Developing Spatial Simulator Engine for Urban Air Pollution Monitoring-Application on MODIS data

Presented By : **Debajyoti Maity**
B2230023

Collaboration With: **Subhojit Mandal**
(Research Scholar IIIT Sri City)

Under the guidance of : **Dr. Swalpa Kumar Roy**

Outline

- Introduction
- Study Area
- Objective
- Problem Statement
- Literature Survey
- Proposed Method
- Experiments & Result
- Conclusion

Introduction

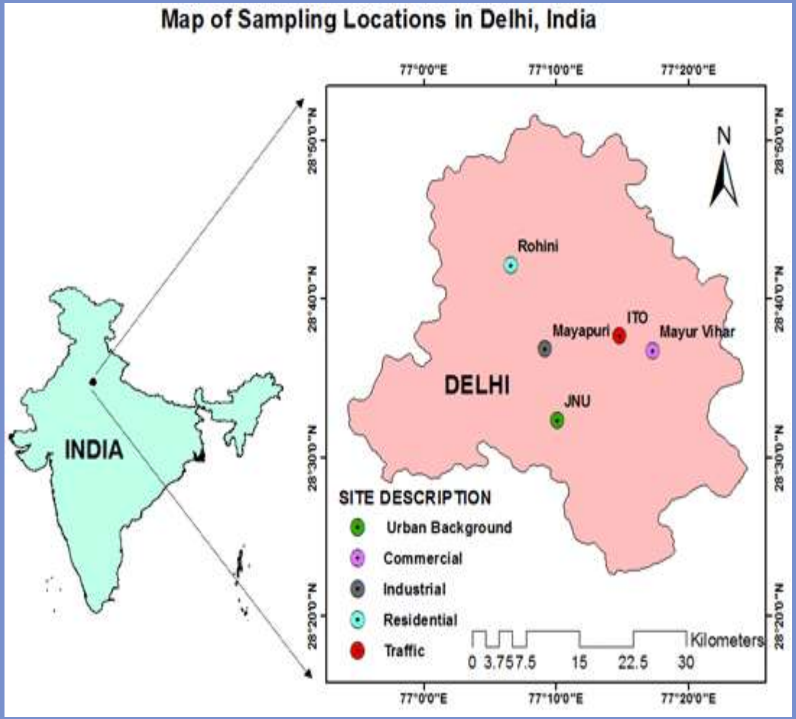
- Now, decades into the 21st century, the specter of air pollution continues to loom as a pressing global challenge.
- As populations surge and industries expand, the quality of the air we breathe has become a focal point of concern for communities, policymakers, and scientists alike.
- The consequences of air pollution are far-reaching, touching every aspect of our lives, from the health of individuals to the stability of ecosystems, and even the stability of our climate

Study Area

- We now focus on Delhi, the bustling capital of India, as a unique and critical case study in the context of air quality challenges, serving as a smaller, representative version of the broader air pollution crisis.

- Delhi's rapid growth, high population density, industrial operations, and heavy traffic have turned it into one of the world's most polluted regions, drawing the focus of researchers, environmentalists, and policymakers.

- By examining air pollution in the Delhi region, we aim to uncover insights that can be applied not only to this specific area but also to urban centers around the world facing similar challenges.



Objective

- **Comprehensive Assessment and Mapping**: The study aims to conduct a thorough assessment and mapping of Aerosol Optical Depth (AOD) in urban areas.

- **Spatial and Temporal Distribution**: The research focuses on investigating the spatial and temporal distribution patterns of AOD within urban environments.

- **Insights for Air Quality Monitoring**: The research aims to provide insights and data that can be used for more effective air quality monitoring in urban areas.

- **Urban Planning**: Findings from this study can inform urban planning strategies, helping in the development of cleaner and healthier cities.

- **Temporal Trends**: Temporal patterns of these hotspots are also considered, helping to determine if they exhibit seasonal variations or changes over time.
- **Aerosol Pollution Analysis**: It involves a detailed analysis of the composition and sources of aerosol pollutants contributing to these hotspots. This analysis may include identifying the types of aerosols, such as fine particulate matter (PM2.5) or other pollutants.

Problem Statement

- Indian Data product INSAT 3D has 70% above nan value of the geostationary observation.
  So in our study we consider Modis dataset. Whenever Aod data is not there because of the cloud cover ,we can build a model to predict the missing value with help of spatial contextual information and temporal information.

- How to use contextual data for predict AOD?
  Spatial contextual informations & Temporal contextual informations

- Can the model capability increase if we use ground station monitoring system data era5 , merre 2?

- Urban air pollution monitoring system can we use modis data?

Literature Survey

- For this project we follow this literature
  1)https://arxiv.org/pdf/1607.02976.pdf

- In this paper, one of the primary objectives is to introduce an advanced model called the Generalized Regression Neural Network (GRNN).

- The goal is to demonstrate that this model can effectively and accurately represent the relationship between Aerosol Optical Depth (AOD) and PM2.5 concentrations.
- We also follow the paper
  2)https://www.sciencedirect.com/science/article/abs/pii/S0169809515002197

Description of AOD

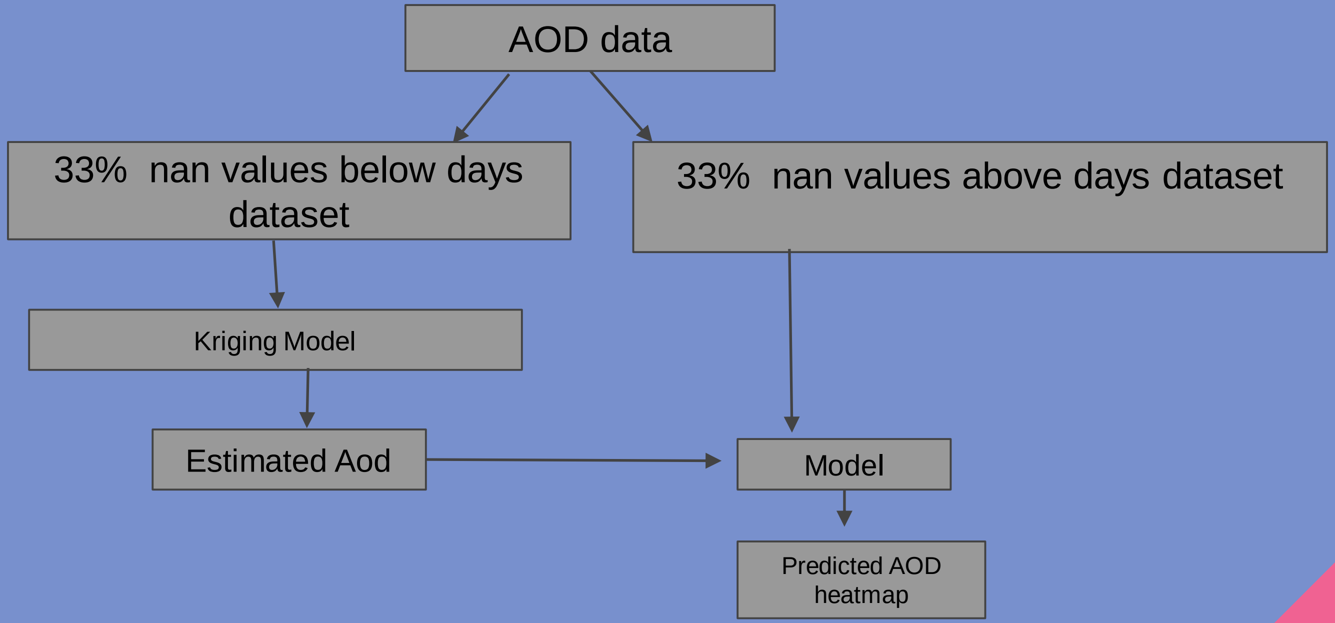
- Aerosol Optical Depth (AOD) is a measure of the amount of aerosol particles present in the Earth's atmosphere and how much they affect the transmission of sunlight through the atmosphere.
- Aerosols are tiny solid or liquid particles suspended in the air, and they can include things like dust, smoke, pollen, pollutants, and sea salt.
- AOD is an important parameter for understanding atmospheric composition, air quality, climate, and visibility. It quantifies the degree to which aerosols scatter and absorb sunlight, affecting the amount of direct and diffuse solar radiation that reaches the Earth's surface.
- A high AOD indicates a higher concentration of aerosol particles and can lead to reduced visibility, altered energy balance, and potential cooling effects on the climate.

