Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
e-ISSN: 2395-3454

Vol. 16 | No. 1 | pp. 155-161 | 2017

Original Research Paper | Open Access

Research on Spatial Correlation Between Air Quality and Land Use Based on GWR Models

Xu Feng
City Institute, Dalian University of Technology, Liaoning Dalian 116600, China

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 22-11-2015
Accepted: 28-01-2016

Key Words:
Air quality
Land use
Spatial autocorrelation
Geographically weighted regression (GWR)

ABSTRACT

Air quality has the strong regional characteristics, and the land use type can reflect human activities in the area to some extent. In order to explore the spatial relationships between air quality indicators ($PM_{2.5}$, PM_{10}, O_3, SO_2, NO_2, CO) and land use indicators (tree, grass, farm, water and building proportion), the ordinary least squares (OLS) and geographically weighted regression (GWR) models are established and tested by comparing R^2 and AICc (Akaika Information Criterion corrected) of the models. The Moran's I statistics on the residuals from OLS and GWR models shows that the GWR models can be a good deal with the spatial autocorrelation. Meanwhile local parameters at different locations from GWR models can be good performance in the spatial heterogeneity of the air quality and land use, providing a scientific basis for rational and effective regional governance.

INTRODUCTION

With the rapid economic development and considerable energy consumption, environment pollutants have more and more serious effects to the quality of human life, which is attracting considerable attention and becomes a serious social problem in China, especially the haze events concerned in recent years (Zheng et al. 2016). Due to the liquidity of the atmosphere, air pollution has distinct regional characteristics. With the special natural geographical condition and the particular atmospheric circulation, the air pollution in one area can transport through a short distance to affect other areas around. The land use and cover changes (LUCC) can affect the atmospheric composition (Yuan 2008), by the effects of air pollution removal rate and the amount of carbon sequestration, and so on. Moreover, the human activity in different land use type has great difference, while human activity also has a significant effect on the atmosphere composition. Currently, numerous studies have been conducted worldwide to analyse the relationships between air quality and other factors using statistical methods, mostly based on ordinary least squares (OLS). Air quality variables from multiple ground stations are used as dependent variables, and explore the relationships with other responding factors. These methods can hide important local variations in the model parameters, and are not able to deal with spatial autocorrelation existing in the variables. While a recently developed technique, geographically weighted regression (GWR), as an exploratory spatial data analysis (ESDA) technique, can take fully account into the spatial heterogeneity of parameters. GWR models make great improvement of model performance and get more reliable spatial relationships through reducing spatial autocorrelation of the parameters (Jun Tu et al. 2008). Existing studies concentrate on exploring relationships between water quality and land use type based on GWR models, and got good results (Jun Tu et al. 2008, Jun Tu et al. 2011, Mehaffey et al. 2005, Schoonover et al. 2005, Sliva et al. 2001, Stutter et al. 2007, Woli et al. 2004). Therefore, this paper proposed a good method to explore the relationships between air quality and land use type based on GWR models. Through modelling the relationships between air quality indexes and land use type indicators, it can explore the different spatial correlation in different areas by the local parameters in GWR models. So it will provide scientific solution for regional air pollution control, but also provide the scientific basis for the air quality forecast exactly.

MATERIALS AND METHODS

Study area: Beijing-Tianjin-Hebei region (BTH), is one of the most populous and developed region of China. In recent years, with the economic booming and rapid urbanization, the air quality problem in BTH has attracted more and more social attention. BTH outbreak a severe haze event in October, 2014, when it happened four consecutive times severe haze in one month, causing a tremendous social repercussion. So BTH as the study area, the haze in October, 2014 as the study case, it selected 13 cities and 79 ground stations as the data source, collecting hourly air compositions ($PM_{2.5}$, PM_{10}, O_3, SO_2, NO_2, CO) as the air quality indicators. To facilitate research, the hourly station data are averaged in one month, which can well reflect the region air quality states because the severely polluted in this month. In order to explore the relationship between air quality and land use type indicators, this study uses MCD12 land use type from MODIS as the base data. For the study needs, we processed the BTH land use type data, reclassifying 12 land use types as 5 main land use types (forest, grass, farm, water, urban) (Fig. 1).

