Supplementary materials for

Synergistic and Trade-off Influences of Combined PM2.5-O3 Pollution in the Shenyang Metropolitan Area, China: A Comparative Land Use Regression Analysis

Tuo Shi*, Xuemei Yuan, Chunjiao Li and Fangyuan Li

College of Life Science, Shenyang Normal University, Shenyang 110034, China

* Correspondence: shi_tuo@synu.edu.cn (T.S.)

S1. Land use regression (LUR) model variables

1.1. Dependent variables

O_3 concentration data from 43 stations in 2020 were obtained from the Liaoning real-time air quality publishing system (http://sthj.ln.gov.cn, Table S1). The monitoring station data are obtained through continuous hourly and daily sampling. According to the regulations on monitoring pollutant concentration data in the Ambient Air Quality Standards (GB 3095-2012), the daily average concentration data should be recorded for at least 20 hours per day, and the monthly average concentration record should include at least 27 days of daily average concentration records. Based on this, invalid data were excluded, and linear interpolation was used to fill small data gaps (i.e., single day missing no more than 3 hours of data). If the missing data were more than 3 hours on a single day, the average daily variation method was used for interpolation, namely, it was replaced with the average value of the corresponding periods in the four days before and after the missing date. When there was a significant amount of incomplete data, they were replaced with the same period from the

previous year. Finally, the annual and seasonal average data (Winter: November to March; Spring: April to May; Summer: June to August; Autumn: September to October) were synthesized and used as dependent variables in the LUR modelling.

Table S1 List of coordinates of monitoring stations.

City (Station counts)	Station coordinate	
	Longitude	Latitude
Shenyang (11)	123.68	41.93
	123.51	41.74
	123.42	41.70
	123.59	41.84
	123.43	41.85
	123.41	41.77
	123.47	41.79
	123.40	41.80
	123.38	41.92
	123.24	41.73
	123.60	41.91
Anshan (7)	123.04	41.12
	123.01	41.09
	123.05	41.14
	122.96	41.10
	122.93	41.08
	123.02	41.08
	123.13	41.02
Fushun (5)	123.80	41.85
	123.92	41.88
	124.04	41.86
	123.92	41.86
	124.10	41.90
Shenfu (1)	123.71	41.84
Benxi (6)	123.83	41.31
	123.76	41.33
	123.72	41.31
	123.77	41.29
	123.80	41.27
	123.81	41.35
Fuxin (5)	121.61	42.00
	121.67	42.03
	121.67	42.07

	121.67	42.02
	121.65	42.05
Liaoyang (4)	123.20	41.20
	123.17	41.25
	123.24	41.26
	123.15	41.30
Tieling (4)	123.81	42.30
	123.86	42.29
	123.85	42.29
	123.72	42.22

1.2. Independent variables

We considered the availability of data and the potential influencing factors and finally selected nine categories of predictor variables (Table S2). Different buffer radii of air quality monitoring stations were used to extract road length, land use area proportion and landscape metrics. Most of the data, such as the digital elevation model (DEM), land use data, meteorological data, vegetation index, and atmospheric composition, were obtained from the Google Earth Engine (https://earthengine.google.com/) platform.

Table S2 Description and abbreviations of environmental variables.

Variable		Abbreviation	Unit
Geographic information			
1	Longitude	X	°E
2	Latitude	Y	°N
3	Elevation	DEM	m
Population			
4	Population count	pop_count	
Road data			
5	Total road length within buffer radius	road_n[a]	m
6	Distance to the nearest road	dist_road	m
7	Distance to the nearest road intersection	dist_intersection	m
Land use			
8	Tree cover ratio within buffer radius	Tree cover_n[a]	%
9	Shrubland cover ratio within buffer radius	Shrubland_n[a]	%
10	Grassland cover ratio within buffer radius	Grassland_n[a]	%
11	Cropland cover ratio within buffer radius	Cropland_n[a]	%
12	Built-up cover ratio within buffer radius	Built-up_n[a]	%

