Concurrent Spatiotemporal Daily Land Use Regression Modeling and Missing Data Imputation of Fine Particulate Matter Using Distributed Space Time Expectation Maximization

Seyed Mahmood Taghavi-Shahri^{a,b}, Alessandro Fassò^c, Behzad Mahaki^{a,1,*}, Heresh Amini^d

^a Department of Epidemiology and Biostatistics, School of Health, Isfahan University of Medical Sciences, Isfahan, 81746, Iran

^b Student Research Committee, School of Health, Isfahan University of Medical Sciences, Isfahan, 81746, Iran

^c Department of Management Information and Production Engineering, University of Bergamo, Dalmine, 24044, Italy

^d Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, 02215, United States

* Corresponding author at: Department of Epidemiology and Biostatistics, School of Health, Isfahan University of Medical Sciences, Hezar-Jerib Ave., Isfahan, IR. Iran; Postal Code: 8174673461; Fax: +983137923232; Tel: +989128077960

E-mail addresses: taghavi_m@razi.tums.ac.ir (S.M. Taghavi-Shahri), alessandro.fasso@unibg.it (A. Fassò), behzad.mahaki@kums.ac.ir (B. Mahaki), heresh@hsph.harvard.edu (H. Amini)

¹ Present address: Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Shahid Beheshti Blvd., Kermanshah, IR. Iran; Postal Code: 6715847141

Highlights

- First Land Use Regression using D-STEM, a recently introduced statistical software
- Assess D-STEM in spatiotemporal modeling, mapping, and missing data imputation
- Estimate high resolution (20x20 m) daily maps for exposure assessment in a megacity
- Provide both short- and long-term exposure assessment for epidemiological studies

Graphical Abstract

Abstract

Land use regression (LUR) has been widely applied in epidemiologic research for exposure assessment. In this study, for the first time, we aimed to develop a spatiotemporal LUR model using Distributed Space Time Expectation Maximization (D-STEM). This spatiotemporal LUR model examined with daily particulate matter $\leq 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$) within the megacity of Tehran, capital of Iran. Moreover, D-STEM missing data imputation was compared with mean substitution in each monitoring station, as it is equivalent to ignoring of missing data, which is common in LUR studies that employ regulatory monitoring stations' data. The amount of missing data was 28% of the total number of observations, in Tehran in 2015. The annual mean of $\text{PM}_{2.5}$ concentrations was $33 \mu\text{g}/\text{m}^3$. Spatiotemporal R-squared of the D-STEM final daily LUR model was 78%, and leave-one-out cross-validation (LOOCV) R-squared was 66%. Spatial R-squared and LOOCV R-squared were 89% and 72%, respectively. Temporal R-squared and LOOCV R-squared were 99.5% and 99.3%, respectively. Mean absolute error decreased 26% in imputation of missing data by using the D-STEM final LUR model instead of mean substitution. This study reveals competence of the D-STEM software in spatiotemporal missing data imputation, estimation of temporal trend, and mapping of small scale (20 x 20 meters) within-city spatial variations, in the LUR context. The estimated $\text{PM}_{2.5}$ concentrations maps could be used in future studies on short- and/or long-term health effects. Overall, we suggest using D-STEM capabilities in increasing LUR studies that employ data of regulatory network monitoring stations.

Keywords

Air pollution; D-STEM; LUR; Missing data; $\text{PM}_{2.5}$

1. Introduction

Air pollution caused 6.5 million deaths and 167 million disability-adjusted life-years (DALYs) in 2015 worldwide. In particular, ambient particulate matter with an aerodynamic diameter smaller than 2.5 μm ($\text{PM}_{2.5}$), caused 4.2 million of deaths and 103 million DALYs worldwide in the same year (Forouzanfar et al., 2016).

Air pollution epidemiological studies need accurate exposure assessment and one frequently used method for individual exposure assessment is land use regression (LUR) modeling. LUR uses geographic independent variables (typically extracted using geographic information system (GIS)) as Potentially Predictive Variables (PPVs) to predict spatial variation of air pollution at unmonitored locations (Amini et al., 2017c; Hoek et al., 2008; Ryan and LeMasters, 2007).

