2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) | 978-1-6654-7692-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSPA57446.2023.10087637

Effect of Climate Change using Predictive Models with Remote Sensing Data

M S Bhumika
PES University, Bangalore
msbhumikaaug21@gmail.com

Niyam Momaya
PES University, Bangalore
niyammomaya@gmail.com

Rohit Nandan
PES University, Bangalore
rohitnandanpanth@gmail.com

K Suhas
PES University, Bangalore
suhas720@gmail.com

Shikha Tripathi
PES University, Bangalore
shikha@pes.edu

***Abstract*—Human-induced climate change is causing unpredictable changes around the world to global economies and people's way of living. We present a study undertaken on the prediction of three parameters that quantify climatic change, namely Land Surface Temperature, Carbon Monoxide, and Precipitation. Since the three parameters are essentially non stationary in nature, the differencing method has been looked at to fine-tune the output of the prediction models. The two seasonal autoregressive models chosen for the study are Seasonal Auto Regressive Integrated Moving Average (SARIMA) and TBATS, due to non-stationarity of the data. The focus is to predict climate change using predictive models for Indian sub-continent. Region-wise dataset was extracted from Google earth engine using a reducer function. A comparison of the forecasting models using root mean square error and mean absolute scaled error suggests that SARIMA performs better in the case of Land Surface Temperature and Precipitation whereas TBATS performs better for Carbon Monoxide Concentration.**

***Keywords*— climate change, auto-regressive, non-stationarity**

I. INTRODUCTION

Rapid environmental changes are causing food and water scarcity, while some species are pushed into extinction due to environmental collapse. The effects of a changing climate can be seen in the areas that we value and rely on, such as water, energy, wildlife, and human health. If climate change is not monitored effectively then it would lead us to a serious situation. This paper aims to quantify the effects of human-induced climate change via remote sensing techniques. Future scenarios that are affected by climate change are not inevitable. We are aware of a multitude of issues and solutions, and continuing research on these impacts keeps generating new ideas for better life in future.

Remote sensing is a technique for identifying and monitoring physical features by measuring a region's reflected and emitted radiation at a distance. By using remote sensing, we plan to evaluate climate parameters. The paper presented by Sofia L. Ermida et al [1] presents four algorithms, namely (Multivariate Adaptive Regressive Spline (MARS), Wavelet Neural Network (WNN), Adaptive Neuro fuzzy Inference System (ANFIS), and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) that can be used to predict the Land Surface Temperature (LST). The input to the training model is Landsat images. Results show that MARS, DENFIS, and ANFIS can be used to predict the LST values of any region and the ANFIS model showed the highest performance metrics during training. The WNN algorithm is not very efficient when it comes to computing LST values. The author(s) are using Google Earth Engine for extracting datasets and computing LST based on the values extracted from the LANDSAT 7/8 satellite. In the paper by Banerjee et al [2], two daily precipitation gridded products were also used to construct temporal variations and spatial connections in the research region. Station-based periodical precipitation and rainy weather data were gathered from the India Meteorological Department (IMD) for the time frame of 1983 to 2008 and implemented. The CHIRPS dataset was utilized in order to obtain monthly precipitation data across a period of 20 years. In the paper by M Mustafa et al [3], The authors have provided a code repository that allows computing LSTs from Landsat 4,5,7, and 8 within Google Earth Engine. NDVI (Normalized Difference Vegetation Index) is a factor in the optical band. The key to computing LST values lies in emissivity. Generally, the NDVI is often used to determine the percentage of vegetation cover usually termed as Fraction of Vegetation Cover (FVC). The LST values are calculated using the SMW (Single Mono Window) algorithm and observed that the LSTs computed were accurate in comparison with the in-situ LST values.

