.98	<.0001	24
AR1,1	0.39307	0.12112	3.25	0.0012	1
AR1,2	0.35209	0.08416	4.18	<.0001	2
AR2,1	-1.04997	0.06176	-17.00	<.0001	12
AR2,2	-0.06423	0.04096	-1.57	0.1169	24

Constant Estimate	0.000019
Variance Estimate	0.009298
Std Error Estimate	0.096425
AIC	-1360.3
SBC	-1318.52
Number of Residuals	767

Correlations of Parameter Estimates									
Parameter	MU	MA1,1	MA1,2	MA2,1	MA2,2	AR1,1	AR1,2	AR2,1	AR2,2
MU	1.000	0.488	0.486	0.025	-0.028	0.068	-0.014	0.042	0.036
MA1,1	0.488	1.000	0.996	-0.000	-0.058	0.157	-0.035	0.051	0.048
MA1,2	0.486	0.996	1.000	-0.002	-0.060	0.070	0.048	0.050	0.048
MA2,1	0.025	-0.000	-0.002	1.000	-0.748	0.012	-0.029	0.792	0.442
MA2,2	-0.028	-0.058	-0.060	-0.748	1.000	0.005	-0.015	-0.646	-0.232
AR1,1	0.068	0.157	0.070	0.012	0.005	1.000	-0.937	0.001	-0.004
AR1,2	-0.014	-0.035	0.048	-0.029	-0.015	-0.937	1.000	-0.030	-0.037
AR2,1	0.042	0.051	0.050	0.792	-0.646	0.001	-0.030	1.000	0.867
AR2,2	0.036	0.048	0.048	0.442	-0.232	-0.004	-0.037	0.867	1.000

The current model is characterized by a relatively low white noise probability when compared to our best model i.e ARIMA(212)(111)s, and has similar AIC and SBC values. Therefore, it is possible that this model could provide a good fit for our data by effectively capturing the underlying patterns present in the dataset.

ARIMA(2,0,2)(0,1,1)s



Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.01139	0.0014982	7.60	<.0001	0
MA1,1	-0.38092	0.17036	-2.24	0.0254	1
MA1,2	0.22904	0.09228	2.48	0.0131	2
MA2,1	0.95691	0.01791	53.42	<.0001	12
AR1,1	0.11841	0.16207	0.73	0.4650	1
AR1,2	0.69590	0.15128	4.60	<.0001	2

Constant Estimate	0.002114
Variance Estimate	0.009509
Std Error Estimate	0.097512
AIC	-1359.99
SBC	-1332.12
Number of Residuals	768

Correlations of Parameter Estimates						
Parameter	MU	MA1,1	MA1,2	MA2,1	AR1,1	AR1,2
MU	1.000	0.016	0.002	-0.208	0.018	-0.002
MA1,1	0.016	1.000	-0.788	-0.070	0.979	-0.947
MA1,2	0.002	-0.788	1.000	0.119	-0.850	0.913
MA2,1	-0.208	-0.070	0.119	1.000	-0.076	0.102
AR1,1	0.018	0.979	-0.850	-0.076	1.000	-0.971
AR1,2	-0.002	-0.947	0.913	0.102	-0.971	1.000

The ARIMA(2,0,2)(0,1,1)s model has shown exceptional performance in various statistical tests, including the white noise test. Furthermore, the distribution of errors in the model appears to be consistent with the characteristics of white noise. Moreover, when compared with other models that were previously tested, the current model has demonstrated comparatively lower AIC and SBC values, indicating its superior fit to the data. Collectively, these results suggest that the ARIMA(2,0,2)(0,1,1)s model should be favored as it is capable of accurately capturing the underlying patterns in the data.

We have stored the results of the model as work.out for our model comparison code.
We have stored the results of the model as Model3 for our model comparison code.

5. MODEL COMPARISON

Finally we perform model comparison to figure out the accuracy and model fitting with different p,d,q parameters. We have selected MAPE, AIC and SBC values for model accuracy and model fitting comparison output.