Many LUR studies that based on data of regulatory network monitoring stations, use annual or seasonal mean of air pollutants with a subset of GIS generated PPVs in a multiple regression for spatial modeling and mapping (Huang et al., 2017; Kashima et al., 2018; Zou et al., 2015) that ignore temporal trend of air pollution. However, some other studies use relatively advanced models, such as linear mixed effect models, with satellite, meteorological, and land use predictors for spatiotemporal estimation of air pollution concentrations (Just et al., 2015; Kloog et al., 2011; Meng et al., 2016). However, the common issue is treating of air pollution missing data, and perhaps limited capability in accounting for temporal and spatial autocorrelations.

D-STEM (Distributed Space Time Expectation Maximization) is a comprehensive statistical software for analyzing and mapping of environmental data that was introduced in 2013, which is programmed and run in Matlab environment (The MathWorks). The underlying D-STEM general model accounts for spatial and temporal autocorrelations and measurement errors, spatial and temporal and spatiotemporal predictors, fixed and random effect coefficients, concurrent missing imputation and

model fitting using Expectation Maximization (EM) algorithm. In addition, D-STEM supports multivariate modeling even in heterogenous stations (for example, simultaneous modeling of meteorological variables and air pollutants measured in the same or different stations), and data fusion capability for joint modeling (and calibration) of remote sensing data measured by a satellite along with air pollution concentrations measured in the ground stations (Finazzi, 2013; Finazzi and Fasso, 2014).

D-STEM software has been previously used in several air pollution studies in Europe (Calculi et al., 2015; Fassò, 2013; Fassò et al., 2016; Finazzi, 2013; Finazzi et al., 2013), but primary focus of those studies have been on statistical theory. Moreover, the resolution of their output maps has been about 1 km, which might be course for epidemiological studies that need individual-level exposure estimates. In addition, none of those studies used a pool of GIS generated PPVs for extensive independent variables selection. Overall, D-STEM has not been used so far for LUR modeling and for exposure estimation with very fine spatial resolution.

To date, several LUR studies have been published in Tehran to predict air pollution. These have been for particulate matter with an aerodynamic diameter smaller than 10 μm (PM_{10}) and sulfur dioxide (SO_2) (Amini et al., 2014), nitrogen oxides (NO , NO_2 , and NO_x) (Amini et al., 2016), $\text{PM}_{2.5}$, carbon monoxide (CO), and nitrogen dioxide (NO_2) (Hassanpour Matikolaei et al., 2017), and alkylbenzenes (benzene, toluene, ethylbenzene, *m*-xylene, *p*-xylene, and *o*-xylene (BTEX), and total BTEX) (Amini et al., 2017b). In fact, multiple regression has been used in majority of these studies for modeling of annual or seasonal averages of air pollutants. However, the study of $\text{PM}_{2.5}$, CO , and NO_2 (Hassanpour Matikolaei et al., 2017), used panel regression to predict hourly concentrations with limited capability in accounting for spatial and temporal autocorrelations, and not supporting spatiotemporal missing imputation (Hassanpour Matikolaei et al., 2017).

In the present study, we aimed to utilize D-STEM with a relatively large pool of GIS generated PPVs for producing high resolution daily estimations of airborne PM_{2.5} in the Middle Eastern megacity of Tehran, Iran. We further aimed to assess some characteristics of the recently introduced D-STEM software. In addition, we aimed to compare performance of air pollution missing data imputation under two scenarios: (1) using mean substitution in each monitoring station, and (2) using D-STEM concurrent spatiotemporal missing imputation and LUR modeling.

2. Materials and Methods

2.1 Study area and data

The study area of this research is the megacity of Tehran, Iran which has 9 million urban residents and, due to diurnal migration, a daytime population of more than 10 million. Alborz Mountains are located in north of the city and there is a desert in the south. The populated area covers approximately 613 km² and its elevation ranges from 1,000 to 1800 meters above sea level from south to north. The prevailing winds blow from west and north. The mean daily temperature ranges from about -15 °C in January to 43 °C in July with an annual average of 18.5 °C. The precipitation ranges from 0 millimeters (mm) in September to 40 mm in January, with an annual average of 150 mm. The weather of most days is sunny, and the mean cloud cover is about 30%. Geographic location of Tehran is in zone 39N of Universal Transverse Mercator (UTM) (Amini et al., 2016; Amini et al., 2014). Figure S1 of Supporting information shows the study area and location of monitoring stations.