Description of PM 2.5

- ❖ PM2.5 is a mass concentration measurement.
- ❖ PM2.5 is a point measurement.
- ❖ PM2.5 is a measure of the mass of particles in a specific size range within a given volume of air near the surface.

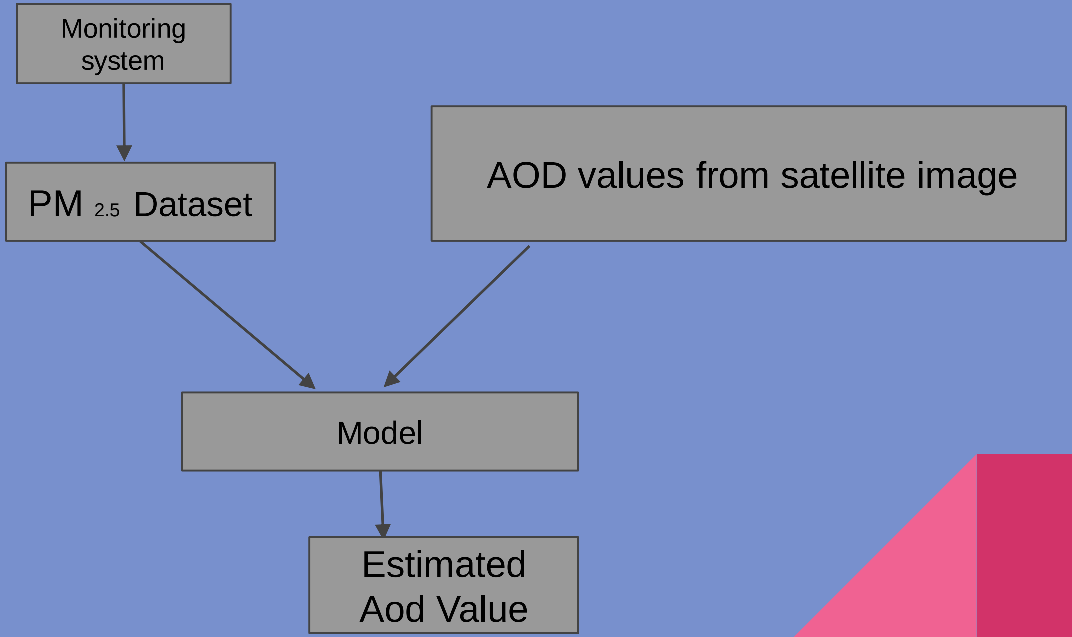
Proposed method-1

- ❖ Indian Data product INSAT 3D has 70% above nan value of the geostationary observation.
- ❖ So in our study we consider Modis dataset.



Proposed Method-2

- Can the model capability increase if we use monitoring system data era5 ,merra 2 ?
- Urban air pollution monitoring system can we use Modis data ?



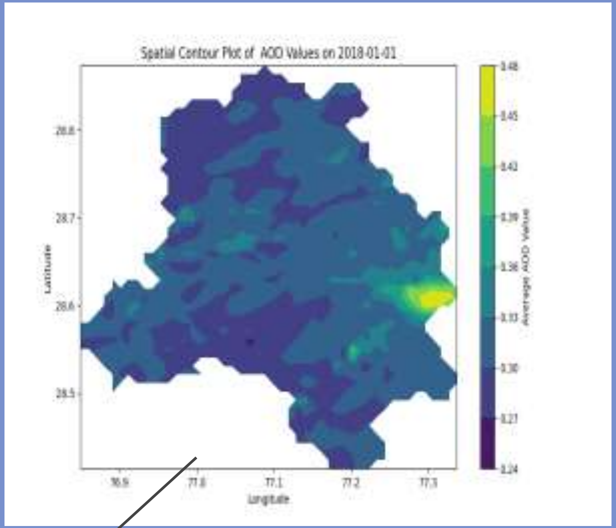
Proposed Method 3

Meteorological Data

Features

1)Temperature
2)Precipitation
3)Humidity
4)Wind Speed and Direction
5)Pressure
6)Cloud Cover
7)Solar Radiation
8)Weather Phenomena
9)Air Quality

Ground Observations:
PM 2.5 & Locations



Satellite Image :
Aod heat map

Model:
f

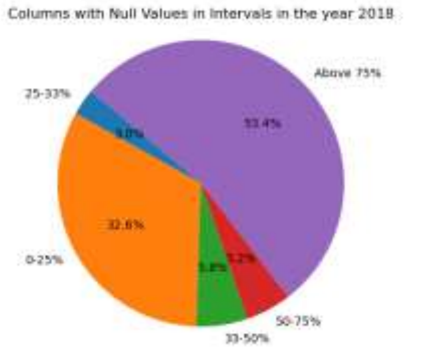
Next trend AOD Prediction

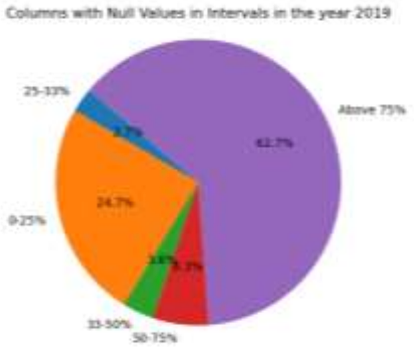
Introduction Of Modis Dataset

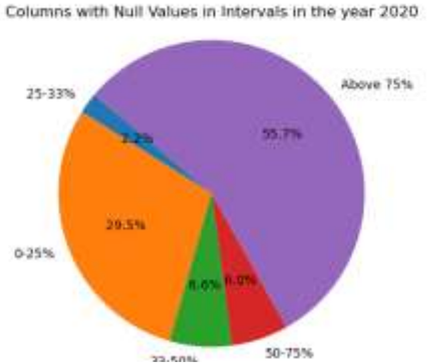
- MODIS(Moderate Resolution Imaging Spectroradiometer) Aqua and Terra satellite data, which provide a wealth of high-resolution and multi-spectral information.

