MODEL DEVELOPMENT

`

4 Chapter 4

4.1 Results and Discussions

4.1.1 Statistical Distribution Analysis (Objective 1)

Normality test was carried out on time series data with KS statistic, Shapiro-wilk test and Jerque-bera test. The normality test results (Table 4-1 to Table 4-3) showed that the nature of the $PM_{2.5}$ dataset is not normal across the different agricultural climate zones of India.

Table 4-1: Jerque bera test

Sl.	Station Name	Statistics	p - value
1	Gaya	26352.07	0
2	Howrah	23894.81	0
3	Aurangabad	34571.29	0
4	Shillong	9087.63	0
5	Thiruvananthapuram	30495.83	0
6	Talcher	20338.43	0
7	Delhi	21342.89	0
8	Lucknow	55558.44	0
9	Amritsar	42599.20	0
10	Siliguri	22019.28	0
11	Ahmedabad	36304.95	0
12	Vijaywada	16395.21	0
13	Hyderabad	953.17	0
14	Varanasi	30553.29	0
15	Kanpur	30440.94	0
16	Patna	9077.11	0
17	Guwahati	9501.70	0
18	Agra	52506.72	0
19	Nagpur	26394.50	0
20	Jodhpur	32282.83	0
21	Chennai	110593.70	0
22	Mumbai	8069.81	0
23	Solapur	29087.92	0
24	Bengaluru	202634.40	0
25	Jamshedpur	2535.89	0
26	Mandideep	43518.65	0

Table 4-2: KS statistics

Sl.	Station Name	Statistics	p -value
1	Gaya	1	0
2	Howrah	1	0
3	Aurangabad	1	0
4	Shillong	1	0
5	Thiruvananthapuram	1	0
6	Talcher	1	0
7	Delhi	1	0
8	Lucknow	1	0
9	Amritsar	1	0
10	Siliguri	1	0
11	Ahmedabad	1	0
12	Vijaywada	1	0
13	Hyderabad	1	0
14	Varanasi	1	0
15	Kanpur	1	0
16	Patna	1	0
17	Guwahati	1	0
18	Agra	1	0
19	Nagpur	1	0
20	Jodhpur	1	0
21	Chennai	1	0
22	Mumbai	1	0
23	Solapur	1	0
24	Bengaluru	1	0
25	Jamshedpur	1	0
26	Mandideep	1	0

Table 4-3: Shapiro wilk test

Sl.	Station Name	Statistics	p -value
1	Gaya	0.834	0
2	Howrah	0.803	0
3	Aurangabad	0.878	0
4	Shillong	0.764	0
5	Thiruvananthapuram	0.810	0
6	Talcher	0.806	0
7	Ito	0.843	0
8	Lucknow	0.810	0
9	Amritsar	0.816	0
10	Siliguri	0.807	0
11	Ahmedabad	0.837	0
12	Vijaywada	0.914	0
13	Hyderabad	0.953	0
14	Varanasi	0.840	0
15	Kanpur	0.810	0
16	Patna	0.856	0
17	Guwahati	0.805	0
18	Agra	0.794	0
19	Nagpur	0.859	0
20	Jodhpur	0.872	0
21	Chennai	0.813	0
22	Mumbai	0.865	0
23	Solapur	0.890	0
24	Bengaluru	0.803	0
25	Jamshedpur	0.940	0
26	Mandideep	0.821	0

Data analysis based on seasonality revealed that $PM_{2.5}$ concentrations range from 195.5 $\mu g/m^3$ during the winter to 113.8 $\mu g/m^3$ during the monsoon. Many authors reported similar results earlier [134]. During the monsoon season, the lowest mean concentration was observed in the WCPG region and the highest mean concentration in the WD. The MGP has the highest average concentration during the winter, while the EH area has the lowest. Notably, the highest average concentration is found in the MGP region during the pre-monsoon season, and the highest average concentration is found in the TGP region during the post-monsoon season.

Northwestern India has higher PM2.5 concentrations, particularly during the monsoon and post-monsoon seasons, according to the seasonal plots (Figure 4-1). The Indo-Gangetic plain region had the highest concentration during the winter and pre-monsoon season.

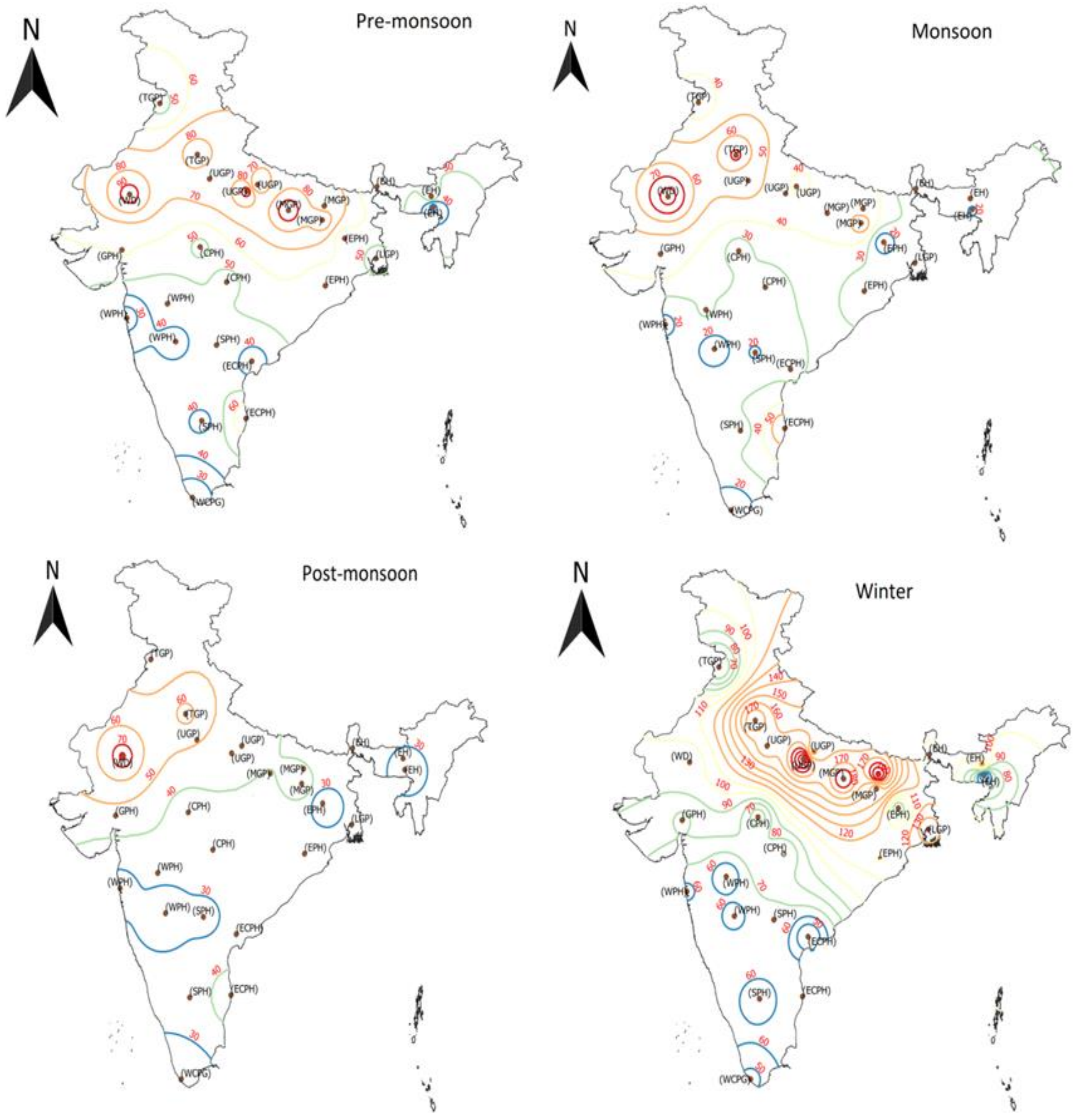


Figure 4-1: Seasonal Plots

`

To find out what kind of distribution is predominated in the PM2.5 data of different agricultural climate zones, we conducted a thorough statistical distribution analysis. Understanding the statistical distribution in data is crucial for effective forecasting. Seven common distributions were applied to each dataset to determine the best fit model for air pollution studies. These distributions include- Gamma, Beta, Log Normal, Weibull, Logistic, Normal, and Laplace. The minimum sum of square error criteria was used to find the best-fit distribution (Figure 4-2 (A-Z)).

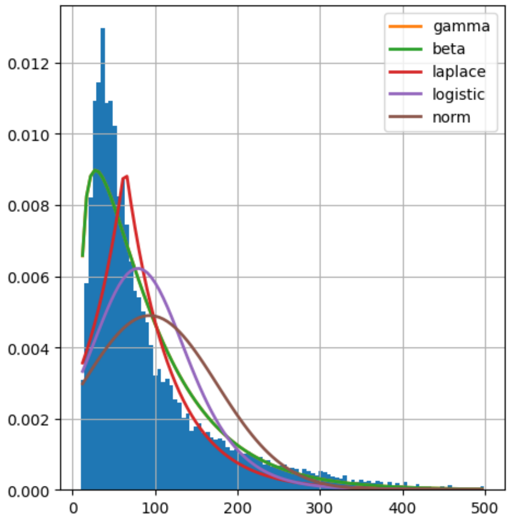


Fig. A (Agra)

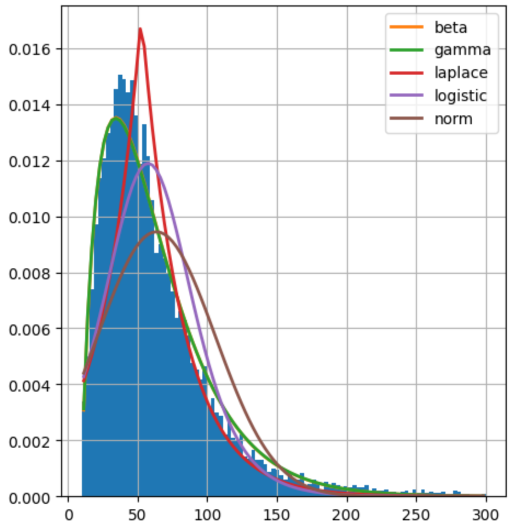


Fig. B (Ahmedabad)

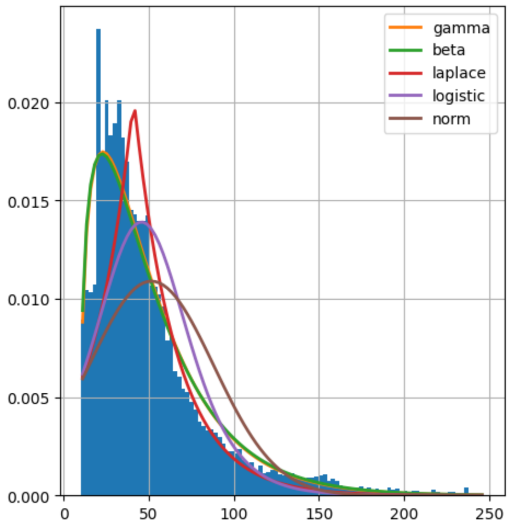


Fig. C (Amritsar)

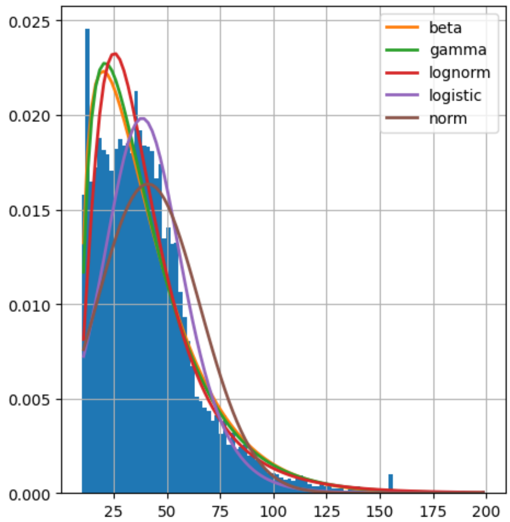


Fig. D (Aurangabad)

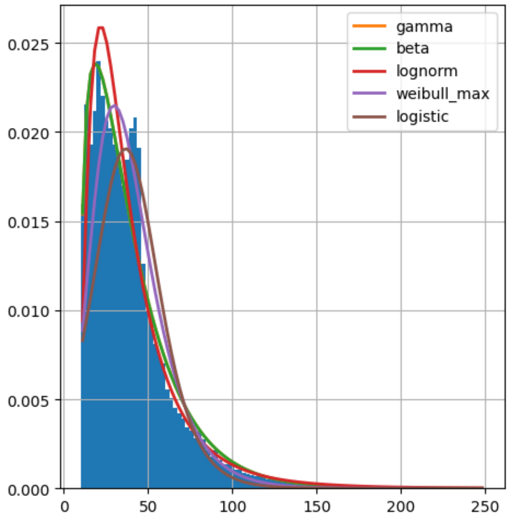


Fig. E (Bengaluru)

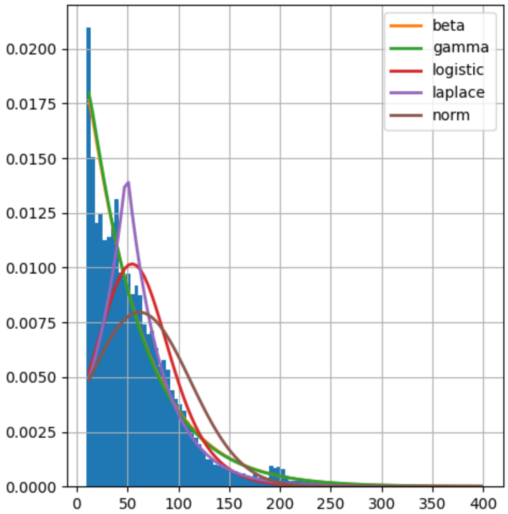


Fig. F (Chennai)

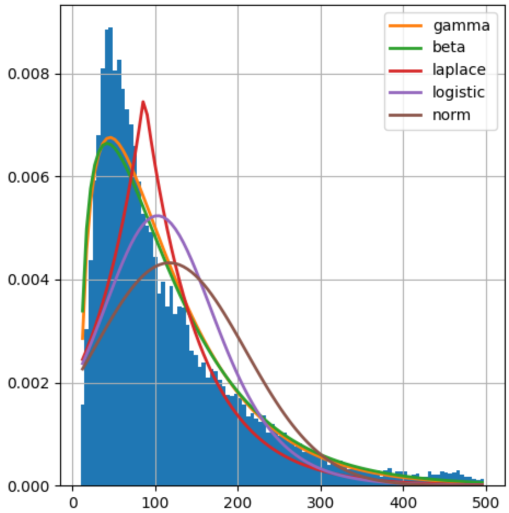


Fig. G (Delhi)

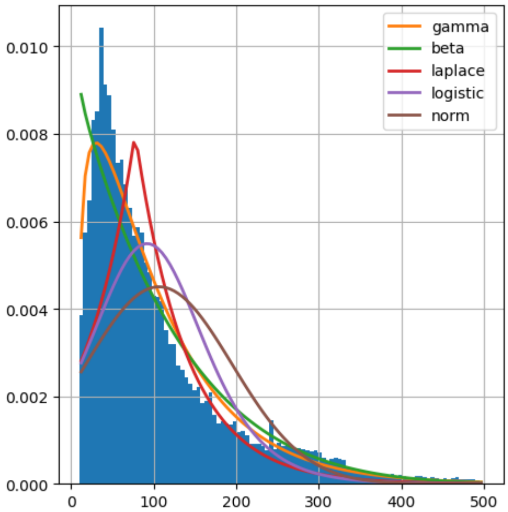


Fig. H (Gaya)

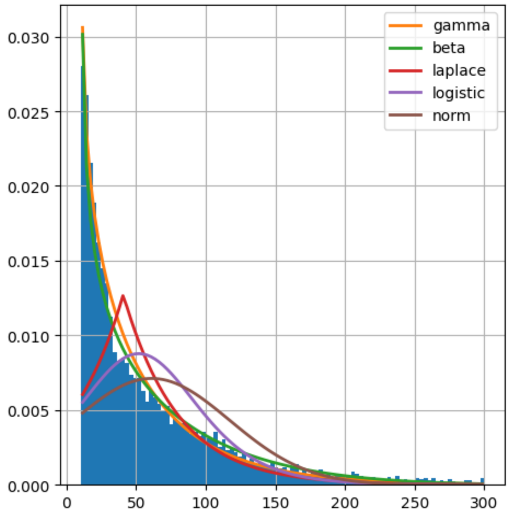


Fig. I (Guwahati)

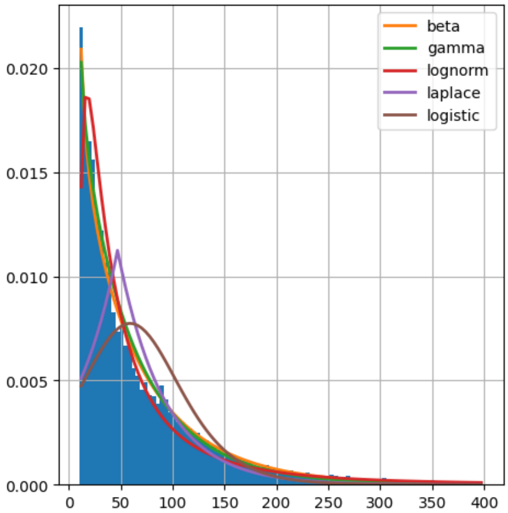


Fig. J (Howrah)

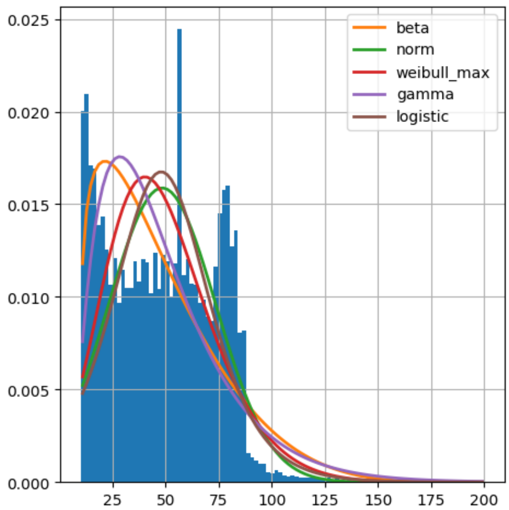


Fig. K (Hyderabad)

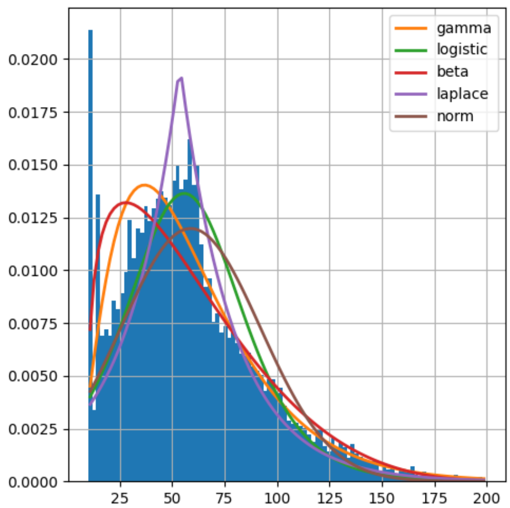


Fig. L (Jamshedpur)

