CHAPTER 4

ESTIMATION OF PM10 OVER BRAHMAPUTRA VALLEY USING GEOGRAPHIC WEIGHTED REGRESSION MODEL

4.1 INTRODUCTION

Estimation of PM10 surface concentration using Geographic Weighted Regression (GWR) model by integrative assimilation of surface PM10, MODIS AOD 550 nm and ERA-interim AT reanalysis data, and investigation in the heterogeneity of PM10-PBLH-AOD 550 nm relationship over BV forms the core of this chapter.

Geographic Weighted Regression (GWR) is globally applied to estimate PM10 surface concentration. GWR is a spatial regression technique that evaluates a local model of the variable, predicted by fitting a regression equation to every feature in the dataset. GWR builds local R^2 values for each observation in the dataset, generates parameter estimates, and calculates standard errors. GWR is also effective in modelling spatially varying relationships [1] and investigates heterogeneity in data relationships across space by creating multiple equations [2, 3]. The spatially varying relationships among variables (dependent and explanatory) are captured by generating local regression coefficients (β) [4].

PM10, one of the major contributors to air pollution, is related to meteorological variables and PBLH over a region. An increase in AT enhances photochemical reactions of precursors gases in the atmosphere leading to an increase in PM10 concentration. While PBLH plays a critical role in dictating dispersion of pollutants by turbulence mixing (mechanical and thermal), vertical diffusion, convective transport, and entrainment of aerosols within the mixing height [5, 6]. PM10 varies spatially owing to proximity to sources, short atmospheric lifetime, and surface topography juxtaposed with meteorological parameters [7-9]. AOD 500 nm, quantifies the number of columnar aerosol particles (size ranging from 0.1 to 2 mm) based on optical sensitivity to visible channels [10], which is an important aerosol parameter that is related to PM10 surface concentration.

BV, as mentioned in the preceding chapters, has an inhomogeneous surface that often develops local circulations and modifies ambient synoptic weather to create unique local meteorological conditions [11, 12]. Micro-meteorology and boundary layer dynamism over such complex terrain play a significant role in the spatial distribution of PM10 surface concentration.

Most of the studies in the BV are conducted to study the properties and characteristics of PM10 over the valley (please refer to Table 2.2 Chapter 2). There is a dearth of studies on

the estimation of PM10 surface concentration, heterogeneity of PM10-meteorology-AOD 550 nm relation, and spatial distribution pattern of PM10 over BV. GWR, known for its efficiency in capturing spatial heterogeneity in relationship and estimation of surface concentration of particulate matter, and widely applied for varieties of terrains and climatic zones, has not yet been applied for the complex terrain of BV. PM10 surface concentration is found to be high by researchers [8, 12] for some select sites over BV. One of the limiting factors for this dearth of study is the non-availability of ground-based data for PM10 and the other independent variables.

Ground-based stations are the most reliable source for PM10 surface concentration measurements over a large geographical area. For estimation of PM10 surface concentration, the inadequate number, unavailability at a high temporal scale, and disproportionate distribution of ground-based stations over BV have been major limiting factors. Although the Pollution Control Board of Assam monitors the PM10 surface concentration, the number of stations measuring PM10 is far from sufficient to adequately represent the PM10 surface concentration for the entire valley. The limitations in ground-based data can be addressed by the effective utilization of satellite-derived and reanalysis data. The satellite retrieved aerosol products, and ERA-interim ECMWF reanalysis meteorological data are known for accuracy, spatial, and temporal coverage. They have been extensively used as predictive variables in regression models to estimate PM10 surface concentration over a large geographic area [13-18].

This study (a) estimates PM10 surface concentration using GWR model based on MODIS AOD 550 nm and reanalysis derived AT, PBLH explanatory variables over the BV for the period 2016-2018; (b) explores the relationship between PM10 and the explanatory variables of AOD 550 nm, AT and PBLH height across the valley (c) analyse the spatial distribution of PM10 surface concentration, AOD 550 nm, AT and PBLH. The scope includes evaluation of GWR model performance, model validation and efficacy to represent PM10 surface concentration for BV. To the best knowledge of the authors, the current study is the first attempt to apply the GWR model for estimating PM10 concentration over BV.

4.2 DATA, SOURCES AND PRE-PROCESSING

The technical details of the datasets used are provided in the Table 4.1. The salient features of the data and sources are discussed in the following sub-sections.

Table 4.1 Technical specifications and the sources of the dataset used.

Parameters	Spatial resolution	Temporal resolution	Data Source	Available at
PM10	point	8-hourly (6:00 h LST to 14:00 h LST)	SPCB	https://www.pcbassam.org/
AT, BLH	0.5º x 0.5º	hourly	ERA-interim	https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets
AOD 550 nm	10 x 10 km	10:30 h LST and 13:30 h LST	MODIS onboard Terra and Aqua	https://ladsweb.modaps.eosdis.nasa.gov/search/

AT= Air temperature, BLH= boundary layer height, SPCB= State Pollution Control Board of Assam

4.2.1 PM10 ground-based data

PM10 data were collected from the State Pollution Control Broad of Assam (SPCB), for the period 2016-2018. SPCB has generated data using Respirable Dust Sampler (Envirotech APM 460 NL) based on the gravimetric method. The instrument is reportedly operated for 8-hourly sampling (6:00 h to 14:00 h LST) on alternate days in a week. The monthly mean of PM10 data of 10 stations across different districts of Assam for the period 2016 to 2018 was used for this study. As the measurements are recorded 8-hourly twice or thrice a week, overestimation of the daytime PM10 concentrations was a distinct possibility, as the 8-hourly sampling period does not capture the variability of the pollutant on the finer temporal scale [19].