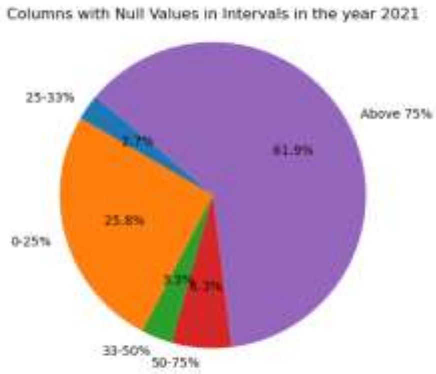
- Terra and Aqua are two Earth observing satellites launched as part of Earth Observing System(EOS) program. They are design to collect data about various aspects of Earth’s Atmosphere ,Oceans and Climate.

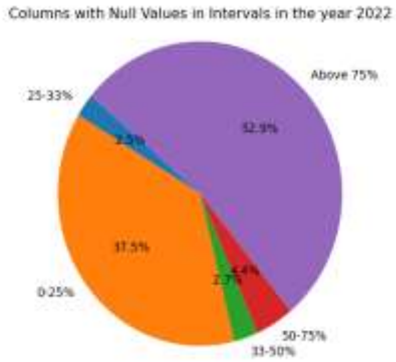
Missing Value Plot

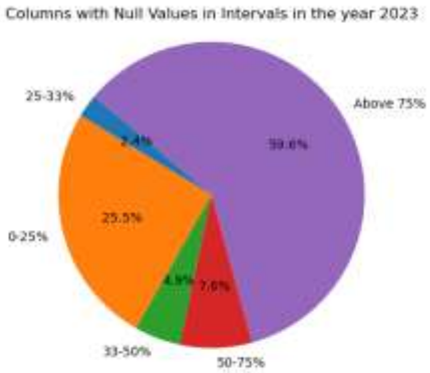




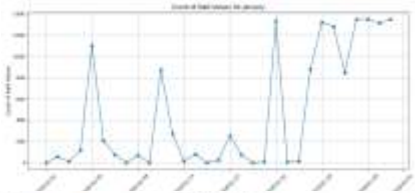


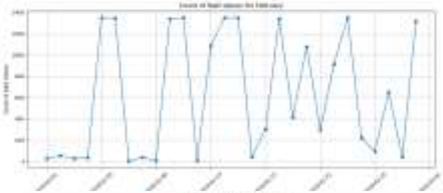


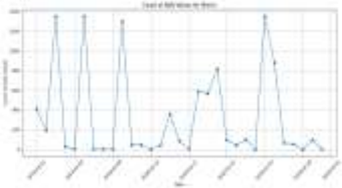


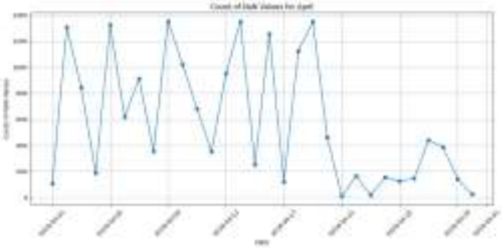


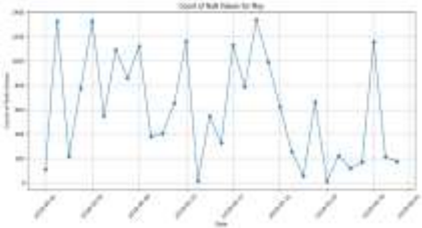
Nan Values plot every month in 2018

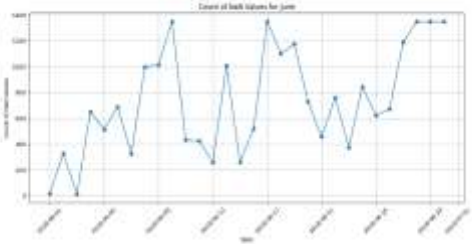


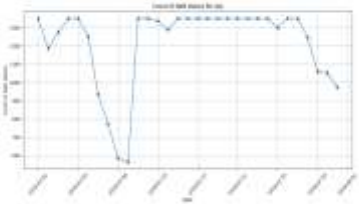


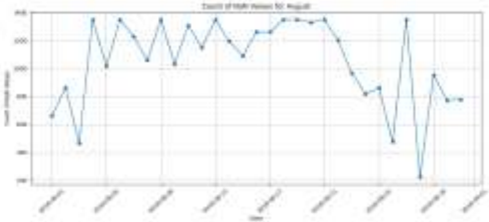


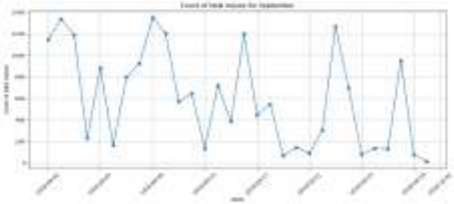


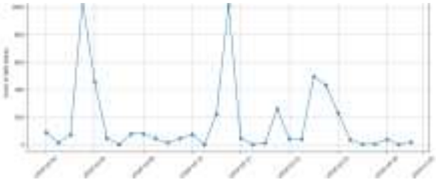


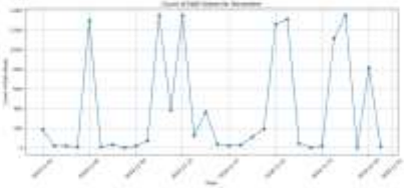


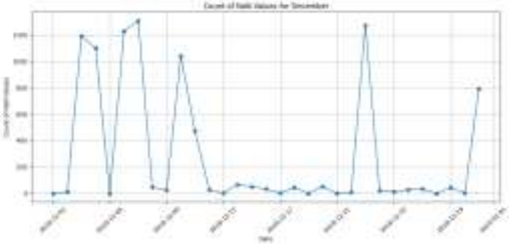












LINEAR KRIGING

- Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field.

- Kriging weights are calculated such that points nearby to the location of interest are given more weight than those farther away.

- Kriging can be understood as a two-step process: first, the spatial covariance structure of the sampled points is determined by fitting a variogram.

- Weights derived from this covariance structure are used to interpolate values for unsampled points or blocks across the spatial field.