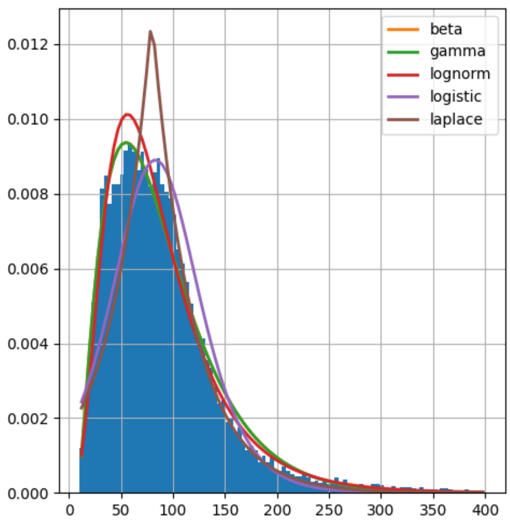


Fig. M (Jodhpur)

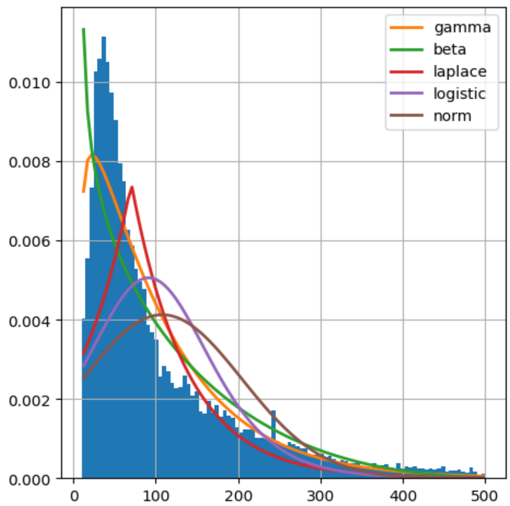


Fig. N (Kanpur)

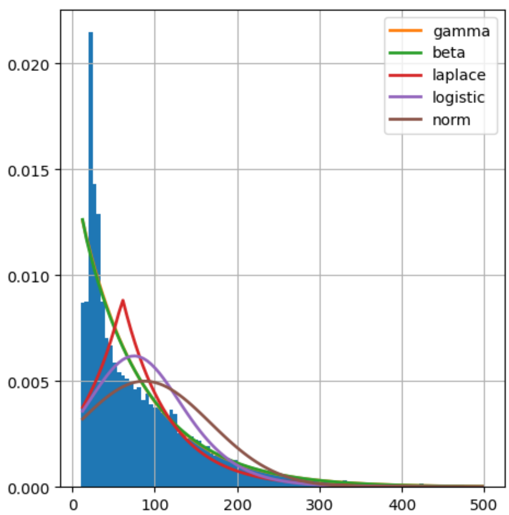


Fig. O (Lucknow)

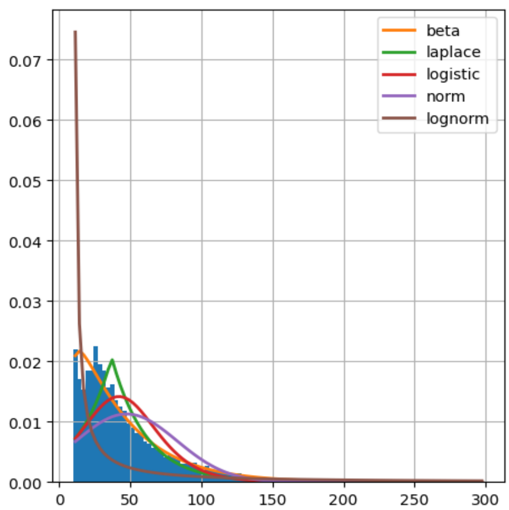


Fig. P (Mandideep)

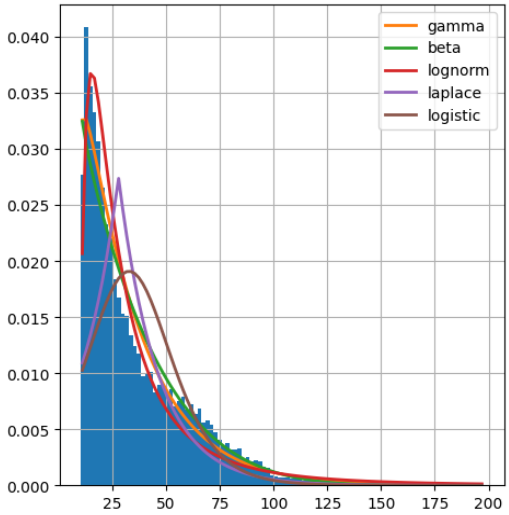


Fig. Q (Mumbai)

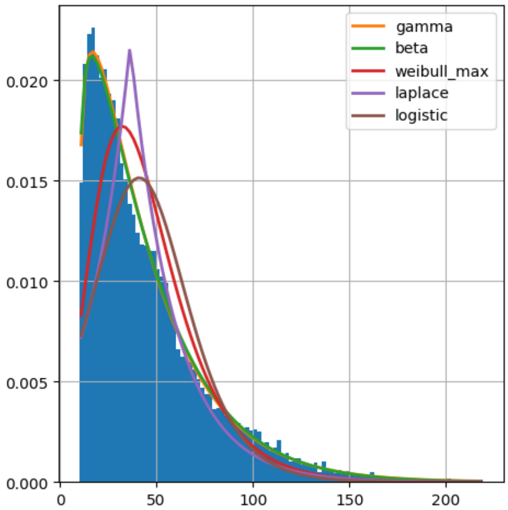


Fig. R (Nagpur)

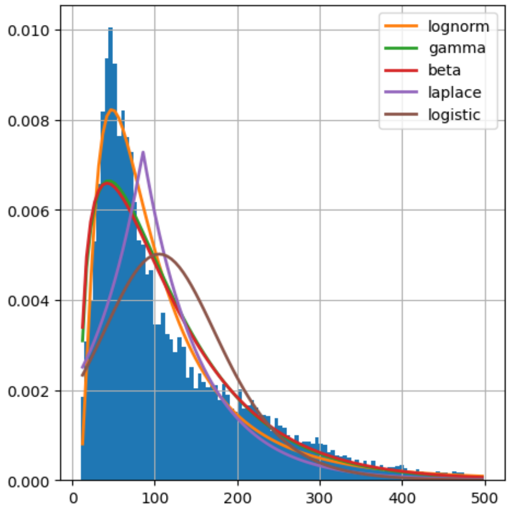


Fig. S (Patna)

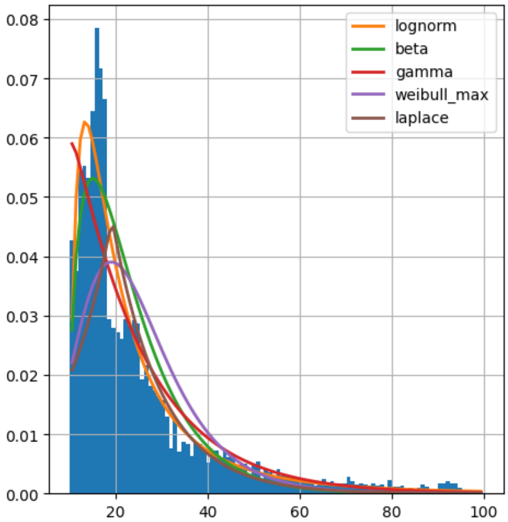


Fig. T (Shillong)

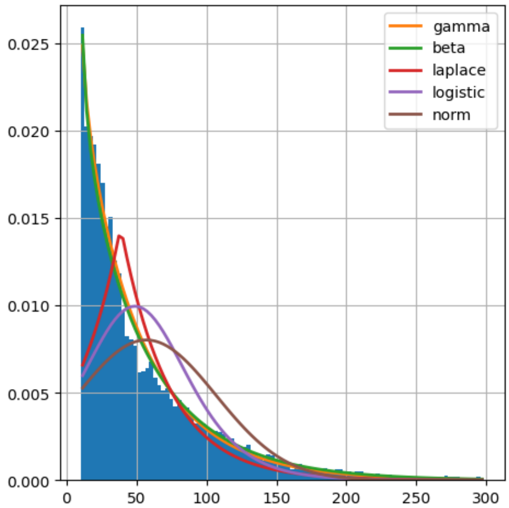


Fig. U (Siliguri)

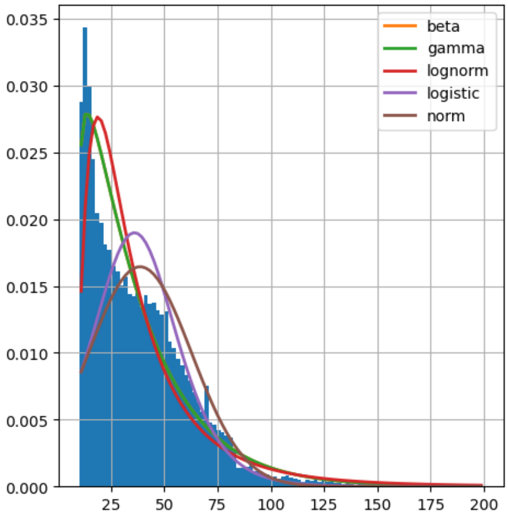


Fig. V (Solapur)

`

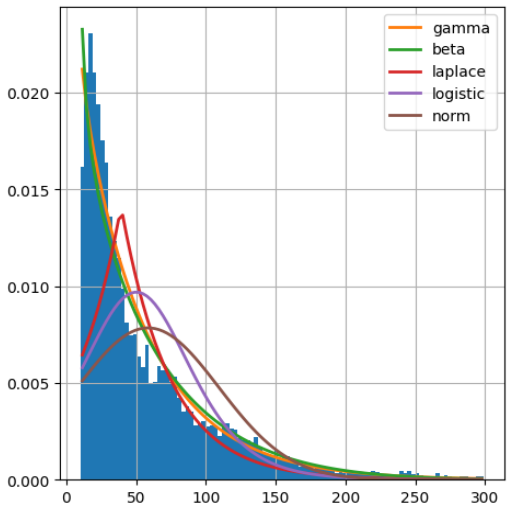


Fig. W (Talcher)

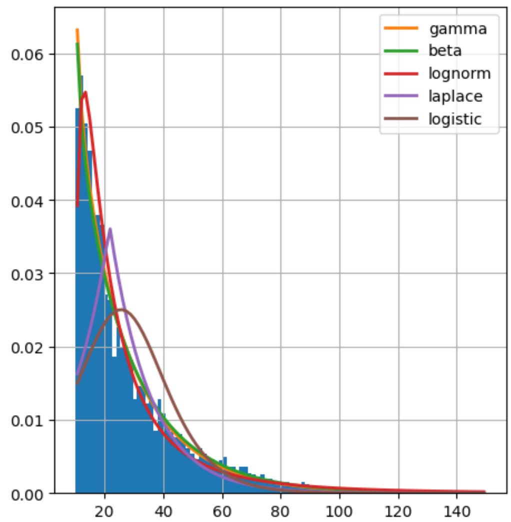


Fig. X (Thiruvananthapuram)

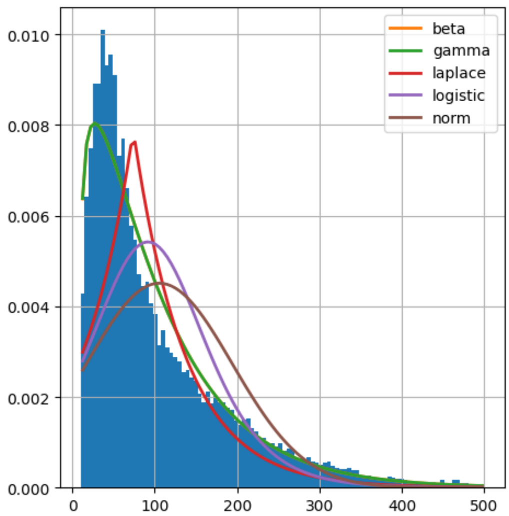


Fig. Y (Varanasi)

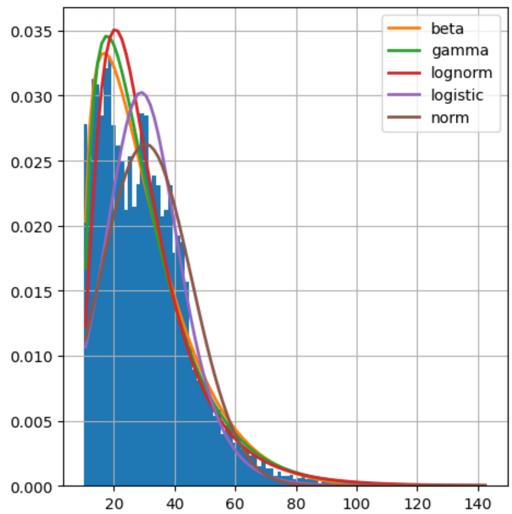


Fig. Z (Vijaywada)

Figure 4-2:(A-Z) Distribution Plot

At 24 out of 26 sites, the gamma and beta distributions were predominant (Table 4-4). At 14 locations, the concentration of $PM_{2.5}$follows a gamma distribution, at 10 places a beta distribution, and at the rest of the two locations a lognormal distribution. It is important to note that the Gamma distribution is the best fit distribution for the sites at EPH, TGP and UGP. LGP, ECPH, and GPH all solely follow the beta distribution. A log normal distribution was only observed at Patna in the MGP and Shillong in the EH area. Mixed distribution fitting results were found in the remaining regions.

`

Table 4-4: Best fit Statistical distribution parameters values

Agro Climatic Zone	Site	Best fit Distribution	Location parameter	Scale Parameter	Scale/ Location
Eastern Himalayan Region	Guwahati	gamma	1.00E+01	5.55E+01	5.55E+00
	Shillong	lognorm	9.18E+00	1.05E+01	1.14E+00
	Siliguri	gamma	1.00E+01	4.79E+01	4.79E+00
Lower Gangetic Plain Region	Howrah	beta	1.00E+01	5.34E+03	5.34E+02
Middle Gangetic Plain Region	Gaya	gamma	9.88E+00	7.62E+01	7.71E+00
	Patna	lognorm	4.89E+00	8.32E+01	1.70E+01
	Varanasi	beta	9.93E+00	2.85E+13	2.87E+12
Upper Gangetic Plain Region	Agra	gamma	9.94E+00	6.55E+01	6.59E+00
	Kanpur	gamma	9.97E+00	8.55E+01	8.57E+00
	Lucknow	gamma	1.00E+01	7.81E+01	7.81E+00
Trans-Ganga Plain Region	Amritsar	gamma	9.95E+00	2.84E+01	2.85E+00
	Delhi	gamma	9.96E+00	7.18E+01	7.21E+00
Eastern Plateau and Hills	Jamshedpur	gamma	3.94E+00	2.20E+01	5.58E+00
	Talcher	gamma	1.00E+01	5.01E+01	5.01E+00
Central Plateau and Hills	Mandideep	beta	9.95E+00	5.30E+03	5.33E+02
	Nagpur	gamma	1.00E+01	2.88E+01	2.88E+00
Western Plateau and Hills	Aurangabad	beta	9.67E+00	4.45E+02	4.60E+01
	Mumbai	gamma	1.00E+01	2.43E+01	2.43E+00
	Solapur	beta	1.00E+01	4.53E+08	4.54E+07
Southern Plateau and Hills	Bengaluru	gamma	9.92E+00	2.19E+01	2.20E+00
	Hyderabad	beta	9.78E+00	1.91E+02	1.95E+01
Eastern Coastal Plain and Ghats	Chennai	beta	1.00E+01	2.63E+03	2.63E+02
	Vijayawada	beta	9.82E+00	1.73E+02	1.76E+01
Western Coastal Plain and Ghats	Thiruvananthapuram	gamma	1.00E+01	2.03E+01	2.03E+00
Gujrat Plain and Hills	Ahmedabad	beta	9.67E+00	6.43E+13	6.64E+12
Western Dry Region	Jodhpur	beta	8.59E+00	1.62E+12	1.88E+11

`

An inter station correlation of $PM_{2.5}$ concentrations were done and results are described in Figure 4-3.

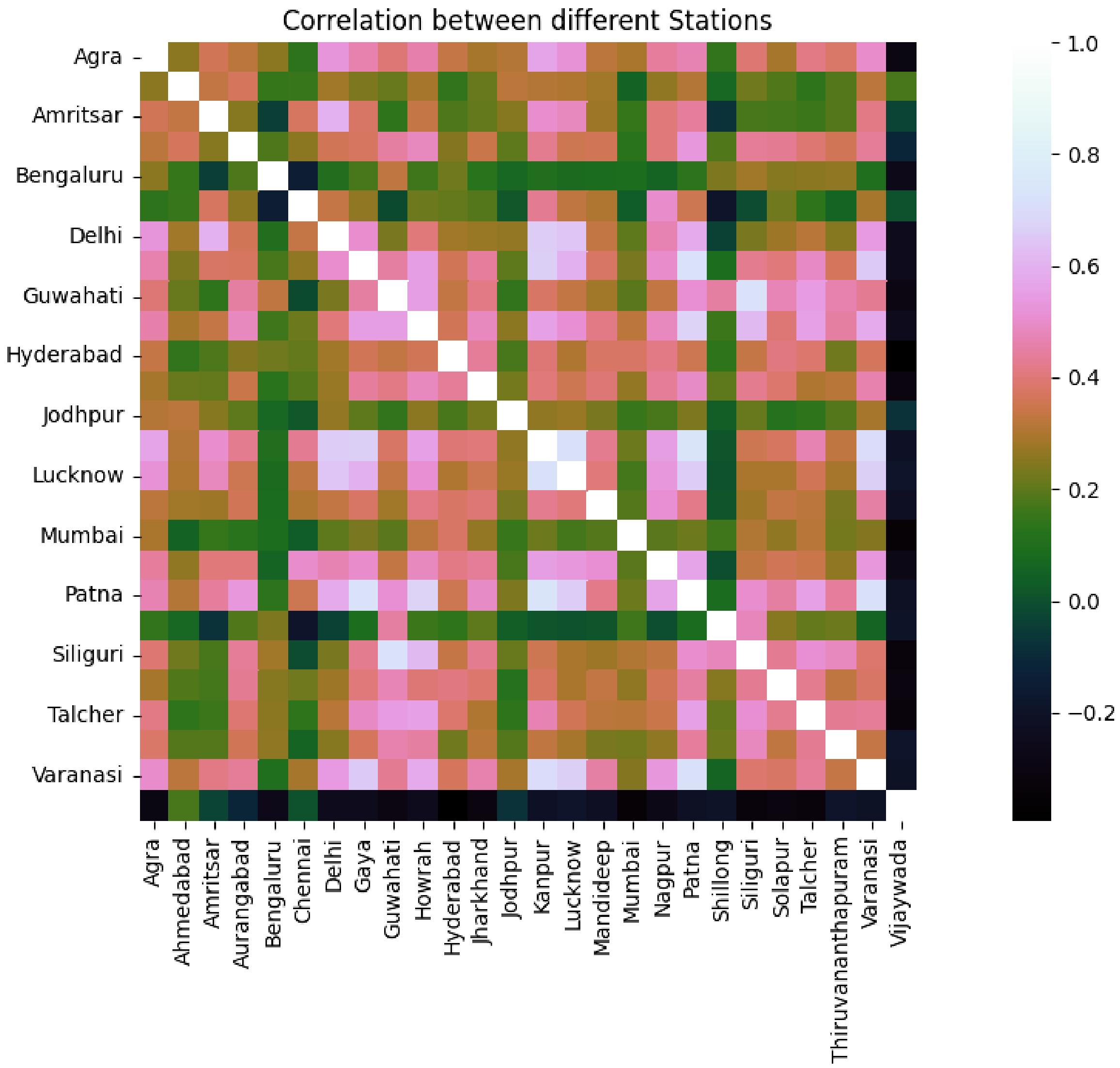


Figure 4-3: Inter Station Correlation

4.1.2 Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX)

We employed the statistical technique known as Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) as a baseline for comparing the results. SARIMAX was performed in selected cities covering all four directions of

India. Delhi the national capital of India is one of the highly polluted cities of the world. SARIMAX result (Table 4-5) for Delhi is presented below.

