Investigating Temperature Distribution of Kathmandu City Using Markov Chain Monte Carlo (MCMC) Approach

Ram Krishna Pudasaini (Roll: 24)

Abstract

The primary goal of this research study is to understand the temperature distribution of Kathmandu City at a 2 meter range from the surface and forecast the temperature of Kathmandu city. By analyzing daily temperature data from 1982 to 2023, different trends in temperature distribution are explored. By using MCMC methodologies, specifically the Markov Chain, the goal is to predict the temperature at 2 meter range for Kathmandu city. Here we train the markov model with data up to 2020 and prediction is done in the testing data from 2021 to 2023. By comparing the observed and predicted temperature for test data the result shows that using the markov chain method the temperature is forecasted with accuracy of 95.64%.

Keywords: MCMC Methods, Markov Chain, Trend Prediction

Introduction

A. Background of Kathmandu City:

At the elevation of approximately 1400 meters above the sea level within the latitude 27.700769, and the longitude 85.300140 lies the Kathmandu city, the capital city of Nepal. With a rich cultural heritage, vibrant city life, and rapid urbanization, Kathmandu City serves as the political, economic, and cultural hub of Nepal. However, like many urban centers, Kathmandu faces various environmental challenges, including air pollution, rapid population growth, and climate change impacts.

B. Importance of Studying Temperature Distribution:

Temperature is a key climatic variable that influences various aspects of urban life, including public health, energy consumption, agriculture, and infrastructure planning. Understanding the distribution of temperature within Kathmandu City is crucial for enhancing the city's resilience, improving quality of life and identifying extreme temperature events and minimizing risk associated with temperature extremes. Governments can also use this knowledge for initiating climate adaptation strategies plans in advance.

C. Overview of Markov Chain Monte Carlo (MCMC) Approach:

The Markov Chain Monte Carlo (MCMC) approach is a powerful statistical technique used for sampling from complex probability distributions. By iteratively generating samples from a target distribution, MCMC methods allow researchers to estimate parameters, make predictions, and conduct inference in a wide range of applications, including climate science.

Markov chain is the probabilistic model describing a sequence of possible events where the probability of each event depends only on the state attained in the previous event. In simpler terms, the next step in the sequence depends only on the current state, not on the states that came before it. In this study, we employ MCMC techniques to analyze the temperature distribution of Kathmandu City and forecast the temperature for 2024 using available time series data.

D. Objectives of the Study:

The primary objectives of this study are as follows:

- To analyze distribution of temperature within Kathmandu City based on available time series data from 1982 to 2023.
- To perform the statistical analysis of the temperature variable.
- To understand the correlations of temperature with other variables
- To understand the monthly average temperature distribution of Kathmandu city.
- To identify the extreme temperature points of Kathmandu city by analyzing the historical data and extreme temperature threshold.
- To investigate and forecast the long-term temperature trends in Kathmandu City using Markov Chain Monte Carlo (MCMC) methods.

Literature Review

A. Previous Studies on Temperature Distribution in Kathmandu:

Previous studies have investigated the temperature distribution in Kathmandu City using various methods and datasets. Several research papers have focused on analyzing temperature trends, seasonal variations, and extreme events based on meteorological station data and remote sensing observations. For example, studies have examined the urban heat island effect in Kathmandu City and its impacts on local climate and human health[1]. Additionally, researchers have explored the [2] Grid based temperature and relative humidity distribution map of the Kathmandu valley. These studies provide valuable insights into the climatic characteristics of Kathmandu City and serve as a basis for further analysis in this study.

B. Challenges and Limitations in Studying Urban Temperature Distribution:

Studying urban temperature distribution poses several challenges and limitations, including data availability, spatial heterogeneity, and methodological uncertainties. Obtaining high-quality temperature data at fine spatial and temporal resolutions can be challenging, especially in rapidly urbanizing areas like Kathmandu City. Additionally, urban areas exhibit complex temperature patterns influenced by factors such as land use, vegetation cover, and anthropogenic heat emissions, which may not be adequately captured by existing models and observational networks. Moreover, methodological challenges related to spatial interpolation, model validation, and uncertainty propagation require careful consideration when analyzing urban temperature distribution. Despite these challenges, advances in data collection methods, modeling techniques, and computational resources have improved our ability to study and understand urban temperature dynamics, highlighting the importance of continued research in this area.