TWO APPROACH OF KRIGING

- we saw how to apply a model of local spatial dependence (i.e.a variogram model) to prediction by kriging.
- To avoid information overload, we deferred discussing the kriging equations, and in particular in what sense kriging is an optimal local predictor.
- Note: It is not necessary to understand this topic completely in order to correctly apply kriging. The derivation is necessarily mathematical and in places requires knowledge of matrix algebra or differential calculus. Still, everyone who uses kriging should be exposed to this at least once.
- For further details

click

https://www.esri.com/content/dam/esrisites/en-us/events/conferences/2020/federal-gis/kriging-an-intro-to-concepts.pdf

Spatial Estimation

$Z(U_\alpha)$: Data values at U_α

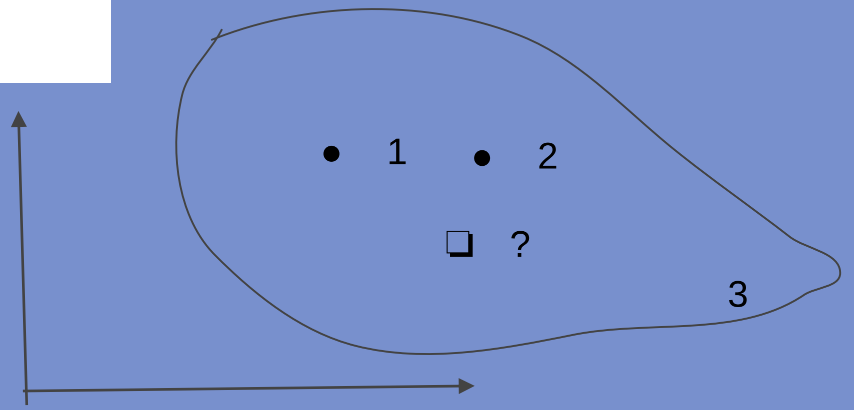
$Z^*(U^\circ)$: Estimate

λ_α : Data weights

m_z : Global mean

$$Z^*(U^\circ) = \sum_\alpha \lambda_\alpha Z(U_\alpha) + (1 - \sum_\alpha \lambda_\alpha) m_z$$

Written in latex



Simple Kriging : Some Details

There are three equations to determine the three weight:

$$\lambda_1 \cdot C(u_1, u_1) + \lambda_2 \cdot C(u_1, u_2) + \lambda_3 \cdot C(u_1, u_3) = C(u_0, u_1)$$
$$\lambda_1 \cdot C(u_2, u_1) + \lambda_2 \cdot C(u_2, u_2) + \lambda_3 \cdot C(u_2, u_3) = C(u_0, u_2)$$
$$\lambda_1 \cdot C(u_3, u_1) + \lambda_2 \cdot C(u_3, u_2) + \lambda_3 \cdot C(u_3, u_3) = C(u_0, u_3)$$

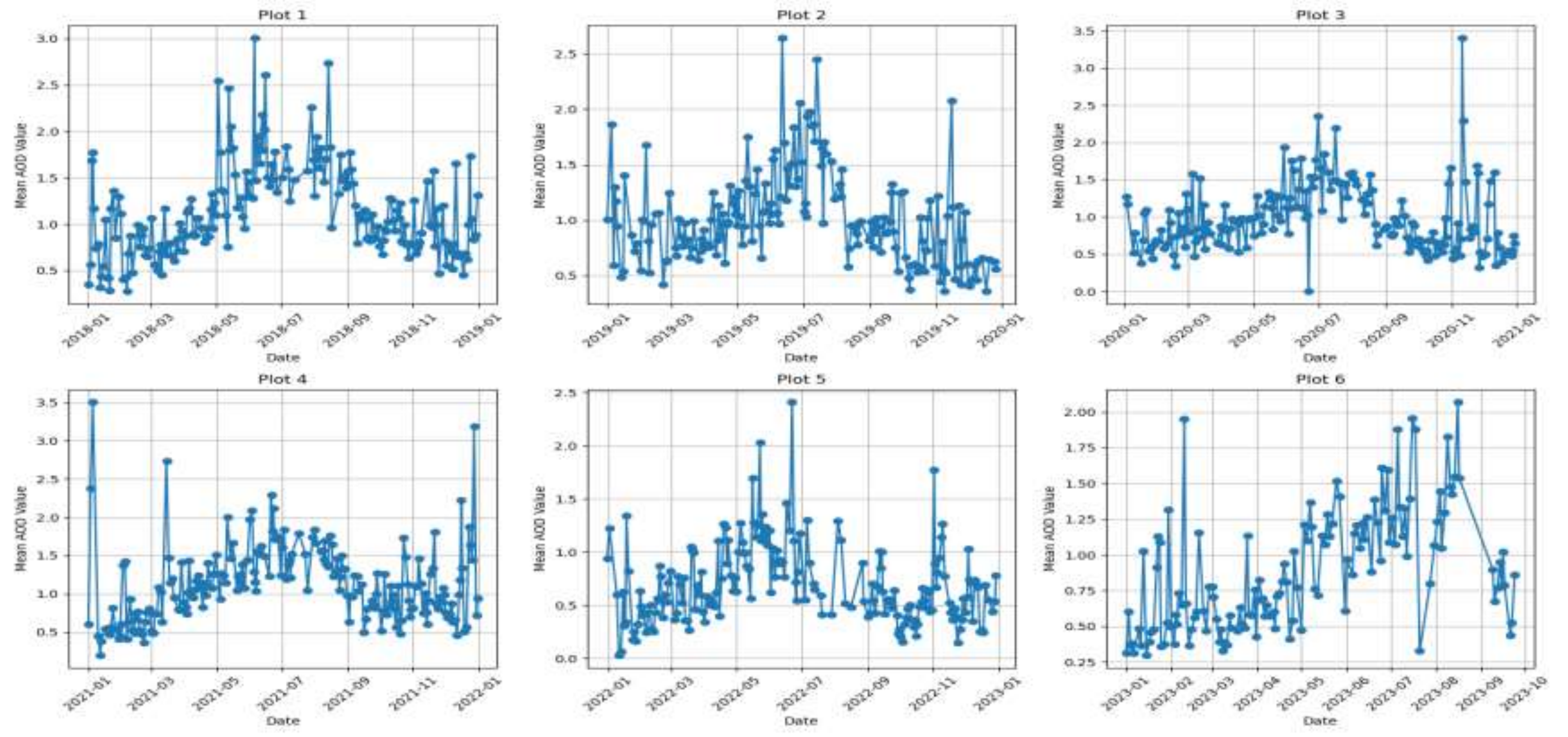
$$\begin{bmatrix} C(u_1, u_1) & C(u_1, u_2) & C(u_1, u_3) \\ C(u_2, u_1) & C(u_2, u_2) & C(u_2, u_3) \\ C(u_3, u_1) & C(u_3, u_2) & C(u_3, u_3) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C(u_0, u_1) \\ C(u_0, u_2) \\ C(u_0, u_3) \end{bmatrix}$$