The code for model comparison via macros is attached below:

```
%let nhold=24;
%include "C:\Users\admin\Desktop\Husky\sem2\dataset\macros2.sas" / source2;

%accuracy_prep(indsn=STSM.'GLOBALFULL'n,series='LandAndOceanAverageTemperature
'n,timeid='dt'n,
 numholdback=&nhold);

ods noproctitle;
ods graphics / imagemap=on;
ods select none;

proc arima data=Work.preProcessedData plots
    (only)=(series(corr crosscorr) residual(corr normal)
                forecast(forecast) );
        identify var=LandAndOceanAverageTemperature(12);
        estimate p=(1 2) q=(1) (12) method=ML outstat=work.outstat;
        forecast lead=960 back=12 alpha=0.05 id=dt interval=month  out=work.out;
        estimate p=(1 2) (12) q=(1 2) (12) method=ML;
        forecast lead=960 back=12 alpha=0.05 id=dt interval=month  out=work.model2;
        estimate p=(1 2) (12 24) q=(1 2) (12 24) method=ML outstat=work.outstat;
```

```
        forecast lead=960 back=12 alpha=0.05 id=dt interval=month  out=work.model3;
        outlier;
        run;
quit;

ods select all;

%accuracy(indsn=work.out,series='LandAndOceanAverageTemperature'n,timeid='dt'n,
 numholdback=&nhold);
%accuracy(indsn=work.model2,series='LandAndOceanAverageTemperature'n,timeid='dt'n,
 numholdback=&nhold);
%accuracy(indsn=work.model3,series='LandAndOceanAverageTemperature'n,timeid='dt'n,
 numholdback=&nhold);


data work.allmodels;
 set work.out
 set work.model2
 set work.model3
run;

proc print data=work.allmodels label;
 id series model;
run;
```

The output for model comparison are:

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
'LandAndOceanAverageTemperature'n	work.out	780	0.52%	0.079434	0.010365	0.10181

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
'LandAndOceanAverageTemperature'n	work.model2	780	0.51%	0.077226	.009847661	0.099235

Series	Model	Holdback Periods	MAPE	MAE	MSE	RMSE
'LandAndOceanAverageTemperature'n	work.model3	780	0.61%	0.093210	0.029144	0.17072

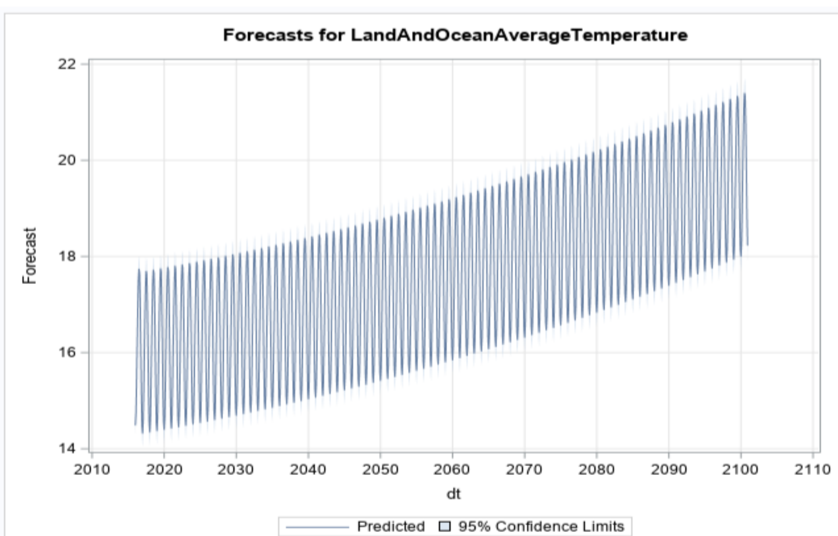
From above comparison of MAPE and RMSE values, we can generalize that model workout, i.e. ARIMA(2,0,2)(0,1,1)s has lowest MAPE values and lowest AIC and SBC values as well, which makes it the best model.

So, we used ARIMA(2,0,2)(0,1,1)s to forecast temperature for next 100 years to analyze temperature variation in the upcoming years.

6. Model Result:

	dt	LandAndOceanAverageTemperature	FORECAST
764	01AUG2014	17.607	17.463698034
765	01SEP2014	16.975	16.947960192
766	01OCT2014	16.029	16.104626708
767	01NOV2014	14.899	15.042736548
768	01DEC2014	14.41	14.323028276
769	01JAN2015	14.255	14.207748918
770	01FEB2015	14.564	14.544580263
771	01MAR2015	15.193	15.173363348
772	01APR2015	15.962	16.026982686
773	01MAY2015	16.774	16.740246738
774	01JUN2015	17.39	17.365110759
775	01JUL2015	17.611	17.668730607
776	01AUG2015	17.589	17.543440304
777	01SEP2015	17.049	16.972784785
778	01OCT2015	16.29	16.133444842
779	01NOV2015	15.252	15.224634006
780	01DEC2015	14.774	14.57578223
781	01JAN2016	.	14.485767271
782	01FEB2016	.	14.738470042
783	01MAR2016	.	15.340529133
784	01APR2016	.	16.173406677

	dt	LandAndOceanAverageTemperature	FORECAST
1780	01APR2099	.	19.701510024
1781	01MAY2099	.	20.456755694
1782	01JUN2099	.	21.063690152
1783	01JUL2099	.	21.339159411
1784	01AUG2099	.	21.268181244
1785	01SEP2099	.	20.680236637
1786	01OCT2099	.	19.796771714
1787	01NOV2099	.	18.798762338
1788	01DEC2099	.	18.164900266
1789	01JAN2100	.	17.998656245
1790	01FEB2100	.	18.26684973
1791	01MAR2100	.	18.914350067
1792	01APR2100	.	19.764199071
1793	01MAY2100	.	20.5194817
1794	01JUN2100	.	21.126453116
1795	01JUL2100	.	21.401959333
1796	01AUG2100	.	21.331018124
1797	01SEP2100	.	20.743110476
1798	01OCT2100	.	19.859682511
1799	01NOV2100	.	18.861710094
1800	01DEC2100	.	18.22788498