The PM10 ground-based stations in the western part include Bongaigaon (BNG; 90.58°E, 26.41°N), Nalbari (NBL; 91.43°E, 26.47°N), and Guwahati (GHY; 91.75°E, 26.17°N); in the central part include Nagaon (NGN; 92.5°E, 26.22°N) and Tezpur (TZU; 92.83°E, 26.7°N); and in the eastern part include Golaghat (GLGT; 93.85°E, 26.5°N), Lakhimpur (LMP; 94.05°E, 27.16°N), Sivasagar (SVG; 94.63°E, 26.98°N), Dibrugarh (DBR; 94°.84 E, 27.36°N), and Tinsukia (TSK; 95.36° E, 27.5°N) (Figure 4.1).

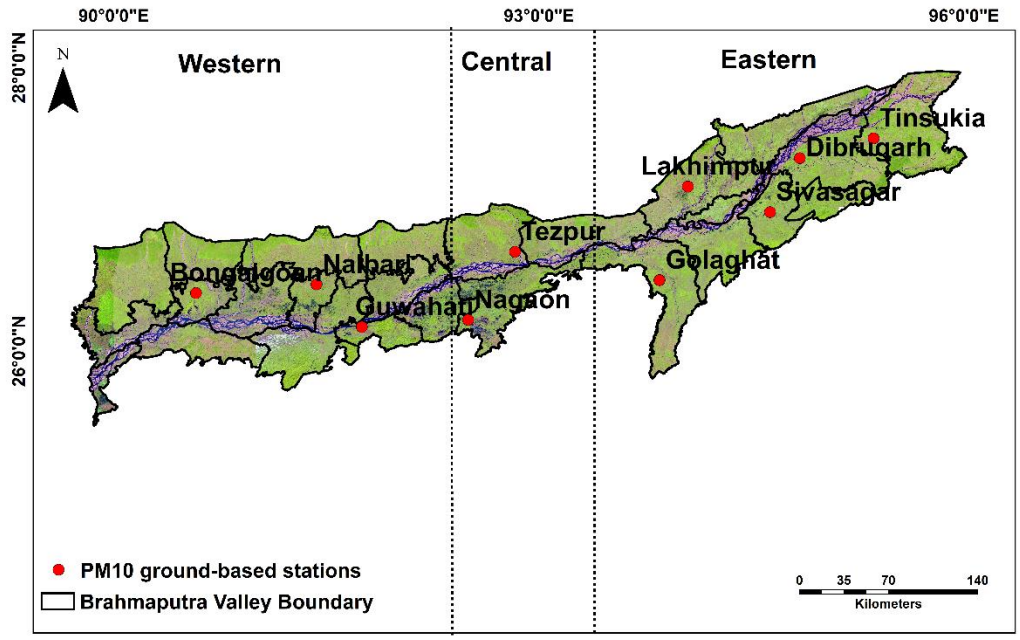


Figure 4.1 Ground-based stations of PM10 over Brahmaputra valley.

4.2.2 ERA-interim meteorological data

ERA-Interim ECMWF is one of the best sources of global data for many climatological parameters [20]. Reanalysis systems assimilate measurements of in-situ and space-based into weather prediction models. ERA-Interim fields of AT and BLH are available at 1, 6 and 12 hourly forecast times. AT data used for the study were at 0.5º x 0.5º spatial resolution and hourly average of 0:00 h to 12:00 h UTC or 5:30 h to 17:30 h LST, 2016-2018 (Table 4.1).

Within the scientific community ECMWF spectral model is highly regarded for its surface and climatological data, as it contains a sophisticated cloud scheme [21]. ECMWF reanalysis models can resolve many of the topographically induced circulations and effects of complex terrain. It may, therefore, explicitly capture the effects of surface features [22], thereby reducing the uncertainty of ERA-Interim surface and climatological data over complex terrain.

4.2.3 MODIS AOD 550 nm

MODIS AOD 550 nm onboard Terra and Aqua is extracted for the corresponding10 sampling stations at the native spatial resolution of 10 x 10 km. The other technical details of MODIS AOD 550 nm are already discussed in Section 2.2.1 Chapter 2.

4.2.4 Data pre-processing

Collocation of PM10 and MODIS AOD

Surface measurement provides 8-hourly data of PM10, while MODIS AOD 550 nm is 10 x10 km spatial resolution data. To understand the relationship between columnar AOD and surface concentration of PM10, both the measurements must be collocated in space and time. For collocating the PM10 data with MODIS AOD retrievals in space and time, the MODIS AOD data at the spatial resolution of 10 x 10 km was averaged to a window of 50 x 50 km (5 x 5 pixel box at 10 km resolution) centered at the location of PM10 ground location [16, 23]. Furthermore, MODIS Aqua and Terra AOD 550 nm data was averaged for the aforementioned PM10 concentration days, and finally monthly mean was calculated.

4.3 METHODOLOGY

The schema of the methodology followed is illustrated in Figure 4.2. The method followed for model fitting, data preparation, data integration and model performance assessment are discussed in the following subsection.