Hourly data of fine particulate matter (PM_{2.5}), which have been measured using fixed beta-gauge monitors, were obtained from Tehran Air Quality Control Company (AQCC) and Iranian Department of Environment (DOE) for year 2015. Daily PM_{2.5} was calculated by 24-hours averaging of hourly measured

PM_{2.5} concentrations when days had complete records in all hours. Daily mean of PM_{2.5} for a set of 19 fixed measuring sites and for 365 days of year 2015 was used as response variable in this study.

A pool of 212 potentially predictive variables (PPVs) in six classes and 73 sub-classes were used for selecting model independent variables. These variables were in raster format for whole study area with resolution of 5×5 meters. The six GIS generated classes were Distance (60 variables), Product terms (52 variables), Land Use (50 variables), Traffic Surrogates (26 variables), Population Density (22 variables), and Geographic Location (2 variables). The full list of our PPVs could be found elsewhere (Amini et al., 2016; Amini et al., 2014).

2.2 Model Specification

D-STEM spatiotemporal LUR model for PM_{2.5} is given by the following equation where Y is response variable, which is spatiotemporally indexed at location s and time t, provided X's the GIS generated spatial predictors that are assumed temporally constant over time and spatially indexed at location s:

$$Y_{PM_{2.5}}(s, t) = \sum_{i=1}^p \{ \beta_i \times X_i(s) \} + \sum_{j=1}^q \{ \alpha_j \times \omega_j(s, t) \times X_j(s) \} + z(t) + \varepsilon(s, t)$$

While the left-hand side of the above equation defines the response variable, the right-hand side of this equation consists of four parts. First, sum of p summand: β_i multiplied by $X_i(s)$, which are a fixed effect coefficient and the corresponding GIS generated spatial predictor. Second, sum of q summand: α_j multiplied by $\omega_j(s, t)$ and multiplied by $X_j(s)$, which are a scale changing coefficient and the corresponding spatiotemporal random effect coefficient, and the corresponding GIS generated spatial predictor, respectively. Third part is $z(t)$ which is a latent temporal state variable that concurrently estimated using a sub-model. Forth part is $\varepsilon(s, t)$ which denotes measurement error that assumed normally distributed with zero mean and uncorrelated over space and time.

Latent spatiotemporal random effect coefficients $\omega_j(s, t)$ assumed to be independent of each other and normally distributed with zero mean and unit variance that uncorrelated over time but correlated over space with exponential spatial auto-covariance function $C_{\theta_j}(d) = \exp(\frac{-d}{\theta_j})$ where d is the distance between two generic spatial points and θ_j is auto-covariance parameter.

Since the latent spatiotemporal random effect coefficients $\omega_j(s, t)$ assumed to have zero means, we are not allowed to incorporate any GIS generated spatial predictor in random effect part, except if we already used it in fixed effect part. In the current representation, β could be interpreted as global effect of GIS generated predictor, and α could be used for testing spatiotemporal variation in the effect of corresponding predictor.

Finally, latent temporal state variable (third part of model) estimated concurrently using following sub-model:

$$z(t) = G \times z(t-1) + \eta(t)$$

Where G is coefficient of temporal state transition and η is temporal state innovation that assumed normally distributed with zero mean.

The estimated parameters of the D-STEM spatiotemporal LUR model are as follows: fixed effect coefficients, scaling changing coefficients of the corresponding spatiotemporal random effect coefficients, parameters of the underlying auto-covariance functions of the corresponding spatiotemporal random effect coefficients, temporal state transition coefficient, variance of the temporal state innovation, and variance of the measurement error.

2.3 Model fitting, assessing, and mapping

A forward algorithm, based on maximization of spatial Leave-One-Out Cross-Validation (LOOCV) R-squared and consistency with prior knowledge, was used to select suitable predictors from pool of PPVs. This algorithm considers sign consistency of fixed effect coefficients with a priori assumed direction of effects. Effect directions defined as positive, negative, or arbitrary (unknown) based on knowledge about emission of air pollutions.

The final daily LUR model was assessed in three forms: a) spatiotemporal, b) spatial (temporally averaged to assess spatial variations), and c) temporal (spatially averaged to assess temporal variations). Also, Normalized Mean Bias Factor (NMBF) was used for checking whether models underestimate or overestimate observed data (Yu et al., 2006).

