Supplementary Materials: Optimization of $PM_{2.5}$ Estimation Using Landscape Pattern Information and Land Use Regression Model in Zhejiang, China

Shan Yang [1,2], Haitian Wu [1,2], Jian Chen [1,2,*], Xintao Lin [1,2] and Ting Lu [1,2]

1. Data Pretreatment prior to Ordinary Kriging for LUR-based Data-mining

The median and mean of the $PM_{2.5}$ concentration shown in the histogram were nearly equal (Figure S1). In the normal QQ diagram, the data fell on the line of the normal distribution (Figure S2). These results indicated that the predicted $PM_{2.5}$ concentrations obeyed a normal distribution. It was found that when performing a trend analysis from south to north, there was a gradual increase in the $PM_{2.5}$ concentration, while going from east to west, the $PM_{2.5}$ concentration gradually increased at first and then decreased (Figure S3). Therefore, these trends can be eliminated under normal conditions. After removing the trends, the remaining model-construction data or short-range variation components were subjected to statistical analysis. Right before the construction of the last surface, the trends will be automatically added back to generate the final spatial distribution of the $PM_{2.5}$ concentration. Through a semivariable function, the spatial autocorrelation between the sample points can be examined (Figure S4). Compared with the east-west direction, the south-north direction had a higher semivariable value and was parallel to the coastline. Thus, the heterogeneity should be considered when simulating the concentration of $PM_{2.5}$.

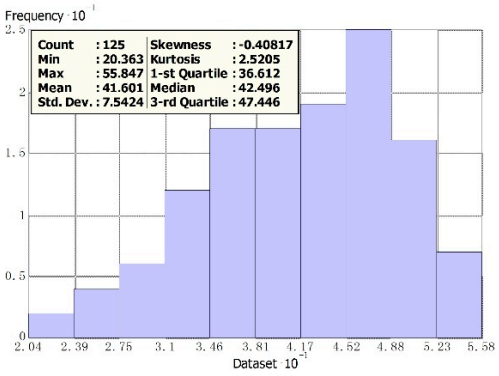


Figure S1. Histogram.

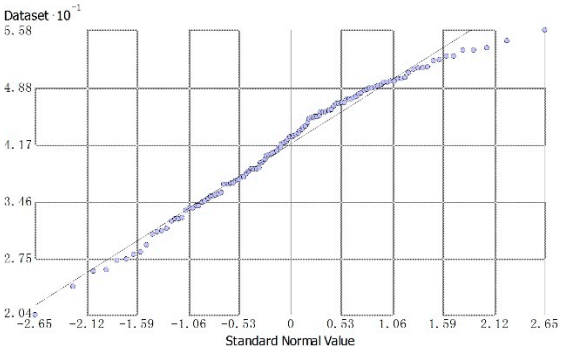
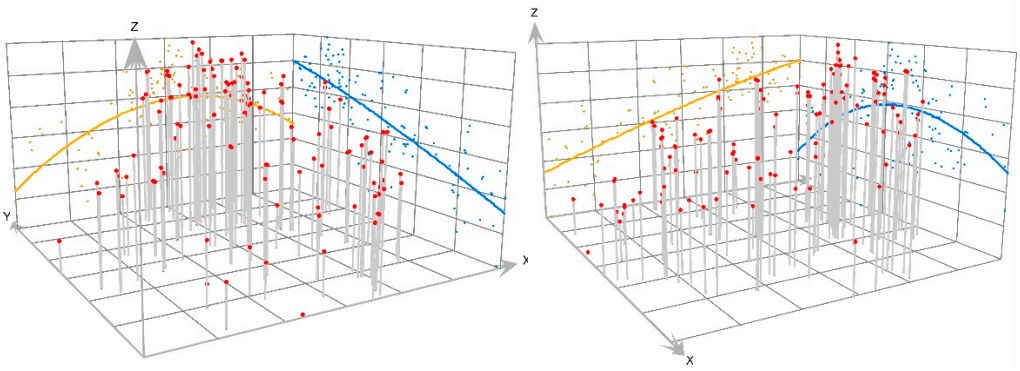
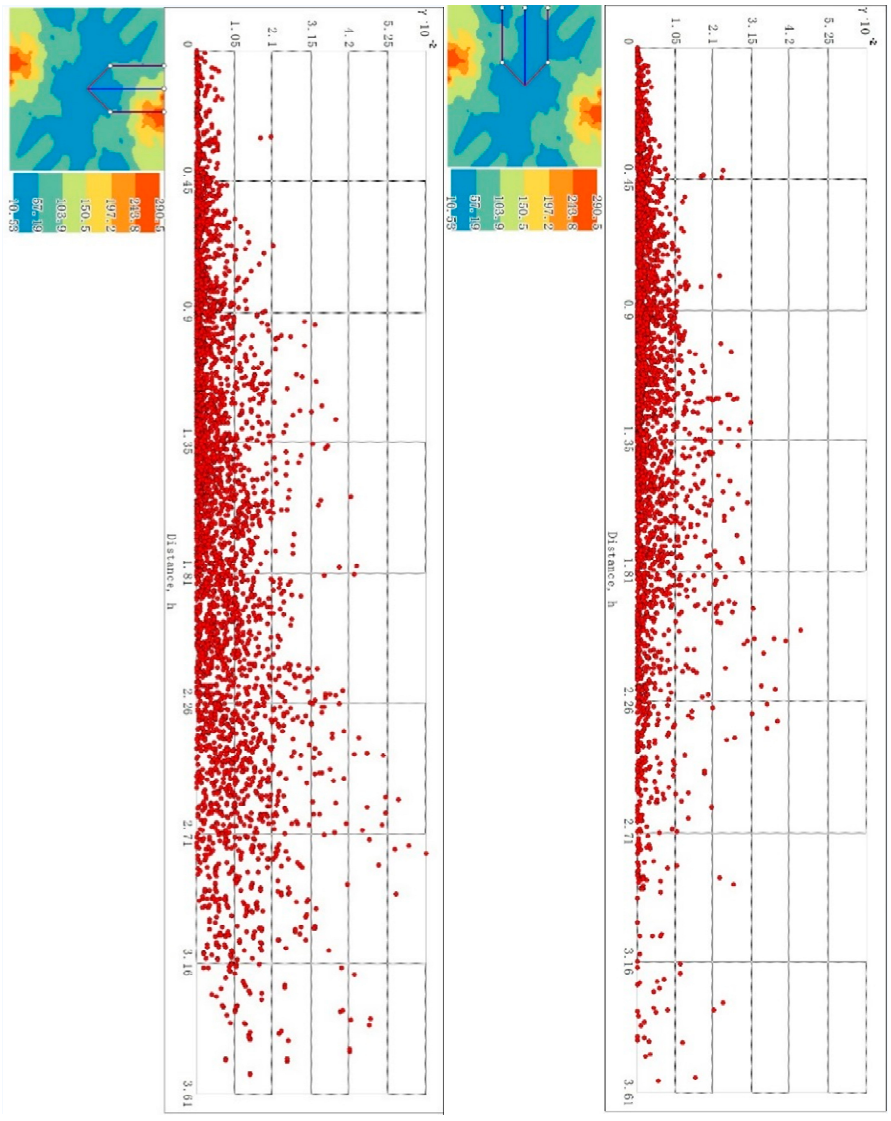