Distance of the information: $\mathrm{C}(\mathrm{u}, \mathrm{u}_i)$

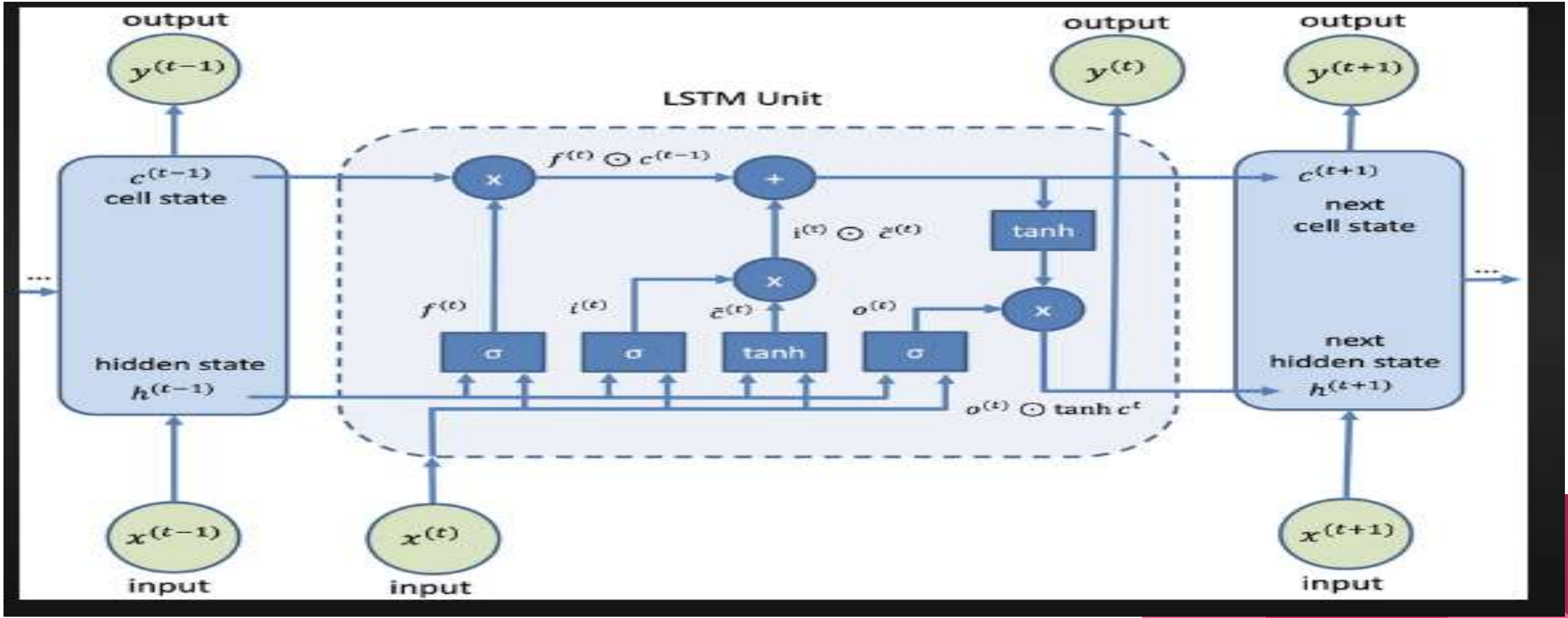
Configuration of the data: $\mathrm{C}(\mathrm{u}_i, u_j)$