The power of missing data imputation was compared between mean substitution in each monitoring station, and final spatiotemporal LUR model by 10 independent simulations. In each simulation 10% of spatiotemporal observed data was randomly selected and was replaced by missing data. Afterwards, both models were independently employed for missing imputation. The predicted values of the randomly missed data and original observed values (that randomly missed), was used to compute Mean Absolute Error (MAE) in missing imputation of each model. Finally, MAEs that computed in these independent simulations and percent of MAEs change between two models, was averaged and reported.

Mapping was done using D-STEM for each day, by providing GIS generated maps that correspond to predictors of the final daily LUR model. Seasonal maps were generated by averaging of daily output maps in the corresponding time periods. Finally, we applied a quantification limit (QL) that its lower bound was the minimum of within monitoring station averages of observed concentration in the corresponding time period divided by square root of two, and its upper bounds was 120% of the

maximum of within monitoring station averages of observed concentration in the same time period. Grid cells in the final output maps that had values outside QL interval were set to nearest QL bound and frequency of these corrections was calculated and reported (Amini et al., 2016; Amini et al., 2014; Henderson et al., 2007)

3. Results and Discussion

3.1 Description of data

Number of PM_{2.5} daily observations at 19 monitoring stations was 5016. Frequency of missing data varies from 12% to 47% in monitoring stations with an average of 28% over all stations. Frequency of available data for each day varied from 2 stations to 19 stations with an average of 14 stations. Data availability pattern of daily PM_{2.5} in monitoring stations of Tehran in 2015 is presented in Figure S2 of Supporting information.

Annual mean of observed PM_{2.5} concentrations (spatially and temporally average of monitoring stations) was 33 µg/m³ that is more than three times of World Health Organization (WHO) guideline (10 µg/m³) and about three times of US Environmental Protection Agency (EPA) guideline (12 µg/m³) (US EPA, 2013; WHO, 2006), with a range of 21 µg/m³ to 48 µg/m³ in different monitoring stations. This is in line with the findings of Brauer et al. (2016) and Shaddick et al. (2018) where they reported very high population-weighted exposure estimates for PM_{2.5} in the Middle East (Brauer et al., 2016; Shaddick et al., 2018). Faridi et al. (2018) also reported very similar annual mean for PM_{2.5} in Tehran (Faridi et al., 2018). Daily mean of Tehran PM_{2.5} concentrations in 2015 (spatially average of monitoring stations) was greater than WHO daily standard (25 µg/m³) in 246 days (67% of year), and was greater than US EPA daily standard (35 µg/m³) in 121 days (33% of year), and eventually was greater than 55 µg/m³ in 31 days. As a result, the air quality index (AQI) in Tehran in 2015 was unhealthy for at least sensitive groups in one third of

year, and was unhealthy for all groups in about one month (8.6% of year that exceeded 55 $\mu\text{g}/\text{m}^3$), based on US EPA breakpoints (US EPA, 2013; WHO, 2006). The very high values of air pollution have been confirmed in large Tehran Study of Exposure Prediction for Environmental Health Research (SEPEHR) measurement campaigns (Amini et al., 2017a).

3.2 Description of the LUR model

Natural logarithm of $\text{PM}_{2.5}$ was standardized (using mean and standard deviation of its all spatiotemporal values) and utilized in spatiotemporal LUR model as response variable. Also, standardized version of GIS generated PPVs was employed in the model. Summary statistics for these variables are presented in Table S1 of Supporting information.

Table 1 presents estimated parameters of spatiotemporal daily LUR model. This model accounts for both temporal and spatial autocorrelations, and moreover supports inclusion of predictors for concurrent missing imputation and LUR model fitting. Among 212 GIS generated predictors that produced a huge possible combination of predictors, eventually 8 predictors were selected by our described algorithm to include in the spatiotemporal final daily LUR model.

Table 1. Estimated parameters of spatiotemporal final daily LUR model for $\text{PM}_{2.5}$ in Tehran 2015.

Category	Term	Estimate	Std. Error	p-value
Fixed effect Coefficients	LNDIST to AMB	-0.170	0.009	< 0.001
	PD.1000	+0.177	0.012	< 0.001
	LNDIST to HZRFAC	+0.219	0.011	< 0.001
	ELEV	-0.282	0.012	< 0.001
	SEN.400	-0.119	0.008	< 0.001
	DIST to SCSC	+0.218	0.012	< 0.001