Variable		Abbreviation	Unit
13	Bare vegetation cover ratio within buffer radius	Bare vegetation_n[a]	%
14	Permanent water bodies cover ratio within buffer radius	Permanent water bodies_n[a]	%
15	Herbaceous wetland cover ratio within buffer radius	Herbaceous wetland_n[a]	%
Building height			
16	Average building height within buffer radius	BH_n[a]	m
Meteorological factor			
17	Wind speed	WS	m/s
18	Surface net solar radiation	SNSR	J/m^2
19	Air temperature of 2 m	TEMP	℃
20	Relative humidity	RH	%
21	Surface pressure	SP	Pa
22	Land surface temperature	LST	℃
Vegetation index			
23	Normalized difference vegetation index	NDVI	
Atmospheric composition			
24	Aerosol optical depth	AOD	
25	Total atmospheric column of ozone	O_3_C	mol/m^2
26	Tropospheric formaldehyde column number density	HCHO_TC	mol/m^2
27	Tropospheric nitrogen dioxide column number density	NO_2_TC	mol/m^2
Landscape metrics			
28	Aggregation index within buffer radius	AI_n[a]	%
29	Interspersion and Juxtaposition index within buffer radius	IJI_n[a]	%
30	Splitting index within buffer radius	SPLIT_n[a]	%
31	Mean fractal dimension index within buffer radius	FRAC_MN_n[a]	
32	Perimeter-area fractal dimension within buffer radius	PAFRAC_n[a]	
33	Large patch index within buffer radius	LPI_n[a]	
34	Shannon's diversity index within buffer radius	SHDI_n[a]	

[a]**n= 50, 100, 200, 500, 1000, 2000, 5000**

The sources and processing methods of different types of variables were as follows:

(1) Location attribute. Location attributes include longitude (X), latitude (Y) and elevation. Elevation data were obtained from the DEM of the Space Shuttle Radar Topographic Mapper SRTM V4, with 90 m resolution.

(2) Land use data. The data were selected from the European Space Agency

Worldcover 2020 10 m product. The research area contained 8 land use types (Figure 1).

(3) Demographic data. Demographics were obtained from the LandScan data set provided by Oak Ridge National Laboratory, with 30 arc-seconds resolution (https://landscan.ornl.gov/).

(4) Road data. Road vector data were obtained from Open Street Map (https://www.openstreetmap.org/). Road density was defined as the total length of the roadways within the buffer zone. There were many different road types in the original data, and this research selected the length of motor vehicle roads for analysis. The distance between the monitoring station and the nearest road and intersection was calculated using the road vector in ArcGIS 10.8.

(5) Building height data. The building height raster data came from the data set CNBH-10m (https://zenodo.org/records/7923866) of China building height information, with 10 m resolution.

(6) Meteorological data. Daily meteorological data from ERA5 were from the ECMWF 5th generation global climate atmospheric reanalysis. The data include the 10 m wind eastward (u) and northward (v) components, 2 m dew point temperature, 2 m air temperature (TEMP), surface pressure (SP), and surface net solar radiation (SNSR) with a spatial resolution of 0.25°×0.25°. Wind speed (WS) is calculated from its u and v component, and relative humidity is estimated according to the empirical rules for dew point temperature and air temperature. The daily summary data were further averaged on a seasonal and

annual basis. The land surface temperature (LST) data were from the MOD11A2V6.1 product, which provides a surface temperature product with 1-km grid resolution.

(7) Vegetation index. The normalized differential vegetation index (NDVI) was obtained from MOD13A2V6.1 product which provided a vegetation index at 1-km resolution every 16 days.

(8) Atmospheric composition data. The 1-km multi-angle atmospheric correction (MAIAC) aerosol optical depth (AOD), which was obtained by Resolution Imaging Spectroradiometer (MODIS) aboard the Aqua and Terra satellites NASA, was regarded as an index of local aerosol conditions. The 1-km resolution images of ozone column density (O_3_C), tropospheric nitrogen dioxide (NO_2_TC) and formaldehyde column density (HCHO_TC) were obtained from the Sentinel-5P satellite tropospheric monitor.