Table 4-5: SARIMAX Model

```
SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                52608
Model:               SARIMAX(3, 0, 1)   Log Likelihood             -254688.616
Date:                Tue, 10 Oct 2023   AIC                         509387.232
Time:                        20:29:29   BIC                         509431.585
Sample:                             0   HQIC                        509401.092
                              - 52608
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          2.0699      0.001   1751.207      0.000       2.068       2.072
ar.L2         -1.2842      0.002   -823.438      0.000      -1.287      -1.281
ar.L3          0.2141      0.001    223.460      0.000       0.212       0.216
ma.L1         -0.9703      0.001   -931.464      0.000      -0.972      -0.968
sigma2       938.6782      0.971    966.710      0.000     936.775     940.581
===================================================================================
Ljung-Box (L1) (Q):                   2.14   Jarque-Bera (JB):          36306755.56
Prob(Q):                              0.14   Prob(JB):                         0.00
Heteroskedasticity (H):               0.66   Skew:                             0.27
Prob(H) (two-sided):                  0.00   Kurtosis:                       131.70
===================================================================================
```

Table 4-6 gives the result of multi-step error for the SARIMAX model. Figure 4-4 and Figure 4-5 shows the Autocorrelation plots of the SARIMAX model. Figure 4-6 is the decomposition plot of the time series data.

Table 4-6: SARIMAX Result

ERROR	1-Hr	2-Hr	3-Hr	4-Hr	5-Hr	6-Hr	7-Hr	8-Hr	Cumulative
RMSE	29.66	48.18	55.14	62.44	71.37	77.02	81.78	87.88	66.69
MAE	16.06	27.04	32.48	36.67	42.56	48.64	52.48	56.04	39
MAPE	16.37	27.05	31.94	35.23	42.57	53.36	61.32	60.36	41.02

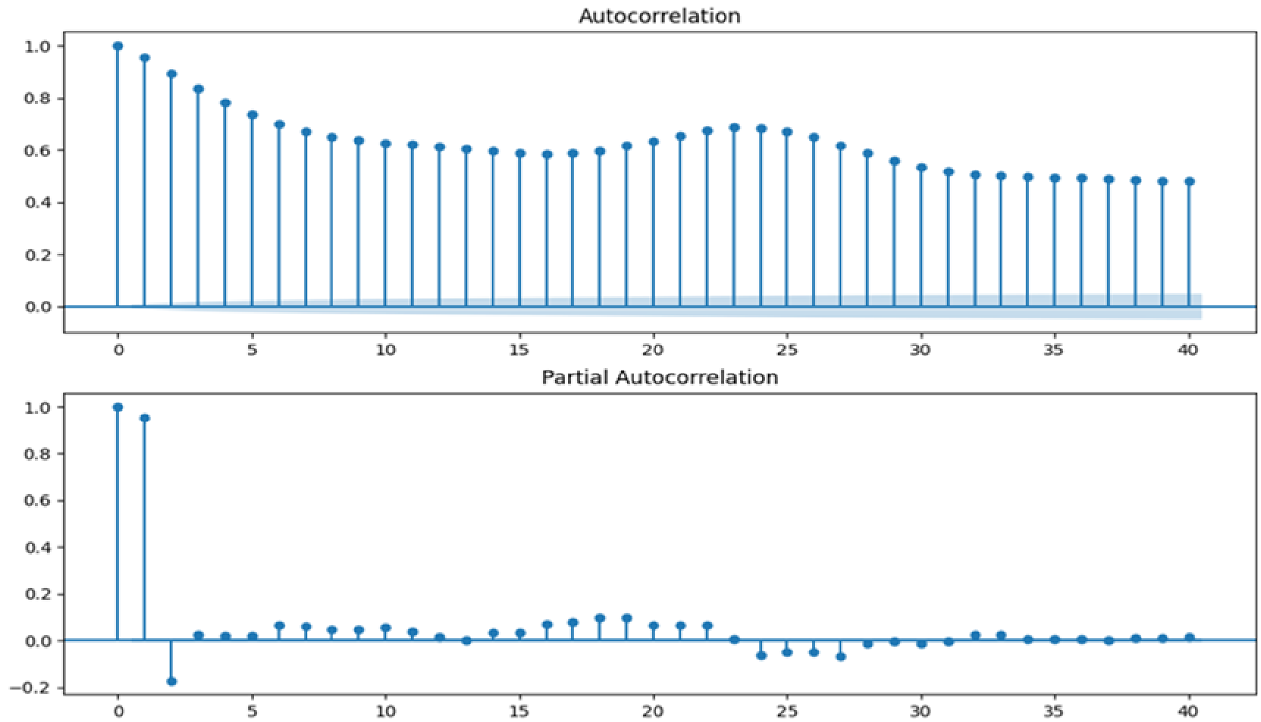


Figure 4-4: Autocorrelation 1

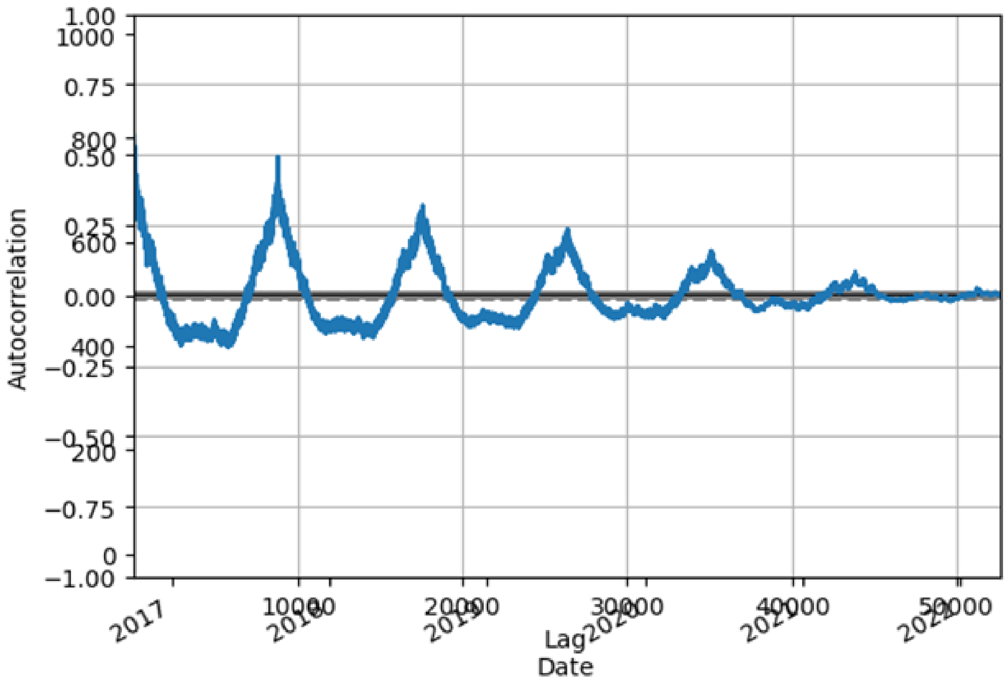


Figure 4-5: Autocorrelation 2

`

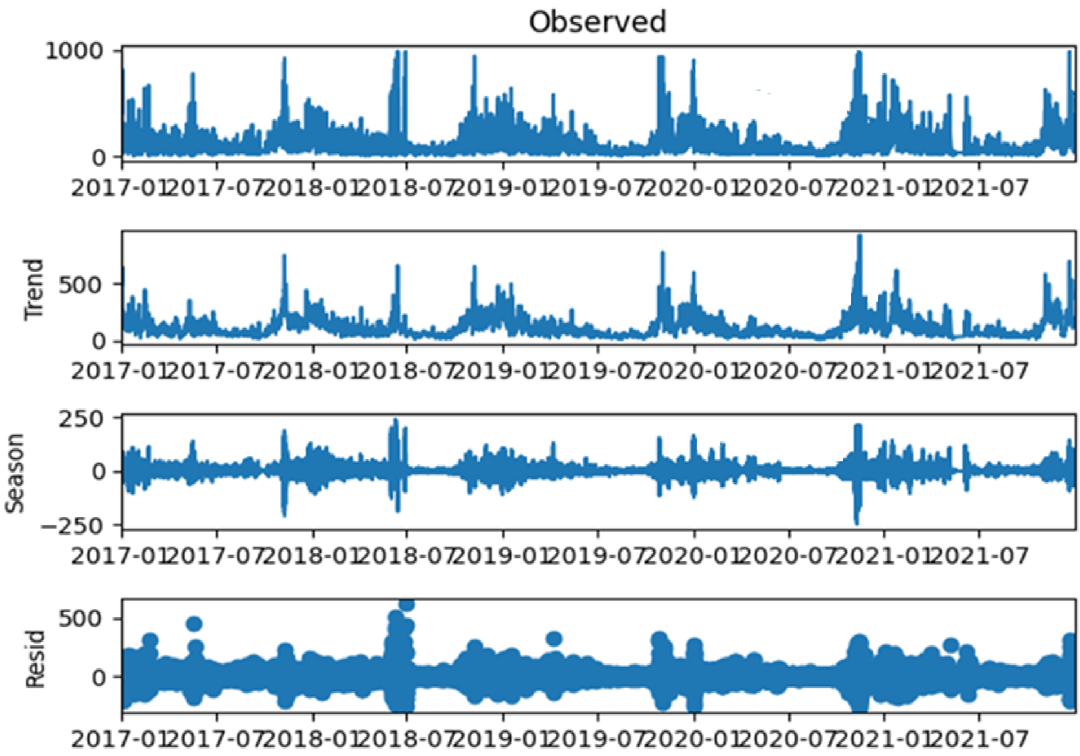


Figure 4-6: Data decomposition

4.1.3 Comparative study of encoder-decoder based deep learning models (Objective 3):

Several Seq2Seq forecasting models based on encoder-decoder architecture have been tested. A total of seven deep learning models were used for the prediction of 8 hour ahead forecasting using previous sequence data of different combinations. Previous sequence of 8 hour, 16 hour, 24 hour, 48 hour and 72 hour were used to forecast future 8 hour values. All models have used the same Guwahati city dataset. The models were namely: stacked LSTM, CNN-LSTM, ConvLSTM-LSTM, stacked BLSTM, ConvLSTM-BLSTM, BLSTM-ConvLSTM, and BConvLSTM-BLSTM.As can be seen from the model architectures, the encoder and decoder parts of each model used the same number of layers. Moreover, it should be noted that all models have used the same hyperparameter configuration. Better results have been found in all models using prior 24- and 72-hour data. Details of the results were presented in Table 4-7 to Table 4-13.

Table 4-7 BConvLSTM-BLSTM

Model Name: BConvLSTM-BLSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	16.67	9.7	21.76	17.28	9.84	20.83	15.93	9.24	20.39	16.71	9.7	20.76	18.1	10.49	24.17
2	24.84	14.16	31.67	27.07	14.82	30.22	22.97	13	29.26	23.51	13.09	28.64	24.15	13.86	31.81
3	29.95	17.26	39.72	33.14	18.08	38.28	26.73	15.23	34.57	27.06	15.26	35.44	27.5	15.9	36.67
4	32.88	19.49	46.21	36.42	20.01	43.81	28.99	16.69	38.22	29.33	16.78	39.78	29.24	17.09	39.42
5	34.87	21.03	51.01	37.76	21.03	46.71	30.7	17.79	41.13	31.26	17.97	42.75	30.44	17.93	41.06
6	36.3	22.33	55.67	37.79	21.34	48.31	31.65	18.45	43.48	32.86	18.89	45.62	31.12	18.44	42.51
7	37.29	23.29	60.08	36.86	21.14	49.92	32.11	18.9	45.25	33.38	19.39	47.75	31.71	18.89	44.08
8	37.82	23.76	62.43	34.97	20.48	50.77	32.57	19.32	47.44	33.43	19.63	50.25	32.22	19.37	45.61
Cumulative	32.07	18.88	46.07	33.34	18.34	41.11	28.23	16.08	37.47	28.97	16.34	38.87	28.42	16.5	38.16

Table 4-8 BConvLSTM-LSTM

Model Name: BConvLSTM-LSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	16.38	9.66	23.97	20.26	12.79	33.94	16.05	9.31	21.92	17.38	10.38	23.06	19.75	12.48	39.24
2	24.86	14.51	35.59	26.17	16.17	40.9	22.82	13.09	31.25	23.34	13.52	29.73	23.95	14.69	43.01
3	30.23	17.8	45.41	30.47	18.56	44.66	26.81	15.44	36.43	26.72	15.55	34.36	26.75	16.2	45.34
4	33.94	20.32	52.9	32.47	19.87	48.37	28.96	16.83	40.75	28.91	16.82	37.82	28.9	17.34	47.66
5	36.72	22.16	58.23	33.74	21.08	56.41	30.49	17.85	44.33	30.4	17.69	40.35	30.27	18.19	49.93
6	38.69	23.72	62.91	34.08	21.1	54.39	31.35	18.55	46.93	31.47	18.41	42.5	31.02	18.82	52.13
7	40.15	24.85	66.5	34.01	21.17	55.19	31.72	18.96	48.55	31.89	18.73	43.98	31.63	19.43	54.07
8	40.86	25.34	68.94	33.65	20.92	54.22	32.49	19.47	50.27	31.79	18.83	44.85	31.9	19.68	54.94
Cumulative	33.69	19.8	51.81	30.95	18.96	48.51	28.08	16.19	40.05	28.15	16.24	37.08	28.31	17.1	48.29

Table 4-9 Stacked BLSTM

Model Name: Stacked BLSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	16.93	9.88	26.87	20.58	11.78	26.63	17.63	11.49	38.03	19.5	12.19	36.76	18.46	10.97	28
2	25.27	14.76	37.12	28.87	16.41	37.69	25.27	15.28	46.78	24.06	15	44.29	23.56	13.86	35.76
3	31.13	18.11	45.74	34.25	19.46	46.13	29.74	17.63	52.5	27.38	16.98	50.22	26.25	15.61	40.88
4	34.88	20.7	52.77	37.64	21.37	51.84	31.56	18.9	56.25	29.62	18.31	55.48	27.81	16.77	44.27
5	37.39	22.62	58.54	38.82	22.14	54.21	32.07	19.58	58.25	30.97	19.09	58.95	28.83	17.49	46.19
6	39.2	24.23	64.23	38.38	22.12	54.54	32.43	20.14	59.4	31.68	19.6	61.2	29.35	17.97	47.52
7	40.3	25.26	67.91	37.1	21.58	53.08	32.93	20.5	59.25	32.16	19.89	62.16	29.61	18.25	48.18
8	40.9	25.9	70.96	35.52	20.93	53.55	33.15	20.44	55.93	32.59	20.29	63.9	29.89	18.52	49.76
Cumulative	34.17	20.18	53.02	34.4	19.47	47.21	29.78	18	53.3	28.82	17.67	54.12	26.97	16.18	42.57

`

Table 4-10 CNN-LSTM

Model Name: CNN-LSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	14.67	9.23	22.43	17.04	11.22	35.53	15.29	11.35	50.14	16.72	10.9	35.51	17.1	11.18	33.26
2	19.7	12.18	29.58	20.93	13.89	45.15	18.33	13.78	57.87	19.65	12.78	40.4	19	12.31	34.71
3	23.14	15.08	39.06	25.45	17.3	56.96	20.68	15.45	61.18	22.4	15.08	48.33	20.83	13.85	39.87
4	27.71	17.49	45.68	28.36	19.06	59.09	23.32	16.79	61.98	26.22	16.84	50.68	23.9	15.37	41.08
5	31.26	19.59	52.29	31.16	21	64.01	27.17	18.26	63.44	30.61	18.72	53.88	26.96	16.69	43.14
6	31.35	20.32	54.19	31.64	21.77	63.71	28.48	19.48	62.6	33.65	21.06	54.08	28.62	18.13	44.27
7	31.89	20.92	56.79	31.29	21.16	64.5	30.82	20.53	64.15	35.43	21.74	54.88	31.22	19.62	46.95
8	32.14	20.13	58.39	31.67	20.5	65.11	32.15	20.39	63.63	35.2	21.11	53.9	30.13	18.49	46.94
Cumulative	27.19	16.87	44.8	27.69	18.24	56.76	25.18	17	60.62	28.32	17.28	48.96	25.22	15.7	41.28

Table 4-11 ConvLSTM-BLSTM

Model Name: ConvLSTM-BLSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	17.6	10.15	26.78	16.9	10.37	31.54	16.19	9.52	22.52	16.86	10.23	28.59	21.11	13.05	41.67
2	25.43	14.69	36.2	23.98	14.25	40.07	23.13	13.3	30.1	23.05	13.7	37.44	25.86	15.04	38.44
3	30.18	17.62	43.58	28.42	16.95	47.61	27.65	15.63	35.25	26.9	16.18	46.02	28.29	16.64	41.25
4	33.75	20	49.74	31.22	18.88	53.04	30.35	17.15	39.75	29.52	17.93	53.88	29.29	17.53	43.76
5	36.34	21.8	54.26	33.47	20.55	57.69	31.34	17.97	42.44	30.75	18.89	56.85	29.83	18.11	45.47
6	38.31	23.32	58.5	35.18	21.53	60.61	31.72	18.47	44.23	31.39	19.29	55.93	30.37	18.57	46.67
7	39.78	24.4	62.02	35.07	21.41	59.71	31.77	18.7	45.24	31.16	19.26	54.39	31.14	19.17	47.9
8	40.53	25.01	65.12	33.1	20.16	55.02	32.2	19.05	46.82	31.33	19.4	55.06	32.25	19.61	48.48
Cumulative	33.58	19.63	49.52	30.26	18.01	50.66	28.54	16.22	38.29	28.05	16.86	48.52	28.71	17.21	44.21