Methodology:

A. Data Collection

The climate data for Kathmandu utilized in this study was sourced from the NASA Langley Research Center (LaRC) POWER Project. This project is funded through the NASA Earth Science/Applied Science Program and provides access to comprehensive climate data through the Data Access Viewer API. The specific dataset used in this analysis is the NASA/POWER CERES/MERRA2 Native Resolution Daily Data.

Data Source:

Project: NASA Langley Research Center (LaRC) POWER Project

Data Set: NASA/POWER CERES/MERRA2 Native Resolution Daily Data

Data Access Viewer API:

The Data Access Viewer API allows for programmatically accessing and retrieving climate data based on specified geographic coordinates. For this study, the latitude and longitude coordinates corresponding to Kathmandu were used to extract the relevant climate data.

Temporal Coverage:
The dataset encompasses a temporal span from January 1, 1982, to December 31, 2023. This extended time range enables a comprehensive analysis of climate patterns and trends over several decades, facilitating a robust understanding of long-term climate variability in the Kathmandu region.

Data Parameters:
The climate data includes a wide range of parameters that provide insights into various aspects of the local climate conditions but for our study purpose we only considered a few variables: 2 meter surface temperature (T2M), Humidity (QV2M), Precipitations (PRECTOTCORR), and Surface Pressure (PS) that represents the daily data of Kathmandu city.

B. Algorithms

Markov Chain Algorithms
The Markov Chain algorithm, named after Russian mathematician Andrey Markov, is a stochastic process that follows the Markov property, which states that the future state of the system depends only on the current state and not on the sequence of events that preceded it.

Using the same principle, a Markov chain based on the historical data of Kathmandu City is created. Here are the mandatory steps that is performed for predicting the temperature of Kathmandu city using Markov Chain algorithm:

Loading and Discretizing Data: First, the historical temperature data is loaded to a DataFrame. The temperature data is then discretized into bins to facilitate the Markov Chain modeling. The number of bins is specified by the num_bins variable.

Training the Markov Chain Model: The train_markov_chain function is used to train the Markov Chain model based on the discretized temperature data. This function computes the transition matrix, which represents the probabilities of transitioning between different temperature states.

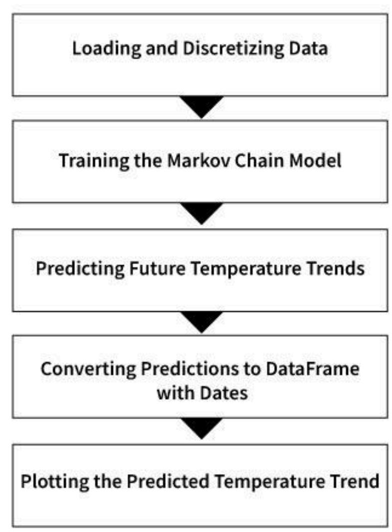
Predicting Future Temperature Trends: The predict_temperature_trend function generates a sequence of future temperature states using the trained Markov Chain model. The initial state for prediction is randomly chosen from the set of possible temperature states.

Converting Predictions to DataFrame with Dates: The predicted temperature trend is converted into a DataFrame with corresponding dates, starting from January 1, 2024. The length of the prediction is set to 365 days.

Plotting the Predicted Temperature Trend: Finally, the predicted temperature trend is plotted as a line graph, with dates on the x-axis and temperatures on the y-axis. This provides a visualization of the forecasted temperature changes over time.

Metropolis Hasting Algorithm Overview
The Metropolis-Hastings algorithm is a Markov chain Monte Carlo (MCMC) method used to generate samples from a target probability distribution.

Block Diagram



Key Finding and Results

Dataset Representations

Head representations of the data.