Methods: This study builds OLS and GWR models, in which air quality indicators are as dependent variables, and land use type indicators are independent variables. Different land use types have geographical correlation in the natural environment, for example the transition between forests and grasslands, the dependencies between water and farmland, and so on. GWR models, however, need to assurance factors in the model non-collinear; otherwise they will cause the model not to work. So each GWR model used only one land use indicator to analyse its association with one air quality indicator in order to avoid the potential multicollinearity, to explain the effect of different land use types on different air quality indicators, and also better describes the different types of human activities to air pollution, which can avoid mixing phenomenon. There are 5 land use type indicators and 6 air quality indicators. Therefore, the relationship for 30 (5 times 6) pairs of air quality and land use type indicators were analysed by building 30 GWR models and 30 OLS models.

The theory of geographically weighted regression: Spatial data has two important properties, spatial autocorrelation and non-stationary, which makes it difficult to meet the assumptions and requirements of conventional regression techniques such as OLS. Spatial autocorrelation is the situation in which the value of a variable at a location is related to the value of the same variable at the location nearby (Zhang et al. 2004). While the spatial non-stationary means that the relationships between the independent and dependent variables are not constant over space (Fotheringham et al. 2002). So it is important to explore the relationship of spatial variables considering the spatial autocorrelation and non-stationary.

Geographically weighted regression, as an exploratory spatial data analysis (ESDA) technique, is an extension of the traditional standard regression framework by allowing local rather than global parameters to be estimated

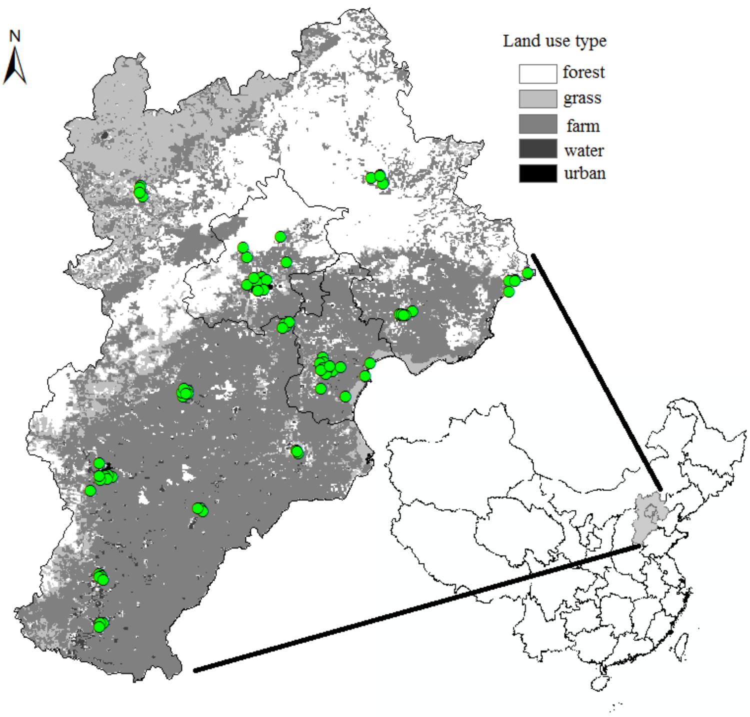

Fig. 1: Land use type and air quality ground stations in research area.

(Fotheringham et al. 2001). As a local statistic, GWR models can produce a set of local parameter estimates showing how a relationship varies over space. For each data point, GWR model will produce the local R^2 and local residual, and also to examine the spatial pattern of the local estimates to get some understanding of hidden possible causes of this pattern. In contrast, the traditional regression method such as OLS is a type of global statistics (Eq. 1), which assumes the relationship under study is constant over space, so the parameter is estimated to be the same for all the study area (Yaolin Liu et al. 2015). The OLS model is defined as:

$$y_i = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon \quad ...(1)$$

Where y is the dependent variable, β_0 is the intercept, β_i is the parameter estimate (coefficient) for independent variable xi, p is the number of independent variables, ε is the error term.

While GWR model is used the local parameters (Eq. 2), which varies according to different position:

$$y_i = \beta_0(u_j, v_j) + \sum_{i=1}^{p} \beta_i(u_j, v_j)x_{ij} + \varepsilon_{ij} \quad ...(2)$$

GWR model is calibrated by weighting all observations around a sample point using a distance decay function, assuming the observations closer to the location of the sample point have higher impact on the local parameter estimates for the location (Eq. 3). The distance decay function has distance threshold weighted method, inverse distance weighted method, Gaussian function method, bi-square function method and so on. Gaussian and bi-square function method, are two popular weighted functions. Regression coefficient for each regression point can be estimated:

$$\hat{\beta} = (\mathrm{x}^T W(u_i, v_i)\mathrm{x})^{-1} \mathrm{x}^T W(u_i, v_i) y \quad ...(3)$$

That $\hat{\beta}$ stands as the estimation of regression coefficients and $W(u_i, v_i)$ is a diagonal matrix which its diagonal elements are the weight of each n set of local parameters around regression point (i).