To determine the increase in temperature by the end of 2100, we subtract the average temperature in 2015 (represented as "AvgTemp(2015)") from the average forecasted temperature for 2100 (represented as "AvgForecastedTemp(2100)"). This calculation gives us the difference between the two temperatures, which is equal to the projected increase in temperature over the time period.

Using the given values, we can perform the calculation as follows:

Increase in temperature = AvgForecastedTemp(2100) - AvgTemp(2015)

Increase in temperature = 21.12564223 - 15.0023833

= 6.123258

Therefore, the increase in temperature by the end of 2100 is equal to 6.123258 degrees.

7. CONCLUSION AND RECOMMENDATIONS

In conclusion, our time series forecasting project has revealed a sobering outlook on the future of our planet's temperature. Our analysis has shown that if current conditions persist, we can expect a significant increase of 6.123258 degrees Fahrenheit in both land and ocean temperatures by the year 2100. This rise in temperature will have far-reaching and severe impacts on countries around the world.

The impact of this rise in temperature will be widespread and varied, from the melting of polar ice caps and rising sea levels to more frequent and severe natural disasters such as droughts, heatwaves, hurricanes, and flooding. These changes will have a profound effect on many sectors, including agriculture, tourism, energy, and transportation, and could result in significant economic and social disruptions.

However, there is hope. Through continued efforts to reduce global warming pollutants and promote sustainable energy and transportation, we can work towards mitigating the impacts of climate change. By taking steps towards green infrastructure, investing in clean energy, and promoting responsible environmental policies, we can work to minimize the adverse effects of global warming on our planet and create a sustainable future for generations to come.

Overall, our time series forecasting project has demonstrated the pressing need for action to address the issue of global warming. While the challenges ahead are significant, we are confident that with the right steps and policies, we can work towards creating a better, more sustainable future for all.

Mitigating the Effects of Global Warming:

To limit the impact of global warming and the rise in temperature that we have forecasted, a range of actions can be taken at different levels, including individual, community, and government levels. Here are some possible actions:

Reduce energy consumption: Individuals can reduce energy consumption in their homes and workplaces by turning off lights and electronics when not in use, using energy-efficient appliances and equipment, and insulating their homes to reduce heating and cooling needs.

Promote renewable energy: Governments can promote the use of renewable energy sources such as solar, wind, and hydro power, by offering incentives and subsidies to individuals and companies that invest in clean energy.

Encourage sustainable transportation: Governments can encourage the use of public transportation, carpooling, biking, and walking, by investing in public transportation infrastructure and creating bike lanes and pedestrian walkways.

Promote sustainable agriculture: Governments can promote sustainable agriculture practices that reduce the use of chemicals and promote soil conservation, while individuals can support local farmers and buy organic and locally grown food.

Implement policies to reduce greenhouse gas emissions: Governments can implement policies such as carbon taxes, cap-and-trade systems, and regulations that limit greenhouse gas emissions from industries, power plants, and vehicles.

Raise awareness: Education and awareness campaigns can be implemented to raise public awareness about the impact of global warming and the actions that can be taken to reduce it.

These actions can work together to limit the impact of global warming and help create a more sustainable future for all.

8. APPENDIX AND REFERENCES

1. Intergovernmental Panel on Climate Change (IPCC): The IPCC is a United Nations body that provides regular scientific assessments of the state of knowledge on climate change, its impacts, and potential response options. Their website has links to their latest reports, including the IPCC Sixth Assessment Report released in 2021. https://www.ipcc.ch/
2. National Oceanic and Atmospheric Administration (NOAA): NOAA is a US government agency that conducts research and provides information on weather, climate, oceans, and coasts. Their website has a section on climate that includes data and information on climate science, impacts, and trends. https://www.noaa.gov/topics/climate
3. US Environmental Protection Agency (EPA): The EPA is a US government agency responsible for protecting human health and the environment. Their website has a section on climate change that includes information on the science of climate change, impacts, and actions being taken to address it. https://www.epa.gov/climate-change
4. https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data
5. https://berkeleyearth.org/global-temperature-report-for-2022/