Table 4-12 ConvLSTM-LSTM

Model Name: ConvLSTM-LSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	16.15	11.25	46.1	16.64	10.2	30.8	30.65	22.63	82.21	17.06	10.64	31.95	17.96	10.79	24.82
2	24.01	15.62	52.49	25.68	15.18	40.99	31.62	20.26	51.45	22.1	13.55	38.24	22.84	13.52	30.39
3	29.21	18.68	58.27	31.33	18.51	48.04	32.08	20.15	44.05	24.99	15.38	43.37	25.57	15.16	34.39
4	32.58	20.71	62.07	35.07	20.6	52.74	32.06	20.38	46.57	26.74	16.58	47.57	27.05	16.12	37.46
5	34.57	21.96	64.31	36.96	21.85	55.46	32.71	21	48.36	28.27	17.41	50.27	28.23	16.94	39.86
6	35.78	22.93	66.61	37.24	22.14	56.54	33.16	20.74	50.27	28.96	17.99	52	29.26	17.58	41.68
7	36.65	23.58	68.71	36.41	21.61	56.16	34.95	21.69	52.63	29.61	18.4	52.81	30.08	18.12	43.27
8	37.56	24.2	70.94	34.78	20.81	54.96	36.91	22.79	61.02	30.6	18.84	53.05	30.58	18.49	44.4
Cumulative	31.59	19.87	61.19	32.47	18.86	49.46	33.07	21.2	54.57	26.39	16.1	46.16	26.75	15.84	37.03

`

Table 4-13 Stacked LSTM

Model Name: Stacked LSTM															
Input Sequence	1by8			2by8			3by8			6by8			9by8		
Hour	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	16.06	9.41	23.18	16.1	9.13	20.5	15.41	9.7	29.24	16.45	9.97	22.24	18.05	10.6	25.37
2	24.65	14.36	34.48	26.24	14.51	33.96	23	14.22	43.96	22.41	13.15	29.01	23.27	13.42	30.59
3	30.54	17.95	43.9	33.36	18.32	43.2	27.59	16.44	47.87	25.56	15.18	36.38	26.38	15.36	34.99
4	34.56	20.5	50.88	37.77	20.87	50.12	30.08	17.77	50.37	27.96	16.65	41.09	28.42	16.73	38.61
5	37.08	22.13	55.23	40.11	22.21	53.75	31.73	18.73	51.81	29.57	17.65	44.14	29.6	17.71	41.6
6	38.65	23.51	58.89	41.39	22.91	56.26	32.48	19.31	52.73	30.61	18.29	46.45	30.24	18.3	44.04
7	39.52	24.41	61.63	40.58	22.95	57.25	32.43	19.41	53.2	30.96	18.65	48.36	30.81	18.81	46.11
8	40.26	24.99	63.83	39.58	22.53	57.39	32.7	19.6	53.92	31.11	18.85	49.88	31.45	19.21	47.5
Cumulative	33.62	19.66	49	35.39	19.18	46.55	28.76	16.9	47.89	27.26	16.05	39.7	27.61	16.27	38.6

In comparison to other models, BConvLSTM-LSTM and BConvLSTM-BLSTM combination model performs better (Figure 4-7).

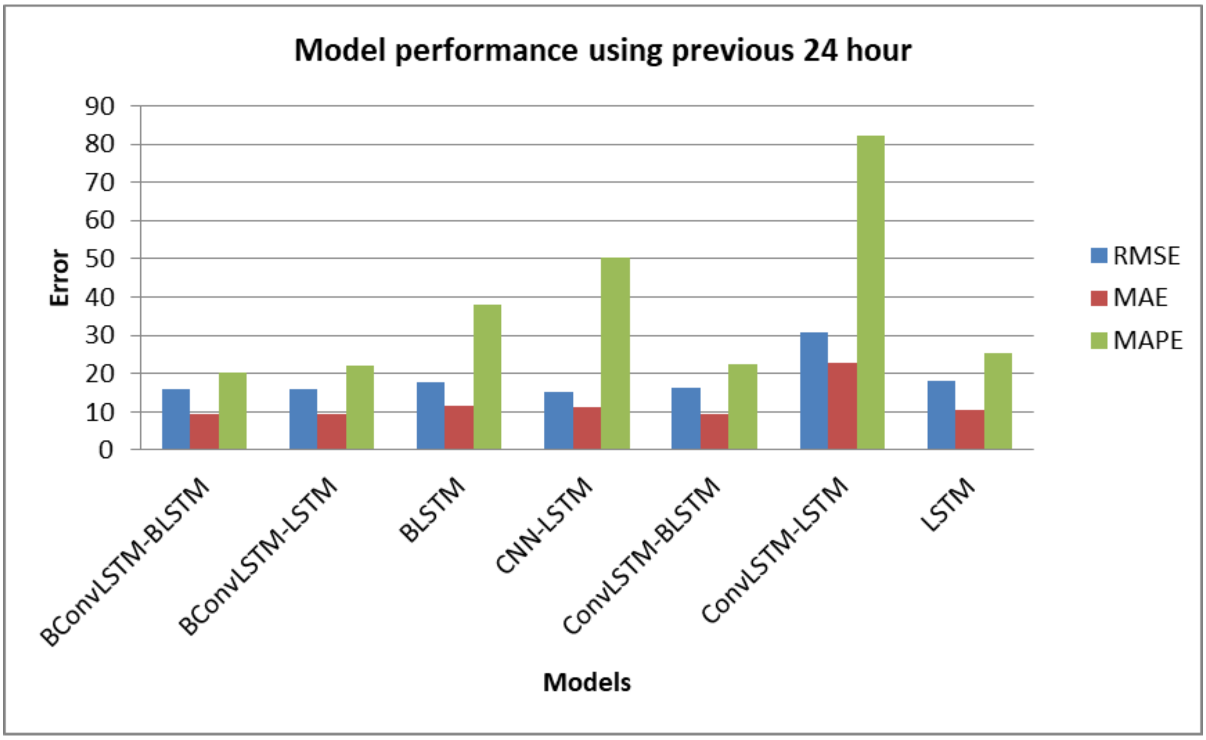


Figure 4-7 Model Comparison

`

4.1.4 Proposed Model and performance evaluation (Objective 2):

Statistical error functions RMSE, MAE, and MAPE were utilized to evaluate the model performance. Other error functions FAC2 and VG are also applied for further evaluation of the model performance. When mean differences are greater between data sets, scale dependent error functions such as RMSE and MAE are not the best choice to use for comparison. MAPE is a unit less function that is scale-independent and ideal for model comparison. In MAPE data with significant variances can influence the extreme values. Detail result of the model is presented in Table 4-14. The RMSE performances of the model varies from the lowest performance of 7.09 in Shillong, in the EH region, to the highest performance of 53.81 in Patna, in the MGP in 8 hour consecutive forecast. Likewise, Shillong and Patna yielded a minimum of 5.41 and a maximum of 34.09 in terms of MAE, respectively. In case of MAPE, Hyderabad recorded the lowest value of 18.6 percent and Chennai recorded the highest value of 52.7 percent. Nine cities had less than 10 µg/m3 of RMSE value in first step prediction, while, Howrah in LGP, Shillong in EH, and Nagpur and Mandideep in CPH had value between 10 and 15 µg/m3. As for MAE, the values of less than 10 µg/m3 are found in 15 of the 26 cities, and the total of 24 cities has MAE values that are less than or equal to 15 µg/m3. Patna and Talcher are the only two places with MAE values higher than 15 g/m3., Patna in MGP has MAE of 20.52 µg/m3 and Talcher in the EPH has an MAE value of 20.29 µg/m3. Similar to this, 20 sites have more than 35% MAPE values, and 11 sites have values lower than 30% MAPE values. All other sites have MAPE values under 40%, with the exception of Chennai in the ECPH and Jodhpur in the WD zone. Overall, it was found that the western, southern, and central regions had relatively better model performance than the northern and eastern regions for predicting concentrations up to 8 hours in advance. The results of $PM_{2.5}$ concentrations for one hour advance and eight hours average demonstrate the robustness of the suggested model framework's prediction ability. Since the majority of reports on regulatory air quality occurred within this temporal range, it is crucial to evaluate the performance of prior 1-hour and 8-hour average prediction. RMSE, MAE, and MAPE values for 1-hour ahead prediction were found to be lowest at 5.81 (at Shillong), 3.92 (at Aurangabad), and 10.8 (at Howrah), respectively. At Patna and Talcher, the maximum RMSE was 41.384 and 29.04, respectively, for 1-hour and 8-hour average periods. For the 1-hour ahead and 8-hour

`

average forecasting horizons, 18 sites and 13 sites, respectively, recorded less than 20 RMSE values overall. Once more, the 1-hour and 8-hour average MAE values were found to be at their lowest and maximum in Shillong and Patna, respectively. MAE values of three sites - Delhi, Talcher, and Patna found to be greater than 20 µg/m3, suggesting that the results were more consistent throughout India.

Model results were further analyzed in terms of various agroclimatic zones (Table 4-15, Figure 4-12). Regarding one-hour ahead prediction across agroclimatic zones, SPH (7.6) had the best model performance in terms of RMSE, followed by CPH (8.4) and WPH (9.4). Three zones have RMSE values between 20 and 30 and another three have between 30 and 40 µg/m3,respectively, and seven zones have RMSE values less than 20 µg/m3. A comparable error trend was seen for MAE, with the maximum value (16.3) at MGP and the lowest value (5.4) at SPH. Following 8 hours of advance prediction, the same pattern was found at SPH as minimum RMSE and MAE values of 11.1 and 8.0, respectively, and at MGP maximum RMSE (40.4) and MAE (25.8) were found. The results for MAPE values, however, showed a slight variation, with the minimum error value for 1-hour ahead prediction being observed at LGP (10.8%) and cumulative 8-consecutive-hour ahead prediction at EH (21.3 %).

Spatial variability in model performance is evident in the observed forecasting results. The Heatmap of the result (Figure 4-11) makes this clear. The regions primarily along the central, south, southeast, and southwest coasts of India exhibit superior model result for multi-step hourly forecasts. A relatively low model performance was noted in most of North India, with the exception of Shillong, Agra, Jamshedpur, and Amritsar. It is important to note that, with the exception of Patna and Talcher, model performance is best in India for the first step. Although many researchers attempted to use machine learning with deep learning architecture for a few of India's pollution hotspots, cross-country study of model efficiency has not yet been thoroughly studied in India. We then performed a statistical distribution analysis on the results at different sites with the corresponding observed values. Each location in India revealed that the best fit distribution matched the original training data exactly.

Table 4-14: Result of 8 step ahead prediction

Agro Climatic Zone: 1 Eastern Himalayan Region									
Station Name: Shillong									
ERROR	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour	7 hour	8 hour	Cumulative
RMSE	5.81	6.38	6.87	7.14	7.32	7.48	7.68	7.82	7.09
MAE	4.42	4.83	5.17	5.46	5.57	5.73	5.97	6.14	5.41
MAPE	20.97	22.96	23.69	24.57	25.1	25.75	26.57	27.18	24.6
FAC2	1.05	1.04	1.03	1.03	1.03	1.03	1.03	1.03	1.03
VG	1.07	1.08	1.09	1.09	1.1	1.1	1.11	1.11	1.09
Station Name: Guwahati									
RMSE	20.7	28.99	32.73	34.07	34.41	35.23	35.58	36.33	32.62
MAE	12.43	17.28	20.12	21.49	21.9	22.39	22.69	23.64	20.24
MAPE	21.34	29.08	34.79	38.83	40.89	42.38	43.6	46.5	37.18
FAC2	1.06	1.1	1.14	1.18	1.2	1.21	1.22	1.25	1.17
VG	1.08	1.15	1.2	1.23	1.25	1.27	1.28	1.3	1.22
Station Name: Siliguri									
RMSE	12.21	17.79	22.32	25.68	28.18	29.92	30.94	31.69	25.68
MAE	8.03	11.68	14.78	17.11	18.87	20.07	20.77	21.37	16.58
MAPE	11.43	16.4	20.52	23.86	26.42	28.25	29.75	31.43	23.51
FAC2	1	1.02	1.04	1.05	1.06	1.07	1.08	1.1	1.06
VG	1.02	1.05	1.07	1.09	1.11	1.12	1.14	1.15	1.09
Agro Climatic Zone: 2 Lower Gangetic Plain Region									
Station Name: Howrah									
RMSE	9.57	15.4	20.18	23.88	26.75	29.17	31.28	32.96	24.84
MAE	6.37	9.82	12.89	15.37	17.29	18.89	20.32	21.56	15.31
MAPE	10.8	15.16	19.25	22.68	25.6	28.15	30.6	33.01	23.15
FAC2	1.01	1.03	1.05	1.07	1.09	1.1	1.12	1.13	1.08
VG	1.02	1.04	1.06	1.08	1.1	1.12	1.14	1.16	1.09
Agro Climatic Zone: 3 Middle Gangetic Plain Region									
Station Name: Gaya									

`

RMSE	23.62	28.06	30.3	31.29	31.96	32.48	32.99	34.31	30.79
MAE	15.2	18.65	20.53	21.42	21.93	22.28	22.55	23.21	20.72
MAPE	27.86	34.45	38.4	40.47	41.41	42.1	43.13	44.94	39.1
FAC2	1.15	1.21	1.24	1.26	1.26	1.27	1.28	1.3	1.25
VG	1.12	1.16	1.19	1.21	1.21	1.22	1.23	1.24	1.2
Station Name: Patna									
RMSE	41.38	47.59	51.88	54.37	56.17	57.89	58.86	59.77	53.81
MAE	20.52	27.58	32.36	35.37	37.26	38.92	39.92	40.81	34.09
MAPE	14.57	20.56	25.11	28.15	30.08	31.51	32.39	33.05	26.93
FAC2	1.04	1.07	1.09	1.11	1.12	1.13	1.13	1.12	1.1
VG	1.04	1.07	1.1	1.13	1.14	1.15	1.16	1.17	1.12
Station Name: Varanasi									
RMSE	25.78	30.97	34.11	36.42	38.8	39.96	41.21	42.46	36.6
MAE	13.27	18.19	21.26	23.36	25.04	26	26.79	27.65	22.7
MAPE	18.97	25.54	30.47	34.37	37.51	39.66	41.5	43.9	33.99
FAC2	1.06	1.08	1.11	1.14	1.17	1.19	1.2	1.23	1.15
VG	1.07	1.11	1.15	1.18	1.2	1.22	1.23	1.25	1.18
Agro Climatic Zone: 4 Upper Gangetic Plains Region									
Station Name: Agra									
RMSE	15.61	21.68	24.96	26.7	28	28.62	28.95	29.02	25.82
MAE	9.19	13.23	15.59	16.95	17.71	18.17	18.51	18.82	16.02
MAPE	16.47	23.15	27.67	30.79	32.69	33.88	34.72	35.58	29.37
FAC2	1.05	1.08	1.11	1.14	1.16	1.17	1.18	1.19	1.13
VG	1.05	1.09	1.12	1.14	1.15	1.16	1.17	1.17	1.13
Station Name: Kanpur									
RMSE	24.95	33.07	37.35	39.82	41.27	42.29	43.25	45.19	38.89
MAE	14.16	19.92	23.36	25.4	26.59	27.45	28.29	29.54	24.34
MAPE	21.02	28.31	33.8	37.85	40.44	42.14	43.65	45.99	36.65
FAC2	1.1	1.15	1.19	1.23	1.25	1.26	1.27	1.29	1.22
VG	1.08	1.12	1.16	1.19	1.21	1.22	1.24	1.25	1.18
Station Name: Lucknow									

RMSE	18.73	25.74	29.37	32.13	34.05	35.04	35.89	36.89	31.51
MAE	11.17	15.67	18.46	20.38	21.65	22.44	23.24	24.04	19.63
MAPE	20.38	26.5	31.43	34.7	37.09	38.62	40.29	42.96	34
FAC2	1.05	1.08	1.12	1.14	1.16	1.18	1.19	1.22	1.14
VG	1.08	1.12	1.15	1.18	1.2	1.21	1.23	1.25	1.18
Agro Climatic Zone: 5 Trans-Ganga Plains Region									
Station Name: Amritsar									
RMSE	10.04	14.84	18.32	20.51	21.64	22.24	22.85	23.52	19.73
MAE	6.72	10.14	12.6	14.18	14.92	15.25	15.55	16.01	13.17
MAPE	16.4	24.19	30.26	34.39	36.35	37.19	38.1	40.31	32.15
FAC2	1.05	1.08	1.12	1.15	1.16	1.16	1.17	1.2	1.14
VG	1.05	1.09	1.14	1.17	1.18	1.19	1.2	1.22	1.15
Station Name: Delhi									
RMSE	18.73	28.27	34.62	38.89	41.94	43.95	45.36	46.86	38.41
MAE	11.8	18.24	22.71	25.69	27.87	29.39	30.54	31.66	24.74
MAPE	15.37	23.07	29.3	33.71	36.69	38.66	40.21	42.33	32.42
FAC2	1.05	1.09	1.14	1.17	1.19	1.2	1.21	1.23	1.16
VG	1.04	1.08	1.12	1.15	1.18	1.19	1.21	1.22	1.15
Agro Climatic Zone: 6 Eastern Plateau and Hills									
Station Name: Jharkhand									
RMSE	11.68	15.76	18.09	19.46	20.16	20.59	20.82	21.3	18.74
MAE	7.69	11.06	13.01	14.04	14.79	15.35	15.64	16.1	13.46
MAPE	17.01	24.19	28.35	31.18	33.36	34.86	36.04	37.99	30.37
FAC2	1.06	1.09	1.12	1.14	1.16	1.17	1.18	1.21	1.14
VG	1.06	1.1	1.12	1.14	1.16	1.17	1.18	1.19	1.14
Station Name: Talcher									
RMSE	29.04	38.41	41.76	42.9	42.58	42.2	41.9	42.58	40.41
MAE	20.29	27.47	30.35	31.16	30.85	30.91	30.84	31.36	29.15
MAPE	25.07	36.09	40.65	41.94	42.14	42.23	41.87	42.88	39.11
FAC2	1.08	1.14	1.16	1.16	1.16	1.16	1.15	1.15	1.15
VG	1.1	1.2	1.24	1.26	1.26	1.26	1.26	1.27	1.23