The paper by Huang et al [4] is a landmark work, providing a solid reference in the field of non-linear time series analysis. The author(s) have provided a method having an adaptive basis, which is the intrinsic mode functions (IMFs). The IMFs are used in conjugation with the Hilbert transform. Rhif et al[5], have proposed the Wavelet Transform as a method for performing multivariate/univariate analysis on non-stationary time series. The author(s) also talk about the other methods of performing analysis on non-stationary signals, that is the EMD/first difference, and the short time Fourier transform. Of the two, only the Wavelet and STFT are in the time frequency domain. The paper by Xike Zhang [6] presents a novel hybrid model that is used for forecasting daily Land Surface Temperature using Long Short-Term Memory (LSTM). The LSTM model is based on Ensemble Empirical

Mode Decomposition (EEMD) and this helps in reducing the difficulty of modeling and increasing efficiency. To validate this model, the hybrid EEMD - LSTM model is compared with various other models and it is found that the EEMD - LSTM model performs the best and is concluded that it's suitable for temperature forecasting.

In the existing literature most of the papers are focused on a small city in a particular region and there has not been any prior work done in this area for the Indian Subcontinent. The existing papers have considered a single model to predict the climate parameter whereas this paper presents a multi-model comparison using a comparative approach and incorporates multiple parameters to assess and evaluate the effects of climate change with validation. In this paper, study on the prediction of three parameters that quantify climatic change, namely Land Surface Temperature, Carbon Monoxide (CO), and Precipitation has been undertaken. Due to non-stationarity of the data two seasonal autoregressive models chosen for the study are Seasonal Auto Regressive Integrated Moving Average (SARIMA) and Trigonometric Seasonality-Box Cox Transformation-ARMA errors-Trend-Seasonality (TBATS). The time series data extracted from the Earth Engine satellites are fed to the two seasonal prediction models and forecasts were generated. It is observed that SARIMA forecasts for LST and Precipitation is closer to the recorded data whereas TBATS outperforms in the case of CO concentration.

The roadmap of the paper is as follows. Section II details the used methodology. Section III outlines the dataset extraction procedure, time series chart generation and the prediction model implementation. Section IV analyzes the forecasts of both the models for the 3 chosen parameters based on common forecast evaluation metrics. The conclusion and future work are in Section V.

II. PROPOSED TECHNIQUE

LST data obtained from MODIS, CO concentration obtained from Sentinel-5P and precipitation data obtained from CHIRPS are fed to the prediction models to produce a forecasted output as seen in Fig.1.

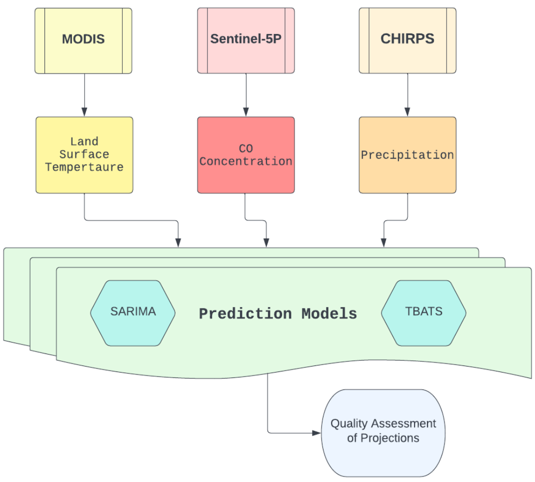

Fig.1 Proposed methodology

The three parameters derived from remote sensing data, is a 2D matrix of pixels, with each pixel's intensity value representing the Land Surface Temperature (LST), CO concentration, or the precipitation value. A reducer essentially takes a mean of all these pixel values, and outputs a dataset in the form (batch_size * feature(s)). As the data is non-linear in nature, first/second difference methods, the Empirical Mode signal Decomposition (EMD) or the Wavelet Transform can be used. As the Wavelet Transform reduces resolution in both the time and frequency domains, the EMD and the first/second difference could be considered. Of the two, the first/second difference offers marginal performance drawbacks over the EMD but is much faster, computationally. Hence first/second difference is used in this work.

III. IMPLEMENTATION

The Google Earth Engine is utilized for geospatial dataset processing and visualization in the fields of science. This platform has been utilized for computing parameters including Land Surface Temperature, Carbon Monoxide concentration, and Precipitation as well as obtaining requisite datasets.