(9) Landscape metrics. The research selected seven landscape metrics affecting air quality and indices considering aspects such as the degree of polymerization, shape, diversity, area and edge. These landscape metrics were calculated based on 10 m land use data and using the 'landscapemetrics' package (https://cran.r-project.org/web/packages/landscapemetrics/index.html) in R language.

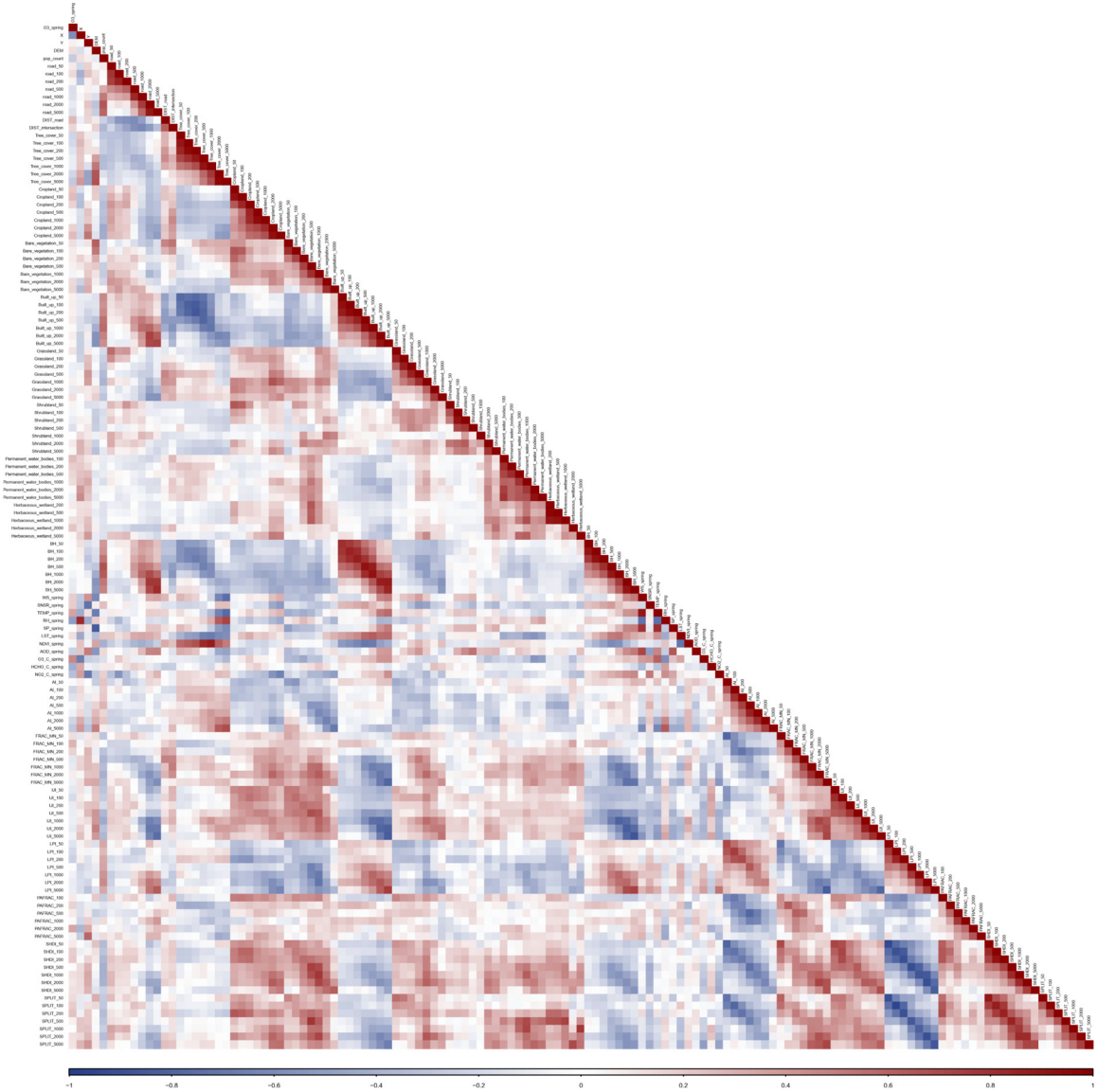


Figure S1. Correlation matrix of O_3 concentrations with independent variables in the spring land use regression (LUR) model, with grid colors indicating correlation coefficients.

