

DECLARATION BY THE CANDIDATE


I, Pranjol Goswami, hereby declare that the subject matter in this thesis entitled, "Time series characterization and prediction of ambient PM_{2.5} concentrations in India: A deep learning approach", is a presentation of my original research work. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgment of collaborative research and discussions.

The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University.

This thesis is being submitted to the Tezpur University for the degree of Doctor of Philosophy in Environmental Science.

Place: Tezpur University

Date: 28/10/2024


Signature of the candidate



TEZPUR UNIVERSITY



CERTIFICATE OF THE SUPERVISOR

This is to certify that the thesis entitled "**Time series characterization and prediction of ambient PM_{2.5} concentrations in India: A deep learning approach**" submitted to the School of Sciences of Tezpur University in part fulfillment for the award of the degree of Doctor of Philosophy in **Environmental Science** is a record of research work carried out by **Mr. Pranjol Goswami** under my supervision and guidance.

All help received by him from various sources have been duly acknowledged.

No part of this thesis have been submitted elsewhere for award of any other degree /diploma

Supervisor

Amit Prakash
28/10/2024

(Amit Prakash)

Designation: Assistant Professor

School: School of Sciences

Department: Environmental Science

Assistant Professor

Dept. of Environmental Science

Tezpur University,

Date: *28/10/2024*

DECLARATION BY THE CANDIDATE

I, Pranjol Goswami, hereby declare that the subject matter in this thesis entitled, "Time series characterization and prediction of ambient PM_{2.5} concentrations in India: A deep learning approach", is a presentation of my original research work. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgment of collaborative research and discussions.

The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University.

This thesis is being submitted to the Tezpur University for the degree of Doctor of Philosophy in Environmental Science.

Place: Tezpur University

Signature of the candidate

Date:



TEZPUR UNIVERSITY

CERTIFICATE OF THE SUPERVISOR

This is to certify that the thesis entitled "**Time series characterization and prediction of ambient PM_{2.5} concentrations in India: A deep learning approach**" submitted to the School of **Sciences** of Tezpur University in part fulfillment for the award of the degree of Doctor of Philosophy in **Environmental Science** is a record of research work carried out by **Mr. Pranjol Goswami** under my supervision and guidance.

All help received by him from various sources have been duly acknowledged.

No part of this thesis have been submitted elsewhere for award of any other degree /diploma

Supervisor

(Amit Prakash)

Designation: Assistant Professor

School: School of Sciences

Department: Environmental Science

Date:

Acknowledgements:

I would like to express my sincere and heartfelt gratitude to all individuals, institutions and open source resources for support and guidance to complete my Ph.D.

First of all I would like to thank and express my sincere gratitude to my research supervisor, Dr. Amit Prakash, for giving me the opportunity to join Ph.D. under his guidance. I am extremely grateful for his invaluable guidance, and support throughout the Ph.D. course. His lessons and expertise have become instrumental in shaping the direction of my research .

I would like to thank and extend my gratitude towards doctoral committee (DC) members: Prof. A.K. Das and Dr. A. N. Jha for support, guidance and valuable suggestions in my research work.

I would also like to thank to all faculty members of the Department of Environmental Science, Tezpur University: Prof. R.R. Hoque, Prof. A. Devi, Prof. K. Marimuthu, Dr. N. Gogoi, Dr. S. S. Bhattacharya, Dr. S. Handique, Dr. N. Gogoi, Dr. S. Kalita and Dr. P Deka for their valuable suggestions and cooperation throughout my research work. Their willingness to engage in thoughtful discussions greatly enriched the quality of my findings.

I am also grateful to the research scholars who volunteered their time and shared their insights, without whom this research would not been possible.

I extend my thanks to my colleagues who offered valuable suggestions and assistance during the various stages of this research. Their input has been invaluable.

I would also like to acknowledge the Central Pollution Control Board and other State Pollution control Boards of India for providing data for my research work. My acknowledgement also goes to open source programmes Python, Keras and CUDA.

I am grateful to the Department of Science and Technology, DST, New Delhi, for their financial support in purchasing a high-performance computing system as part of the DST Project. I could not have finished my research without this machine.

I convey my sincere gratitude to Tezpur University for giving me NOC for doing Ph.D. I also acknowledge the Department of Environmental Science, Tezpur University for providing computing systems, for my research work.

Lastly, I want to acknowledge the enduring support of my family and friends. Their encouragement and understanding have been a constant source of strength.

