# Appendix

**Table A1.** Descriptive statistics of the independent variables (median value) stratified by urban population size.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CL | | | | | WB | | | | |
| City group | Large | Medium | Small | Other | All | Large | Medium | Small | Other | All |
| PLAND | 45.02 | 44.04 | 29.22 | 35.75 | 41.81 | 4.90 | 3.46 | 2.59 | 1.94 | 3.57 |
| LPI | 18.68 | 21.16 | 9.72 | 16.17 | 17.50 | 2.97 | 2.14 | 2.12 | 1.58 | 2.21 |
| CIRCLE\_AM | 99.74 | 99.66 | 98.98 | 99.40 | 99.59 | 98.80 | 98.50 | 98.18 | 97.40 | 98.50 |
| NLSI | 0.043 | 0.048 | 0.054 | 0.034 | 0.044 | 0.055 | 0.071 | 0.084 | 0.057 | 0.064 |
| ENN\_AM | 87.78 | 94.31 | 137.21 | 130.24 | 96.67 | 360.75 | 558.35 | 736.65 | 505.09 | 431.2 |
|  | BL | | | | | FT | | | | |
| City Group | Large | Medium | Small | Other | All | Large | Medium | Small | Other | All |
| PLAND | 33.66 | 25.19 | 19.80 | 19.24 | 25.94 | 7.05 | 11.62 | 18.61 | 3.82 | 8.78 |
| LPI | 17.25 | 14.49 | 12.93 | 12.43 | 15.00 | 1.33 | 3.04 | 4.62 | 0.94 | 1.95 |
| CIRCLE\_AM | 99.43 | 99.16 | 99.07 | 99.15 | 99.22 | 97.75 | 97.95 | 98.50 | 97.20 | 97.88 |
| NLSI | 0.035 | 0.040 | 0.040 | 0.030 | 0.037 | 0.054 | 0.051 | 0.056 | 0.048 | 0.052 |
| ENN\_AM | 108.55 | 125.85 | 147.99 | 191.34 | 123.2 | 273.04 | 195.34 | 154.85 | 405.95 | 233.5 |
|  | GL | | | | |  |  |  |  |  |
| City Group | Large | Medium | Small | Other | All |  |  |  |  |  |
| PLAND | 1.29 | 2.47 | 9.94 | 11.72 | 2.82 |  |  |  |  |  |
| LPI | 0.30 | 0.81 | 2.97 | 4.89 | 0.84 |  |  |  |  |  |
| CIRCLE\_AM | 96.68 | 96.85 | 97.98 | 98.44 | 97.20 |  |  |  |  |  |
| NLSI | 0.064 | 0.063 | 0.057 | 0.045 | 0.060 |  |  |  |  |  |
| ENN\_AM | 573.38 | 320.09 | 282.90 | 323.11 | 387.9 |  |  |  |  |  |

**Note A1. Production process of the adjusted natural environmental variables**

Hourly data from surface meteorological station measurements in China that reflected the surface meteorological data for each prefecture were acquired from the China Meteorological Data Service Center (CMDSC) for the same study period. The real-time data were quality controlled by inspection of permitted scope and gross error elimination. All the meteorological data consisted of eight measurements per day, and daily average meteorological data were then calculated using all the available measurements. The quarterly average and the annual mean were then calculated based on the daily average estimation. The elevation data were acquired from the “hole-filled” Shuttle Radar Topography Mission (SRTM) data (~30 m at the equator). In addition, by the use of ANUSPLIN 4.3 (Hutchinson 2004), which is a professional meteorological data interpolation package, and elevation-independent covariates, we interpolated the surface meteorological station data in the study period for each prefecture to obtain the temperature (Tem.) and precipitation (Pre.) variables.

Dilution rate (Dil.) (product of mixing layer height and wind speed averaged over the mixing layer) is a widely used meteorological parameter. We calculated this from National Aeronautics and Space Administration (GMAO/MERRA) hourly data.

The distance from coast factor (Coast) was obtained from the Natural Earth vector dataset (available at: http://www.naturalearthdata.com/), which contains the most common global natural surface elements. The coastal variable was considered to test our theory that coastal prefectures are more likely to have high air quality due to their mild and rainy climate, luxuriant vegetation cover, and superior dilution and diffusion conditions for PM2.5.

**Table A2.** The description of the natural environmental variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Abbreviations | Range | Unit | Description |
| Meteorological | Tem. | - | °C | Two-year average temperature |
| Pre. | >0 | 1,000 mm | Two-year average precipitation |
| Dil. | >0 | m2/s2 | Two-year harmonic average product of wind speed and mixing layer height |
| Topological | Coast | >0 | Km2 | The distance from urban area to coast |

**Note A2. Production process of the socio-economic factors**

Population density was calculated from the latest gridded population dataset for the year 2010, for which the relative error is 4.5–13.6% (Fu et al. 2014). As nighttime light data can be considered as a proxy for economic activity in China, we used a Visible Infrared Imaging Radiometer Suite (VIIRS) dataset ranging from June 2014 to May 2016. The high sensitivity of VIIRS (Welsch et al. 2001) data provides a means for detecting human activity at night, and VIIRS data have been extensively used for providing uniform, spatially explicit, continuous, and timely measurements of demographic- and economic-related activities (Shi et al. 2014). To monitor economic activity, we used the stray-light corrected radiance composite images (i.e., in “vcmcfg” mode), which give the monthly average brightness of 15-arc-second (~500 m at the equator) pixels in units of digital number. Prior to averaging, the data were filtered to exclude the data impacted by stray lights, lightning, lunar illumination, and cloud cover. Cloud cover was determined using the VIIRS Cloud Mask (VCM) product.