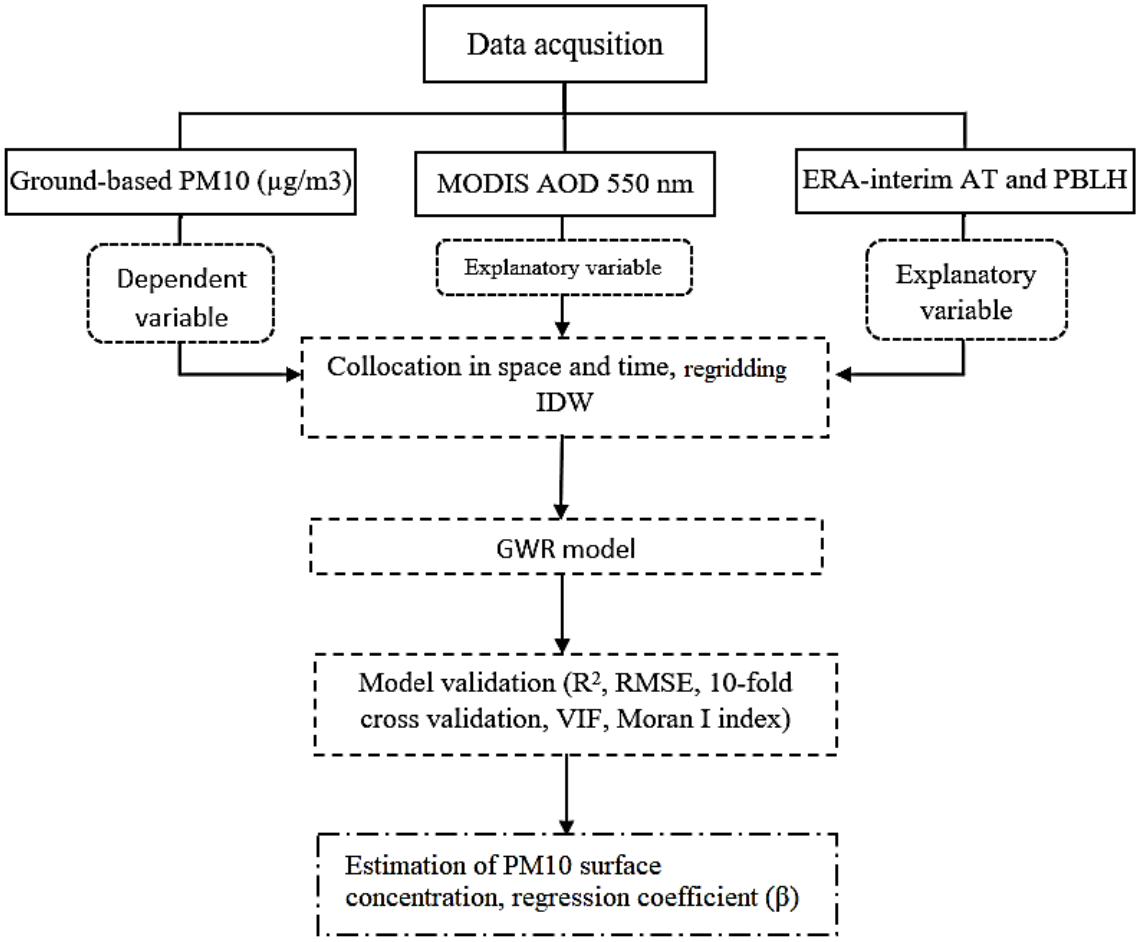


Figure 4.2 Schema of methodology.

4.3.1 GWR model

This study utilized GWR model to generate local regression coefficients for the dependent variable PM10 and the select set of independent variables- AOD 550 nm, AT and PBLH- over the 10 sampling stations.

The GWR model structure is

$$PM10_{ij} = \beta_{0,ij} + \beta_{1,ij} \times AOD_{ij} + \beta_{2,ij} \times PBLH_{ij} + \beta_{3,ij} \times AT_{,ij} + \varepsilon \quad (1)$$

where $PM10_{ij}$ ($\mu g/m^3$) is the monthly mean of surface PM10 concentration at a location i_{1-10} on month j, while $\beta_{0,ij}$ denotes the location-specific intercept on month j. $\beta_{1,ij}$–$\beta_{3,ij}$ are the location specific slopes on month j. AOD_{ij} (unitless) is the AOD 550 nm averaged from

Terra and Aqua MODIS AOD products at a location $i_{1\text{-}10}$ on month j. $PBLH_{ij}$ (unit: m), AT_{ij} (unit: °C) are meteorological parameters at location i on month j.

Meteorological factors AT, WS, RH, Ps, and PBLH coupled with AOD 550 nm were first fitted as predictive variables in the model. Due to the existence of strong multi-collinearity (VIF>7.5) among the explanatory variables, the model showed poor performance, simulating highly spatially clustered values of predicted PM10. After training the model with different predictive variables, finally AT, PBLH, and AOD 550 nm stand out as the key predictive variables for PM10 estimation over the valley. Since vertical mixing within the boundary layer can substantially change particle extinction properties [14], and high air temperature can intensify photochemical reactions producing more particulate matter [6, 24], they are the key predictive variables for the model fitting [Eq. (1)]. MODIS AOD 550 nm as the sole predictive variable did not represent well in estimating the PM10 near-surface concentration in the model fitting. It is probably due to AOD measures the columnar aerosol concentration while PM10 represents the near-surface mass concentration of particulate matter.

4.3.2 VIF multicollinearity test and Local R^2

Variance inflation factors (VIF) are calculated to detect the collinearity or redundancy among the predictor variables used in a GWR model [1]. If two (or more) explanatory variables are found with high multicollinearity (VIF> 7.5) in the models, it indicates that those variables are possibly telling the same relationship. VIF helps to find the redundant variable that can explain the unique aspect of the dependent variable.

The local R^2 value indicates the model performance and the spatial heterogeneity of the GWR model [14, 16].

4.3.3 Moran I spatial autocorrelation test

Moran I spatial autocorrelation test is an appropriate statistical test that measures the level of spatial autocorrelation in the regression residuals. If the residuals exhibit a random spatial pattern, it follows normal distribution indicating spatial heterogeneity or random spatial pattern. Based on residual, the over/under prediction of a model can be examined.

4.3.4 10-fold cross-validation

The 10-fold cross-validation (CV) method [25] was applied for evaluating the GWR model performance. This method is widely used for the validation of model fitting [14, 16-17]. The original dataset was randomly divided into 10 folds with approximately 10% of the total data points in each fold. In each step of CV, the model was fitted with nine folds (90% of the total data set) to predict one-fold of the dataset. This step was repeated 10 times until every single fold was tested. Furthermore, R^2 and RMSE were calculated for model fitting and cross-validation.