One of the important issues in GWR model is the determination of weight matrix. In practice, the results of GWR are relatively insensitive to the choice of weighting function (e.g. Gaussian or bi square Kernel or any other continues weighting functions) but they are sensitive to the selection of the bandwidth of the particular weighting function (Fotheringham et al. 2002). The difference between Gaussian and bi square Kernel is that, in Gaussian the weight never gets zero (nonzero) but close to zero no matter how far away from the regression point (Eq. 4). However in bi square (clear cut range) the weight of the data point out of the bandwidth set to zero (Eq. 5).

$$w_{in} = \exp\left(\frac{-d^2_{in}}{b^2}\right) \quad ...(4)$$

$$w_{in} = \begin{cases} (1-(\frac{d_{in}^2}{b^2}))^2 \therefore d_{in} < b \\ 0 \end{cases} \quad ...(5)$$

Where w_{in} is the Euclidian distance between regression point and nth data point, b considered as bandwidth, which has significant role in the performance of GWR models. One remarkable aspect of GWR is the capability of this method in the case of existence of not well distributed data points (spares data around regression point) (Ma et al. 2014). Therefore, there are two type of methods for selecting bandwidth which are known as fixed (which is set fixed for all regression point) or adaptive (kernel has the larger bandwidth when the data are spares, and narrower when the data are plenty). Adaptive bandwidth is the optimal bandwidth selected so that there are the same number of data points with nonzero weights for each regression point. There are several strategies in determination of optimal adaptive bandwidth (Guo et al. 2008) most famous one are cross validation approach (based on minimum sum of squared error) or Akaike Information Criterion (AICc) (Eq. 6).

$$AIC_C = 2n\log_e(\delta) + n\log_e(2\pi) + n\left(\frac{n + tr(H)}{n - 2 - tr(H)}\right) \quad ...(6)$$

$$H_{(i,:)} = X_i(\mathrm{X}^T w_i X)^{-1} X^T w_i \quad ...(7)$$

In the above equation, n is the sample size, δ is the estimation of the standard error of the error term and $tr(H)$ is the trace of the hat matrix (maps the vector of response to fitted value) that elements in each row of eat can be calculated through Eq. 7. In this study we use AICc (the lowest value if AICc assigned to the best model) for selecting the optimal bandwidth also to determine the best collection of the predictors in the GWR model.

Model comparison: Because OLS and GWR models are built in this paper, it can evaluate advantage of the GWR models over OLS models, by comparing the R^2 and AICc of models. R^2 can measure models' relationship for variations, the value of which is between 0 and 1. Usually, the higher R^2 indicates that the independent variations can explain more information of dependent variations, the lower R^2 indicates less (Yaolin Liu et al. 2015). AICc is used to compare the performance of regression models. When comparing the models, AICc is standard of model goodness. Usually, when the AICc difference of two models is more than 3, the model of less AICc is considered the optimal one. Comparing the AICc value between GWR models and AICc models is one of the good methods to evaluate the advantage of local models (GWR) over global models (OLS).

Result of spatial autocorrelation comparison: To comparison the model skill of spatial autocorrelation problem, the global Moran's I of model residual is analysed. Moran's I is the common index of spatial autocorrelation, the value of which is between -1 and 1. The value 1 means best spatially positive correlation (high-high, or low-low cluster), while the value -1 means best spatially negative correlation (high-low, or low-high cluster), and the value 0 means spatially random distribution. If spatial autocorrelation exists in models, the hypothesis of model random distribution will be rejected, and the models built will be discarded.

RESULTS AND DISCUSSION

Due to isotropy of air diffusion, research before shows the change of meteorological in 10km can be ignored. So in this study, the 10km buffer is built for every ground station, to analyse models by comparing the proportion of land use type in different station buffer by overlaying analysis.

Indicators statistics comparison: Through statistics analysis of air quality and land use type indicators (Table 1), it shows the indicators have large difference in space. For land use type, the main indicator is farm, the proportion of which is from 5.7% to 90.4%, and the second is urban, the proportion of which is from 0 to 85.8%. Also, the diversity of air quality around different stations is still large, for example, the value of PM_{10} is from 7.58 to 300.14.