`

Agro Climatic Zone: 7 Central Plateau and Hills									
Station Name: Mandideep									
RMSE	8.81	11.6	12.86	13.57	14.09	14.39	14.63	14.9	13.25
MAE	5.86	8.03	9.11	9.76	10.19	10.44	10.58	10.8	9.35
MAPE	19.31	23.54	26.38	28.52	29.91	30.8	31.34	32.78	27.82
FAC2	1.05	1.08	1.11	1.14	1.15	1.16	1.16	1.16	1.13
VG	1.04	1.08	1.11	1.12	1.13	1.14	1.14	1.15	1.11
Station Name: Nagpur									
RMSE	7.9	9.83	11.34	12.32	12.87	13.17	13.36	13.62	11.95
MAE	5.76	7.24	8.18	8.82	9.2	9.38	9.48	9.69	8.47
MAPE	19.31	23.54	26.38	28.52	29.91	30.8	31.34	32.78	27.82
FAC2	1.07	1.08	1.1	1.12	1.13	1.14	1.15	1.16	1.12
VG	1.06	1.08	1.1	1.12	1.13	1.13	1.14	1.14	1.11
Agro Climatic Zone: 8 Western Plateau and Hills									
Station Name: Aurangabad									
RMSE	5.93	8.53	9.76	10.32	10.67	10.92	11.06	11.25	9.95
MAE	3.92	5.67	6.65	7.18	7.53	7.78	7.9	8.09	6.84
MAPE	11.49	16.04	19.36	21.74	23.38	24.44	25.13	26.25	20.98
FAC2	1.04	1.06	1.09	1.11	1.12	1.13	1.14	1.15	1.11
VG	1.02	1.04	1.06	1.07	1.08	1.08	1.08	1.09	1.07
Station Name: Mumbai									
RMSE	13.47	17.69	19.77	20.4	19.42	18.22	17.45	17.29	18.08
MAE	10.19	13.59	15.25	15.85	15.35	14.54	13.91	13.64	14.04
MAPE	24.6	33.34	37.41	38.77	37.92	35.8	34.71	35.05	34.7
FAC2	1.07	1.1	1.11	1.12	1.12	1.12	1.12	1.13	1.11
VG	1.09	1.16	1.2	1.22	1.2	1.18	1.16	1.17	1.17
Station Name: Solapur									
RMSE	8.74	10.97	11.75	12.14	12.47	12.61	12.71	12.85	11.85
MAE	5.93	7.54	8.19	8.6	8.85	8.98	9.06	9.2	8.29
MAPE	18.14	22.48	24.63	25.88	26.54	26.9	27.26	28.36	25.02
FAC2	1.05	1.08	1.09	1.1	1.1	1.1	1.1	1.12	1.09

`

VG	1.06	1.08	1.09	1.1	1.11	1.11	1.11	1.12	1.1
Agro Climatic Zone: 9 Southern Plateau and Hills									
Station Name: Bengaluru									
RMSE	8.97	11.42	12.26	12.42	12.53	12.58	12.58	12.97	12.03
MAE	6.23	7.97	8.6	8.91	9.08	9.17	9.19	9.27	8.55
MAPE	20.4	26.85	29.67	31.26	32.31	32.9	33.27	33.75	30.05
FAC2	1.07	1.11	1.14	1.16	1.17	1.18	1.19	1.2	1.15
VG	1.07	1.1	1.12	1.13	1.14	1.14	1.14	1.15	1.12
Station Name: Hyderabad									
RMSE	6.25	8.7	9.86	10.56	11	11.23	11.37	11.49	10.2
MAE	4.6	6.43	7.31	7.89	8.25	8.46	8.61	8.79	7.54
MAPE	11.27	15.75	17.99	19.45	20.34	20.9	21.41	22.03	18.64
FAC2	1.02	1.03	1.04	1.05	1.05	1.06	1.06	1.07	1.05
VG	1.02	1.05	1.06	1.07	1.07	1.07	1.08	1.08	1.06
Agro Climatic Zone: 10 Eastern Coastal Plains and Hills									
Station Name: Chennai									
RMSE	23.05	27.73	29.8	30.49	30.63	30.7	31.1	31.12	29.44
MAE	12.96	16.15	17.59	18.17	18.42	18.63	18.99	19.26	17.52
MAPE	35.74	47.36	52.64	54.49	55.48	56.36	58.52	60.94	52.69
FAC2	1.21	1.3	1.35	1.37	1.38	1.39	1.41	1.44	1.36
VG	1.2	1.29	1.33	1.35	1.35	1.36	1.38	1.4	1.33
Station Name: Vijaywada									
RMSE	7.01	8.22	8.8	9.14	9.34	9.64	9.78	9.88	9.02
MAE	4.93	6.05	6.6	6.84	6.94	7.21	7.42	7.52	6.69
MAPE	21.07	25.4	27.23	28.28	28.83	30.16	31.19	31.94	28.01
FAC2	1.06	1.08	1.08	1.09	1.1	1.11	1.11	1.13	1.09
VG	1.07	1.1	1.11	1.12	1.13	1.14	1.14	1.14	1.12
Agro Climatic Zone: 11Western Coastal Plains and Ghats									
Station Name: Thiruvananthapuram									
RMSE	12.35	15.27	15.99	16.31	16.83	16.9	17.05	17.55	16.11
MAE	8.6	10.87	11.58	11.83	12.24	12.54	12.63	12.87	11.64

`

MAPE	27.21	33.6	35.27	35.15	36.8	37.79	39.51	41.22	35.82
FAC2	1.04	1.06	1.06	1.05	1.05	1.06	1.08	1.11	1.06
VG	1.14	1.19	1.2	1.21	1.22	1.23	1.25	1.26	1.21
Agro Climatic Zone: 12 Gujarat Plains and Hills									
Station Name: Ahmedabad									
RMSE	10.6	15.43	18.28	19.76	20.64	21.36	21.77	22.09	19.1
MAE	7.33	10.53	12.5	13.58	14.35	14.92	15.27	15.57	13
MAPE	18.65	26.02	31.16	34.68	37.21	38.9	39.95	41.46	33.5
FAC2	1.08	1.12	1.17	1.2	1.22	1.24	1.25	1.28	1.2
VG	1.05	1.1	1.13	1.16	1.17	1.19	1.19	1.2	1.15
Agro Climatic Zone: 13Western Dry Region									
Station Name: Jodhpur									
RMSE	21.6	28.61	32.08	33.94	35.68	36.9	37.59	38.03	33.47
MAE	14.87	20.19	22.97	24.38	25.51	26.6	27.23	27.61	23.67
MAPE	26.05	36.42	42.81	46.36	48.49	50.21	51.54	52.39	44.28
FAC2	1.11	1.19	1.24	1.28	1.29	1.3	1.31	1.32	1.26
VG	1.1	1.18	1.23	1.26	1.29	1.3	1.32	1.32	1.25

Line plot of Delhi city result for one step ahead prediction is shown in Figure 4-8:

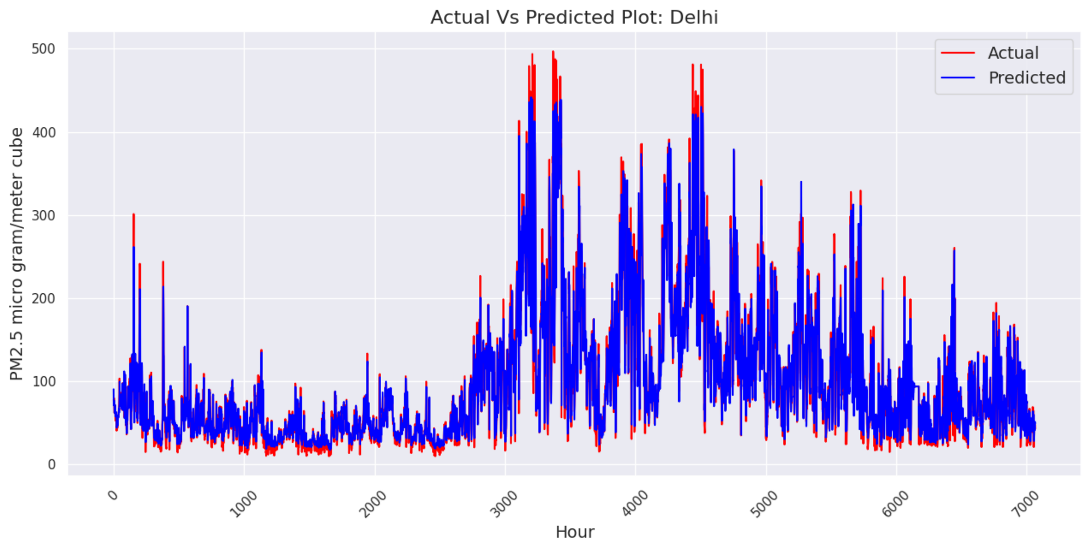


Figure 4-8: Actual Vs Predicted Plot for Delhi

`

A subplot of the above figure is shown on Figure 4-9 for better visualization of the graph.

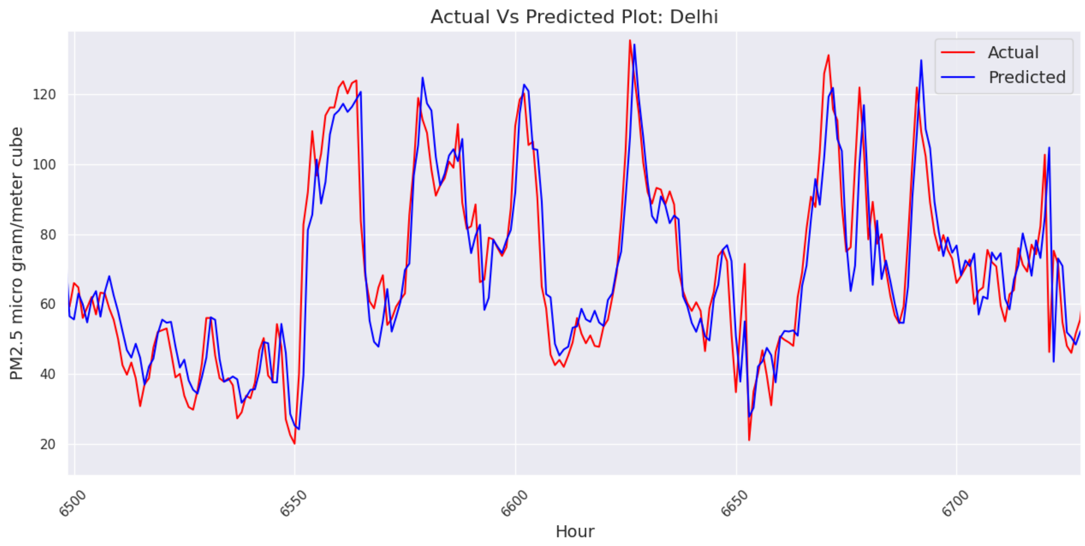


Figure 4-9: Sub plot of Figure 4-8

Figure 4-10 represents scatter plot of the Delhi result for one step ahead prediction:

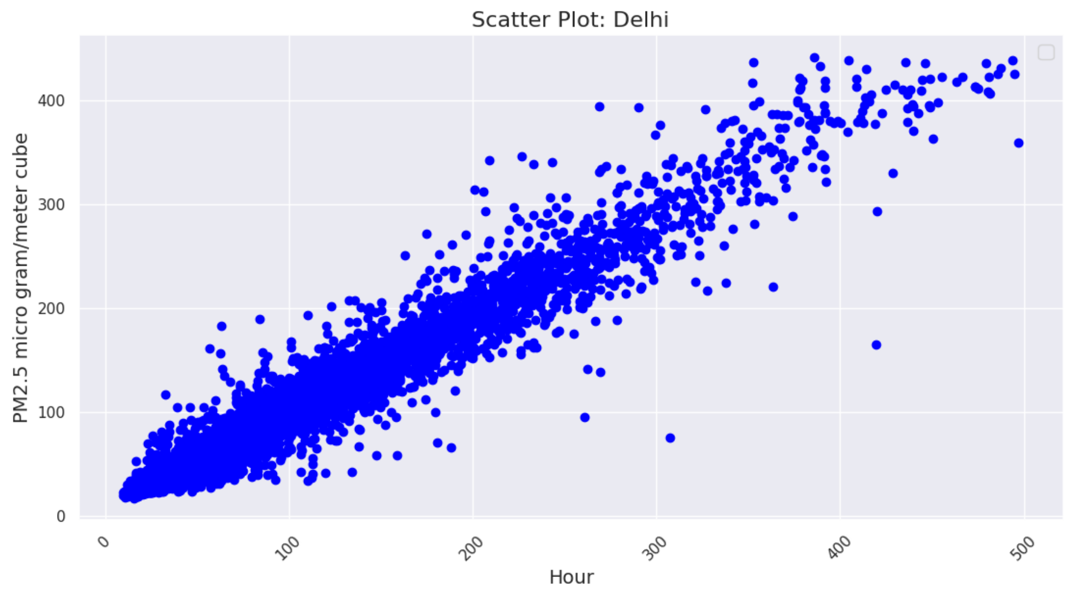


Figure 4-10: Scatter plot of Delhi

Heat Map of all the Stations for 8-step ahead prediction in terms of MAE is shown in Figure 4-11.

`

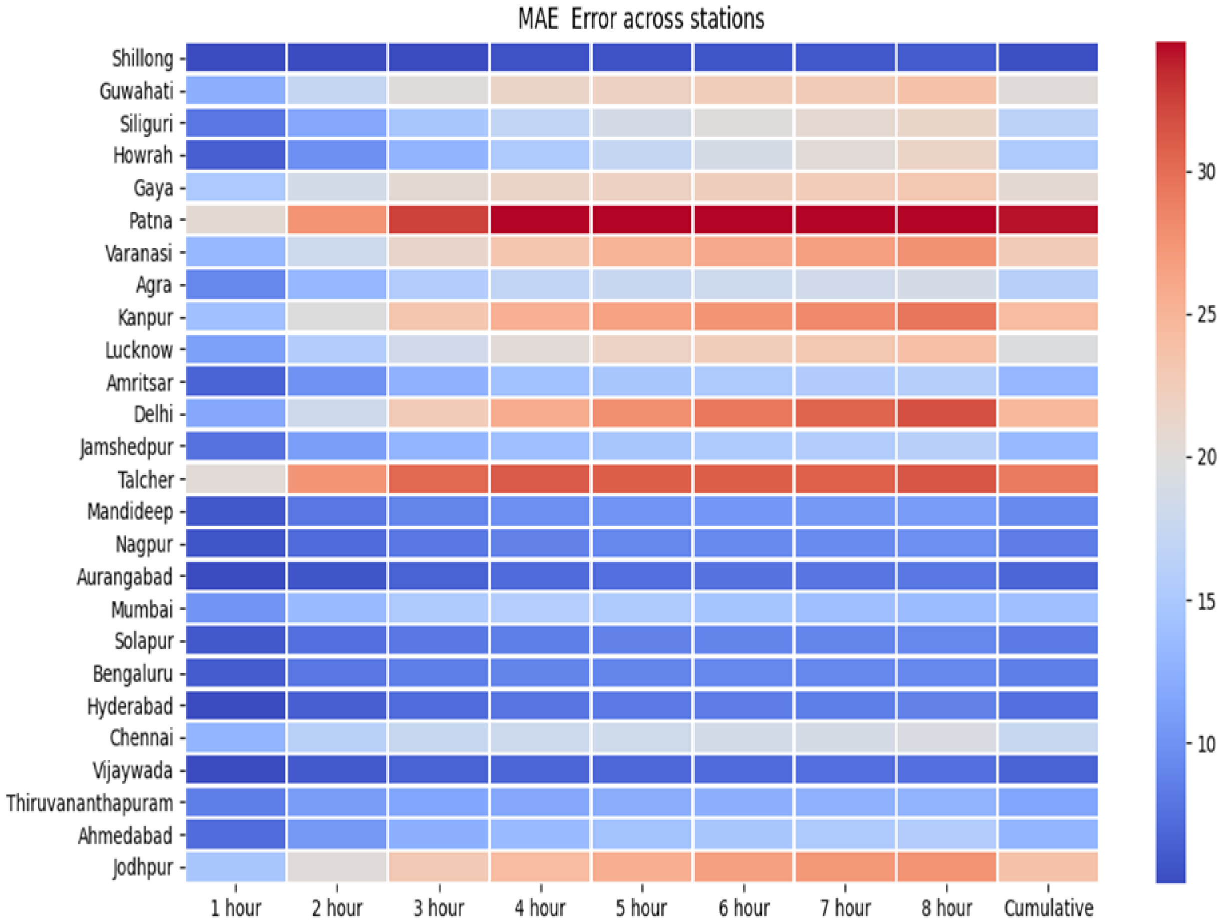


Figure 4-11: Heat Map of MAE Error

`

Table 4-15: Model performance agro-climatic zone wise

Zone	Error Metrics	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour	7 hour	8 hour	Cum
CPH	RMSE	8.4	10.7	12.1	12.9	13.5	13.8	14.0	14.3	12.6
	MAE	5.8	7.6	8.6	9.3	9.7	9.9	10.0	10.2	8.9
	MAPE	18.0	23.5	26.8	29.1	30.6	31.5	32.0	33.0	28.1
ECPH	RMSE	15.0	18.0	19.3	19.8	20.0	20.2	20.4	20.5	19.2
	MAE	8.9	11.1	12.1	12.5	12.7	12.9	13.2	13.4	12.1
	MAPE	28.4	36.4	39.9	41.4	42.2	43.3	44.9	46.4	40.4
EH	RMSE	12.9	17.7	20.6	22.3	23.3	24.2	24.7	25.3	21.8
	MAE	8.3	11.3	13.4	14.7	15.4	16.1	16.5	17.1	14.1
	MAPE	17.9	22.8	26.3	29.1	30.8	32.1	33.3	35.0	28.4
EPH	RMSE	20.4	27.1	29.9	31.2	31.4	31.4	31.4	31.9	29.6
	MAE	14.0	19.3	21.7	22.6	22.8	23.1	23.2	23.7	21.3
	MAPE	21.0	30.1	34.5	36.6	37.8	38.5	39.0	40.4	34.7
GPH	RMSE	10.6	15.4	18.3	19.8	20.6	21.4	21.8	22.1	19.1
	MAE	7.3	10.5	12.5	13.6	14.4	14.9	15.3	15.6	13.0
	MAPE	18.7	26.0	31.2	34.7	37.2	38.9	40.0	41.5	33.5
LGP	RMSE	9.6	15.4	20.2	23.9	26.8	29.2	31.3	33.0	24.8
	MAE	6.4	9.8	12.9	15.4	17.3	18.9	20.3	21.6	15.3
	MAPE	10.8	15.2	19.3	22.7	25.6	28.2	30.6	33.0	23.2
MGP	RMSE	30.3	35.5	38.8	40.7	42.3	43.4	44.4	45.5	40.4
	MAE	16.3	21.5	24.7	26.7	28.1	29.1	29.8	30.6	25.8
	MAPE	20.5	26.9	31.3	34.3	36.3	37.8	39.0	40.6	33.3
SPH	RMSE	7.6	10.1	11.1	11.5	11.8	11.9	12.0	12.2	11.1
	MAE	5.4	7.2	8.0	8.4	8.7	8.8	8.9	9.0	8.0
	MAPE	15.8	21.3	23.8	25.4	26.3	26.9	27.3	27.9	24.3
TGP	RMSE	14.4	21.6	26.5	29.7	31.8	33.1	34.1	35.2	29.1
	MAE	9.3	14.2	17.7	19.9	21.4	22.3	23.0	23.8	19.0
	MAPE	15.9	23.6	29.8	34.1	36.5	37.9	39.2	41.3	32.3
UGP	RMSE	19.8	26.8	30.6	32.9	34.4	35.3	36.0	37.0	32.1
	MAE	11.5	16.3	19.1	20.9	22.0	22.7	23.3	24.1	20.0
	MAPE	19.3	26.0	31.0	34.4	36.7	38.2	39.6	41.5	33.3
WCPG	RMSE	12.4	15.3	16.0	16.3	16.8	16.9	17.1	17.6	16.1
	MAE	8.6	10.9	11.6	11.8	12.2	12.5	12.6	12.9	11.6
	MAPE	27.2	33.6	35.3	35.2	36.8	37.8	39.5	41.2	35.8
WD	RMSE	21.6	28.6	32.1	33.9	35.7	36.9	37.6	38.0	33.5
	MAE	14.9	20.2	23.0	24.4	25.5	26.6	27.2	27.6	23.7
	MAPE	26.1	36.4	42.8	46.4	48.5	50.2	51.5	52.4	44.3
WPH	RMSE	9.4	12.4	13.8	14.3	14.2	13.9	13.7	13.8	13.3
	MAE	6.7	8.9	10.0	10.5	10.6	10.4	10.3	10.3	9.7
	MAPE	18.1	24.0	27.1	28.8	29.3	29.0	29.0	29.9	26.9