To validate the GWR model performances, the predicted PM10 (μg/m^3) concentrations were fitted against the observed values. Statistical indicators coefficient of determination (R^2) and root mean square error (RMSE) were calculated to assess the goodness of fit between the predicted and observed PM10 concentration and prediction accuracy of the GWR model respectively. Furthermore, R^2 and RMSE were calculated for model fitting and cross-validation.

4.4 RESULTS AND DISCUSSION

4.4.1 GWR model performance in estimating PM10 surface concentration

The scatter plots of the model fitting and cross-validation of the GWR model are shown in Figure 4.3. For R^2 and RMSE values of the model fitting between the estimated and observed PM10 are 0.62 and 22.74 μg/m^3 respectively (Figure 4.3a). Compared to the model fitting, cross-validation (CV) R^2 values decreased by 0.09 (R^2= 0.53), and RMSE decreased by 0.14 (RMSE= 22.60 μg/m^3) (Figure 4.3b) which suggested that the model is slightly over fitted. The overall R^2 of model fitting suggests that the GWR model can explain 62% of the total variability of data while estimating PM10 surface concentrations. Also, the observed t-values of the dataset are statistically significant ($t>2$, $p<0.05$), which explains the tightness of fit of each explanatory variable with the dependent variable.

PM_{10} μg/m^3 surface concentration (Figure 4.4b) estimated by the GWR model was slightly underestimated as compared to the observed PM_{10} μg/m^3 surface concentration (Figure 4.4a) across the sampling stations.

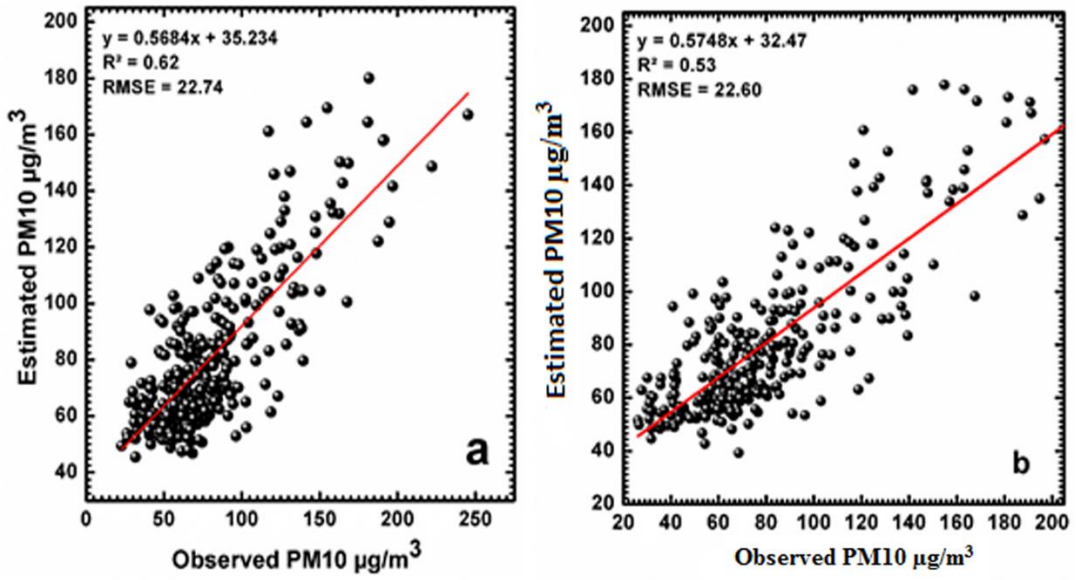


Figure 4.3 Scatter plot of model fitting cross-validation (a) model fitting results of the GWR model; (b) 10-fold cross-validation results of the GWR model. The red line is the linear regression of the scatter plot. R^2, RMSE shows the prediction accuracy of the model fitting and cross-validate results.