Written in latex

TIME SERIES PLOT AFTER KRIGING

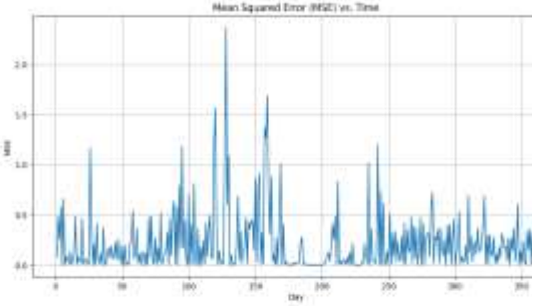


Architecture of LSTM

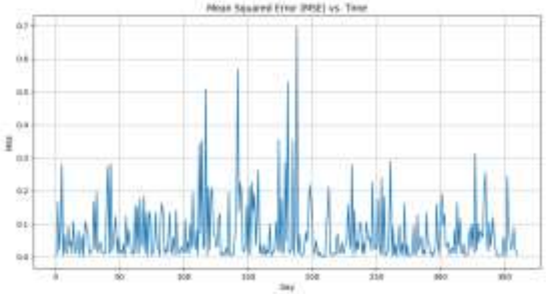


MEAN SQUARE ERROR PLOT

2018 YEAR MSE PLOT

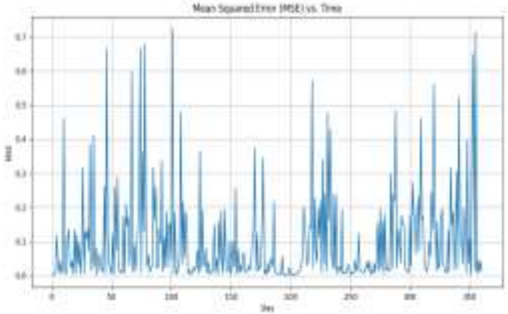


2019 YEAR MSE PLOT

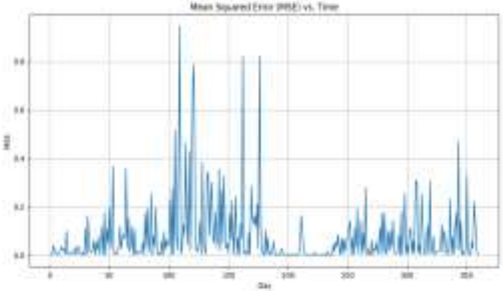


2020 YEAR MSE PLOT

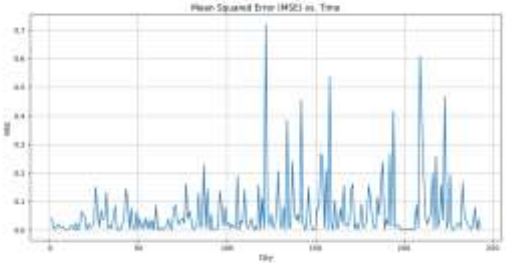
2021 YEAR MSE PLOT



2022 YEAR MSE PLOT



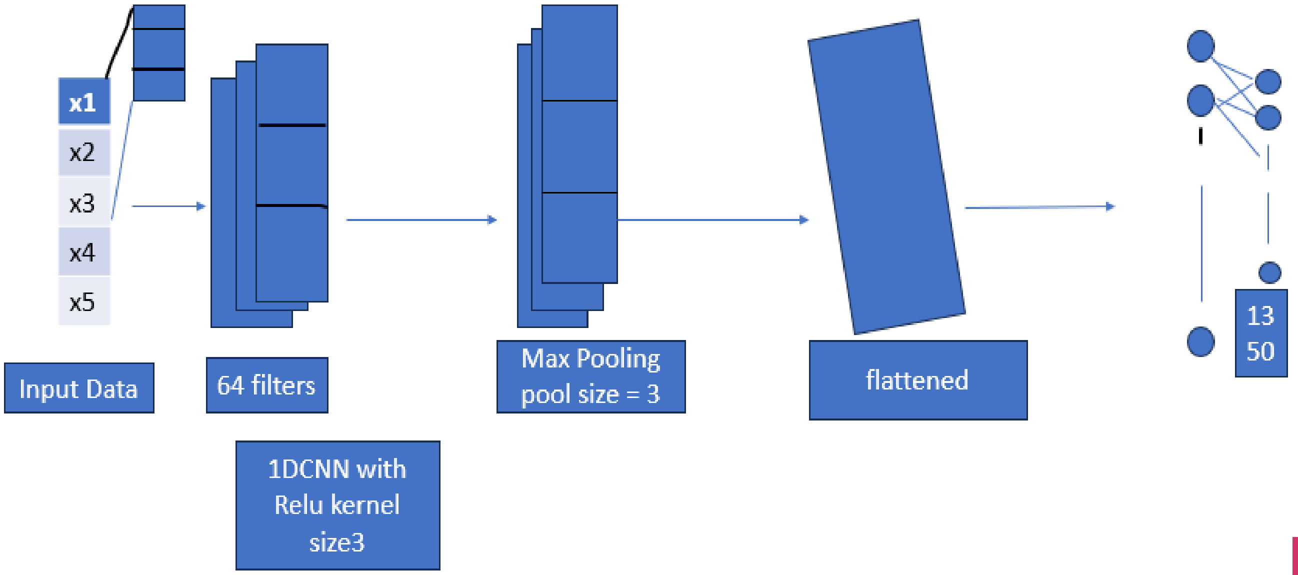
2023 YEAR MSE PLOT



MSE VALUE FOR EVERY YEAR FOR FIT THE LSTM MODEL

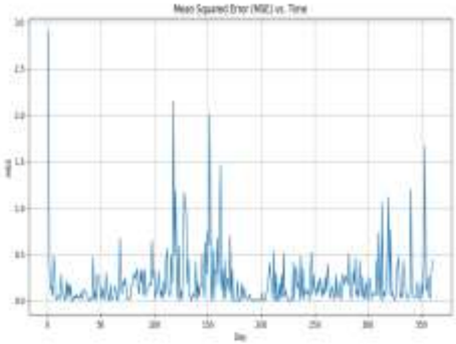
- THE MEAN SQUARE ERROR OF TEST DATA IN 2018 YEAR IS 0.27799022791550865
- THE MEAN SQUARE ERROR OF TEST DATA IN 2019 YEAR IS 0.11780011346698749
- THE MEAN SQUARE ERROR OF TEST DATA IN 2020 YEAR IS 0.12765933786346556
- THE MEAN SQUARE ERROR OF TEST DATA IN 2021 YEAR IS 0.16513408452772274
- THE MEAN SQUARE ERROR OF TEST DATA IN 2022 YEAR IS 0.04598796115728515
- THE MEAN SQUARE ERROR OF TEST DATA IN 2023(TILL NOW) YEAR IS 0.09604098549890211