The datasets used are MODIS for LST computation, Sentinel-5P for Carbon Monoxide concentration and CHIRPS for precipitation:

a) MODIS - MOD11A1.061: Terra Land Surface Temperature Daily Global 1km.
b) Sentinel-5P - OFFL CO: Offline Carbon Monoxide.
c) CHIRPS - CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation.

Fig. 2 depicts a historical time series chart for LST which demonstrates the seasonal nature of the parameter. It is observed that LST has a dual seasonality, i.e., annual and monthly seasonal periods. Fig. 3 presents the CO concentration values over the selected Region of interest (Indian Subcontinent) from 2019-21. Fig 4 represents the precipitation time series for a period of 15 years.

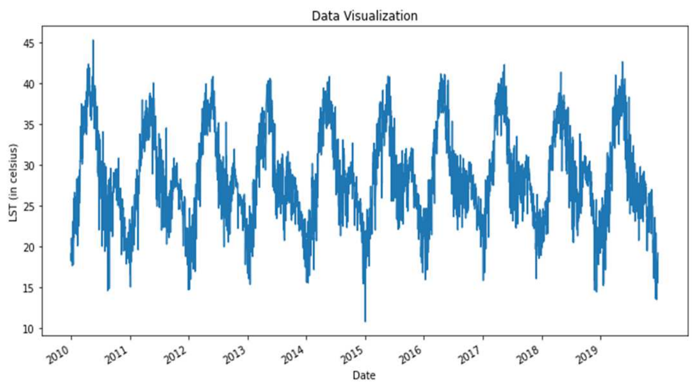

Fig. 2 LST over a duration of 10 years

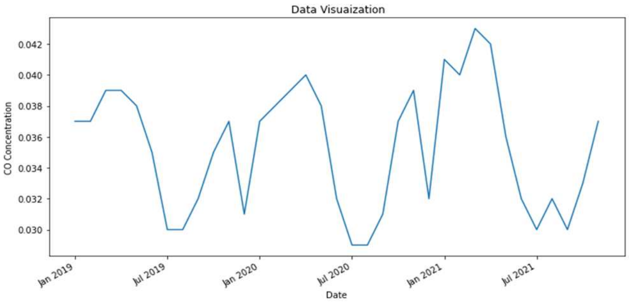

Fig.3 CO concentration

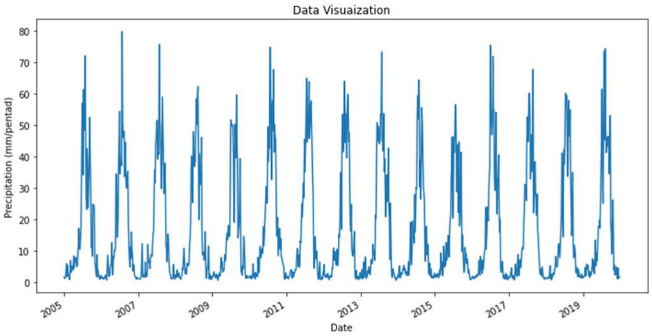

Fig.4 Precipitation

The 2 prediction models are explained in detail in the following subsections.

A. SARIMA

SARIMA stands for Seasonal Autoregressive Integrated Moving Average. It is used to forecast a time series. It adds 3 parameters, namely:
p - Autoregressive Order
d - Difference Order
q - Moving Average Order

SARIMA takes stationary input data whereas the computed dataset is non stationary in nature which is verified using the Augmented Dickey-Fuller test. Hence, the differencing method is used for the required transformation. The degree of differencing is dependent on the order of non-stationarity. The seasonal part of an AR and MA model i.e. p and q values can be inferred by looking at the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plots.SARIMA can be used in three steps namely:

- Define the model
- Fit the defined model
- Make predictions with the fit model

The three univariate time series are then fed to a SARIMA model, which in addition to the regressed time steps, also incorporates seasonality. Auto-regressive models were found to be a better fit than non-linear neural networks, as the data seems to be auto-regressive in nature.

The model is implemented by using the "statsmodels" python module that provides classes and functions for conducting statistical tests and statistical data exploration. The SARIMAX function used here uses the p,d,q terms for modelling the seasonal and autoregressive nature of the input data.