	YEAR	DOY	T2M	T2MWET	TS	T2M_RANGE	T2M_MAX	T2M_MIN	QV2M	RH2M	PRECTOTCORR	PS	WS2M	WS2M_MAX	WS2M_MIN	WS2M_RANGE	WS10M	WS10M_MAX	WS10M_MIN	WS10M_RANGE
0	1982	1	10.93	4.36	8.14	11.92	18.41	6.48	3.78	41.69	0.0	87.60	1.12	3.20	0.16	3.03	1.67	4.32	0.30	4.02
1	1982	2	10.20	3.57	7.24	11.12	17.26	6.15	3.54	41.00	0.0	87.45	1.17	3.52	0.08	3.45	1.66	4.73	0.07	4.66
2	1982	3	11.08	4.43	8.07	11.27	17.80	6.52	3.78	41.19	0.0	87.49	1.01	3.05	0.04	3.02	1.45	4.05	0.01	4.05
3	1982	4	13.26	6.25	10.29	10.85	20.05	9.19	4.15	39.25	0.0	87.71	1.22	3.06	0.30	2.77	1.76	4.08	0.44	3.64
4	1982	5	14.07	8.05	11.19	10.42	20.92	10.49	5.25	46.62	0.0	87.67	1.68	4.02	0.66	3.36	2.51	5.52	1.07	4.45

Extracted Columns

After formatting the date, dropping unnecessary column the head representation looks like below:

	Date	T2M	QV2M	PRECTOTCORR	PS
0	1982-01-01	10.93	3.78	0.0	87.60
1	1982-01-02	10.20	3.54	0.0	87.45
2	1982-01-03	11.08	3.78	0.0	87.49
3	1982-01-04	13.26	4.15	0.0	87.71
4	1982-01-05	14.07	5.25	0.0	87.67

Our column of interest in the dataset is primarily date and temperature but we also want to figure out the statistical significance of these variables and understand the descriptive behavior of these variables in our datasets.

	count	mean	std	min	25%	50%	75%	max
T2M	15340.0	19.202557	5.031258	4.12	14.7975	20.66	23.15	30.11
QV2M	15340.0	9.522055	5.390868	0.98	4.6400	8.00	15.38	19.59
PRECTOTCORR	15340.0	2.502519	5.642385	0.00	0.0000	0.06	2.22	91.66
PS	15340.0	87.268512	0.403068	85.92	86.9500	87.29	87.60	88.44

Fig: Statistical Analysis

From the statistical analysis the mean temperature is 19.20 and the standard deviation is 5.03. The minimum temperature (2 meters above surface) is 4.12 degree centigrade and the maximum temperature of the considered latitude and longitude of Kathmandu valley is 30.11 degree centigrade.

By plotting the correlation matrix we tend to understand the correlation of the temperature with the variables like precipitation , humidity and surface pressure.

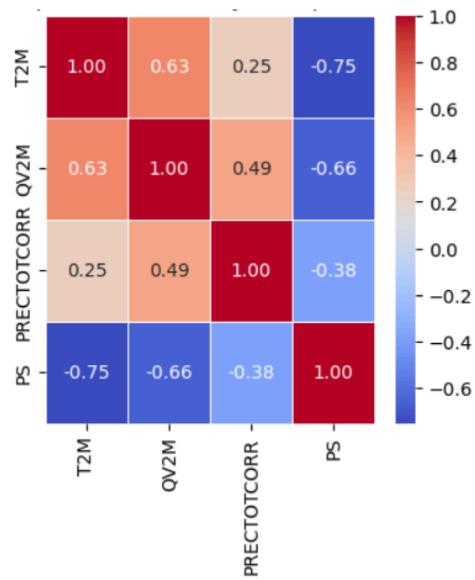


Fig: Correlation Matrix

Here we can see that temperature is positively correlated with humidity and precipitation and negatively correlated with pressure. Thus only two variable humidity and precipitation is influencing the temperature trend.

The trend of the temperature of Kathmandu city over the year 1982 to 2023 is shown in the figure 5. The plot represents the daily temperature from the year 1982 to 2023.