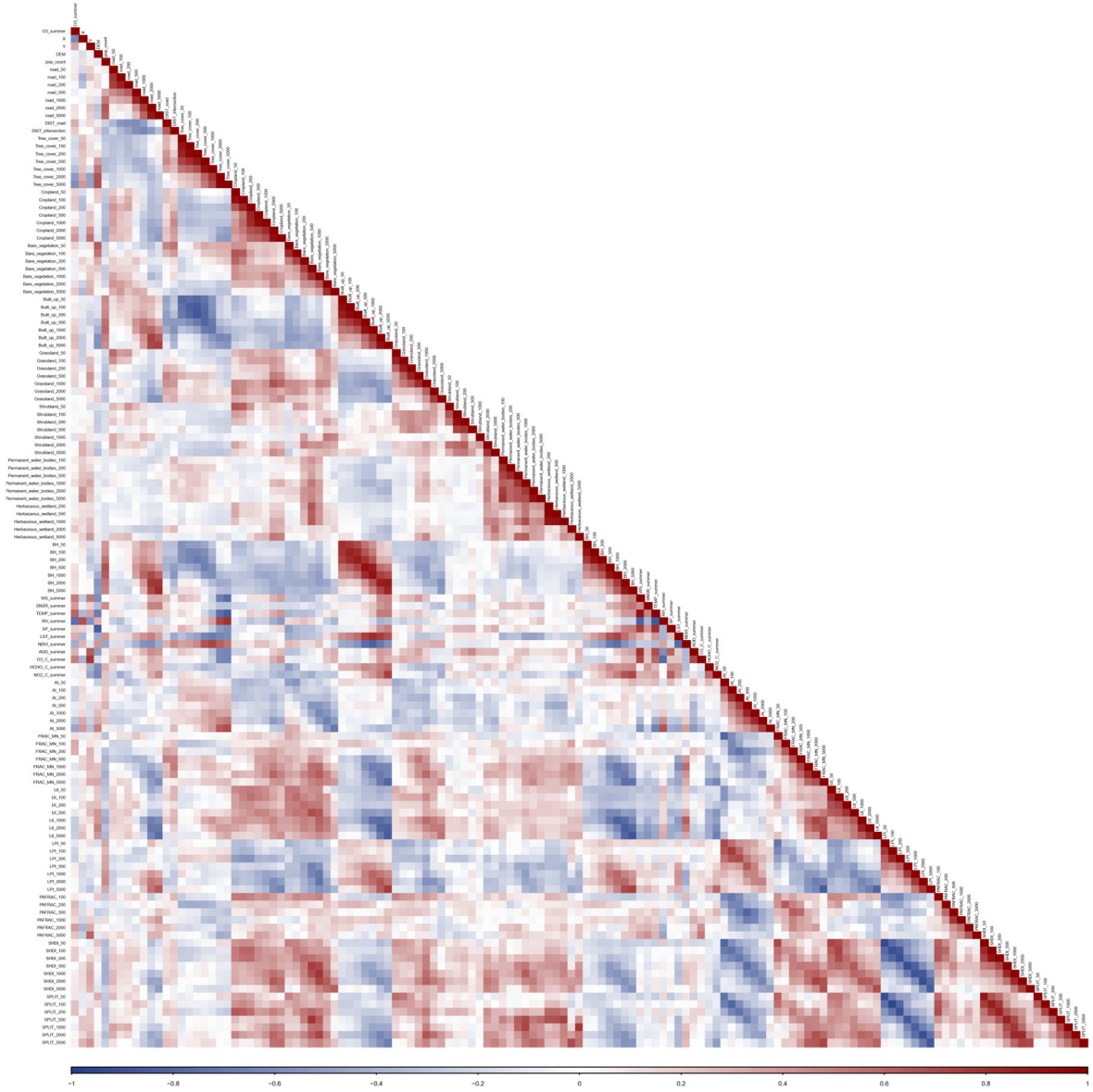


Figure S2. Correlation matrix of O_3 concentrations with independent variables in the summer land use regression (LUR) model, with grid colors indicating correlation coefficients.