B. TBATS

TBATS stands for -

- T- Trigonometric Seasonality
- B- Box-Cox Transformation
- A- ARMA Errors
- T- Trend
- S- Seasonal Components

It is a forecasting method to model time series data with multiple complex seasonal patterns. TBATS model is an extension to exponential smoothing state space models. The trigonometric representation based on Fourier terms of the seasonal components allows for non-integer large periods. It combines a Box-Cox Transformation to handle nonlinear data and an ARMA model to capture the autocorrelation in the residuals. The Trend component explains the long-term change in the mean value of the series and Seasonal component explains the periodical variation in the series. Behind the scenes, TBATS considers multiple alternatives before selecting the best performer.

Finally, using the Akaike's Information Criterion, the best performing model for any input data consisting of multiple seasonal periods (for ex: hourly data might have daily, weekly and monthly seasonalities) is chosen.

The model is implemented using functions available in the python "sktime" package which is a flexible and modular open-source framework for a wide range of time series and machine learning tasks.

IV. Results and Analysis

The devised algorithms are implemented on the various climate parameters and the forecasts are as follows:

A. SARIMA

1) Land Surface Temperature

The dataset for Land Surface Temperature is of 10 years from a period of 2010 - 2019. It is split into a 70:30 training to testing ratio. The SARIMA model performance over the test set is presented in Fig 5 (a) whereas the forecast for the next 4 years is shown in Fig 5 (b). The TBATS model performance over the test set and model forecast is shown in Fig 8 (a) and (b) respectively.

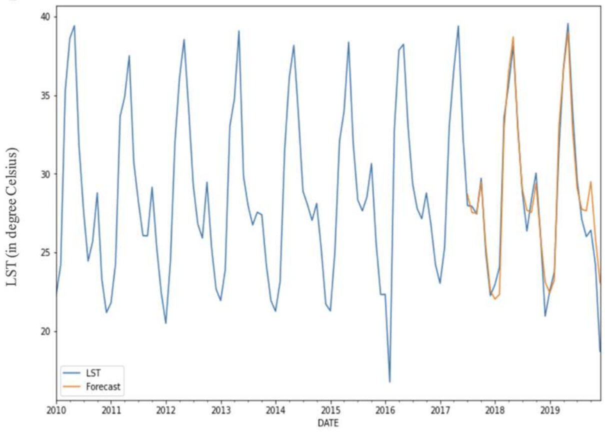

Fig.5 (a) Model output for the test data

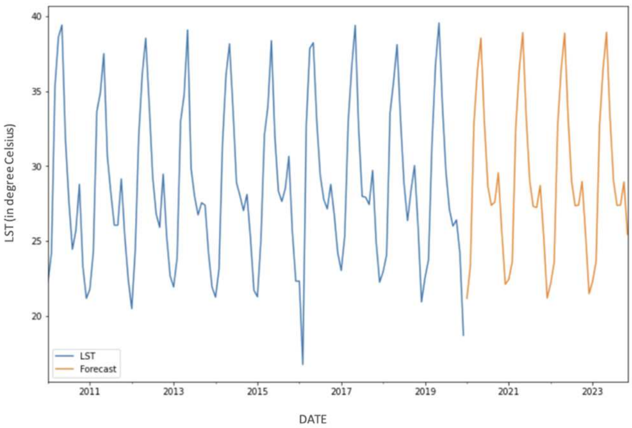

Fig.5 (b) Forecast for 2020-2023

2) *Carbon Monoxide*

The CO concentration dataset is 3 years long from 2018 to 2021 wherein the test set consists of one year data. Fig 6 (a) shows the model performance over the test set whereas Fig 6 (b) shows the forecast over the next 2 years. The TBATS model performance over the test set and model forecast is shown in Fig 9 (a) and (b) respectively.