Model correlation coefficient comparison: Through building OLS and GWR models in ArcGIS with 6 air quality indicators and 5 land use indicators, the R^2 of models can be calculated and analysed (Table 2), which shows the R^2 of GWR models is larger than OLS models. For grass and $PM_{2.5}$, the R^2 of GWR is up to 0.91, which means the grass can explain 91% spatial information of $PM_{2.5}$, while the R^2 of corresponding OLS model is 0.3. Research the GWR models, it's clear that the R^2 of models built by air quality and land use type factors is larger than 0.6, while for OLS models, the R^2 mostly is less than 0.1, even that the R^2 of OLS model built by PM_{10} and urban is down to 0.00058.

Table 1: Indicators statistics.

	Minimum	Maximum	Mean	Standard deviation
Land use type indicators				
Forest%	0.0%	78.8%	14.8%	0.190255
Grass%	0.3%	39.2%	11.9%	0.105638
Farm%	5.7%	90.4%	45.6%	0.211076
Water%	0.0%	46.6%	11.4%	0.134917
Urban%	0.0%	85.8%	16.3%	0.263363
Air quality indicators				
$PM_{2.5}$	10.31	179.04	107.58	36.824079
PM_{10}	7.58	300.14	169.26	62.541755
O_3	24.79	169.26	85.72	24.565055
SO_2	6.46	85.05	29.96	15.756565
NO_2	23.12	96.06	56.37	14.360888
CO	0.75	2.36	1.35	0.359708

AICc of models comparison: The higher R^2 of GWR models shows the improvement for OLS models, which means GWR model has advantage in exploring spatial instability. But it ca not prove whether this instability is from spatial data or accidental factors, so it is necessary to test the improvement of statistical significance. This study introduces AICc index to achieve this test. To comparison the performance of GWR models and OLS models, the AICc of 30 GWR models and OLS models is calculated and analysed (Table 3). The difference between GWR and OLS models is about 50, while commonly the difference value 3 is the optimal model selection standard. So in this study, the performance of GWR models is much better than OLS models. From this statistical view, the spatial relationship between air quality and land use type is more suitable GWR models.

Variation coefficient of models comparison: For GWR model, it takes full account of local feature of variations. For each location, it gives the coefficient of dependent and independent variations, which not only shows the independent variation relationship strength with dependent variation, but also reflects the relation type with dependent variation. When the coefficient is negative, it means the negative relationship between the independent and dependent variations. This coefficient means when all other independent keep constant, the dependent variation expected change caused by the corresponding independent unit change. For every GWR model built by air quality and land use type indicator, the coefficient is calculated in every location (Table 4). The coefficient by every pair of air quality and land use type indicator has large difference in space. For example, the relationship between PM_{10} and grass, it shows positive relationship in some place, while others are negative. However, the global model will hide the spatial heterogeneous. So for air quality and land use type indicators, it is necessary to build local models to consider spatial heterogeneous for the relationship analysis.

Moran's I index of models comparison: For models' expression of performance, accurate, spatial heterogeneity, GWR models show the great advantage. To explore model kills of solving spatial autocorrelation, this study statistics

Table 2: Comparison of R^2 of OLS and GWR Models.

		Forest	Grass	Farm	Water	Urban
$PM_{2.5}$	R^2_G	0.863889	0.910917	0.831725	0.882768	0.82309
	R^2_O	0.160456	0.302437	0.028576	0.150427	0.030902
PM_{10}	R^2_G	0.700571	0.912964	0.71027	0.81219	0.693513
	R^2_O	0.116319	0.254331	0.06114	0.195221	0.000583
O_3	R^2_G	0.454825	0.450106	0.328906	0.436509	0.407028
	R^2_O	0.007537	0.034784	0.0532	0.019901	0.060042
SO_2	R^2_G	0.607528	0.545976	0.641136	0.642367	0.637166
	R^2_O	0.009674	0.008184	0.097806	0.182753	0.189124
NO_2	R^2_G	0.425534	0.339949	0.356579	0.391076	0.436031
	R^2_O	0.063548	0.023435	0.00567	0.075649	0.026556
CO	R^2_G	0.650746	0.63284	0.603005	0.651778	0.601206
	R^2_O	0.023371	0.005375	0.02857	0.166264	0.041825

R^2_G is the R^2 of GWR models, R^2_O is R^2 of OLS models.

Table 3: Comparison of AICc of OLS and GWR Models.