`

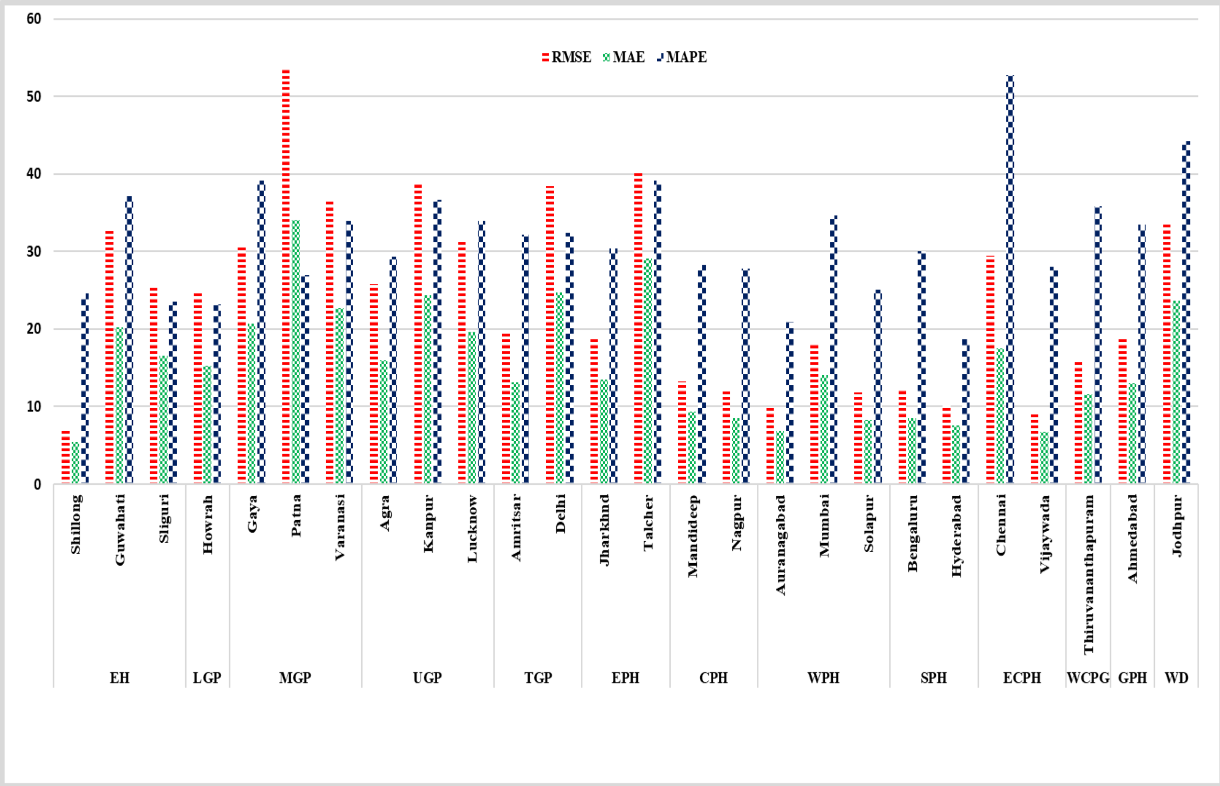


Figure 4-12: Model performance Agro climatic zone wise

4.1.5 Model Performance in terms of Data length and SNR values

Since the forecasting results vary widely among India's various agroecosystems, we have carried out an investigation to find the effect of noise present in the data on model performance. Moreover correlation analysis was done to find effect of data length on model performance. Signal to noise ratio (SNR) measures the percentage of accurate or desired information in each data series compared to inaccurate or undesired information. SNR values were used for this purpose (Table 4-16). The following equation was used to determine SNR values [135] for each pre-processed time series data.

$$SNR = \mu/\sigma$$

where μ and σ are the mean and standard deviation of the time series data respectively.

The data length and model performance error as represented by the trend line (Figure 4-13) and scatter plot (Figure 4-14), reveals no correlation between them, suggesting that the length of the data has little to no impact on the performance of the model. A scatter diagram (Figure 4-15) shows a sharply declining trend in MAE versus SNR for both 1-hour ahead and 8-hour cumulative forecasting. Clearly, for both forecasting horizons, there is a significant decline in the model error as the noise component decreases. It is to be noted that error variance dramatically decreases at SNR value greater than approximately 1.5 and the model's performance greatly increases. One possible explanation for the variation in results among Indian sites could be the degree of noise in the data series. Due to high traffic volumes and population densities, pollution sources differ greatly in northern India, which raises the relative variance in the data set. Fertilizer waste burning and large-scale farming are well-known practices in the Indo Gangetic Plain (IGP) area. With the exception of the monsoon season, when easterlies bring monsoon rains, it should be noted that westerlies are predominant in this area all year long. The dust load over the area is highly unpredictable due to dust storms caused by westerlies and agricultural burning. The uncertainties resulting from weather patterns could be the reason for the model's poor multistep ahead forecasting performance in this area.

Table 4-16: SNR Values

Sl.No.	Agro Climatic Zone	Station Name	SNR
1.	Eastern Himalayan Region	Shillong	1.633
		Guwahati	1.178
		Siliguri	1.203
2.	Lower Gangetic Plain Region	Howrah	1.140
3.	Middle Gangetic Plain Region	Gaya	1.217
		Patna	1.170
		Varanasi	1.229
4.	Upper Gangetic Plains Region	Agra	1.161
		Kanpur	1.134
		Lucknow	1.168
5.	Trans-Ganga Plains Region	Amritsar	1.480
		Delhi	1.320
6.	Eastern Plateau and Hills	Jamshedpur	1.835
		Talcher	1.259
7.	Central Plateau and Hills	Mandideep	1.416
		Nagpur	1.493
8.	Western Plateau and Hills	Aurangabad	1.735
		Mumbai	1.539

`

		Solapur	1.629
9.	Southern Plateau and Hills	Bengaluru	1.556
		Hyderabad	2.115
10.	Eastern Coastal Plains and Hills	Chennai	1.347
		Vijayawada	1.899
11.	Western Coastal Plains and Ghats	Thiruvananthapuram	1.588
12.	Gujarat Plains and Hills	Ahmedabad	1.542
13.	Western Dry Region	Jodhpur	1.688