List of Figures:

Figure 2-1 Agro climatic zones of India	33
Figure 2-2: Windrose Plot of EH	36
Figure 2-3 Windrose Plot of LGP	36
Figure 2-4 Windrose Plot of MGP	36
Figure 2-5 Windrose Plot of UGP	36
Figure 2-6 Windrose Plot of TGP	37
Figure 2-7 Windrose Plot of EPH	37
Figure 2-8 Windrose Plot of CPH.....	37
Figure 2-9 Windrose Plot of WPH.....	37
Figure 2-10 Windrose Plot of SPH	38
Figure 2-11 Windrose Plot of ECPH	38
Figure 2-12 Windrose Plot of WCPG.....	38
Figure 2-13 Windrose Plot of GPH	38
Figure 2-14 Windrose Plot of WD.....	39
Figure 2-15 Box plot of Data	40
Figure 2-16 Average distribution of Data	40
Figure 2-17 Line plot of Delhi data	41
Figure 2-18 Seq2Seq model.....	42
Figure 2-19 RNN	43
Figure 2-20 LSTM Network	44
Figure 2-21 3D CNN	45
Figure 2-22 ConvLSTM	46
Figure 2-23 BConvLSTM.....	47
Figure 3-1: Hybrid Model.....	49
Figure 3-2: Stacked LSTM	52
Figure 3-3: CNN -LSTM	52
Figure 3-4: ConvLSTM-LSTM	53
Figure 3-5: ConvLSTM-BLSTM.....	53
Figure 3-6: BConvLSTM-LSTM.....	54
Figure 3-7: BConvLSTM-BLSTM	54
Figure 3-8: BLSTM-BLSTM.....	55
Figure 3-9: Overlapping Moving Window	56

Figure 4-1: Seasonal Plots	64
Figure 4-2:(A-Z) Distribution Plot.....	69
Figure 4-3: Inter Station Correlation.....	71
Figure 4-4: Autocorrelation 1	73
Figure 4-5: Autocorrelation 2	73
Figure 4-6: Data decomposition.....	74
Figure 4-7 Model Comparison.....	77
Figure 4-8: Actual Vs Predicted Plot for Delhi	85
Figure 4-9: Sub plot of Figure 4-8	86
Figure 4-10: Scatter plot of Delhi	86
Figure 4-11: Heat Map of MAE Error	87
Figure 4-12: Model performance Agro climatic zone wise	89
Figure 4-13: Data length Vs Error Plot.....	91
Figure 4-14: Correlation between Data length and Error	92
Figure 4-15: SNR Vs Model performance.....	93
Figure 4-16: IoA for 26 Stations.....	98
Figure 4-17: Quiver Plot for Pre- Monsoon.....	99
Figure 4-18: Quiver Plot for Monsoon	100
Figure 4-19: Quiver Plot for Post-Monsoon	101
Figure 4-20: Quiver Plot for Winter	102

List of Tables:

Table 2-1: Data Description.....	34
Table 3-1: Hyperparamters Value.....	57
Table 3-2: Hyperband Parameters	58
Table 4-1: Jerque bera test	61
Table 4-2: KS statistics	62
Table 4-3: Shapiro wilk test.....	63
Table 4-4: Best fit Statistical distribution parameters values	70
Table 4-5: SARIMAX Model	72
Table 4-6: SARIMAX Result	72
Table 4-7 BConvLSTM-BLSTM	75
Table 4-8 BConvLSTM-LSTM	75
Table 4-9 Stacked BLSTM	75
Table 4-10 CNN-LSTM.....	76
Table 4-11 ConvLSTM-BLSTM	76
Table 4-12 ConvLSTM-LSTM.....	76
Table 4-13 Stacked LSTM.....	77
Table 4-14: Result of 8 step ahead prediction	80
Table 4-15: Model performance agro-climatic zone wise	88
Table 4-16: SNR Values	90

Abbreviations

ARIMA	Auto Regressive Integrated Moving Average
LSTM	Long Short Term Memory
BLSTM	Bidirectional Long Short Term Memory
ConvLSTM	Convolutional Long Short Term Memory
CNN	Convolutional Neural Network
BConvLSTM	Bidirectional Convolutional Neural Network
3DCNN	Three Dimensional Convolutional Neural Networks
RNN	Recurrent Neural Network
SNR	Signal to Noise Ratio
CPCB	Central Pollution Control Board of India
PM _{2.5}	Particulate Matter of Size 2.5 Micron
EH	Eastern Himalayan Region
LGP	Lower Gangetic Plain Region
MGP	Middle Gangetic Plain Region
UGP	Upper Gangetic Plains Region
TGP	Trans-Ganga Plains Region
EPH	Eastern Plateau and Hills
CPH	Central Plateau and Hills
WPH	Western Plateau and Hills
SPH	Southern Plateau and Hills
ECPH	Eastern Coastal Plains and Hills
WCPG	Western Coastal Plains and Ghats
GPH	Gujarat Plains and Hills
WD	Western Dry Region
WH	Western Himalayan Division
IR	Island Region