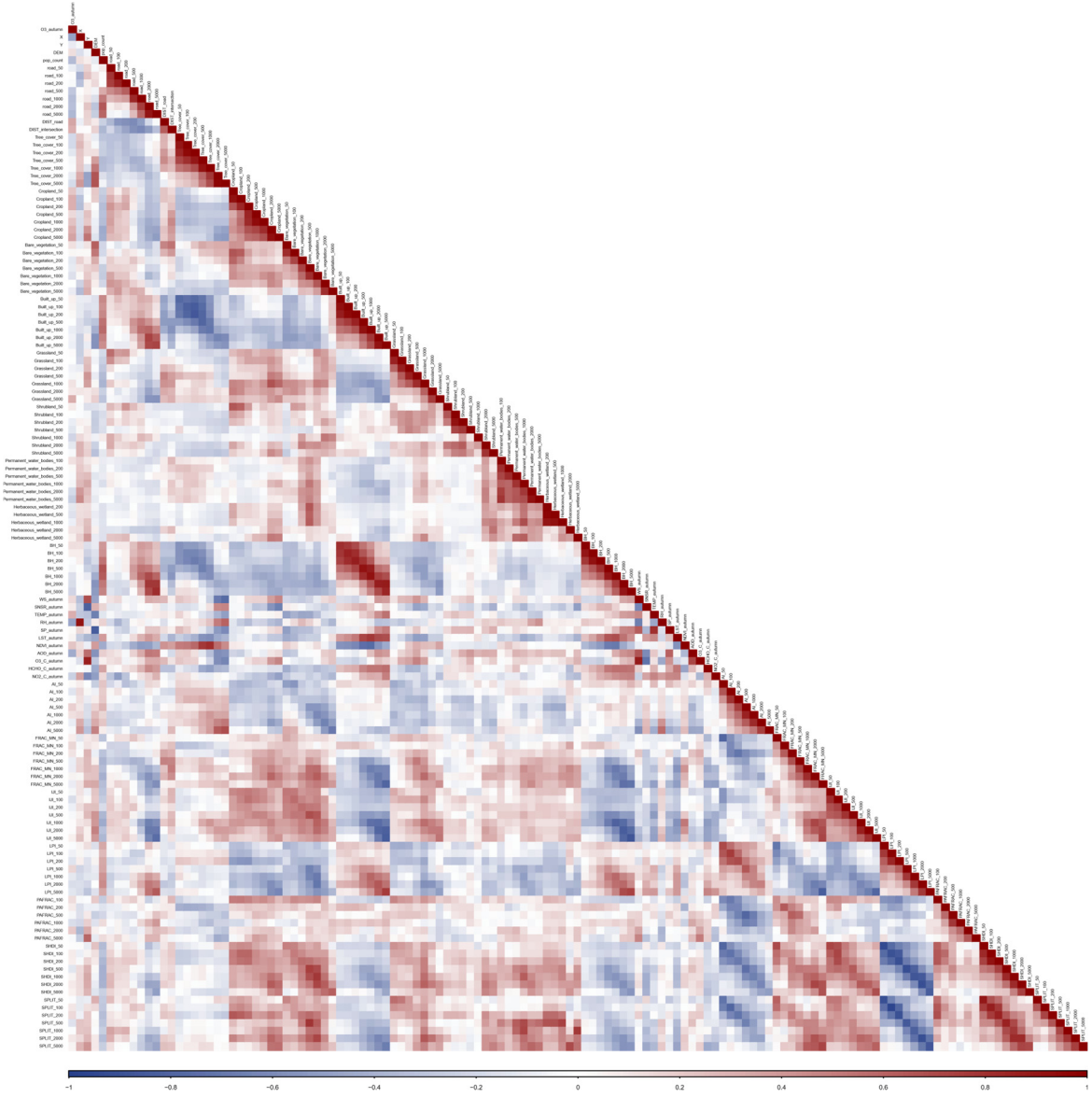


Figure S3. Correlation matrix of O_3 concentrations with independent variables in the autumn land use regression (LUR) model, with grid colors indicating correlation coefficients.

