Time Series Analysis on Climate Variables

Vraj Thakkar (202103052),* Vatsal Shah (202103022),† and Dhruv Shah(202103017)‡
Dhirubhai Ambani Institute of Information & Communication Technology, Gandhinagar, Gujarat 382007, India
SC475, Time Series Analysis

In this project, we will analyse the time series data of an important climate variable which is the **Temperature of the Indian subcontinent** over the past 120 years. The dataset used is publicly available on the official website of the Indian Meteorological Department (IMD)

I. INTRODUCTION

Since the past few decades, the world has witnessed extreme weather conditions which may be due to various factors such as industrialization, globalization, and increased CO_2 emissions, leading to global warming and rising surface temperatures. Understanding these climate patterns has become crucial for the global community to ensure the sustainability of life on our planet.

In this project, we have attempted to analyse the monthly temperature data of the Indian Sub-continent over the time period of 1901-2021 and aimed to derive some meaningful inferences from our analysis.

II. THE DATASET

The Dataset consists of monthly temperature data of the Indian subcontinent between the time period of $1901 - 2021$. In this dataset, each row consists of the temperature in each month of that particular year. Fig. 1 shows a snapshot of the dataset. The last column represents the annual mean temperature over the years.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2	1901	19.32	20.89	24.95	28.22	29.76	29.85	28.24	27.33	27.23	26.33	22.92	20.05	25.42
3	1902	20.17	21.58	25.73	28.15	30	29.47	27.99	27.71	26.76	25.33	22.43	19.77	25.42
4	1903	19.28	20.71	23.92	27.67	29.47	29.53	28.32	27.16	27	25.69	22.01	19.3	25.01
5	1904	19.19	20.32	24.41	28.11	29.17	28.8	27.36	27.26	26.84	25.67	22.16	19.86	24.93
6	1905	18.34	18.37	23.15	26.26	29.73	29.87	28.13	27.65	27.16	26.35	23.24	19.79	24.84
7	1906	19.05	20.45	23.59	28.04	30.55	29.01	27.95	27.19	26.99	25.81	23.04	20.47	25.18
8	1907	20.28	20.75	23.53	27.1	28.95	28.91	28.1	26.89	27.06	26	23.16	19.48	25.02
9	1908	19.2	21.03	24.32	28.72	29.54	29.58	27.45	26.86	26.7	25.43	22.07	19.12	25
10	1909	19.38	20.79	25.08	26.95	29.4	28.45	27.12	26.87	26.83	25.64	23.1	19.96	24.96
11	1910	19.3	21.17	24.5	27.66	29.71	28.59	27.42	27.03	26.72	25.21	21.81	19.01	24.84

FIG. 1: Monthly Temperature Data. *Source: IMD*

We observe an upward trend in the Annual Mean Temperature over the past decades, which seems evident after the 1980s, as seen in Fig.2, which can be used to argue the global warming hypothesis. In further sections, we will attempt to analyse this trend in the time series invoking the concepts of stationarity, AR Models and ARIMA models.

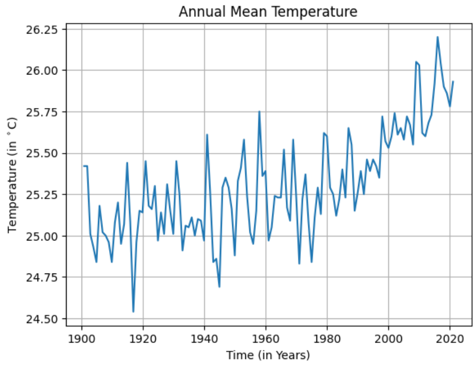

FIG. 2: Annual Mean Temperature

For efficient analysis of our data, we have now created a Monthly Time Series from the dataset using Python. Fig.3 is a part of the dataset from the four year period from $1901 - 1904$. In the subsequent sections, we perform our analysis on both, the annual mean temperature as well as the monthly temperature data.

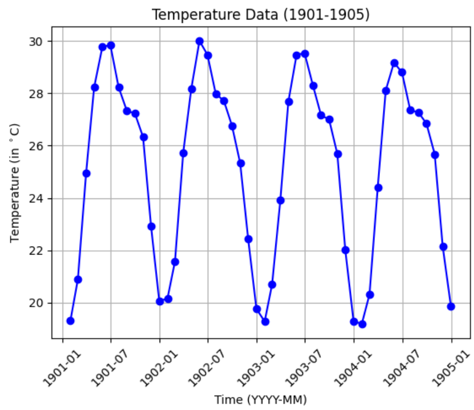

FIG. 3: Monthly Data for the Four Year Period (1901-1904)

From empirical observation of Fig.3, we can say that there is some seasonal nature present in the data. Hence, we will now attempt to analyse the seasonality of the dataset.