	GRS.500	-0.151	0.013	< 0.001
	BGD.500	+0.082	0.009	< 0.001
Scale changing	LNDIST to AMB	0.089	0.018	< 0.001
Coefficients of	SEN.400	0.037	0.019	0.05
Random effect	GRS.500	0.145	0.025	< 0.001
Spatial covariance	LNDIST to AMB	101.30	284.15	-
Parameters	SEN.400	50.78	476.57	-
	GRS.500	25.43	22.35	-
Temporal state Transition coefficient		+0.73	0.04	< 0.001
Variance of Temporal state innovation		0.29	0.02	< 0.001
Variance of Measurement Error		0.224	0.005	< 0.001
<p>The model response was standardized natural logarithm of PM_{2.5} concentrations</p> <p>Also, Predictors were standardized (see Table S1 of Supporting information).</p> <p>Abbreviations:</p> <p>LN PM_{2.5}: natural logarithm of particulate matter with diameter<=2.5 micrometer</p> <p>LNDIST to AMB: natural logarithm of distance to nearest ambulance service</p> <p>PD.1000: total population density in buffer 1000</p> <p>LNDIST to HZRFAC: natural logarithm of distance to nearest hazardous facility</p> <p>ELEV: elevation</p> <p>SEN.400: sensitive areas in buffer 400</p> <p>DIST to SCSC: distance to nearest secondary or high school</p> <p>GRS.500: green space area in buffer 500</p> <p>BGD.500: bridge length in buffer 500 divided by distance to the bridge</p>				

GIS generated predictors that had a positive coefficient in the final spatiotemporal daily LUR model were: total population density in buffer of 1000 m, natural logarithm of distance to nearest hazardous facility (with an arbitrary assumed direction of the effect), distance to the nearest secondary or high school, and bridge length in buffer of 500 m divided by distance to the bridges.

On the other hand, the predictors that had a negative coefficient in the final spatiotemporal daily LUR model were: natural logarithm of distance to nearest ambulance service (assumed sign was arbitrary), elevation (assumed sign was arbitrary), sensitive areas in buffer of 400 m (assumed sign was arbitrary), and green space areas in buffer of 500 m.

Random effect predictors in the final spatiotemporal daily LUR model were as follow: natural logarithm of distance to nearest ambulance service, sensitive areas in buffer of 400 m, and green space areas in buffer of 500 m. The corresponding scale changing coefficients were significant or borderline significant. That means coefficient of these three predictors vary in time and space, and observed data are spatially auto-correlated.

Temporal state transition coefficient was significant ($p < 0.001$) and positive that indicates a positive temporal autocorrelation existed between successive observations. Moreover, variance of the temporal state innovation was significant ($p < 0.001$) that means variations existed in temporal trend of air pollution.

3.3 Assessing fit of the LUR model

Figure 1 shows temporal pattern (spatially averaged) of daily $PM_{2.5}$ for observed and predicted data of the final daily LUR model. Figure 1 confirms that D-STEM succeeds in capturing natural turbulence of $PM_{2.5}$ time series, even without employing any temporal predictors.

Figure 1. Temporal pattern (spatially averaged) of daily $PM_{2.5}$ in Tehran 2015 (observed and predicted).

Spatiotemporal R-squared and spatiotemporal LOOCV R-squared were 78% and 66%, respectively. Spatial R-squared (temporally averaged) and spatial LOOCV R-squared were 89% and 72%, respectively. Temporal R-squared (spatially averaged) and temporal LOOCV R-squared were 99.5% and 99.3%, respectively. Root Mean Square Error (RMSE) criteria are shown in Table S2 of Supporting information for further assessment.

NMBF of the spatiotemporal final daily LUR model was -0.03 that means on average this model slightly underestimated observed data (with a factor of less than 3 percent). This slight underestimation also confirmed in Figure 1 and may be due to some large positive outliers that were not handled appropriately. Unfortunately, this criterion has not been reported in other studies that applied D-STEM (Calculi et al., 2015; Fassò, 2013; Fassò et al., 2016; Finazzi, 2013; Finazzi et al., 2013), so it is not clear

that this slight deficiency is belonging to D-STEM software or it is an accidental observation that happened in our study.