|  |  |  |
| --- | --- | --- |
| **Table A3.** Socio-economic factors used in the study. | | |
| Metric (abbreviation) | Equation | Description |
| Density of population (POPD) | POP/*A* | The average of the urban population per square kilometer |
| Density of economic activity (NTLD) | NTL/*A* | The average economic activity intensity by nighttime light data |
| * POP is the total size of the population in the urban area, and NTL is the summation of the record of VIIRS remote sensing data acquired during the study period in the urban area. | | |

**Note A3. Production process of the straw burning factor**

MODIS fire counts during the period from June 2014 to May 2016 were recorded. Fire locations were produced by the University of Maryland and provided by NASA FIRMS operated by NASA/GSFC/ESDIS, with funding provided by NASA/HQ (available online <https://earthdata.nasa.gov/active-fire-data#tab-content-6> ). For the study period, the radiative power of every fire point within the urban areas was summed to estimate the open biomass burning. As a major source of biomass burning in China, the straw burning in each urban area could be estimated by the ﬁre radiative power summation.

**Table A4.** Collinearity diagnostics of the multivariate models built in this study.

| **Other city** | |  |  | CL\_PLAND | CL\_CIRCLE\_AM | GL\_ENN\_AM | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Tolerance | 0.964 | 0.966 | 0.998 | |
| Variance inflation factor | | 1.037 | 1.036 | 1.002 | |
|  |  | Condition  index | Variance proportions | | | | |
| Dimension | Eigenvalue | (constant) | CL\_PLAND | CL\_CIRCLE\_AM | GL\_ENN\_AM | |
| 1 | 3.318 | 1.000 | 0.006 | 0.020 | 0.006 | 0.030 | |
| 2 | 0.456 | 2.696 | 0.005 | 0.071 | 0.009 | 0.914 | |
| 3 | 0.187 | 4.217 | 0.057 | 0.906 | 0.082 | 0.036 | |
| 4 | 0.039 | 9.206 | 0.932 | 0.003 | 0.903 | 0.019 | |
|  |  |  |  |  |  |  | |
| **Small city** | |  |  | CL\_PLAND | GL\_NLSI |  | |
|  | Tolerance | 0.976 | 0.976 |  | |
| Variance inflation factor | | 1.024 | 1.024 |  | |
|  |  | Condition  index | Variance proportions | | |  | |
| Dimension | Eigenvalue | (constant) | CL\_PLAND | GL\_NLSI |  | |
| 1 | 2.776 | 1.000 | 0.014 | 0.022 | 0.019 |  | |
| 2 | 0.151 | 4.290 | 0.005 | 0.686 | 0.463 |  | |
| 3 | 0.073 | 6.180 | 0.982 | 0.293 | 0.518 |  | |
|  |  |  |  |  |  |  | |
| **Medium city** | |  |  | CL\_PLAND | GL\_NLSI |  | |
|  | Tolerance | 0.963 | 0.963 |  | |
| Variance inflation factor | | 1.038 | 1.038 |  | |
|  |  | Condition  index |  | Variance proportions | | |  |
| Dimension | Eigenvalue | (constant) | CL\_PLAND | GL\_NLSI |  | |
| 1 | 2.722 | 1.000 | 0.010 | 0.016 | 0.027 |  | |
| 2 | 0.229 | 3.449 | 0.007 | 0.209 | 0.612 |  | |
| 3 | 0.050 | 7.412 | 0.983 | 0.775 | 0.361 |  | |
|  |  |  |  |  |  |  | |
| **Large city** | |  |  | CL\_LPI | GL\_NLSI | POPD | |
|  | Tolerance | 0.920 | 0.952 | 0.965 | |
| Variance inflation factor | | 1.087 | 1.051 | 1.037 | |
|  |  | Condition  index |  | Variance proportions | | | |
| Dimension | Eigenvalue | (constant) | CL\_LPI | GL\_NLSI | POPD | |
| 1 | 3.353 | 1.000 | 0.006 | 0.021 | 0.018 | 0.011 | |
| 2 | 0.403 | 2.884 | 0.000 | 0.585 | 0.151 | 0.018 | |
| 3 | 0.193 | 4.167 | 0.007 | 0.077 | 0.580 | 0.361 | |
| 4 | 0.050 | 8.176 | 0.986 | 0.317 | 0.251 | 0.609 | |

**References A.**

Hutchinson, MF. 2004. "ANUSPLIN Version 4.3. Centre for Resource and Environmental Studies, Australian National University."

Shi, Kaifang, Chang Huang, Bailang Yu, Bing Yin, Yixiu Huang, and Jianping Wu. 2014. "Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas." *Remote Sensing Letters* 5 (4):358-66. doi: 10.1080/2150704X.2014.905728.

North American Cartographic Information Society. 2017. "Natural Earth data version v2.0.0." In, edited by Natural Earth data version v2.0.0. http://www.naturalearthdata.com/downloads/.

Welsch, C., H. Swenson, S. A. Cota, F. DeLuccia, J. M. Haas, C. Schueler, R. M. Durham, J. E. Clement, and P. E. Ardanuy. 2001. VIIRS (Visible Infrared Imager Radiometer Suite): a next-generation operational environmental sensor for NPOESS. Paper presented at the IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), 2001.