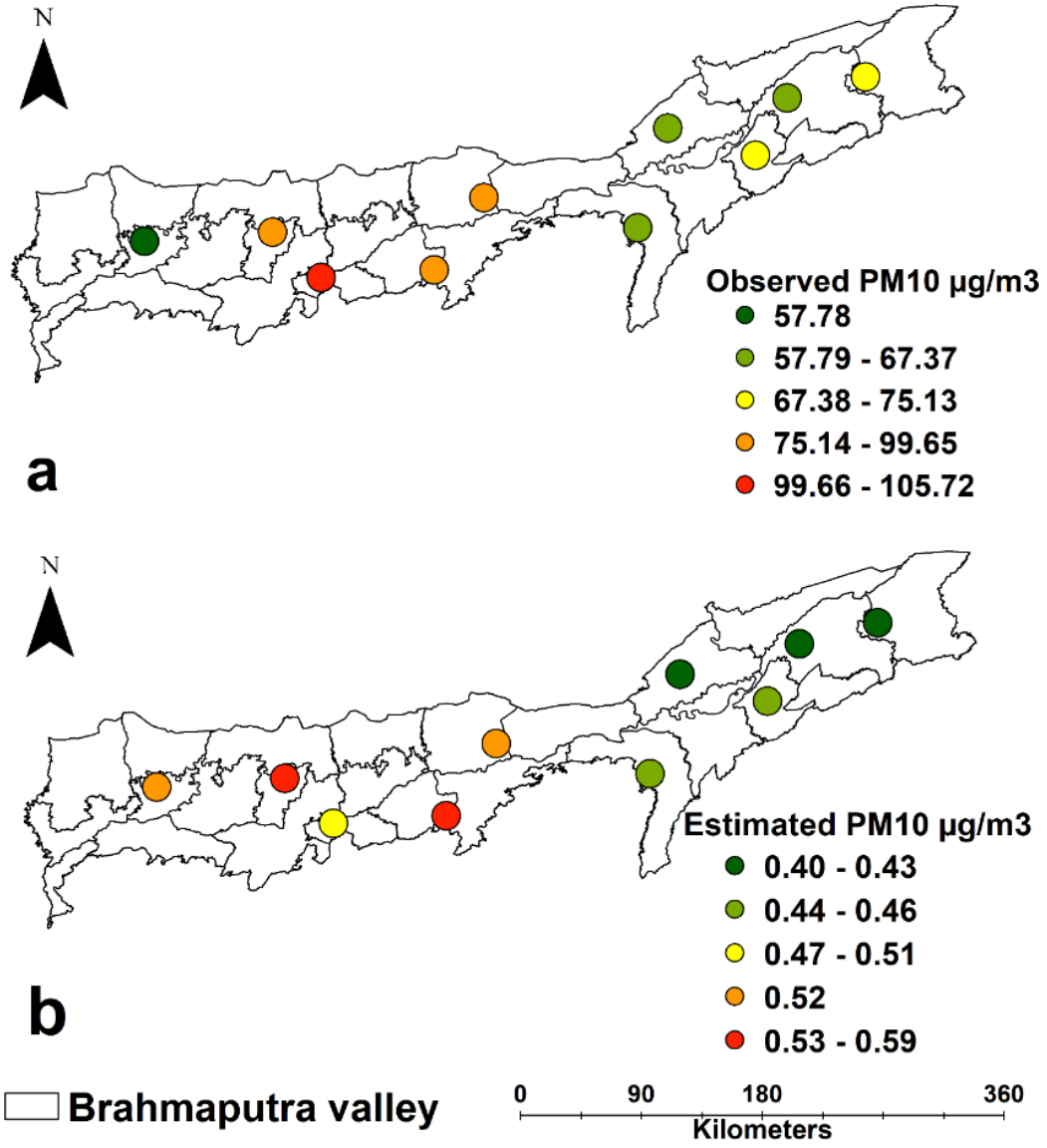
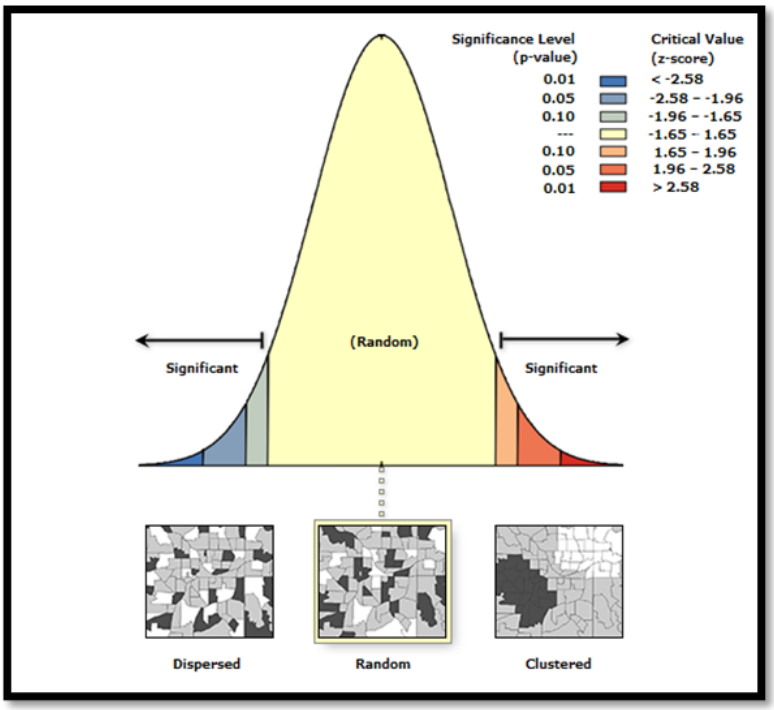


Figure 4.4 shows the variation of (a) observed PM_{10} surface concentration ($\mu g/m^3$) and (b) estimated PM_{10} surface concentration ($\mu g/m^3$) by the GWR model for the ground-based sampling stations, Brahmaputra valley, 2016-2018.

Spatial autocorrelation

The Moran's I index results (z= -1.21, p<0.5) showed no significant spatial autocorrelation for the residual of the GWR model for the present set of data indicating the spatial randomness of distribution (Figure 4.5). The local R^2 values of the GWR model range from 0.164 to 0.547. Low R^2 values were computed for the semi-urban areas whereas high for the urban areas of the western and central parts of the valley (Figure 4.6). The local R^2 values indicate the moderate performance of the model due to the sparsity and non-uniform spatial distribution of the ground-based monitoring network (Figure 4.1).



Global Moran's I Summary

Moran's Index:	-0.019
Expected Index:	-0.003
Variance:	0.00018
z-score:	-1.21
p-value:	0.22

Figure 4.5 Summary of the Moran's I Spatial Autocorrelation used on the GWR residuals. The Moran's I index results (z= -1.21, p<0.5) showed no significant spatial autocorrelation for the residual of the GWR model for the present set of data indicating the spatial randomness of distribution.