Figure S2. Normal QQ diagram.



(a) **(b)**

Figure S3. Trend analysis: **(a)** rotational angle of 0°; **(b)** rotational angle of 90°.

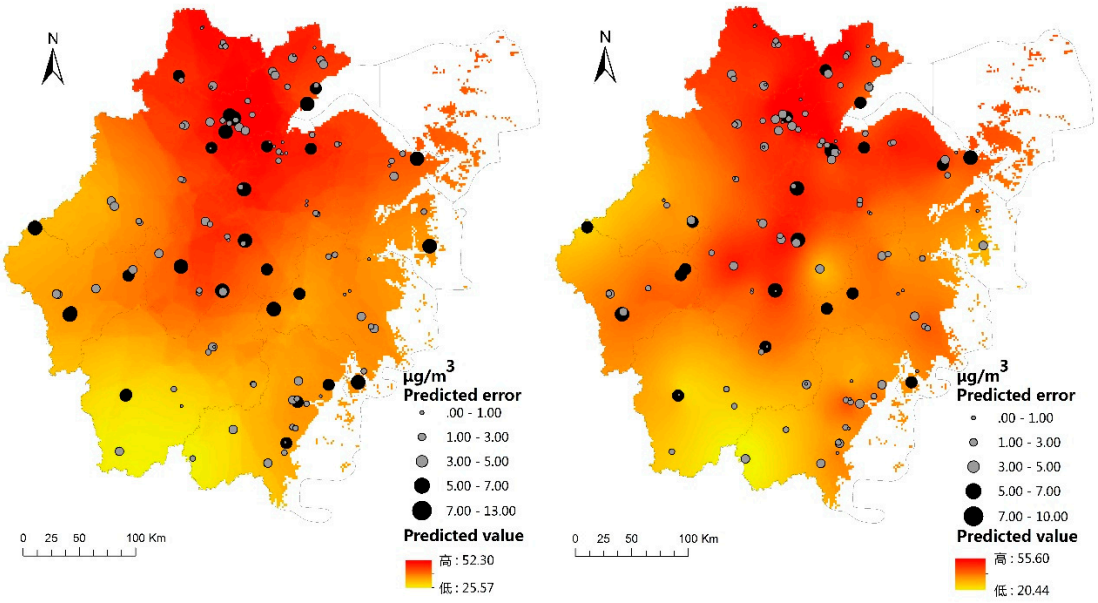


(a) **(b)**

Figure S4. Semi-variogram: **(a)** south-north direction; **(b)** east-west direction.

2. Exploring the Method for the Application of the Geographic Information System (GIS)-based LUR Model including landscape pattern factors

Zou et al. verified that the LUR model was more accurate and showed a higher level of detail and smoother variations than the ordinary kriging of measured values from the monitoring stations[1]. To compare the prediction accuracy of the different LUR-model application methods, the values measured at the stations were utilized to verify the predicted values obtained by the different methods. When directly applying the LUR model to data from the monitoring stations, the RMSE was 3.94 μg/m^3, which was lower than that when applying the model to surface fitting. Additionally, the R^2 obtained when applying the LUR model to point fitting was 0.768, which was higher than the value of 0.74 obtained when it was applied to surface fitting. The number of stations with prediction errors greater than 5 μg/m^3 was far greater when applying the model to surface fitting than when applying it to point fitting (Figure 1). Additionally, the maximum error obtained when applying the model to surface fitting was 2.82 μg/m^3 greater than that when applying it to point fitting (Table 1). Hence, the simulation accuracy was greater when directly applying the land use model to point fitting.



(a) **(b)**

Figure S5. Predicted pollutant concentration and error distribution resulting from different application mechanisms: **(a)** applying model to surface fitting; **(b)** applying model to point fitting.

Table S1. Fitting parameters of different application mechanisms.

	applying model to surface fitting	applying model to point fitting
R^2	0.736	0.768
RMSE	4.280	3.940
Fitting curve	y = 0.681x + 13.996	y = 0.737x + 11.336