*Electronic address: 202103052@daiict.ac.in
†Electronic address: 202103022@daiict.ac.in
‡Electronic address: 202103017@daiict.ac.in

III. ARIMA MODEL FOR ANNUAL MEAN TEMPERATURE DATA

The plot in fig 2 depicting the annual mean temperatures clearly exhibits an upward trend, suggesting a non-stationary process, such as a random walk. Therefore, our initial approach involves detrending the data through first-order differencing to achieve stationarity. Fig.4 shows the detrended time series.

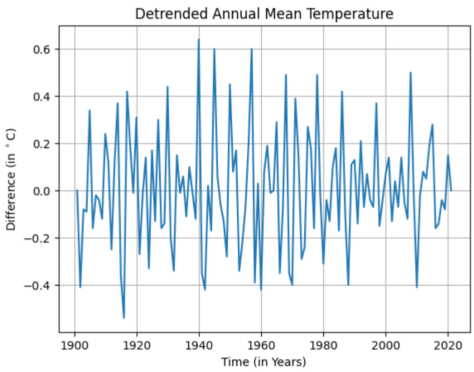

FIG. 4: Detrending the Annual Mean Temperature Using Differencing

Even after detrending the Annual Mean temperature series, we see significant auto-correlations till lag 2, (see fig.5 for ACF).

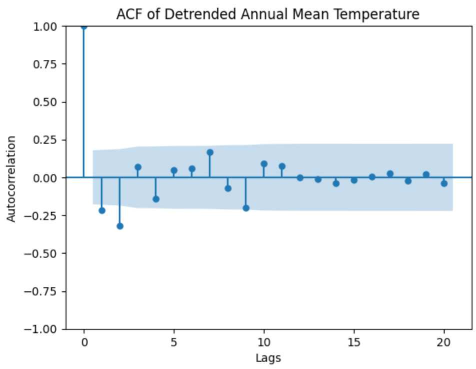

FIG. 5: ACF of the Detrended Time Series

These correlations motivate us to model our data in the form of *Autoregressive (AR)* and *Autoregressive Moving Average (ARMA)* models. Since we observed significant correlations till lag 2, we used an $AR(2)$ model. This can be mathematically represented as:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = \epsilon_t \tag{1}$$

where ϕ_1 and ϕ_2 are the parameters. Using Python libraries to fit the $AR(2)$ model on the dataset, we obtain the parameter values as, $\phi_1 = -0.30$ and $\phi_2 = -0.39$. Now the characteristic equation would be:

$$(1 + 0.30L + 0.39L^2)x_t = \epsilon_t \tag{2}$$

where L is the Lag Operator such that $x_{t-1} = Lx_t$. Checking for the Causality conditions for an $AR(2)$ process, we get $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$. Therefore the process is Causal. Fig.6 shows the AR(2) fit on the time series, we can observe that this gives a low MSE and fits well.

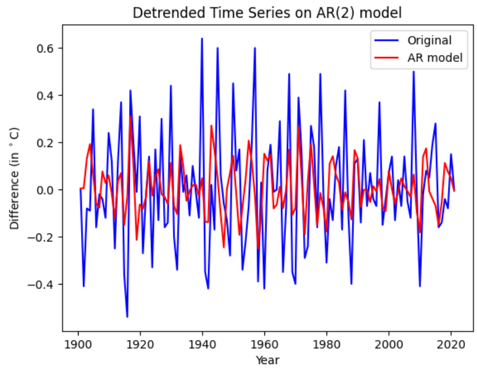

FIG. 6: Fitting the AR(2) Model on the Detrended Annual Mean Temperature Series

IV. ARIMA MODEL FOR MONTHLY TEMPERATURE DATA

We now work on the monthly dataset from 1901 − 2021 and use the same techniques as above to model our data (a part of the dataset is plotted in Fig.3). Here, we observe a seasonality in our data, hence we first remove the trend and the seasonality from the dataset, and then attempt to model the residual data.

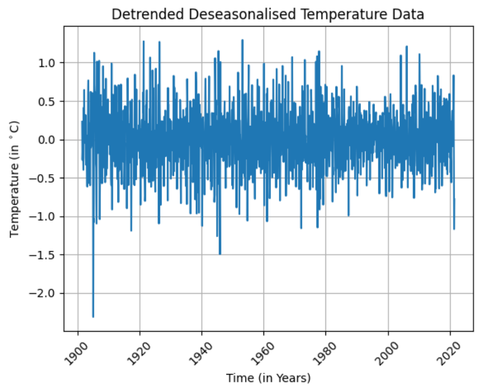

FIG. 7: Detrended Deseasonlised Monthly Temperature Timeseries

When we plot the ACF for the residual data we observe a damped oscillatory ACF, see Fig.8, indicating the presence of periodicity and stationarity in the monthly temperature data.

From fig.8 we also observe significant autocorrelations up to Lag 7, hence it is intuitive to use the *ARIMA(p,d,q)*

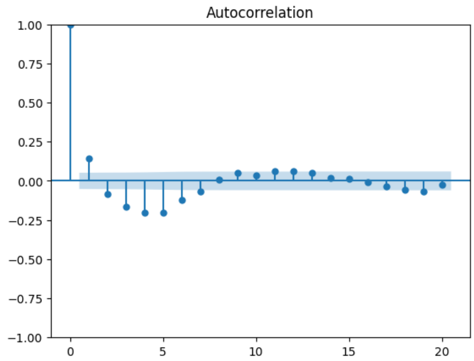

FIG. 8: ACF of the Detrended and Deseasonalised Monthly Data

model. Now we check different orders of AR and MA models and use the model with the least mean squared error.