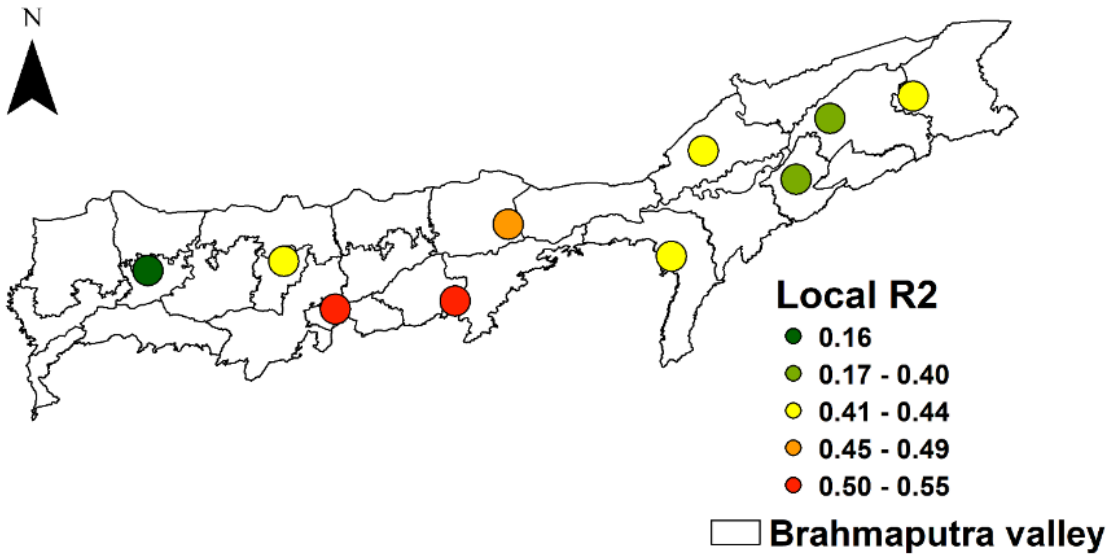


Figure 4.6 Spatial variation of local R^2 value generated from GWR model for the ground-based sampling stations, Brahmaputra valley.

Multicollinearity

Variance inflation factors (VIF) are calculated to detect the collinearity among the predictor variables based on MODIS derived AOD 550 nm, PBLH and AT [1]. The VIF values were below 3, denoting the weak collinearity among the variables in the current model. Hence, variable collinearity does not affect model efficiency.

4.4.2 PM10-AT, PM10-PBLH and PM10-AOD 550 nm relationship

The wide range of spatial variation in regression coefficient (β) values over BV indicated the heterogeneity in the relationship among the variables (Figure 4.7 and Table 4.2). PM10-AOD 550 nm showed a strong positive relationship (β_1= 7.78 to 23.91) over the western part, a weak negative relationship (β_1= -2.91 to -15.68) over the central part, and a strong negative relationship (β_1= -15.47 to -27.27) over the eastern part of the valley. Strong west-east spatial heterogeneity over BV was observed in PM10-PBLH (β_2= -0.01 to 0.005), and PM10-AT relationship (β_3= -2.15 to -9.84) respectively. The observed variation in coefficient values can be attributed to differences in local micrometeorological conditions (atmospheric stability, wind, and turbulence) formed across the valley. Spatial variations in local meteorology are quite usual over a complex terrain [26].