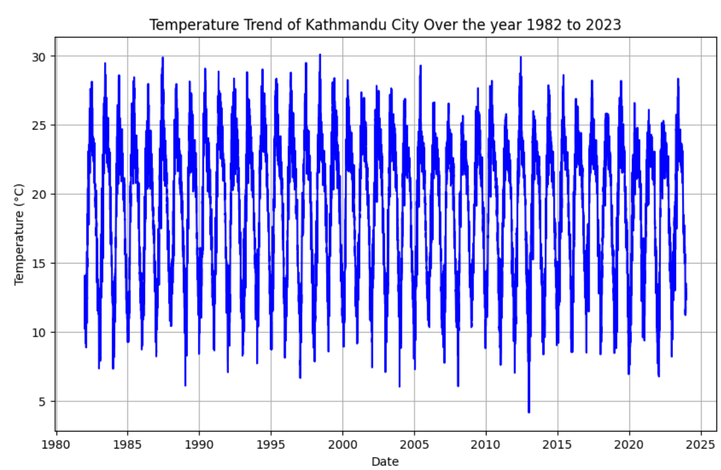


Fig 5: Temperature trend of Kathmandu City

The yearly average mean of 2 meters range temperature of Kathmandu City over the year is given in figure 6.

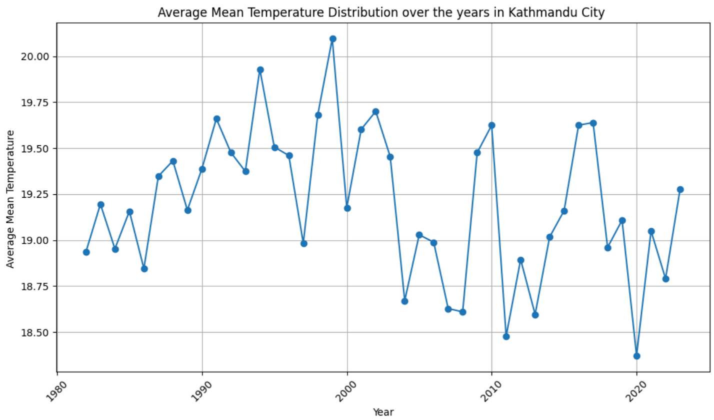


Fig 6: Average Mean temperature over the year.

From the plot the mean temperature is high in 1999 and comparatively lower than other temperatures in 2020.

The percentage change in the temperature trend for the year 1982 to 2023 is given in the figure 7.

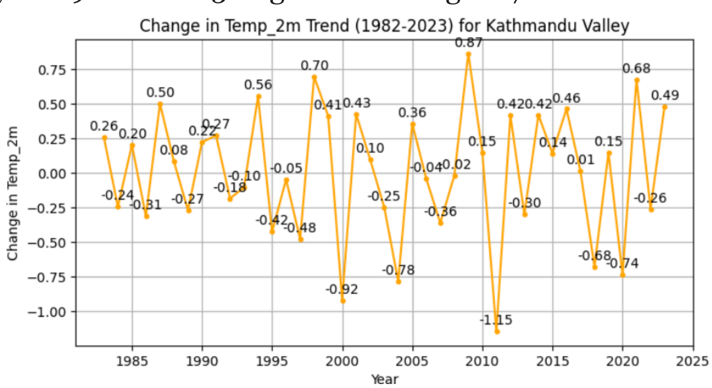


Fig7 : Change in mean temperature trend

From this plot we figure out that moving from 2022 to 2023 the mean temperature of Kathmandu valley has increased by 0.49.

Finally the overall histogram distribution of the temperature is given in figure 8. This plot shows that the temperature distribution over the year is left skewed. Temperatures that range from 20 °c to 25 °c have the higher contribution in the histogram where as the there is still some (around 4°c) low temperature occurrence and high temperature occurrence (around 30°c) in Kathmandu city .