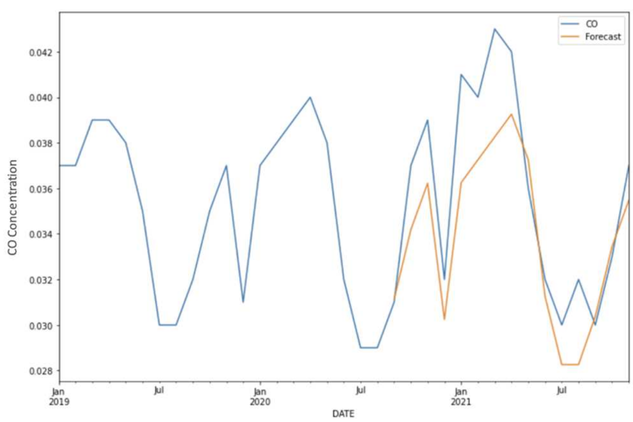

Fig.6 (a) Model output for the test data.

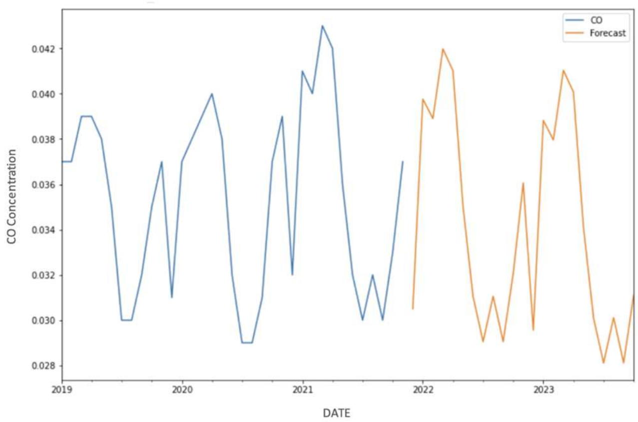

Fig. 6 (b) CO forecast for 2022-23.

3) *Precipitation*

The duration of the dataset for Precipitation is 15 years from a period of 2005 - 2019. The SARIMA model performance over a test set of 6 years is depicted in Fig 7(a). The forecast for the next 2 years is shown in Fig 7(b). The TBATS model performance over the test set and model forecast is shown in Fig 10 (a) and (b) respectively.

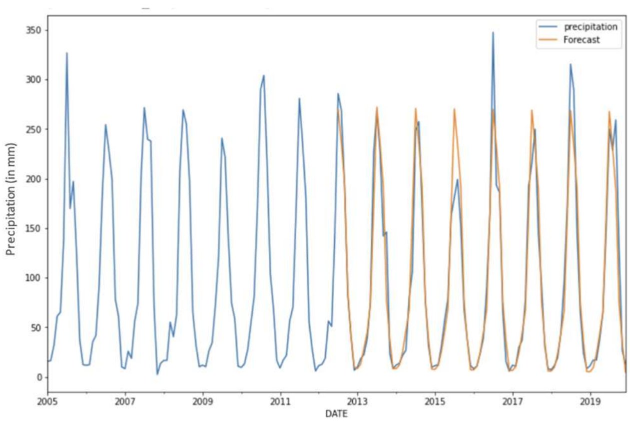

Fig.7 (a) model output for the test data

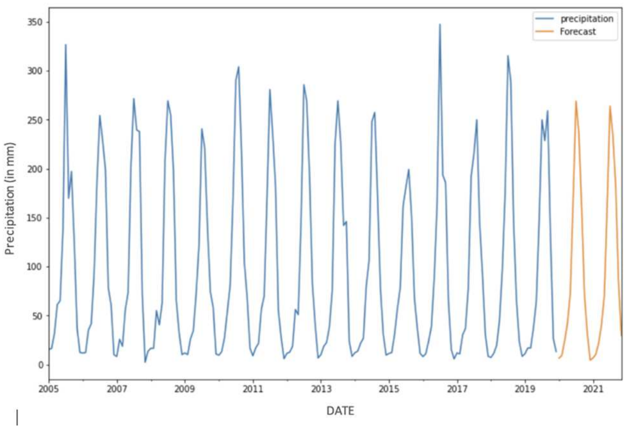

Fig. 7 (b) Forecast for 2020-21

B. TBATS

TBATS considers various combinations of models with and without box-cox transformation, trend, trend damping and ARMA(p,q) process for residual modelling and uses the Akaike's Information Criterion to select the best performer for any given input.