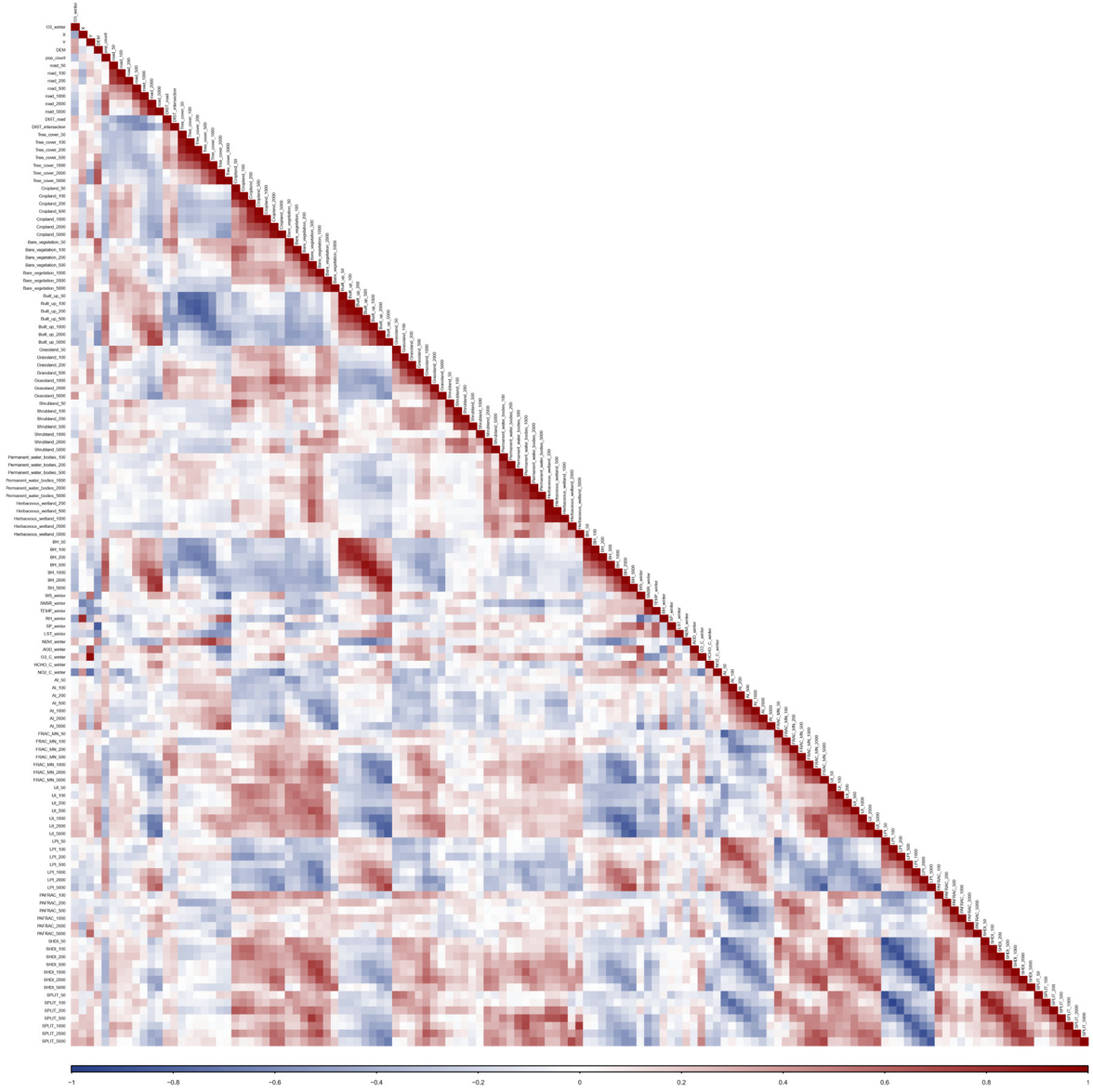


Figure S4. Correlation matrix of O_3 concentrations with independent variables in the winter land use regression (LUR) model, with grid colors indicating correlation coefficients.