		Forest	Grass	Farm	Water	Urban
$PM_{2.5}$	AIC_G	689.7276	676.3697	689.696	670.7183	690.5634
	AIC_O	784.1475	769.5114	795.6739	785.0856	795.4845
PM_{10}	AIC_G	817.0222	805.0993	811.7938	791.639	813.9898
	AIC_O	871.8851	858.4698	876.6701	864.4963	881.6081
O_3	AIC_G	716.7089	722.0154	720.6759	721.4459	718.9293
	AIC_O	733.4039	731.2047	729.6829	732.4136	729.11
SO_2	AIC_G	618.4523	626.8756	615.401	621.601	613.1843
	AIC_O	663.071	663.1897	655.7077	647.8956	647.2773
NO_2	AIC_G	636.0262	640.4247	642.3544	631.79	625.0238
	AIC_O	643.9977	647.3112	648.7354	642.9702	647.0583
CO	AIC_G	13.65005	29.35068	24.2536	19.16916	19.77713
	AIC_O	64.7748	66.21731	64.35312	52.27783	63.2678

AIC_G is AICc of GWR models, AIC_O is AICc of OLS models.

Table 4: Range of variable coefficient in GWR Models.

		Forest	Grass	Farm	Water	Urban
$PM_{2.5}$	Maximum	1120.641	474.1272	218.1134	1889.494	128.709
	Minimum	-1606.68	-1318.68	-171.078	-433.428	-20.75
PM_{10}	Maximum	1029.789	2409.997	164.8404	2229.161	146.5693
	Minimum	-930.46	-1090.53	-162.987	-520.812	-86.8257
O_3	Maximum	574.9709	150.0275	103.5986	1376.306	4.777453
	Minimum	-50.2183	-442.881	-7.67514	-245.287	-117.508
SO_2	Maximum	425.8102	40.54966	68.52988	219.124	152681
	Minimum	-27.9868	7.50402	-143.892	-420.875	-100.832
NO_2	Maximum	217.4729	79.45674	69.57501	532.4127	57.14408
	Minimum	-250.24	-165.692	-57.9322	3.594901	-22.6779
CO	Maximum	5.37195	3.733634	1.635608	9.575728	0.411506
	Minimum	-1.80054	-2.87208	-0.78615	-4.94244	-1.73752

Moran's I index for GWR and OLS models (Table 3). For Moran's I index, the value is more closed to 0, which means that the spatial index has more random distribution. In GWR and OLS models, the residual of GWR models has more strong spatial random distribution, while the residual in OLS models has more strong positive correlated distribution (high-high or low-low). So for air quality and land use type indicators, GWR models are more suitable to solving spatial autocorrelation, to ensure residual random distribution in space. Therefore, from the view of solving spatial autocorrelation, GWR models still show the great advantage, which to some degree also shows that air quality and

Table 5: Comparison Moran's I Index of OLS and GWR Models.

		Forest	Grass	Farm	Water	Urban
$PM_{2.5}$	I_G	-0.09	-0.15	0.28	-0.08	0.22
	I_O	0.74	0.74	0.81	0.67	0.86
PM_{10}	I_G	0.1	-0.17	0.15	-0.05	0.13
	I_O	0.71	0.68	0.71	0.51	0.76
O_3	I_G	-0.11	-0.14	0	-0.12	-0.11
	I_O	0.29	0.3	0.21	0.23	0.19
SO_2	I_G	-0.29	-0.12	-0.32	-0.35	-0.32
	I_O	0.37	0.35	0.23	0.18	0.1
NO_2	I_G	0.03	0.11	0.06	0.16	0.01
	I_O	0.49	0.43	0.46	0.42	0.48
CO	I_G	-0.22	-0.19	-0.02	-0.19	-0.04
	I_O	0.78	0.75	0.7	0.71	0.68

I_G is Moran's I index of GWR modelsÿI_O is Moran's I index of OLS models.

land use type indicators have more local properties, and it is not reasonable to build global model to analysis.

CONCLUSION

This study applied GWR and OLS models for 6 air quality indicators and 5 land use type indicators, based on exploratory spatial data analysis (ESDA) technique. Although these two model types can test spatial correlation to some degree, GWR models show the great advantage on the spatial relationship between air quality and land use type indicators from a set of indexes.