Attempting to fit $MA(q)$ models for $q = 1, 2, ., 6$, we obtain non-invertible MA models hence, we take $q = 0$.

For $AR(p)$ models, we obtain the mean squared error for $p = 1, 2, ..8$ which we plot in Fig.9. We choose the value $p = 2$ due to the observation of a significant dip in the error in going from $p = 1$ to $p = 2$, we do not choose higher orders as they might overfit on our dataset. Therefore, we choose *ARIMA(2,1,0)*.

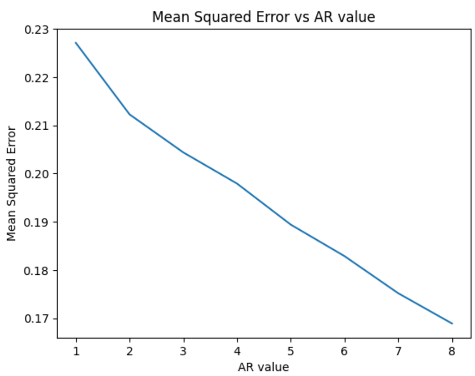

FIG. 9: Mean Squared Error of each Order (p)

Now, the equation of *ARIMA(2,1,0)* process in terms of the lag operator L is given as

$$(1 - \phi_1 L - \phi_2 L^2)(1 - L)x_t = \epsilon_t \tag{3}$$

Using Python to evaluate the coefficients, we get, $\phi_1 = -0.46$ and $\phi_2 = -0.25$. Fig.10 shows the predictions of the *ARMA(2,1,0)* model on the Detrended and Deseasonalised time series.

Checking for the Causality conditions for an $AR(2)$ process, we get $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$. Therefore the process is causal.

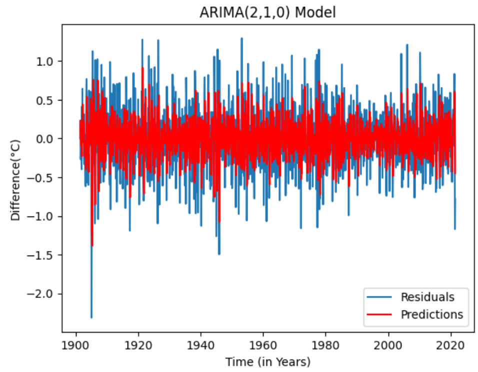

FIG. 10: Fitting the ARIMA(2,1,0) Model on the Detrended, Deseasonalised Time Series

Fig.11 is a snapshot describing the fit of ARIMA(2,1,0) models on the first few years of the monthly data, we observe that it turns out to be a good fit.

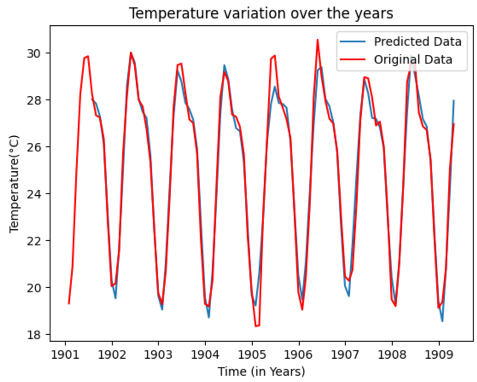

FIG. 11: ARIMA(2,1,0) Model on a section of the entire data

V. SEASONALITY SHIFT ANALYSIS

We made an attempt to understand the phenomenon of seasonal shifts in our data. We tried to observe if there was any trend that the climate conditions we had in a particular month, for instance in May 1901 had shifted to June by the 1990s, in other words we wanted to analyze if the seasons had shifted by a certain period of time over the years. Our approach to the problem was that we took a window of twelve months from January to December of each year and tried to find its correlation with February(current year) to January(next year) of other years. We tried for different combinations of years to see if there was any correlation, to find how much seasonal shift there is between two consequent years. However, all of them showed no significant trends indicating the absence of seasonal shifts. Even if there was a positive trend for a short period of time, it would quickly decay to less positive or negative autocorrelation values. This indicated that there is no seasonal shift on an average in the data and this is also physically evident. The unclear results could also be due to the type of data, as in if we had daily temperatures we would have been able to check if there is any shift even by a few days, but in monthly data, it is trivial that there wouldn't be a shift of month in the seasons.

VI. FUTURE WORK

- Introduce more rigorous concepts of Spectral Analysis to analyse the underlying seasonal nature in the dataset.
- We can use analyse another climate variable such as rainfall or ocean temperature.
- Analyse whether it is correlated to temperature.

(The Video Link for the presentation is here, Google Collab link to the code).