CNN MODEL ARCHITECTURE

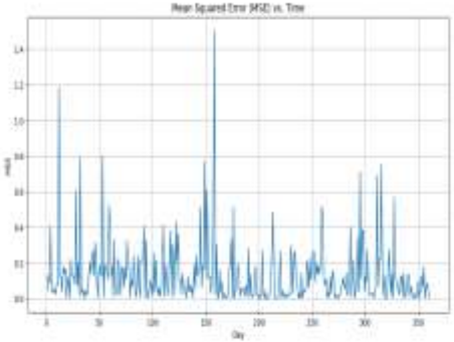


MSE PLOT AFTER FITTING THE CNN MODEL OVER SIX YEAR

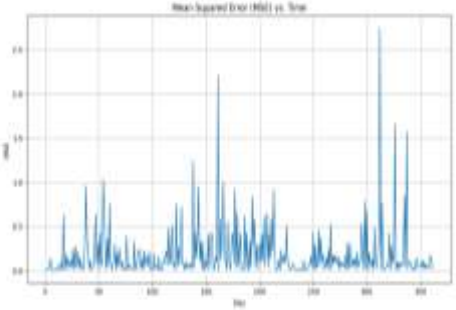
MSE PLOT IN THE YEAR 2018



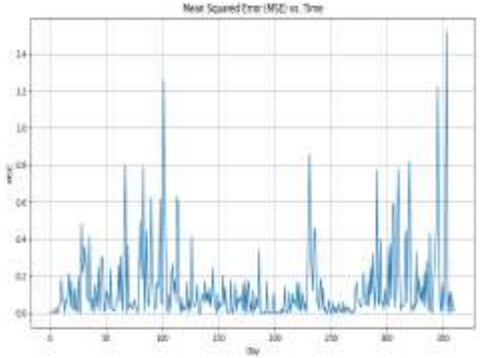
MSE PLOT IN THE YEAR 2019



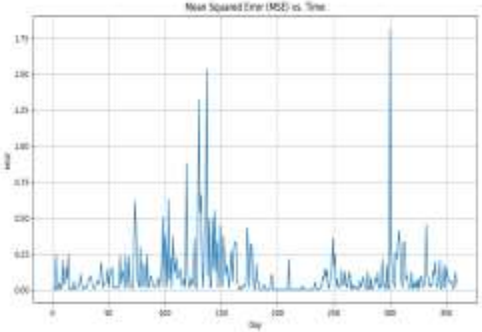
MSE PLOT IN THE YEAR 2020



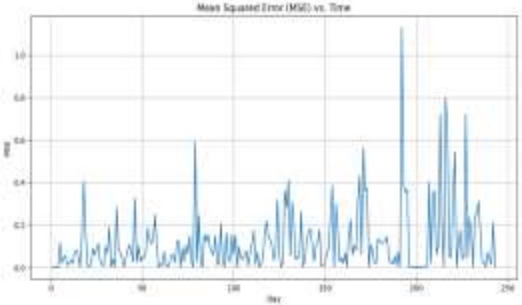
MSE PLOT IN THE YEAR 2021



MSE PLOT IN THE YEAR 2022



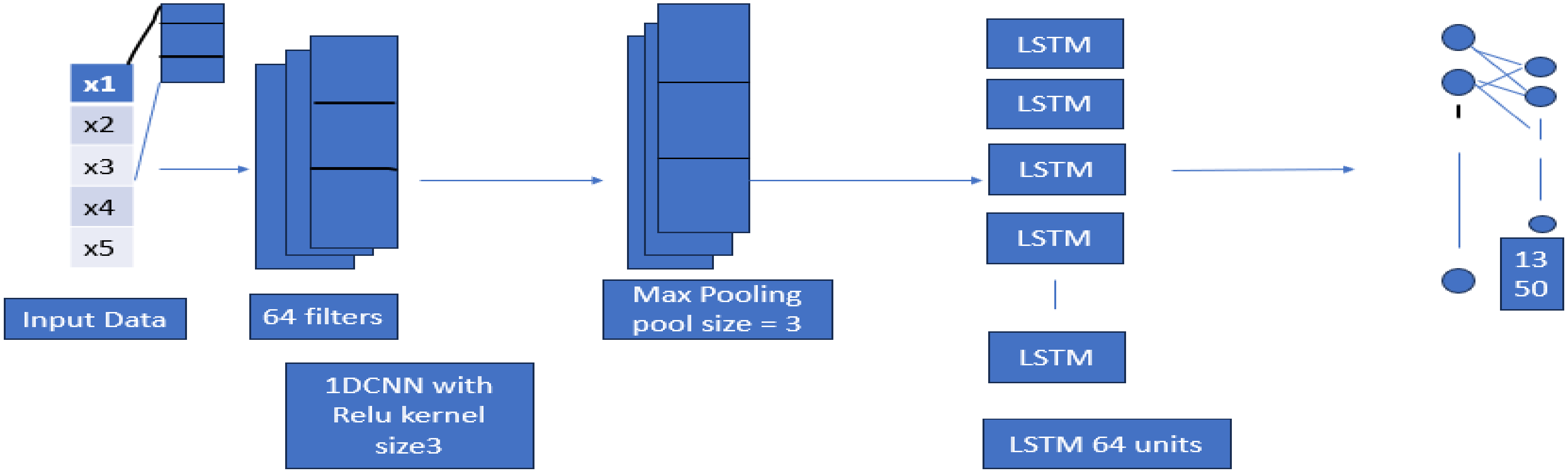
MSE PLOT IN THE YEAR 2023



MSE VALUE FOR EVERY YEAR FOR FIT THE CNN MODEL

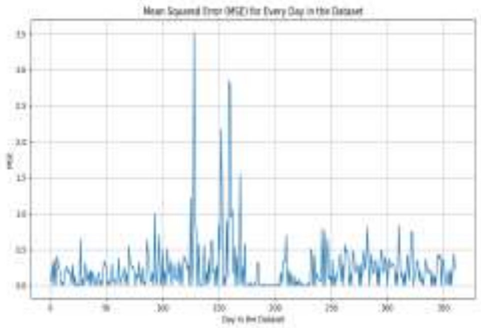
- THE MEAN SQUARE ERROR OF TEST DATA IN 2018 YEAR 0.3632970860844606
- THE MEAN SQUARE ERROR OF TEST DATA IN 2019 YEAR 0.13784187801928338
- THE MEAN SQUARE ERROR OF TEST DATA IN 2020 YEAR 0.24044444372942364
- THE MEAN SQUARE ERROR OF TEST DATA IN 2021 YEAR 0.23934857483316704
- THE MEAN SQUARE ERROR OF TEST DATA IN 2022 YEAR 0.10371696565101717
- THE MEAN SQUARE ERROR OF TEST DATA IN 2023 YEAR 0.13877134106530337