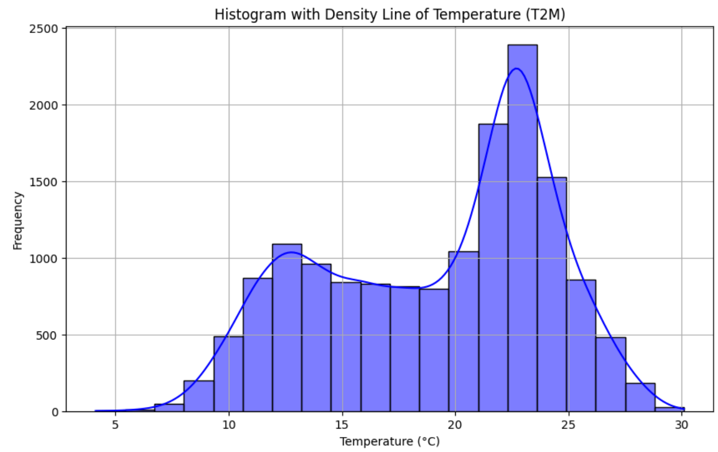


Fig 8: Histogram of Temperature Distribution

In order to figure out the average monthly wise trend of the temperature distribution for a specific year we can use the historical data for the temperature. Here the average monthly distribution of temperature at 2 meter range for the year 2023 is given in figure 9. Similarly we can plot and compare it to any other year which is represented in the figure 10.

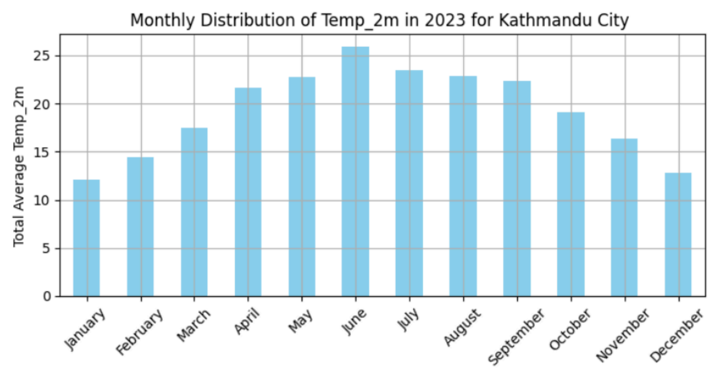


Fig 9: Monthly average temperature of Kathmandu city in the year 2023.

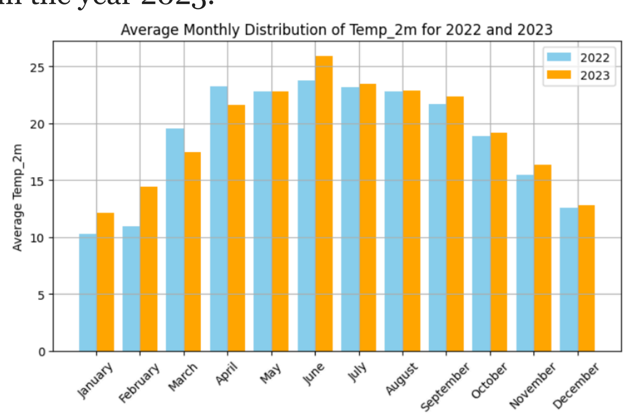


Fig 10: Comparison of Temperature

Similarly, we can compare the temperature to any year from 1982 to 2023. The plot below compares the monthly mean temperature of 2000 and 2023.

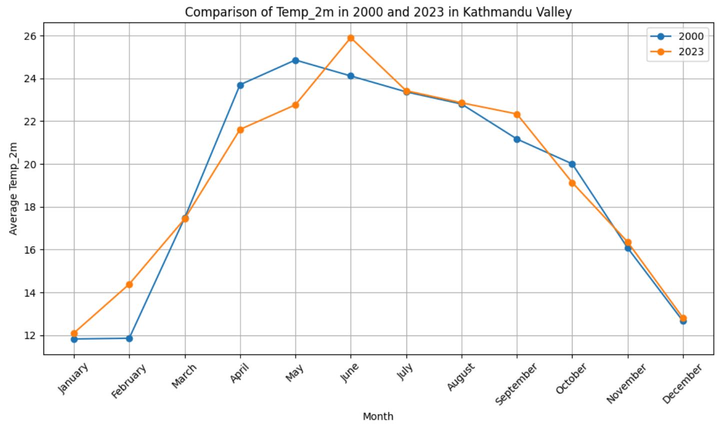


Fig 11: Comparison of Temperature for the year 2000 & 2023

The monthly histogram distribution of the temperature data is represented in figure 12.