However, the temporal prediction of D-STEM LUR model was very well. In fact, temporal R-squared and temporal LOOCV R-squared were greater than 99%. In a study by Kloog et al. (Kloog et al., 2011) where they employed land use and satellite-derived Aerosol Optical Depth (AOD) and multiple stations (spatiotemporal) meteorological predictors in a mixed model, the temporal cross-validated R-squared was 84%. Note that we did not use any spatiotemporal or pure temporal predictors in the D-STEM software. A similar study in Tehran that modelled PM_{2.5} with a panel regression used pure temporal predictors, i.e., one station meteorological variables, just in order to partly capture temporal innovations of air pollution data (Hassanpour Matikolaie et al., 2017). Hence, the use of pure temporal predictors in D-STEM for spatiotemporal modeling of air pollution is completely unnecessary.

The spatial R-squared of our D-STEM final LUR model (and even spatial LOOCV R-squared) was greater than 65% that reported in a recently published study on LUR modelling of PM_{2.5} in Tehran. That study, that used a panel regression for spatiotemporal modeling of fixed monitoring air pollution data, provided GIS generated spatial data, meteorological data of one station, and real time google traffic data as predictors (Hassanpour Matikolaie et al., 2017). Therefore, despite of the fact that predictors of our D-STEM spatiotemporal LUR model were limited to GIS generated spatial data, D-STEM performed better.

3.4 Assessing missing data imputation

Mean Absolute error (MAE) in 10 independent simulations for spatiotemporal missing imputation, was 5.7 µg/m³ with a range of 5.3 µg/m³ to 6.2 µg/m³ in the D-STEM final spatiotemporal daily LUR model, and was 7.7 µg/m³ with a range of 7.1 µg/m³ to 8.8 µg/m³ in Mean substitution in each monitoring station.

Changes between MAEs showed a decreased range from 24.3% to 29.9% with an average of 25.9% in using D-STEM final spatiotemporal daily LUR model instead of Mean substitution in each monitoring station. Details of MAEs and changes between MAEs are presented in Table S3 of Supporting information.

This result is important because missing data exists in almost any air pollution study of fixed air monitoring stations, such as our two previous LUR studies of PM₁₀, SO₂, and nitrogen oxides in Tehran. In those studies we imputed air pollution missing data using EM algorithm in a multivariate way but without considering spatial and temporal autocorrelations (Amini et al., 2016; Amini et al., 2014). Although, some other LUR studies such as one earlier mentioned LUR study of PM_{2.5} in Tehran that modeled using a panel regression, did not describe the missing data issue (Hassanpour Matikolaie et al., 2017).

This study proved that D-STEM missing imputation is better than ignoring of missing data, which is common at least when the amount of missing data is small (Meng et al., 2016). We believe D-STEM is worthy to be applied in missing imputation of (fixed monitoring stations) air pollution data, because it imputes missing data concurrently with modeling, and we see that the underlying temporal model of D-STEM perform very well. Moreover, D-STEM can capture spatial variation of air pollution data by incorporating suitable land use and GIS generated predictors. Furthermore, D-STEM spatiotemporal model that concurrently impute missing data has ability to employ spatiotemporal meteorological predictors, and even remote sensing data.

3.5 Generating output maps

We produced daily PM_{2.5} concentrations maps with resolution of 20×20 meters for whole study area that cover 613 km². Cooler and warmer season maps of PM_{2.5} were produced by averaging of D-STEM generated daily maps in the corresponding periods.

Quantification limits (QL) that calculated based on averaged observed values of air quality monitoring stations in the corresponding periods, had a lower bound of 13.6 $\mu\text{g}/\text{m}^3$ in cooler season, and 15.2 $\mu\text{g}/\text{m}^3$ in warmer season; and had upper bound of 68.1 $\mu\text{g}/\text{m}^3$ in cooler season, and 50.1 $\mu\text{g}/\text{m}^3$ in warmer season.

Values beyond these quantification limits were corrected by setting to the nearest upper or lower limits. Percent of these corrections for lower and upper bounds of the corresponding QL limits were 0.03% and 1.9% in the cooler season map, and were 0.1% and 3.4% in warmer season map, respectively. Hence, the under- or over-estimation of $\text{PM}_{2.5}$ in the D-STEM LUR generated maps were relatively small when compared to other LUR studies (Amini et al., 2016; Amini et al., 2014).

The seasonal estimation maps of $\text{PM}_{2.5}$ concentrations in Tehran in 2015, with resolution of 20×20 meters, are presented in Figure 2. The scatter plot of predicted versus observed values are presented in Figure S3 of Supporting information for cooler and warmer seasons maps of $\text{PM}_{2.5}$ concentrations.