CNN -LSTM HYBRID MODEL ARCHITECTURE

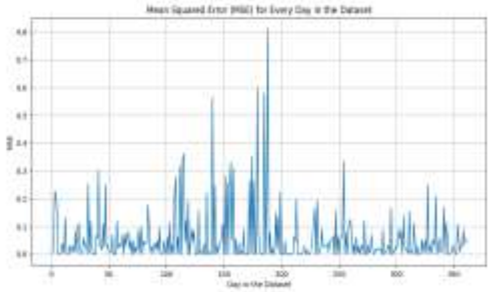


MSE PLOT AFTER FITTING THE CNN-LSTM MODEL OVER SIX YEAR

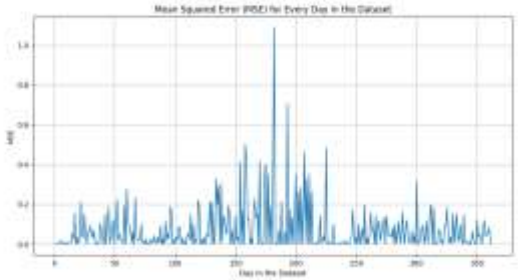
MSE PLOT IN THE YEAR 2018



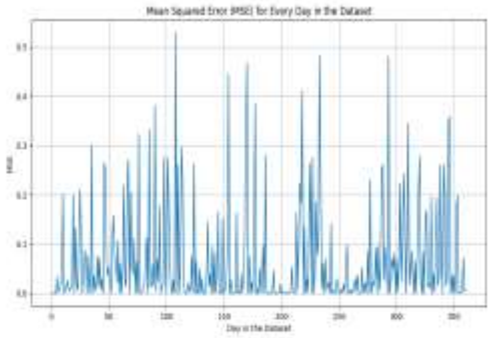
MSE PLOT IN THE YEAR 2019



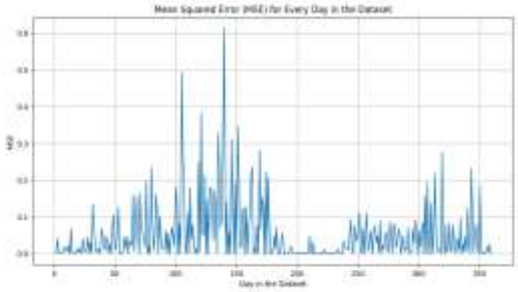
MSE PLOT IN THE YEAR 2020



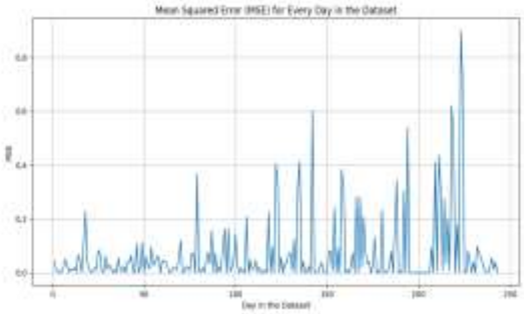
MSE PLOT IN THE YEAR 2021



MSE PLOT IN THE YEAR 2022

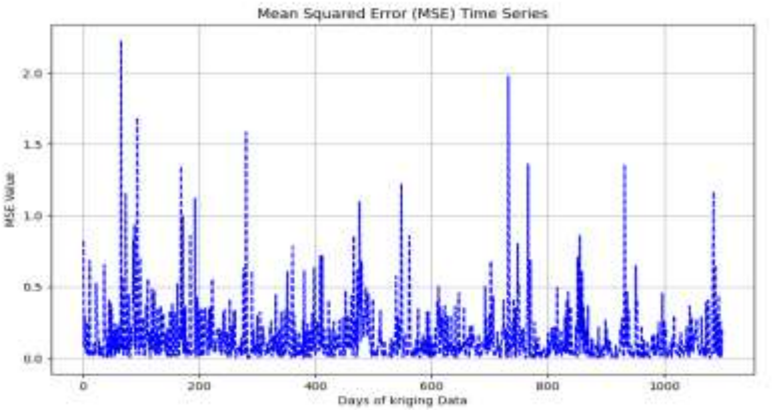


MSE PLOT IN THE YEAR 2023

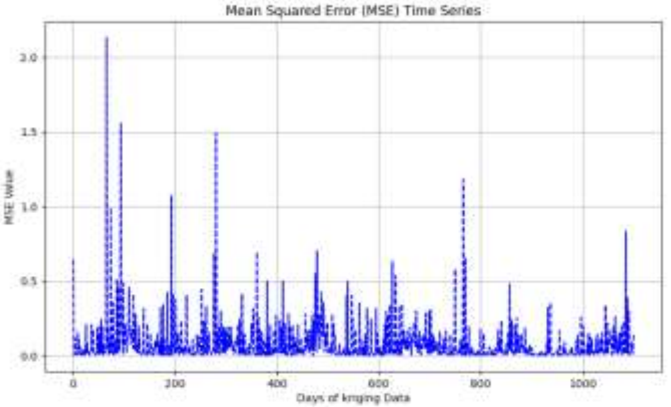


MSE VALUE FOR EVERY YEAR FOR FITTING THE CNN-LSTM HYBRID MODEL

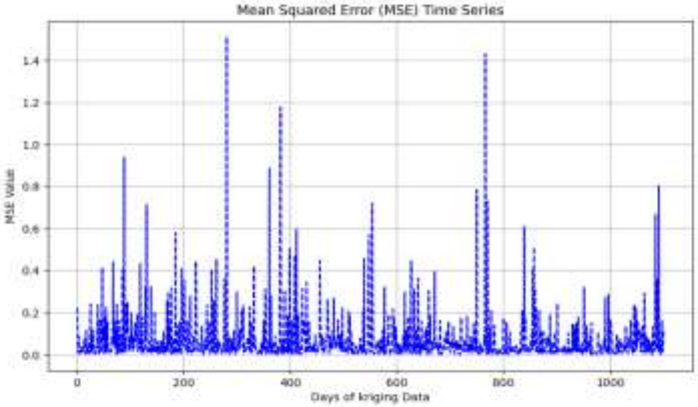
- THE MEAN SQUARE ERROR OF TEST DATA IN 2018 YEAR 0.17620744524437956
- THE MEAN SQUARE ERROR OF TEST DATA IN 2019 YEAR 0.16516667896194457
- THE MEAN SQUARE ERROR OF TEST DATA IN 2020 YEAR 0.10937676890707265
- THE MEAN SQUARE ERROR OF TEST DATA IN 2021 YEAR 0.116533595636971
- THE MEAN SQUARE ERROR OF TEST DATA IN 2022 YEAR 0.05387845607195086
- THE MEAN SQUARE ERROR OF TEST DATA IN 2023 YEAR 0.10629208704792117



CNN

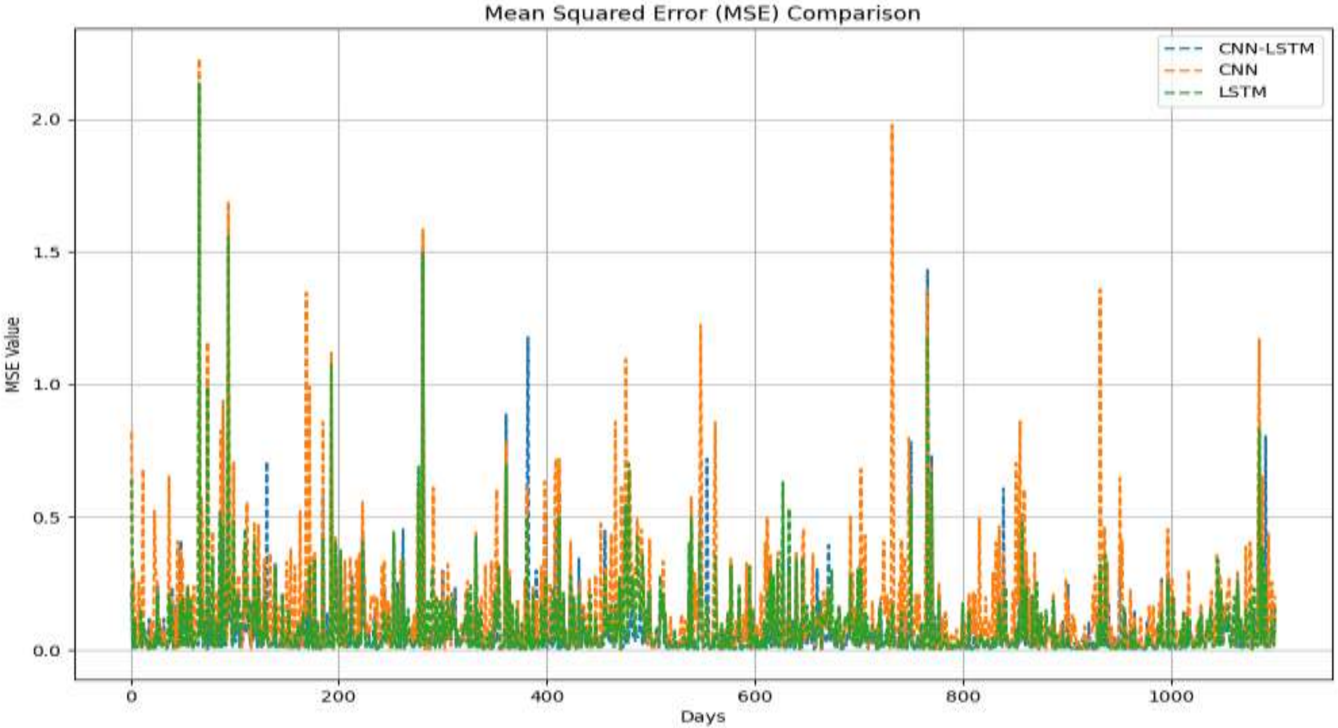


LSTM



CNN-LSTM

MSE PLOT OF THREE MODEL



ERA5 DATASET DOWNLOAD

Click link

era5

- ❖ For further analysis we shall consider this datasets.

CONCLUSION

- ❖ After applying the CNN- LSTM model to the AOD (Aerosol Optical Depth) prediction task, it's evident that the variance of AOD values has undergone a transformation. Prior to model fitting, AOD values exhibited a certain level of variation. However, after fitting the CNN -LSTM model, a new pattern of variation in AOD values has emerged.

- ❖ This observed change suggests a notable improvement in the model's ability to predict AOD values, as it appears to have captured and accounted for different sources of variability, leading to a more accurate and refined prediction.
- ❖ So we follow the proposed method 2 soon ….

THANKING YOU