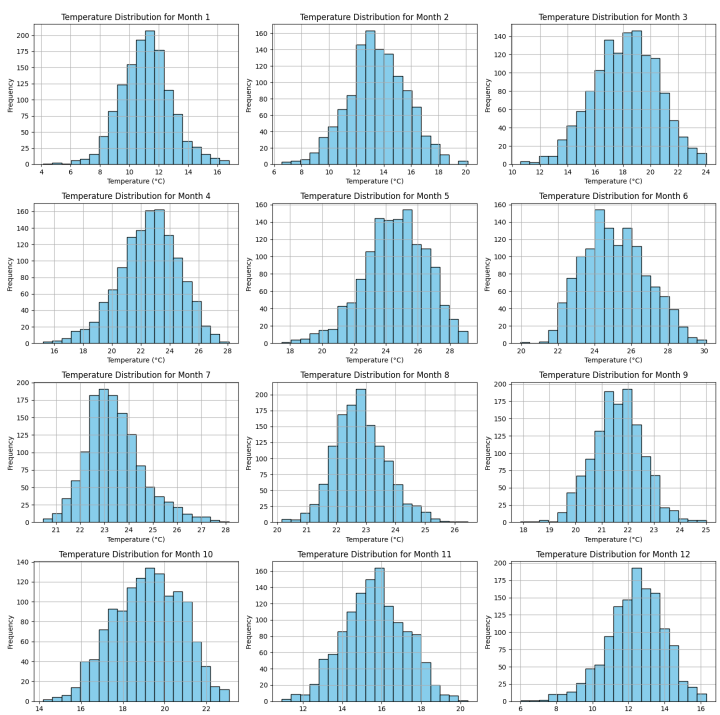


Fig 12: Histogram of month wise temperature distribution of Kathmandu city .

Also in order to figure out the extreme temperature for Kathmandu city we calculate the yearly max temperature with the help of the T2M_MAX variable that represents the maximum temperature at the range of 2 meters and identify the extreme temperature event over the years based on the given temperature threshold.

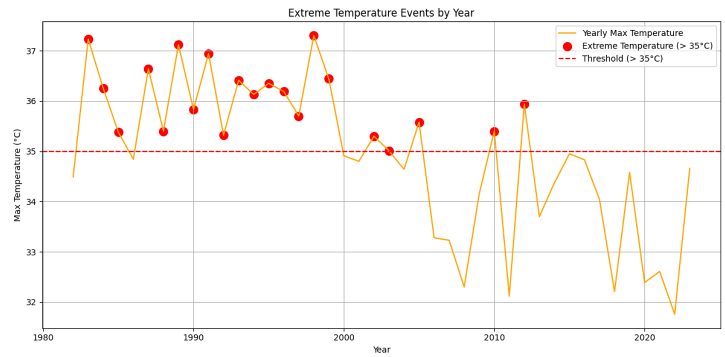


Fig 13: Extreme Temperature Events over year

Implementation and Forecasting Temperature with Markov Chain

In this study Markov Chain model is used to forecast the temperature of Kathmandu city. The dataset is splitted into training and testing dataset with training data consisting data from 1982 to 2020 and testing data consisting data from 2021 to 2023.

```
Training data shape: (14245, 19)
Testing data shape: (1095, 19)
```

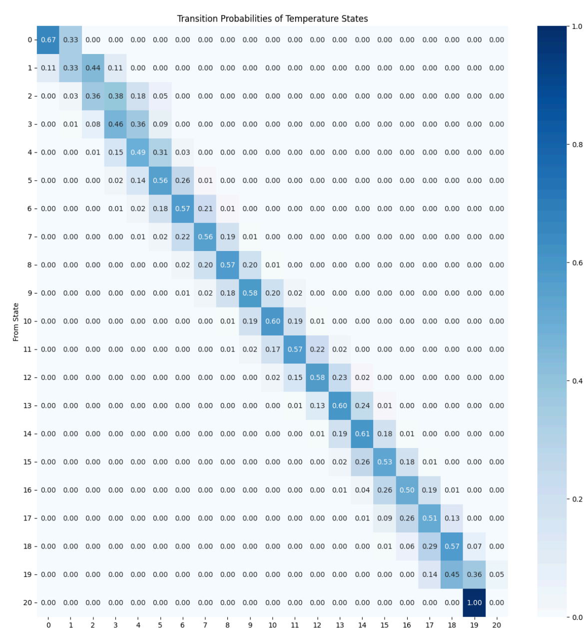


Fig 14: Transition Matrix of 20 x 20

The Markov Model is trained with the Training data and the model is used to predict on the test data. A transition matrix of 20 x 20 is used to figure out the transition from one state to another. Before proceeding to the model training we use the function to discretize the data. This function will provide the discretized data of temperature based on the number of bins used. The total number of bins used to discretize the data is 20.