1) Land Surface Temperature

Fig.8 (a) LST Validation

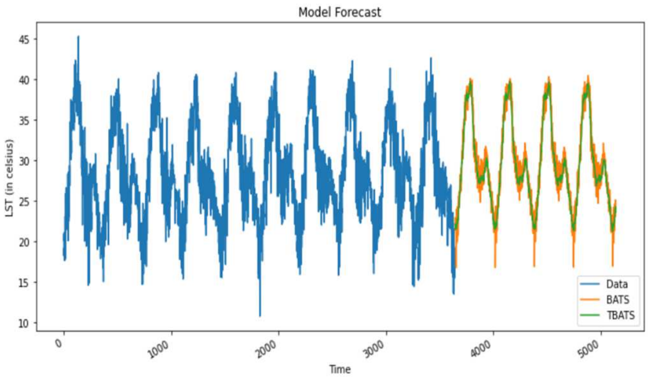

Fig.8(b) LST Forecast

2) Carbon Monoxide

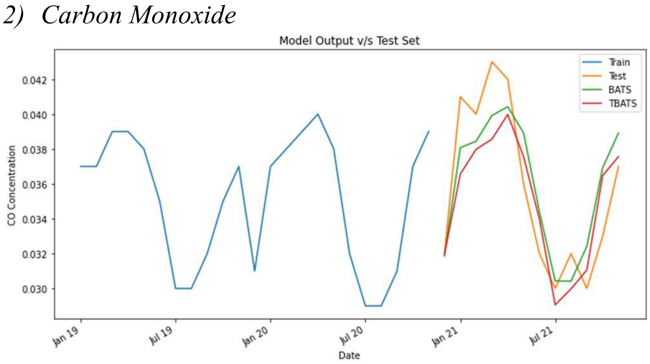

Fig.9 (a) CO Validation

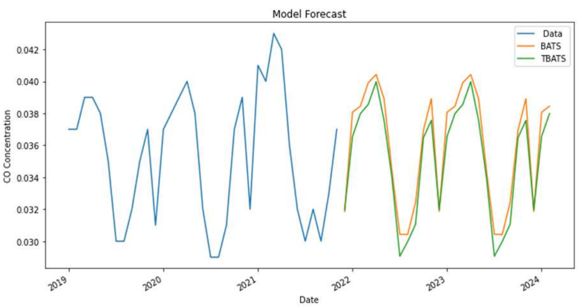

Fig.9(b) CO Forecast

3) Precipitation

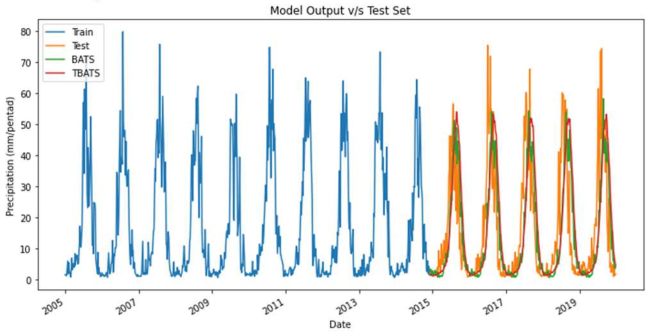

Fig.10 (a) Precipitation Validation

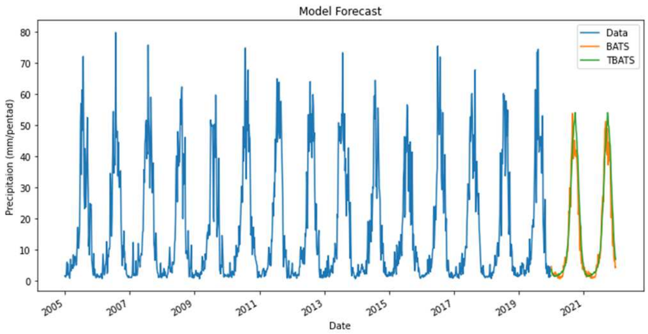

Fig.10(b) Precipitation Forecast

A comparison of the forecasting models using 2 performance metrics namely Root Mean Square Error (RMSE) and Mean Absolute Scaled Error (MASE) is as shown in Table 1.