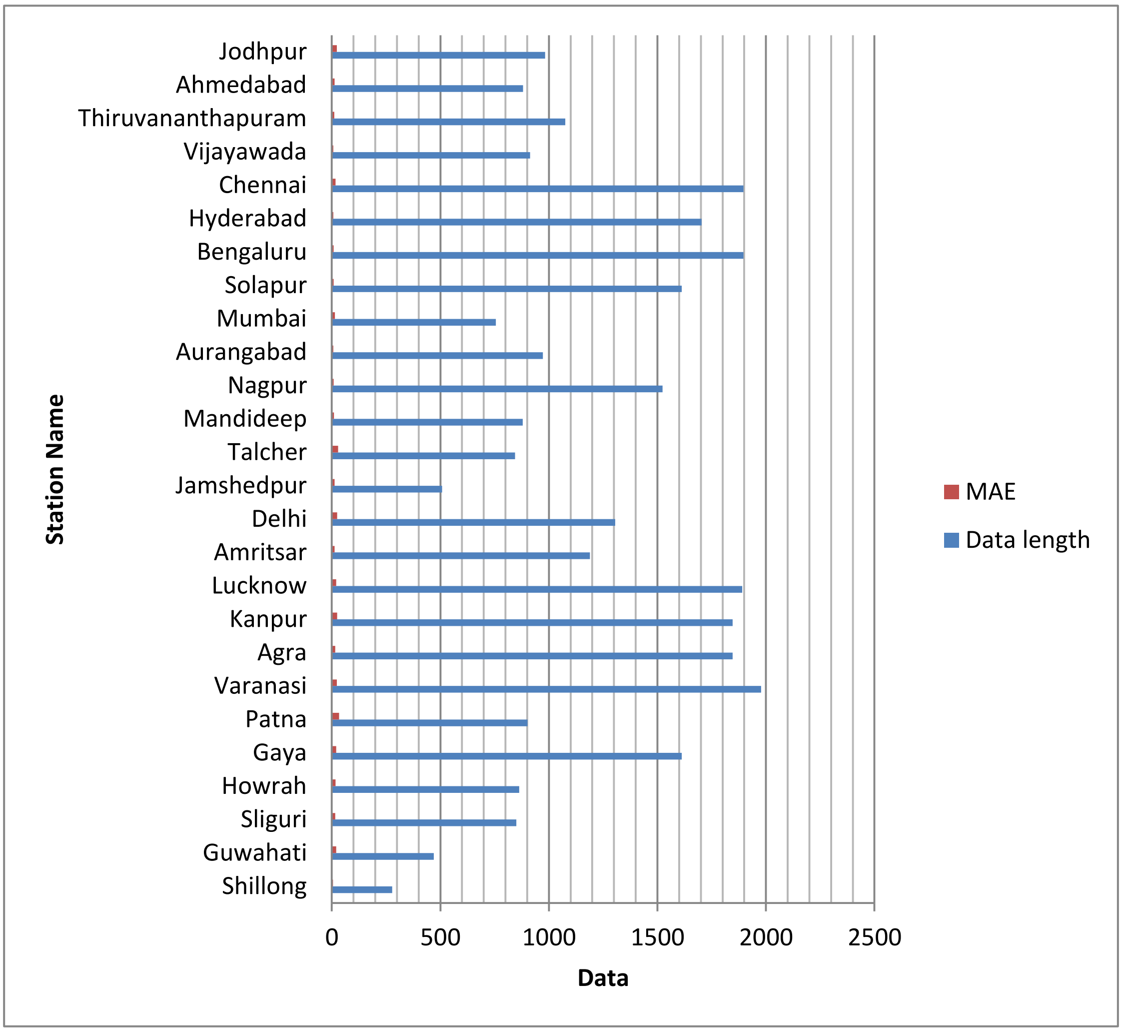


Figure 4-13: Data length Vs Error Plot

`

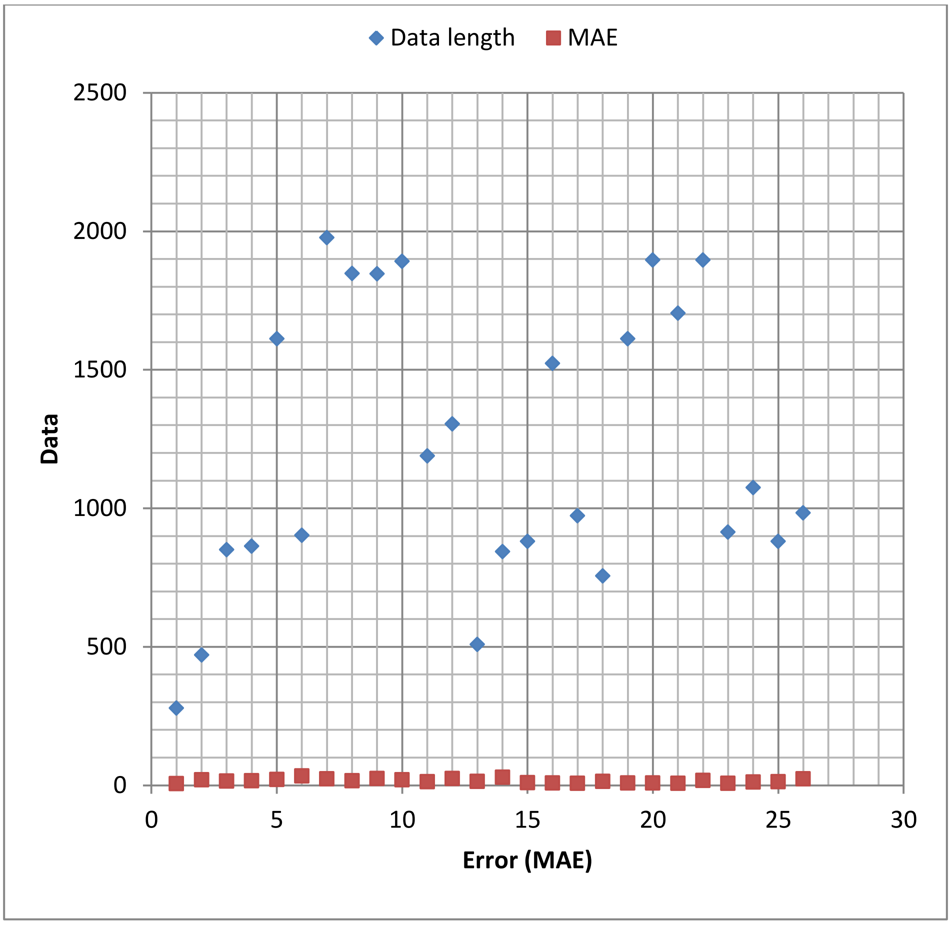


Figure 4-14: Correlation between Data length and Error

`

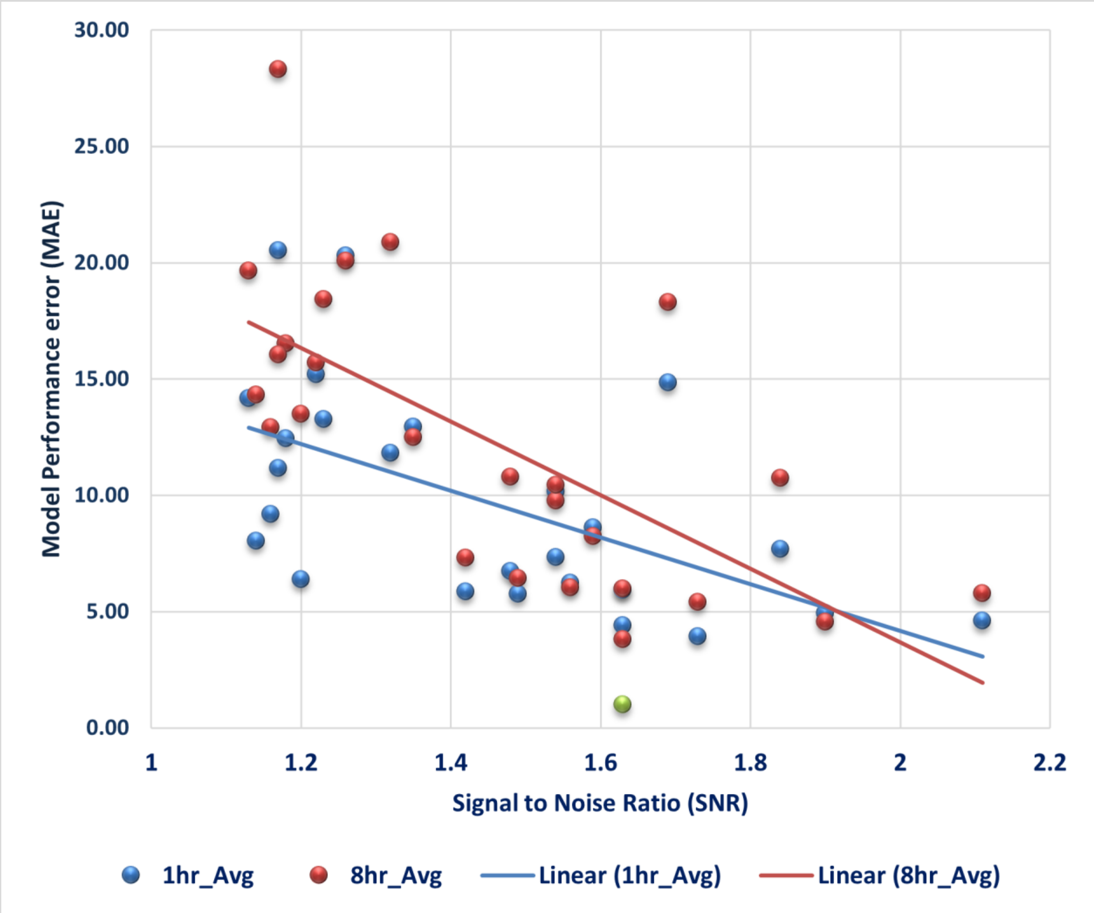


Figure 4-15: SNR Vs Model performance

4.1.6 Index of Agreement

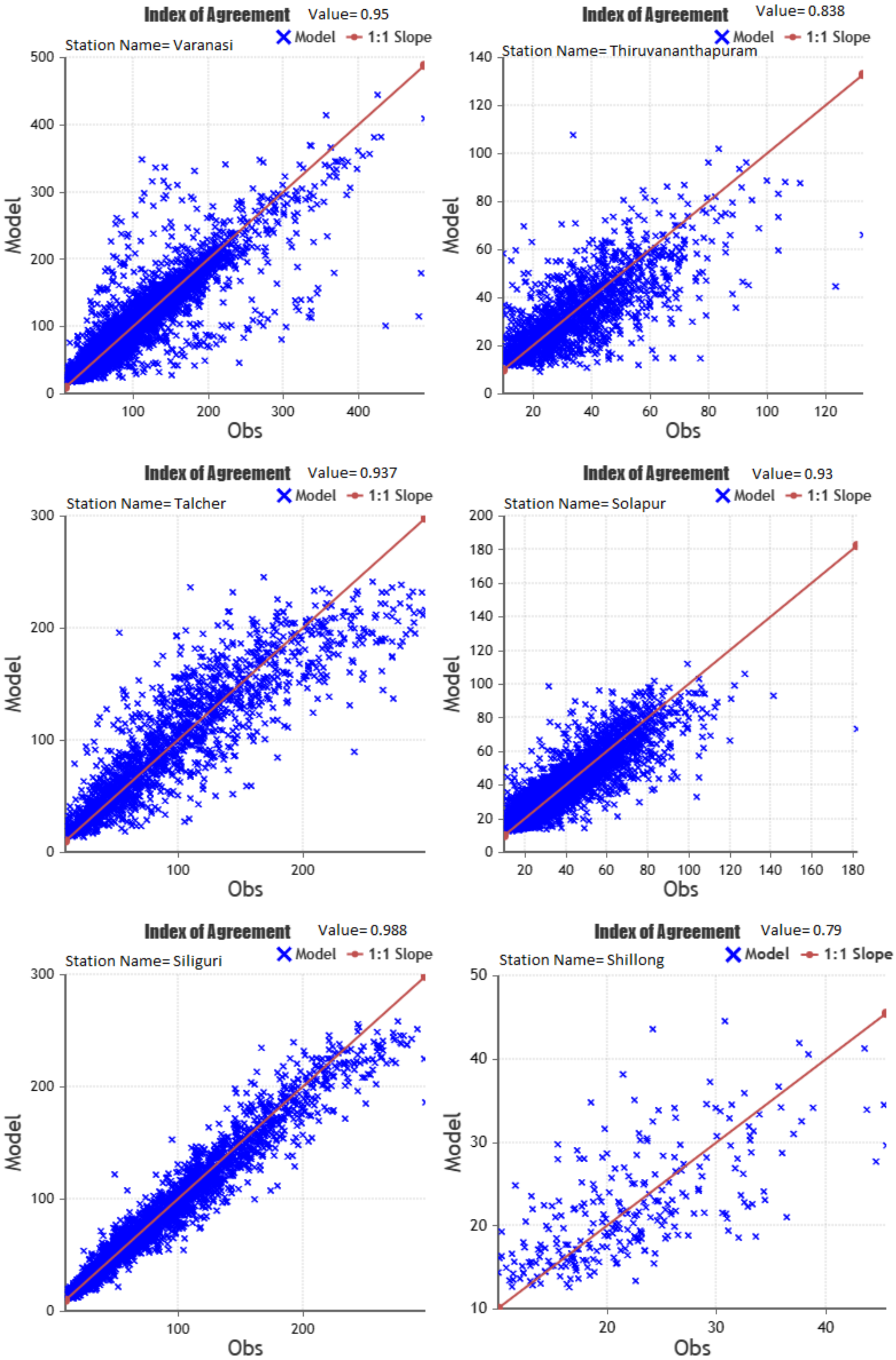
Index of agreement (d) was proposed by [136] as a standardized measure for degree of model prediction error, and its value ranges from 0 to 1. Higher index values suggest that the modelled values have better agreement with the observational values. Due to square differences in the mean square error in the numerator, the index of agreement is sensitive to extreme values even though it offers some improvement over the coefficient of determination.

$$d = 1 - \frac{\sum_{i=1}^{n}(Obs_i - Pred_i)^2}{\sum_{i=1}^{n}\left(|Pred_i - \overline{Obs}| + |Obs_i - \overline{Obs}|\right)^2} \quad , 0 \le d \le 1,$$

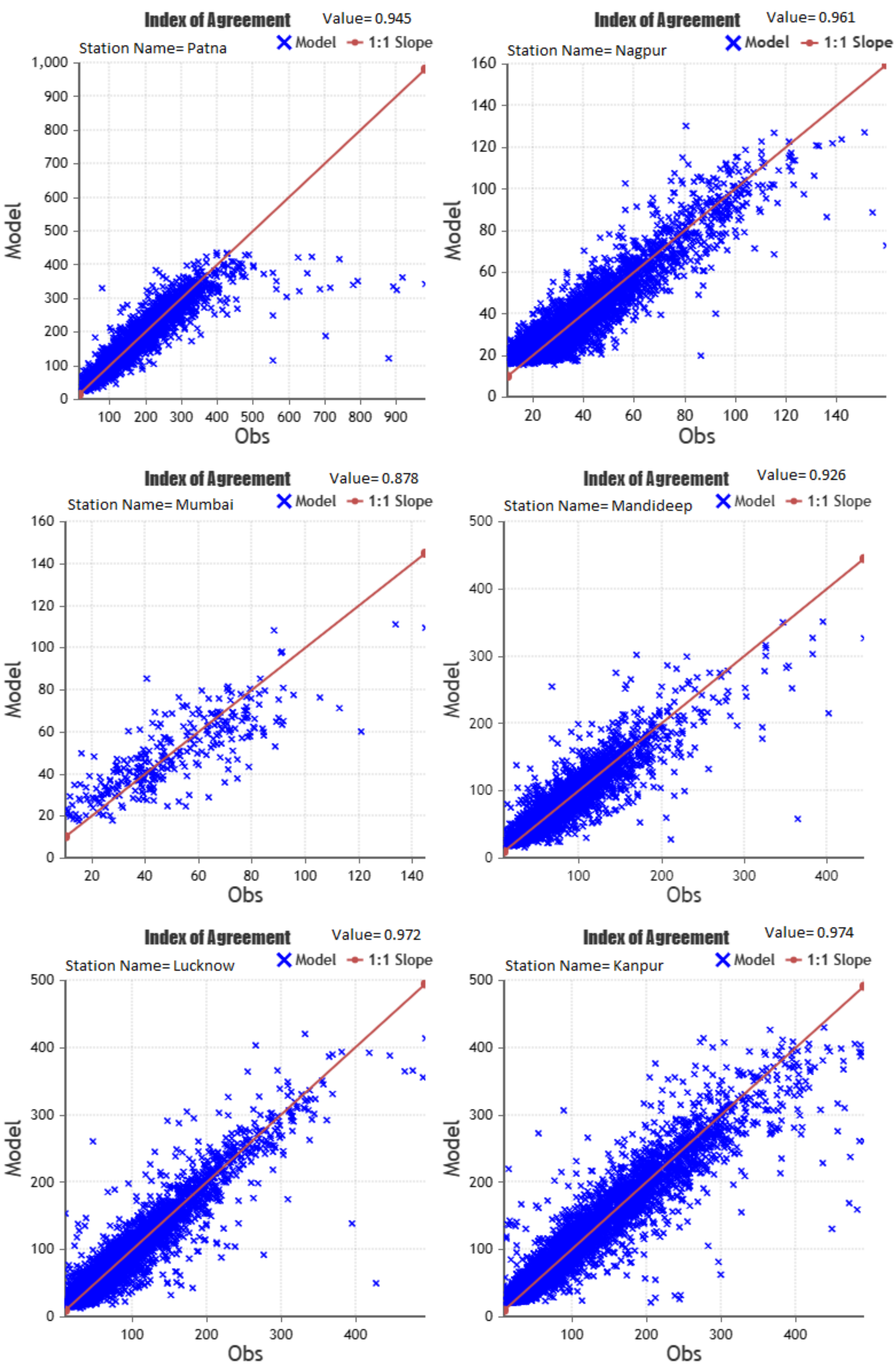
Where Obs_i is observed value, $Pred_i$ is the model predicted value

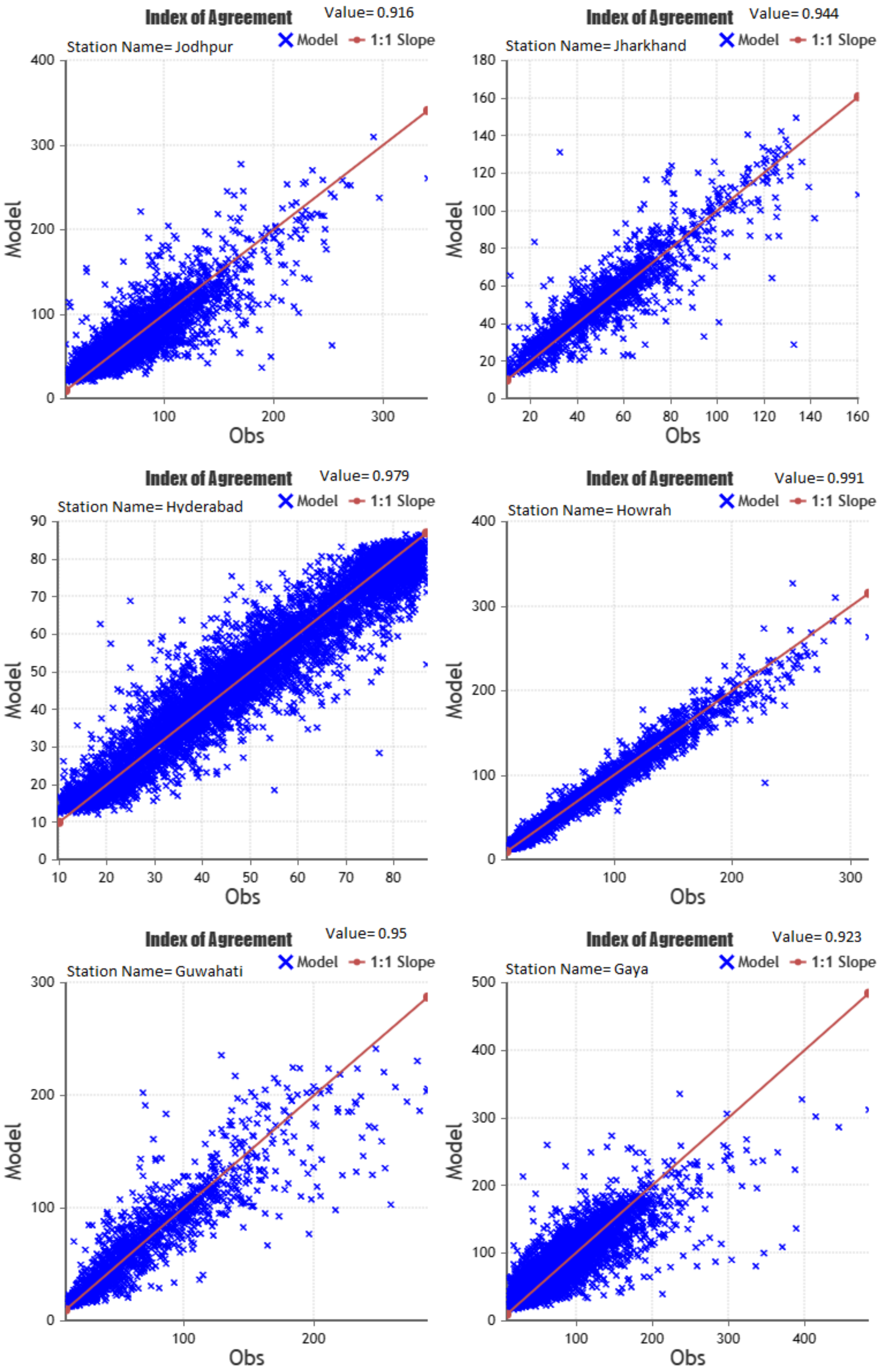
`

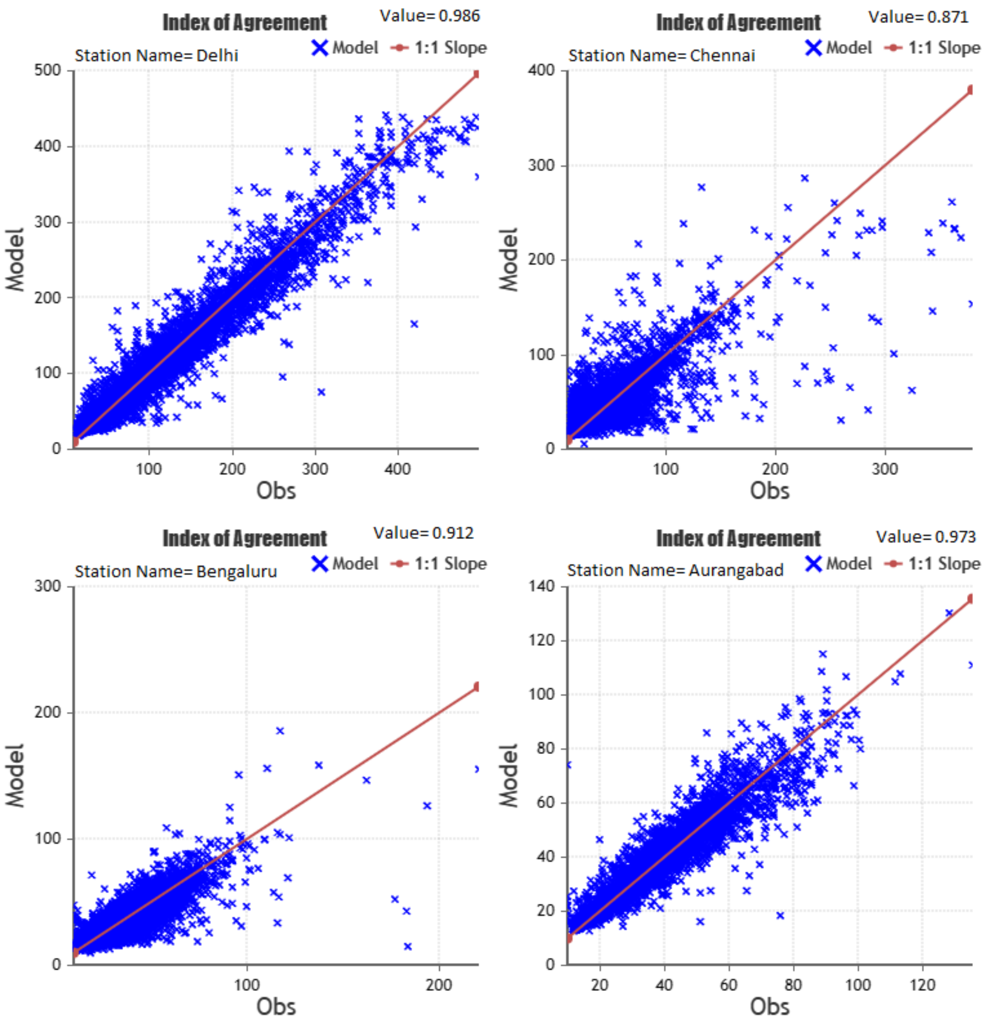
IoA values for all the stations have been calculated [137]. Most of the IoA values are close to 1 which shows better performance of the model. Details of the IoA result are represented in the Figures (Figure 4-16) below



`


Index of Agreement
Value= 0.945
Station Name= Patna
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.961
Station Name= Nagpur
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.878
Station Name= Mumbai
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.926
Station Name= Mandideep
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.972
Station Name= Lucknow
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.974
Station Name= Kanpur
Model
1:1 Slope
Model
Obs


Index of Agreement
Value= 0.916
Station Name= Jodhpur
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.944
Station Name= Jharkhand
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.979
Station Name= Hyderabad
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.991
Station Name= Howrah
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.95
Station Name= Guwahati
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.923
Station Name= Gaya
Model
1:1 Slope
Model
Obs


Index of Agreement
Value= 0.986
Station Name= Delhi
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.871
Station Name= Chennai
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.912
Station Name= Bengaluru
Model
1:1 Slope
Model
Obs
Index of Agreement
Value= 0.973
Station Name= Aurangabad
Model
1:1 Slope
Model
Obs

`

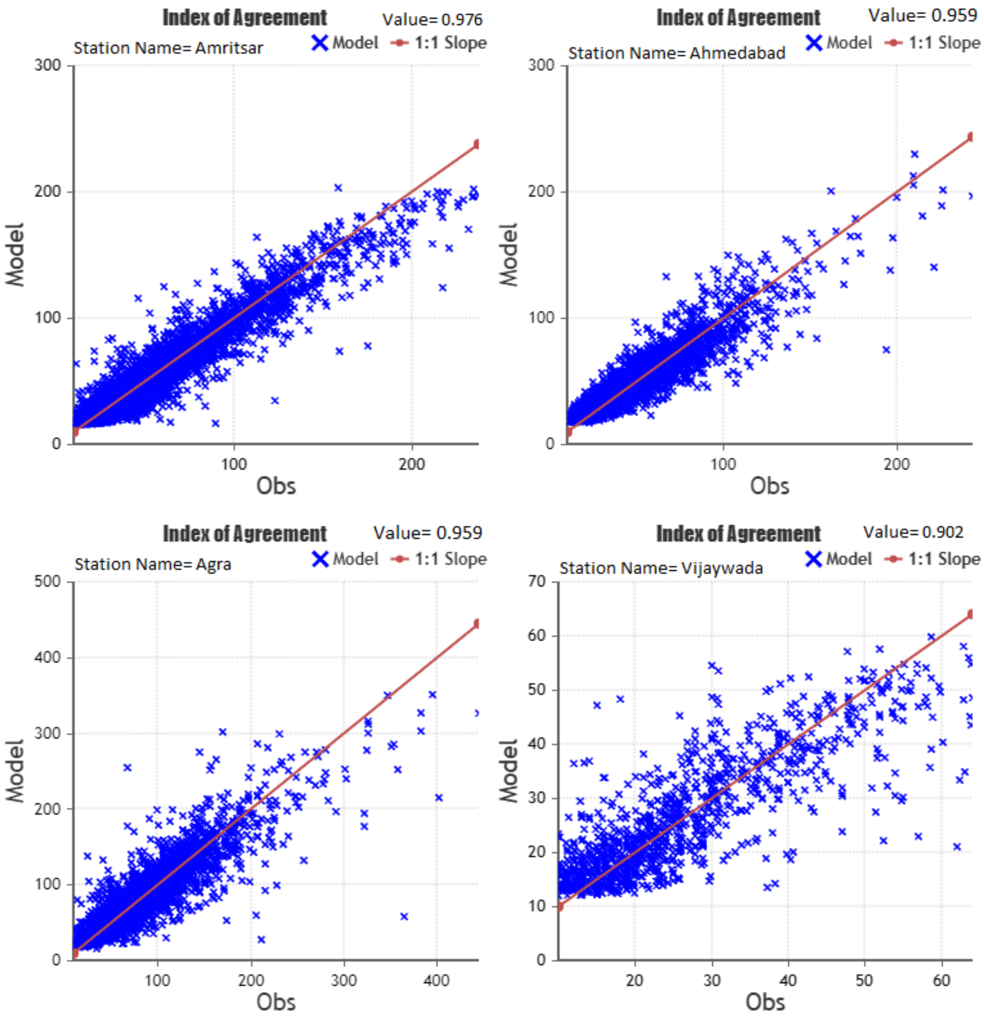


Figure 4-16: IoA for 26 Stations

4.1.7 Seasonal Wind effect :

In order to observe the wind circulation throughout Indian subcontinent Quiver plot was used. A quiver plot is a special point style where the direction of the symbol and the size of the symbol are derived from attributes. Seasonal analysis of wind circulation was done for four seasons – Pre-Monsoon (Figure 4-17), Monsoon (Figure 4-18), Post-Monsoon(Figure 4-19), and Winter(Figure 4-20). A detail of the seasonal wind circulation is presented in figures below.