Function to train the Markov Chain model is **train_markov_chain(data, order=1).** Here we provide discretized data as the input and the chain will use the current state as the initial state, so the order is 1.

After training the model we use the trained model to predict the temperature of the test data. As we already know that the test data includes data from 2021 to 2023. To predict future temperature trend for three year (1095 days) function named predict_temperature_trend() is used. It takes three parameters, first is the transition matrix, second is the initial state and and and third is the steps which represent the day of the year.

The result of the observed and predicted temperature of the three year period is plotted and is represented by the figure 15.

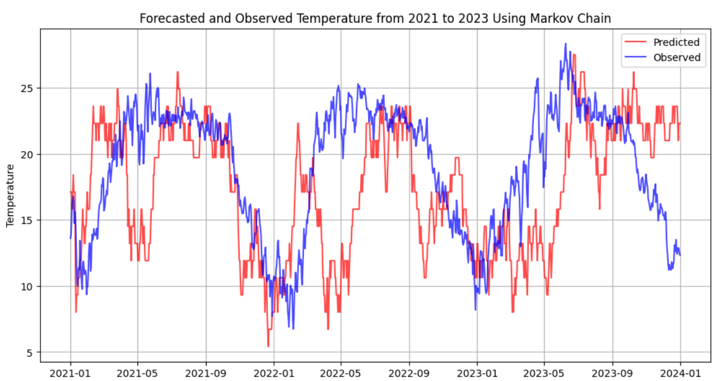


Fig 15: Observed vs Predicted Temperature for the test data. (1095 days).

The accuracy of the forecasted temperatures is 95.64%. The error is calculated by taking the absolute value of forecasted temperature minus observed temperature.

When plotting the Forecasted and Observed Temperature for Kathmandu City Using Markov Chain for the 3 year period from 2021 to 2023 we get the following plot.

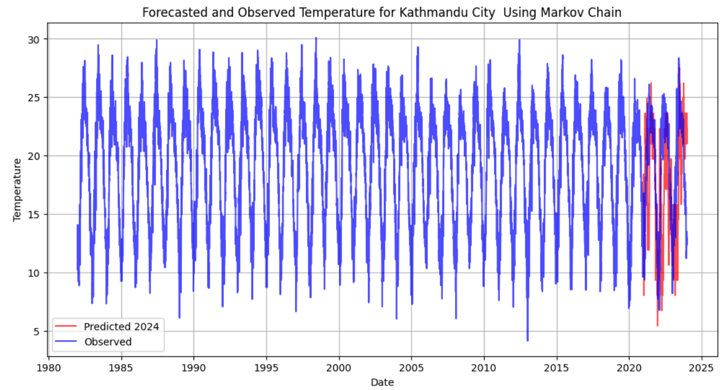


Fig 16: Plot to demonstrate the predicted and observed Temperature of the test data.

The Mean Absolute Error , Mean Square Error and Root Mean Square Error in the prediction is

```
Mean Absolute Error: 4.357662100456621
Mean Square Error: 32.21176660593608
Root Mean Squared Error: 5.675541084860198
```

Conclusion

From this study we conclude that Markov Chain can also be used to study the time series phenomenon where we can analyze the historical past data and predict the future trend. Here we have studied the temperature at 2 meter surface range. Further work can be proceeded to study the trend of other variables like rainfall, humidity that are included in this dataset. One can also apply a different approach like Bayesian approach or LSTM for the prediction of such time series data.

Reference:

[1]:https://www.researchgate.net/profile/Chhabi-Chidi/publication/350133573_Urban_Heat_Island/links/6052faaf92851cd8ce4b75f3/Urban-Heat-Island.pdf

[2]:https://portal.tu.edu.np/publications/122/Grid_Based_Temperature_and_Rel_2023_09_12_10_17_35.pdf