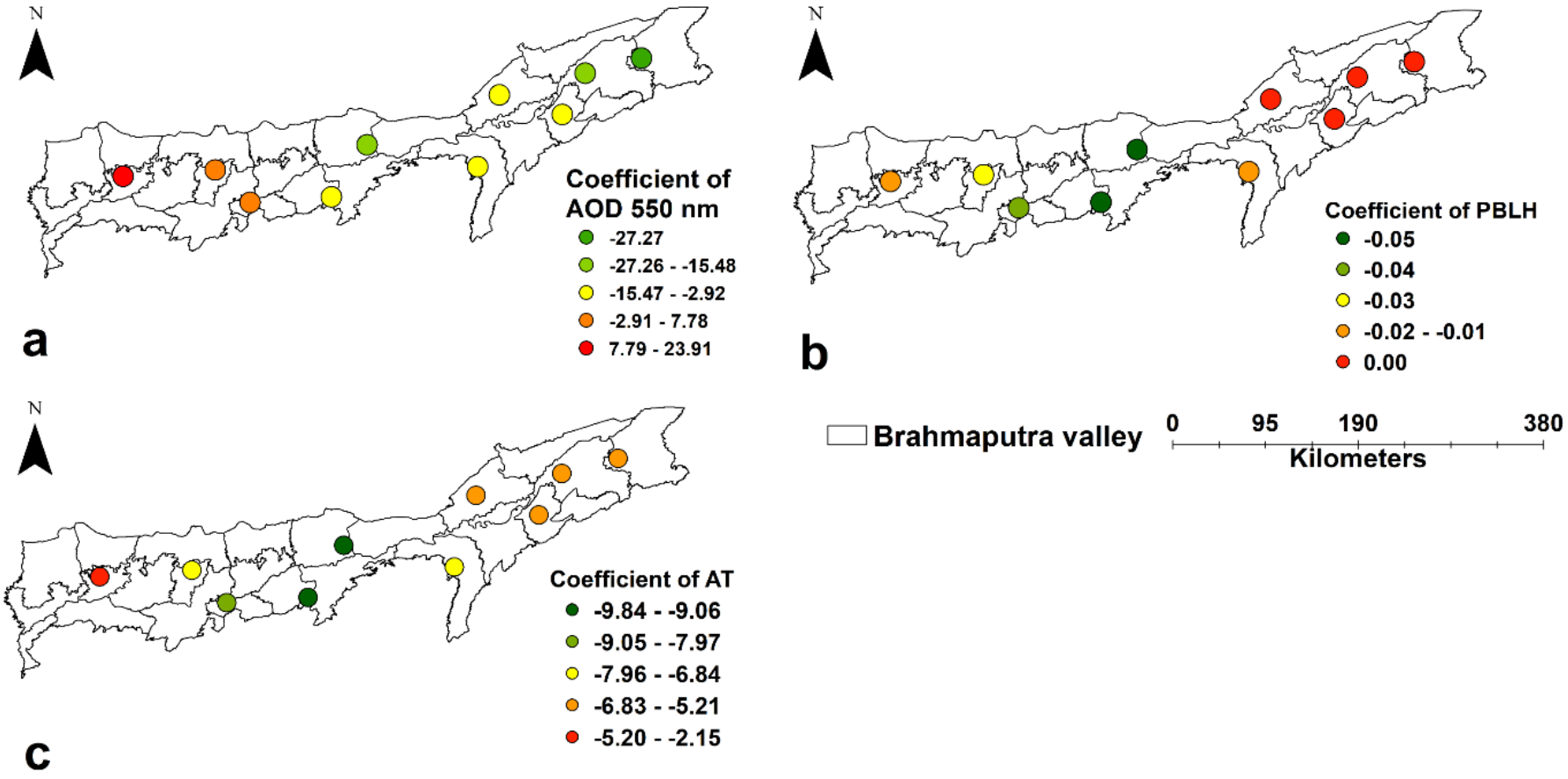


Figure 4.7 Variation in regression coefficient (β) values (a) coefficient of AOD 550 nm (β_1), (b) coefficient of PBLH (β_2), (c) coefficient of AT (β_3) generated by GWR model for the ground-based sampling stations, Brahmaputra valley.

Table 4.2 Regression Coefficient (β) values of predictive variables generated by the Geographic Weighted Regression (GWR) model for the respective PM10 ground-based sampling stations, Brahmaputra valley.

Division	Location	Coefficient AOD 550 nm β_1	Coefficient PBL β_2	Coefficient AT β_3
Eastern	BNG	23.91	-0.01	-2.15
	NBL	7.78	-0.03	-6.86
	GHY	5.94	-0.04	-7.96
Central	NGN	-2.92	-0.05	-9.06
	TZU	-17.68	-0.04	-9.83
	GLGT	-8.08	-0.01	-6.83
Western	LMP	-4.19	0.002	-5.86
	SVG	-8.80	0.0007	-5.20
	DBR	-15.47	0.003	-5.34
	TSK	-27.27	0.004	-5.78

4.4.3 Spatial distribution of PM10 surface concentration

Higher mean value of PM10 surface concentration and AOD 550 nm was observed in the western and central parts of BV than in the eastern valley (Table 4.3). GHY had the highest PM10 concentration (105.72 ± 41.94 µg/m^3) with AOD 550 nm value of 0.51± 0.22. The lowest mean PM10 value was noticed over BNG - 57.78 ± 9.03 µg/m^3 with AOD 550 nm mean value of 0.52 ± 0.26. GHY, being the largest metropolitan city of the BV, displays high loading of carbonaceous aerosols from vehicles and open burning of solid waste disposal [27]. The meteorological parameters PBLH and AT showed less variation in their mean values over the valley. Mean values of PBLH and AT ranged between 492- 567m and 22.96- 24.44 °C. PBLH and AT displayed higher values over NBL and GHY, and lower values over TSK and NGN respectively (Table 4.3).

Table 4.3 Summary of descriptive statistics of the independent and dependent variables used in the Geographic Weighted Regression (GWR) model, 2016 to 2018.

Locations	Statistics	PM10 (µg/m^3)	AOD 550 nm	PBLH (m)	AT (ºC)
BNG	Mean	57.78	0.52	512.00	23.80
	SD	9.03	0.26	197.64	3.82
	Median	58.63	0.50	521.90	24.55
DBR	Mean	63.51	0.40	509.19	23.36
	SD	24.05	0.22	172.48	3.43
	Median	59.58	0.36	493.25	24.10
GLGT	Mean	65.10	0.46	532.99	23.09
	SD	24.93	0.24	151.30	2.77
	Median	60.79	0.45	527.63	24.33
GHY	Mean	105.72	0.51	537.02	24.44
	SD	41.94	0.22	162.53	3.35
	Median	89.46	0.51	553.30	25.35
LMP	Mean	67.37	0.43	529.67	23.55
	SD	28.83	0.25	193.19	3.25
	Median	65.35	0.38	497.55	24.55
NGN	Mean	99.59	0.58	554.48	22.96
	SD	44.80	0.28	165.58	3.19
	Median	90.26	0.55	528.93	23.38
SVG	Mean	75.13	0.46	496.06	23.57
	SD	21.05	0.23	153.86	3.31
	Median	71.23	0.44	498.70	24.65
NBL	Mean	99.65	0.59	567.84	23.84
	SD	39.36	0.25	242.83	3.36
	Median	90.13	0.60	536.70	25.15
TZU	Mean	98.68	0.52	525.40	23.41
	SD	52.97	0.30	156.53	3.56
	Median	95.13	0.42	516.80	24.25
TSK	Mean	73.16	0.40	492.40	23.12
	SD	24.27	0.18	223.25	2.64
	Median	72.59	0.38	496.40	24.35

4.5 CONCLUSION

PM10 $\mu g/m^3$ surface concentration estimated by the GWR model was slightly underestimated as compared to the observed PM10 $\mu g/m^3$ surface concentration across the 10 ground-based stations of BV. The GWR model performed well for the estimation of PM10 surface concentration with R^2 and RMSE values of 0.62 and 22.74 $\mu g/m^3$ respectively.

GWR performed well in capturing the spatial heterogeneity of the relationship between the dependent PM10, and explanatory variables AOD 550 nm, AT, and PBLH over BV. The β values indicated the heterogeneity in the strength of association amongst the variables. The strongest positive relationship in PM10-AOD 550 nm was found to be in the western part of BV.

The monthly mean values of observed PM10 concentration and AOD 550 nm vary strongly with location over the valley depending on the proximity to sources and the effectiveness of dispersal mechanisms. The western part of BV, in comparison to the central and eastern parts, had the highest PM10 surface concentration. This can be attributed to aerosol loading by long-range transported aerosols and coupled with local meteorological conditions. A decreasing gradient of mean values for PM10 and AOD 550 nm is noted from the west to the east of the valley.