320
 321 Figure 2. Estimated cooler and warmer seasons maps of $PM_{2.5}$ in Tehran 2015 with resolution of 20×20
 322 meters that produced by averaging of daily LUR maps over corresponding periods.

323

324 4. Conclusions

325 This was first LUR study that used recently introduced comprehensive statistical software of D-STEM.

326 We used D-STEM in spatiotemporal missing imputation, modeling and mapping of airborne PM_{2.5} in

327 Middle Eastern megacity of Tehran, where concentrations of PM_{2.5} was more than three times of WHO

328 guideline. The estimated maps using D-STEM daily PM_{2.5} LUR model have resolution of 20×20 meters

329 (for whole study area that covers 613 km²), while previous studies that employed D-STEM (in Europe),

330 produced estimation maps with resolution of about 1 km (and all of them were in non-LUR context).

331 Our study demonstrated that D-STEM performed very well in temporal modeling of PM_{2.5} during year

332 2015 in Tehran, even without using any temporal or spatiotemporal predictors. Moreover, D-STEM daily

333 spatiotemporal and spatial prediction and cross-validated prediction were quite good. This could benefit

334 future epidemiological studies with both short-term and long-term exposure estimates. Furthermore, D-

335 STEM build-in missing imputation capability compared with mean substitution in each monitoring

336 station, which is equivalent to ignoring of missing data, proves usefulness of D-STEM LUR model in

337 missing imputation of air pollution data.

338 Nonetheless, further research is needed to explore more advance features of D-STEM software, such as

339 joint modeling of several air pollutants, joint modeling of one or several air pollutants with

340 meteorological variables, and incorporation of remotely sensed data. Overall, we suggest using D-STEM

341 capabilities in increasing LUR studies that employ data of regulatory network monitoring stations.

342

Acknowledgements

This research was part of the first author's Ph.D. thesis in Biostatistics and was supported by Isfahan University of Medical Sciences [grant number 394937]. The authors acknowledge Tehran Air Quality Control Company and Iranian Department of Environment for providing air pollution data.

Declaration of interest

None.

Appendix A. Supporting Information

Figures: Study area and geographical location of air quality monitoring stations, data availability pattern of observed daily $PM_{2.5}$ in each monitoring station, scatter plot of predicted versus observed values for cooler and warmer seasons maps of $PM_{2.5}$ concentrations.

Tables: descriptive statistics for response and predictors, Assessment and validation metrics, Mean Absolute Error (MAE) in 10 independent simulations, D-STEM output of the final model.