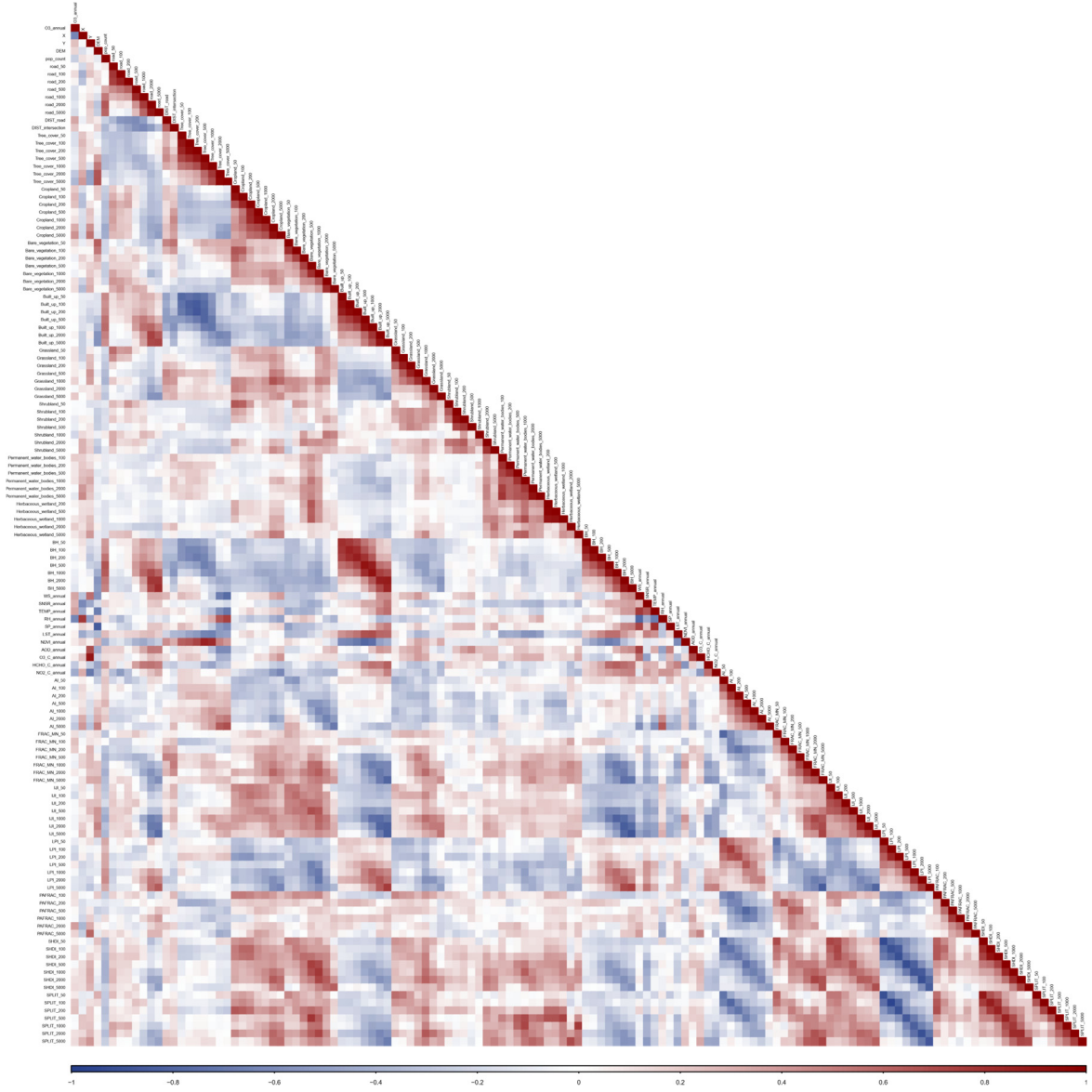


Figure S5. Correlation matrix of O_3 concentrations with independent variables in the annual land use regression (LUR) model, with grid colors indicating correlation coefficients.