1. For R^2, GWR models are all more than 0.6, up to 0.91, while the OLS models are all less than 0.3, most of them are 0.1, even more lower than 0.1. It shows that GWR models are better at variation expression than OLS models.
2. To test statistical significance of models, this study compared the AICc index of models, which shows AICc index of GWR model is all less than OLS models. And most of their difference is about 50, usually the difference 3 as the optimization models standard. It shows that the GWR model has better performance than OLS models.
3. GWR models not only provide global R^2 to show the variations relationship expression, but also give the local parameter for every location, which estimates the spatial relationship between different variations. It shows that most of variations have different relationship intension, and even have negative and positive relationship in different location. So it proves that the local models are more suitable for the spatial relationship between air quality and land use type indicators.
4. To test model skill of solving spatial autocorrelation, it statistics Moran's I for the models residual, the GWR models of which are closed to 0, while the OLS model are far from 0, more closed to 1. It means that the residual of GWR models has more strong random distribution, in which the models are true to show the result and OLS models have more positive correlated distribution, meaning that the result may be led by some positive outside situation, not the factors random showing. So the GWR models can be a good deal with the spatial autocorrelation.

SUMMARY

This paper attempts to introduce ESDA technique on regional air pollution research. To avoid the potential multicollinearity of indicators, considering the GWR model limit on multicollinearity, it applies one air quality indicator and one land use type indicator to build the model, which to some degree can show the effect of one land use type to every air pollutant.

1. Through comparison, the GWR models has great advantage on exploring spatial relationship between air quality and land use type indicators, which means that these indicators have more strong local relationship. However, the existing studies are mostly based on global models.
2. The local parameter for every location provided by GWR models can reflect sufficiently the air quality control should act according to circumstances. The air compositions are strong related to local environment, so the global models are not reasonable to control pollutant.
3. Meanwhile, based on GWR models, the local parameters can be used for regional real-time and effective prediction.
4. While, this paper just introduce air quality and land use type indicators, which may affect the models accurate. In future study, it can introduce other correlation factors

(temperature, relative humidity, atmospheric profiles, etc.) based on the mechanism of air pollution, to improve model accurate and solving spatial heterogeneity, which can provide scientific solution for air pollution control and air quality effective prediction.

REFERENCES

Fotheringham, A.S., Charlton, M.E. and Brunsdon, C. 2001. Spatial variations in school performance: a local analysis using geographically weighted regression. Geogr Environ Model, 5: 43-66.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. 2002. Geographically weighted regression: the analysis of spatially varying relationships. Chichester, Wiley, 269 pp.

Yuan, F. 2008. Land-cover change and environmental impact analysis in the Greater Mankato area of Minnesota using remote sensing and GIS modeling. International Journal of Remote Sensing, 29(4).

Guo, L., Ma, Z. and Zhang, L. 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study. Canadian Journal of Forest Research, 38: 2526-2534.

Jun Tu and Zongguo Xia 2008. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. Science of the Total Environment, 407: 358-378.

Jun Tu 2011. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. Applied Geography, (31): 376-392.

Ma, Z., Hu, X., Huang, L., Bi, J. and Liu, Y. 2014. Estimating ground-level PM2.5 in China using satellite remote sensing. Environ. Sci. Technol., 48: 7436-7444.

Mehaffey, M.H., Nash, M.S., Wade, T.G., Ebert, D.W., Jones, K.B. and Rager, A. 2005. Linking land use cover and water quality in New York City's water supply watersheds. Environmental Monitoring and Assessment, 107: 29-44.

Schoonover, J.E., Lockaby, B.G. and Pan, S. 2005. Changes in chemical and physical properties of stream water across an urban erural gradient in western Georgia. Urban Ecosystems, 8: 107-124.

Sliva, L. and Williams, D.D. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Research, 35: 3462-3472.

Stutter, M.I., Langan, S.J. and Demars, B.O.L. 2007. River sediments provide a link between catchment pressures and ecological status in a mixed land use Scottish River system. Water Research, 41: 2803-2815.

Woli, K.P., Nagumo, T., Kuramochi, K. and Hatano, R. 2004. Evaluating river water quality through land use analysis and N budget approaches in liverstock farming areas. Science of the Total Environment, 329: 61-74.

Yaolin Liu, Long Guo, Qinghu Jiang, Haitao Zhang and Yiyun Chen 2015. Comparing geospatial techniques to predict SOC stocks. Soil and Tillage Research, 148: 46-58.

Zhang, L., Bi, H., Cheng, P. and Davis, C.J. 2004. Modeling spatial variation in tree diameter-height relationships. For Ecol. Manag., 189: 317-29.

Zheng, Y., Zhang, Q., Liu, Y., Geng, G. and He, K., 2016. Estimating ground-level PM 2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmospheric Environment, 124: 232-242.