`

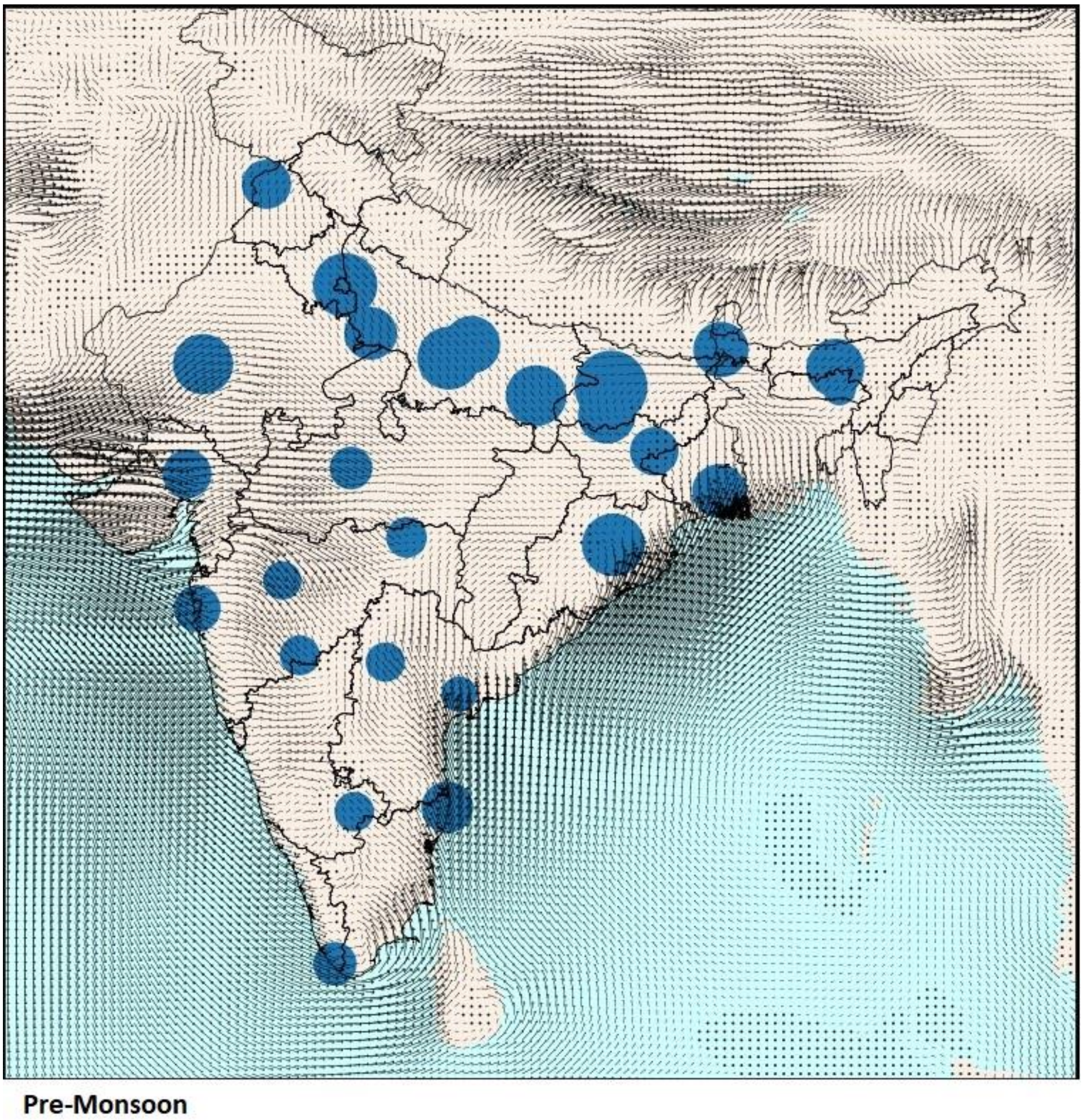


Figure 4-17: Quiver Plot for Pre- Monsoon

`

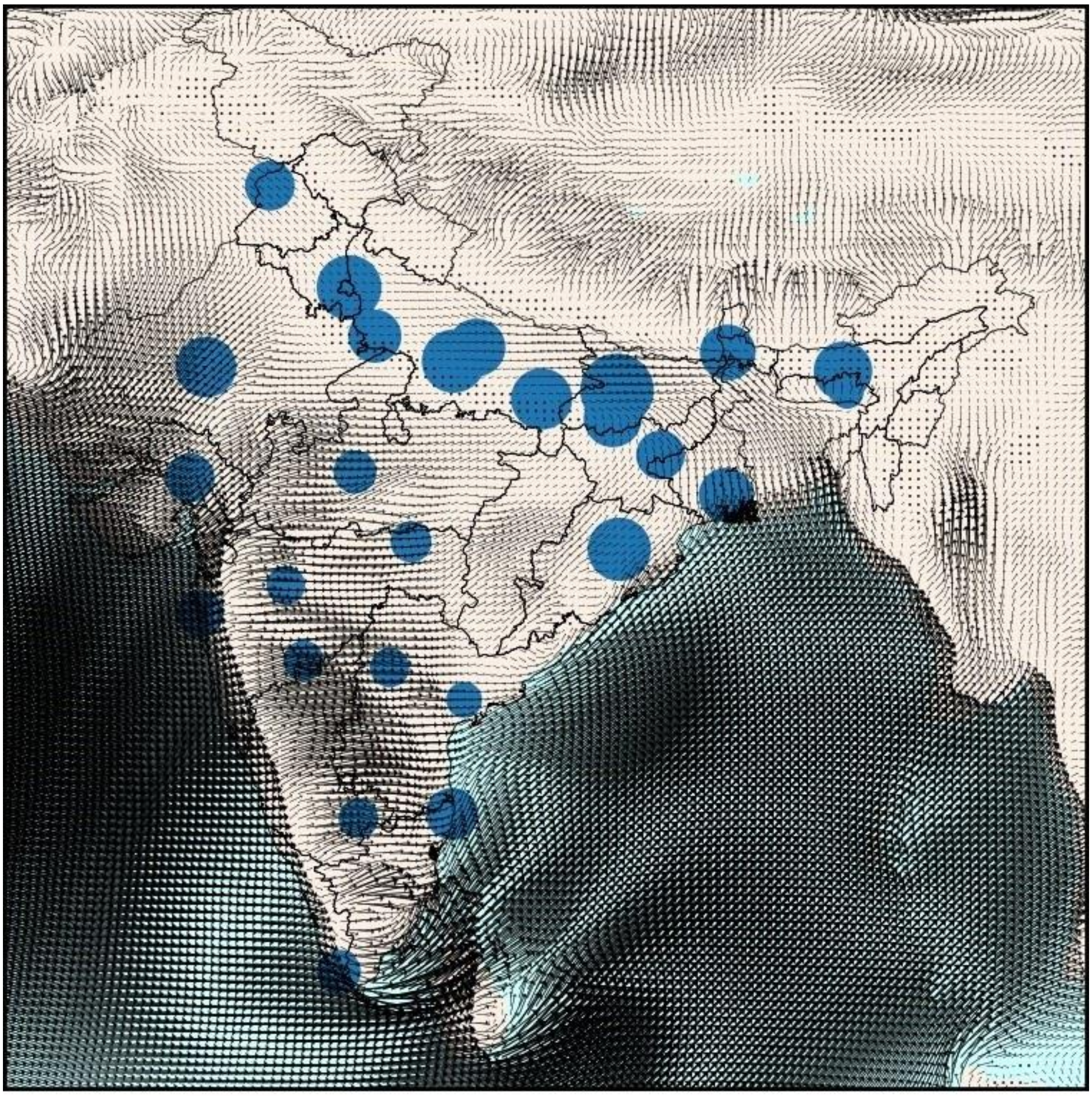


Figure 4-18: Quiver Plot for Monsoon

`

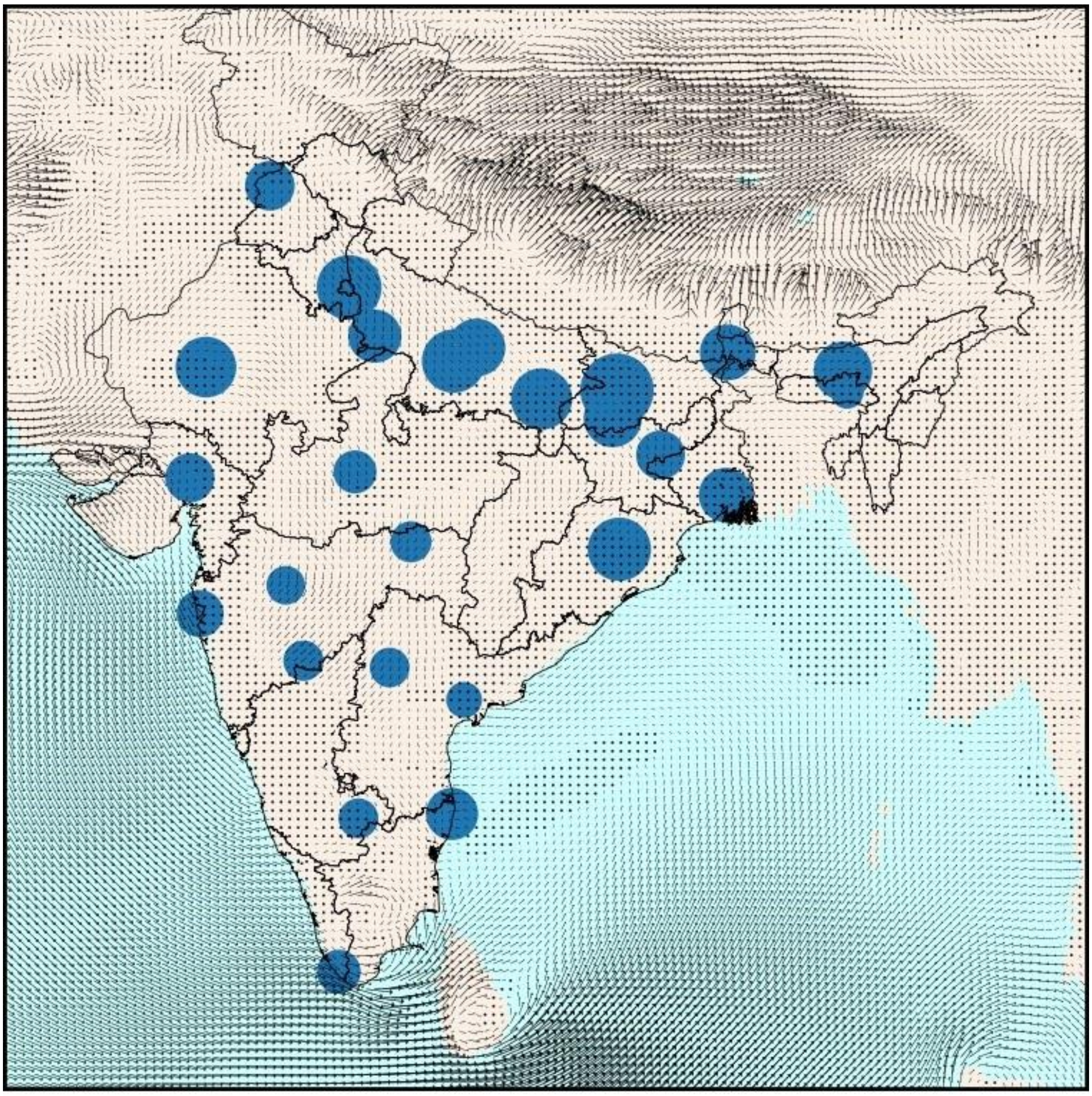


Figure 4-19: Quiver Plot for Post-Monsoon

`

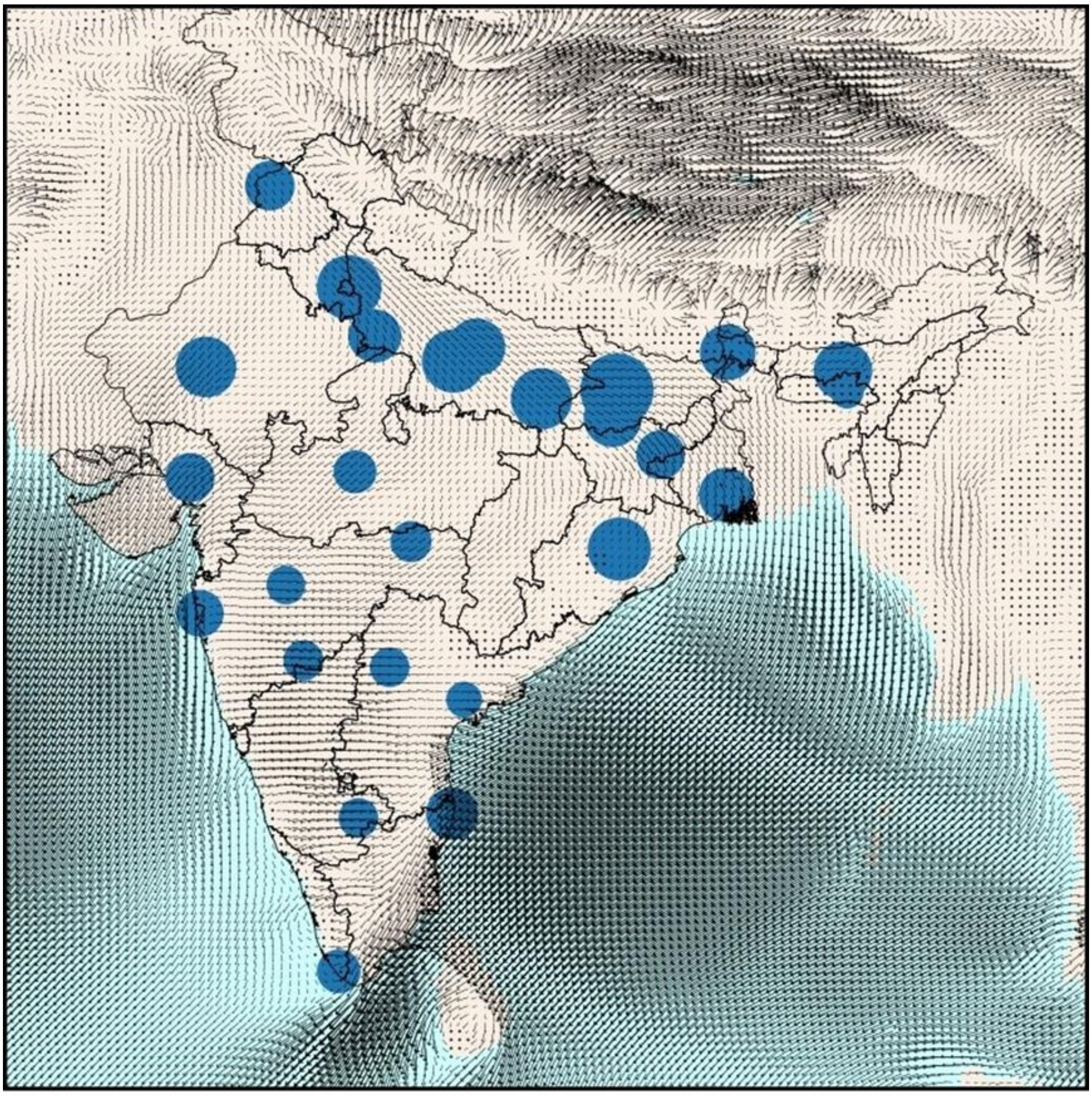


Figure 4-20: Quiver Plot for Winter

`

4.1.8 Comparison with Other Studies

Our model exhibits a significant improvement in terms of RMSE error when compared to multi-output prediction. The best RMSE value of 39 was obtained by [138] using a multi-output auto encoder model to forecast PM2.5 and PM10 concentrations in Beijing City. The model used multiple inputs, including meteorological variables along with pollutant concentrations. [139] have used an ANN model to obtain a correlation coefficient of 0.7301 and an MSE error of 0.0191. The model used single step prediction and the model's applicability for multi-step ahead prediction was not known. Likewise, [140] assessed a basic feed-forward artificial neural network model with multivariate input parameters for the Kolkata region in eastern India in order to predict one-step PM2.5 concentrations during the COVID19 lockdown period. They found that the model's RMSE value was 3.74, while its MAE value was 11.14. Similar to this, [141] tested eight different models on various air pollutants in Kolkata, including the BiLSTM, LSTM-autoencoder, Conv2DLS TM, and stacked LSTM models. They found that the MAE and RMSE values were greater than 10 μg/m3. In the same way, for PM2.5 forecasting in Delhi, [142] has obtained an MAE value of about 15. Additionally, for the Talcher station in India, [138] reported RMSE values of 31, 56, and 68 for 3-hour, 5-hour, and 9-hour ahead prediction using an ANN model. With the help of LSTM and BiLSTM, they have obtained RMSE values of 26, 41, 80 and 42 and 155 and 168, respectively, as opposed to RMSE values of 29.04 and 40.41 for prediction horizons of one hour and eight hours ahead. The proposed model in this study has shown a RMSE range of 7.09 to 53.81 across 26 data stations covering 13 distinct Indian agroclimatic zones. RMSE values below 30 are found in 18 out of 26 locations in India. The findings demonstrate the model's robustness and its applicability to various Indian locations without modification.

<table>
<tr><th>Authors</th><th>Result Obtained</th><th>Our Result</th></tr>
<tr><td>[138] Samal et al. 2021</td><td>The best RMSE value 39</td><td rowspan="3">RMSE values ranges from 7.09 to 53.81 . Out of 26 only two stations have high value of 40.41 and 53.81</td></tr>
<tr><td>[139] Masood et al. 2020</td><td>Correlation coefficient of 0.7301 and an MSE error of 0.0191.</td></tr>
<tr><td>[140] Bera et al. 2021</td><td>RMSE value was 3.74, while its MAE value was 11.14.</td></tr>
<tr><td>[141] Middya, et al. 2022</td><td>MAE and RMSE values were greater than 10 µg/m3</td><td>MAE values ranges from 5.41 to 34.09</td></tr>
<tr><td>[142] Kumar et al. 2020</td><td>MAE value of about 15.</td><td rowspan="2">MAPE values ranges from 18.64 to 52.69
Out of 26 only two stations have high MAPE value of 44.28 and 52.69</td></tr>
<tr><td>[138] Masood et al. 2020</td><td>RMSE values of 31, 56, and 68 for 3-hour, 5-hour, and 9-hour ahead prediction.</td></tr>
</table>