The explanatory variables derived from satellite and reanalysis data sources are found reliable and efficient for PM10 estimation using the GWR model over the terrain of BV, where availability and consistency of ground-based data are limited. GWR model can be developed further for forecasting PM2.5 and PM10 at a higher temporal scale over the valley. As the GWR can aid in modelling the spatial variation in relationships among the variables, it can be applied to analyse the role of individual meteorological parameters influencing PM10 surface concentration.

4.6 REFERENCE

[1] Fotheringham, A. S. and Oshan, T. M. Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems*, 18(4):303-329, 2016.

[2] Fotheringham, A. S., Brunsdon, C., and Charlton, M. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, England, pages 27-46, 2002.

[3] Harris, R., Dong, G., and Zhang, W. Using Contextualized Geographically Weighted Regression to Model the Spatial Heterogeneity of Land Prices in Beijing, China. *Transactions in GIS*, 17(6):901-919, 2013.

[4] Paez, A. and Wheeler, D. C. *Geographically weighted regression*. In International Encyclopedia of Human Geography, 1st edition, Elsevier, Oxford, UK, 2009.

[5] Seinfeld, J. H. and Pandis, S. N. *Atmospheric Chemistry and Physics, from Air Pollution to Climate Change*. John Wiley & Sons, 1998.

[6] Gupta, P. and Christopher, S. A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research: Atmospheres*, 114(D14), 2009.

[7] Kaufman, Y. J., Tanre, D., and Boucher, O. A satellite view of aerosols in the climate system. *Nature*, 419(6903):215-223, 2002.

[8] Pathak, B., Borgohain, A., Bhuyan, P. K., Kundu, S. S., Sudhakar, S., Gogoi, M. M., and Takemura, T. Spatial heterogeneity in near surface aerosol characteristics across the Brahmaputra valley. *Journal of Earth System Science*, 123(4):651-663, 2014.

[9] da Silva Palácios, R., da Silva Sallo, F., and de Souza Nogueira, J. Variability of aerosol optical depth over Cerrado of Mato Grosso, Brazil. *Environment and Ecology Research*, 4:99-105, 2016.

[10] Hoff, R. M. and Christopher, S. A. Remote sensing of particulate pollution from space: have we reached the promised land? *Journal of the Air & Waste Management Association*, 59(6):645-675, 2009.

[11] Gogoi, M. M., Pathak, B., Moorthy, K. K., Bhuyan, P. K., Babu, S. S., Bhuyan, K., and Kalita, G. Multi-year investigations of near surface and columnar aerosols over Dibrugarh, North-eastern location of India: Heterogeneity in source impacts. *Atmospheric Environment*, 45(9):1714-1724, 2011.

[12] Tiwari, S., Dumka, U. C., Gautam, A. S., Kaskaoutis, D. G., Srivastava, A. K., Bisht, D. S., Chakrabarty, R. K., Sumlin, B. J., and Solmon, F. Assessment of $PM_{2.5}$ and PM_{10} over Guwahati in Brahmaputra River Valley: Temporal evolution, source apportionment and meteorological dependence. *Atmospheric Pollution Research*, 8(1):13-28, 2017.

[13] Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., Sarnat, J. A., and Liu, Y. Estimating ground-level $PM_{2.5}$ concentrations in the southeastern US using geographically weighted regression. *Environmental Research*, 121:1-10, 2013.

[14] You, W., Zang, Z., Zhang, L., Li, Z., Chen, D., and Zhang, G. Estimating ground-level PM_{10} concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. *Remote Sensing of Environment*, 168:276-285, 2015.

[15] Fang, X., Zou, B., Liu, X., Sternberg, T., and Zhai, L. Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. *Remote Sensing of Environment*, 186:152-163, 2016.

[16] Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M., He, J., Yan, Y., Wang, F., Min, M., and Zhai, P. Impact of diurnal variability and meteorological factors on the PM2.5-AOD relationship: Implications for $PM_{2.5}$ remote sensing. *Environmental Pollution*, 221:94-104, 2017.

[17] He, Q. and Huang, B. Satellite-based mapping of daily high-resolution ground $PM_{2.5}$ in China via space-time regression modeling. *Remote Sensing of Environment*, 206:72-83, 2018.

[18] van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M. Documentation for the Global Annual $PM_{2.5}$ Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) with GWR, 1998-2016. *Palisades NY: NASA Socioeconomic Data and Applications Center*, 2018.

[19] Pant, P., Lal, R. M., Guttikunda, S. K., Russell, A. G., Nagpure, A. S., Ramaswami, A., and Peltier, R. E. Monitoring particulate matter in India: recent trends and future outlook. *Air Quality, Atmosphere & Health*, 12(1):45-58, 2019.

[20] von Engeln, A. and Teixeira, J. A planetary boundary layer height climatology derived from ECMWF reanalysis data. *Journal of Climate*, 26(17):6575-6590, 2013.

[21] Jakob, C. An improved strategy for the evaluation of cloud parameterizations in GCMs. *Bulletin of the American Meteorological Society*, 84(10):1387-1402, 2003.

[22] ECMWF IFS DOCUMENTATION. Cy43r3 Operational implementation European Centre for Medium-Range Weather Forecasts, England, 11 July 2017.

[23] Mishra, M. K. Retrieval of Aerosol Optical Depth From INSAT-3D Imager Over Asian Landmass and Adjoining Ocean: Retrieval Uncertainty and Validation. *Journal of Geophysical Research: Atmospheres*, 123(10):5484-5508, 2018.

[24] Liu, Y., Franklin, M., Kahn, R., and Koutrakis, P. Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, 107(1-2):33-44, 2007.

[25] Rodriguez, J. D., Perez, A., and Lozano, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 32(3):569-575, 2009.

[26] Arya, S. P. *Introduction to micrometeorology*. International Geophysics series volume 2001.

[27] TERI. *Risk Assessment and Review of Prevailing Laws Standards Policies and Programmes to Climate Proof Cities*. Synthesis Report for Guwahati, 2013.