358 References

- 359 Amini, H., Hosseini, V., Schindler, C., Hassankhany, H., Yunesian, M., Henderson, S.B., Künzli, N., 2017a.
360 Spatiotemporal description of BTEX volatile organic compounds in a Middle Eastern megacity: Tehran
361 Study of Exposure Prediction for Environmental Health Research (Tehran SEPEHR). *Environmental*
362 *Pollution* 226, 219-229.
- 363 Amini, H., Schindler, C., Hosseini, V., Yunesian, M., Künzli, N., 2017b. Land Use Regression Models for
364 Alkylbenzenes in a Middle Eastern Megacity: Tehran Study of Exposure Prediction for Environmental
365 Health Research (Tehran SEPEHR). *Environmental Science & Technology* 51, 8481-8490.
- 366 Amini, H., Taghavi-Shahri, S.-M., Henderson, S.B., Hosseini, V., Hassankhany, H., Naderi, M., Ahadi, S.,
367 Schindler, C., Künzli, N., Yunesian, M., 2016. Annual and seasonal spatial models for nitrogen oxides in
368 Tehran, Iran. *Scientific Reports* 6, 32970.
- 369 Amini, H., Taghavi-Shahri, S.M., Henderson, S.B., Naddafi, K., Nabizadeh, R., Yunesian, M., 2014. Land
370 use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and
371 particulate matter in Tehran, Iran. *Science of the Total Environment* 488, 343-353.
- 372 Amini, H., Yunesian, M., Hosseini, V., Schindler, C., Henderson, S.B., Künzli, N., 2017c. A systematic
373 review of land use regression models for volatile organic compounds. *Atmospheric Environment* 171, 1-
374 16.
- 375 Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R.V., Dentener, F., Dingenen, R.v.,
376 Estep, K., Amini, H., Apte, J.S., 2016. Ambient air pollution exposure estimation for the global burden of
377 disease 2013. *Environmental Science & Technology* 50, 79-88.
- 378 Calculli, C., Fassò, A., Finazzi, F., Pollice, A., Turnone, A., 2015. Maximum likelihood estimation of the
379 multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy.
380 *Environmetrics*.
- 381 Faridi, S., Shamsipour, M., Krzyzanowski, M., Künzli, N., Amini, H., Azimi, F., Malkawi, M., Momeniha, F.,
382 Gholampour, A., Hassanvand, M.S., 2018. Long-term trends and health impact of PM 2.5 and O₃ in
383 Tehran, Iran, 2006–2015. *Environment International* 114, 37-49.
- 384 Fassò, A., 2013. Statistical assessment of air quality interventions. *Stoch Environ Res Risk Assess* 27,
385 1651-1660.
- 386 Fassò, A., Finazzi, F., Ndongo, F., 2016. European population exposure to airborne pollutants based on a
387 multivariate spatio-temporal model. *Journal of Agricultural, Biological, and Environmental Statistics* 21,
388 492-511.
- 389 Finazzi, F., 2013. D-STEM: a software for the analysis and mapping of environmental space-time
390 variables. <https://github.com/graspa-group/d-stem>, GRASPA group, Italy.
- 391 Finazzi, F., Fassò, A., 2014. D-STEM: a software for the analysis and mapping of environmental space-
392 time variables. *Journal of Statistical Software* 62, 1-29.
- 393 Finazzi, F., Scott, E.M., Fassò, A., 2013. A model-based framework for air quality indices and population
394 risk evaluation, with an application to the analysis of Scottish air quality data. *Journal of the Royal*
395 *Statistical Society: Series C (Applied Statistics)* 62, 287-308.
- 396 Forouzanfar, M.H., Afshin, A., Alexander, L.T., Anderson, H.R., Bhutta, Z.A., Biryukov, S., Brauer, M.,
397 Burnett, R., Cercy, K., Charlson, F.J., 2016. Global, regional, and national comparative risk assessment of
398 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a
399 systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 388, 1659-1724.
- 400 Hassanpour Matikolaei, S.A.H., Jamshidi, H., Samimi, A., 2017. Characterizing the effect of traffic density
401 on ambient CO, NO₂, and PM_{2.5} in Tehran, Iran: an hourly land-use regression model. *Transportation*
402 *Letters*, 1-11.

Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental science & technology* 41, 2422-2428.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* 42, 7561-7578.

Huang, L., Zhang, C., Bi, J., 2017. Development of land use regression models for PM_{2.5}, SO₂, NO₂ and O₃ in Nanjing, China. *Environmental research* 158, 542-552.

Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Moody, E., Wang, Y., Lyapustin, A., Kloog, I., 2015. Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily PM_{2.5} Geographical Distribution in Mexico City. *Environmental science & technology* 49, 8576-8584.

Kashima, S., Yorifuji, T., Sawada, N., Nakaya, T., Eboshida, A., 2018. Comparison of land use regression models for NO₂ based on routine and campaign monitoring data from an urban area of Japan. *Science of The Total Environment* 631-632, 1029-1037.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment* 45, 6267-6275.

Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., Liu, Y., 2016. Estimating ground-level PM₁₀ in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution* 208, 177-184.

Ryan, P.H., LeMasters, G.K., 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhalation toxicology* 19, 127-133.

Shaddick, G., Thomas, M.L., Amini, H., Broday, D., Cohen, A., Frostad, J., Green, A., Gumy, S., Liu, Y., Martin, R.V., Pruss-Ustun, A., Simpson, D., Donkelaar, A.v., Brauer, M., 2018. Data integration for the assessment of population exposure to ambient air pollution for global burden of disease assessment. *Environmental Science & Technology*.

The MathWorks, I., MATLAB, Statistics Toolbox, Optimization Toolbox and Mapping Toolbox, Natick, Massachusetts, United States.

US EPA, 2013. National Ambient Air Quality Standards for Particulate Matter.

WHO, 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide.

Yu, S., Eder, B., Dennis, R., Chu, S.H., Schwartz, S.E., 2006. New unbiased symmetric metrics for evaluation of air quality models. *Atmospheric Science Letters* 7, 26-34.

Zou, B., Luo, Y., Wan, N., Zheng, Z., Sternberg, T., Liao, Y., 2015. Performance comparison of LUR and OK in PM_{2.5} concentration mapping: a multidimensional perspective. *Scientific reports* 5.