TABLE I. PERFROMANCE EVALUATION METRICS

Parameter	RMSE	MASE
Land Surface Temperature	1. SARIMA - 0.3742 2. BATS - 3.4117 3. TBATS - 3.1633	1. SARIMA - 0.9213 2. BATS - 1.3239 3. TBATS - 1.2178
Carbon Monoxide	1. SARIMA - 0.4371 2. BATS - 0.0023 3. TBATS - 0.0024	1. SARIMA - 0.9954 2. BATS - 0.8445 3. TBATS - 0.8333
Precipitation	1. SARIMA - 0.3482 2. BATS - 14.0876 3.TBATS - 15.4499	1. SARIMA- 1.231 2. BATS - 1.7793 3. TBATS - 2.1563

The model with the lower RMSE and MASE values is considered to be the best performer for that particular parameter. From TABLE I it is observed that SARIMA performs better prediction for Land Surface Temperature and Precipitation whereas TBATS performs better for Carbon Monoxide Concentration.

V. CONCLUSION AND FUTURE WORK

In this study, two main models were considered for the prediction of the three parameters. SARIMA, a seasonal auto-regressive model and TBATS, an exponential smoothing state space model. Since the input datasets are essentially auto-regressive in nature, SARIMA and TBATS were chosen. The time step of the output prediction is limited by the fact that the input data was not more than a few years in length, however the validation result proves the functionality of the model, given a large dataset. It is observed that SARIMA performs better prediction using Land Surface Temperature and Precipitation whereas

TBATS performs better for Carbon Monoxide Concentration.

Further improvements to the study could involve inclusion of Wavelet Transform based LSTM and SARIMA implementations to the comparison.

REFERENCES

[1] M mustafa, Elhadi & Liu, Guoxiang & Cao, Yun-Gang & Kaloop, Mosbeh & Beshr, Ashraf & Zarzoura, Fawzi. (2020), "Study for Predicting Land Surface Temperature (LST) Using Landsat Data: A Comparison of Four Algorithms", Advances in Civil Engineering. 2020.

[2] Soldatenko, Sergei & Bogomolov, Alexey & Ronzhin, Andrey. (2021). Mathematical Modelling of Climate Change and Variability in the Context of Outdoor Ergonomic.

[3] Sofia L. Ermida, Patrícia Soares, Vasco Mantas, "Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series", *Remote Sens. Environ.* 2020.

[4] Huang Norden E., Shen Zheng, Long Steven R., Wu Manli C., Shih Hsing H., Zheng Quanan, Yen Nai-Chyuan, Tung Chi Chao and Liu Henry H. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*Proc. R. Soc. Lond. A.*

[5] Rhif, Manel & Abbes, Ali & Farah, Imed & Martinez, Beatriz & Sang, Yanfang. (2019). Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. Applied Sciences.

[6] Zhang, Xike & Zhang, Qiuwen & Zhang, Gui & Nie, Zhiping & Gui, Zifan & Que, Huafei. (2018). A Novel Hybrid Data-Driven Model for Daily Land Surface Temperature Forecasting Using Long Short-Term Memory Neural Network Based on Ensemble Empirical Mode Decomposition. International Journal of Environmental Research and Public Health.

[7] Karthikeyan, L. & Kumar, D Nagesh. (2013). Predictability of Nonstationary Time Series using Wavelet and EMD based ARMA Models. Journal of Hydrology.

[8] Agana, Norbert & Homaifar, Abdollah. (2018). EMD-Based Predictive Deep Belief Network for Time Series Prediction: An Application to Drought Forecasting. Hydrology.

[9] Alysha M. De Livera, Rob J. Hyndman and Ralph D. Snyder (2011): Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing, Journal of the American Statistical Association.