1 # Concurrent Spatiotemporal Daily Land Use Regression Modeling and

2 # Missing Data Imputation of Fine Particulate Matter Using Distributed

3 # Space Time Expectation Maximization

4

5 Seyed Mahmood Taghavi-Shahri[a,b], Alessandro Fassò[c], Behzad Mahaki[a,1,*], Heresh Amini[d]

6

7 [a] Department of Epidemiology and Biostatistics, School of Health, Isfahan University of Medical Sciences,
8 Isfahan, 81746, Iran

9 [b] Student Research Committee, School of Health, Isfahan University of Medical Sciences, Isfahan, 81746,
10 Iran

11 [c] Department of Management Information and Production Engineering, University of Bergamo, Dalmine,
12 24044, Italy

13 [d] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston,
14 Massachusetts, 02215, United States

15

16 * Corresponding author at: Department of Epidemiology and Biostatistics, School of Health, Isfahan
17 University of Medical Sciences, Hezar-Jerib Ave., Isfahan, IR. Iran; Postal Code: 8174673461; Fax:
18 +983137923232; Tel: +989128077960

19

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

22

23 [1] Present address: Department of Biostatistics, School of Health, Kermanshah University of Medical
24 Sciences, Shahid Beheshti Blvd., Kermanshah, IR. Iran; Postal Code: 6715847141

Highlights

25

26   -   First Land Use Regression using D-STEM, a recently introduced statistical software

27   -   Assess D-STEM in spatiotemporal modeling, mapping, and missing data imputation

28   -   Estimate high resolution (20x20 m) daily maps for exposure assessment in a megacity

29   -   Provide both short- and long-term exposure assessment for epidemiological studies

30

31

32

Graphical Abstract

33

34

35
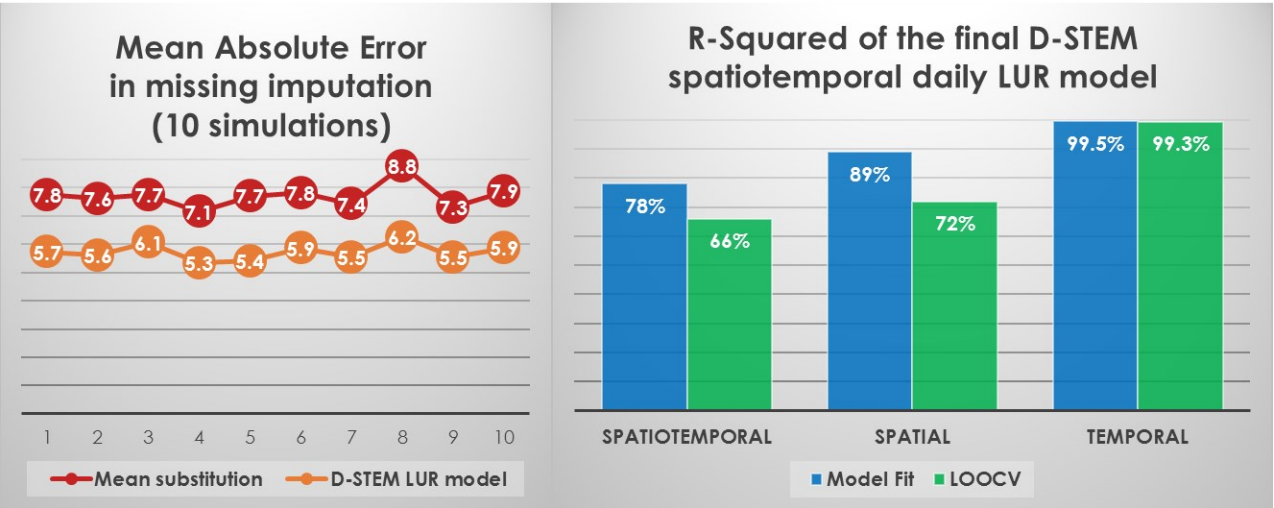
Abstract

36 Land use regression (LUR) has been widely applied in epidemiologic research for exposure assessment.

37 In this study, for the first time, we aimed to develop a spatiotemporal LUR model using Distributed

38 Space Time Expectation Maximization (D-STEM). This spatiotemporal LUR model examined with daily

39 particulate matter ≤ 2.5 µm ($PM_{2.5}$) within the megacity of Tehran, capital of Iran. Moreover, D-STEM

40 missing data imputation was compared with mean substitution in each monitoring station, as it is

41 equivalent to ignoring of missing data, which is common in LUR studies that employ regulatory

42 monitoring stations' data. The amount of missing data was 28% of the total number of observations, in

43 Tehran in 2015. The annual mean of $PM_{2.5}$ concentrations was 33 µg/m^3. Spatiotemporal R-squared of

44 the D-STEM final daily LUR model was 78%, and leave-one-out cross-validation (LOOCV) R-squared was

45 66%. Spatial R-squared and LOOCV R-squared were 89% and 72%, respectively. Temporal R-squared and

46 LOOCV R-squared were 99.5% and 99.3%, respectively. Mean absolute error decreased 26% in

47 imputation of missing data by using the D-STEM final LUR model instead of mean substitution. This

48 study reveals competence of the D-STEM software in spatiotemporal missing data imputation,

49 estimation of temporal trend, and mapping of small scale (20 x 20 meters) within-city spatial variations,

50 in the LUR context. The estimated $PM_{2.5}$ concentrations maps could be used in future studies on short-

51 and/or long-term health effects. Overall, we suggest using D-STEM capabilities in increasing LUR studies

52 that employ data of regulatory network monitoring stations.

Keywords

3

1. Introduction

59

60   Air pollution caused 6.5 million deaths and 167 million disability-adjusted life-years (DALYs) in 2015

61   worldwide. In particular, ambient particulate matter with an aerodynamic diameter smaller than 2.5 μm

62   (PM$_{2.5}$), caused 4.2 million of deaths and 103 million DALYs worldwide in the same year (Forouzanfar et

63   al., 2016).

64   Air pollution epidemiological studies need accurate exposure assessment and one frequently used

65   method for individual exposure assessment is land use regression (LUR) modeling. LUR uses geographic

66   independent variables (typically extracted using geographic information system (GIS)) as Potentially

67   Predictive Variables (PPVs) to predict spatial variation of air pollution at unmonitored locations (Amini et

68   al., 2017c; Hoek et al., 2008; Ryan and LeMasters, 2007).

69   Many LUR studies that based on data of regulatory network monitoring stations, use annual or seasonal

70   mean of air pollutants with a subset of GIS generated PPVs in a multiple regression for spatial modeling

71   and mapping (Huang et al., 2017; Kashima et al., 2018; Zou et al., 2015) that ignore temporal trend of air

72   pollution. However, some other studies use relatively advanced models, such as linear mixed effect

73   models, with satellite, meteorological, and land use predictors for spatiotemporal estimation of air

74   pollution concentrations (Just et al., 2015; Kloog et al., 2011; Meng et al., 2016). However, the common

75   issue is treating of air pollution missing data, and perhaps limited capability in accounting for temporal

76   and spatial autocorrelations.

77   D-STEM (Distributed Space Time Expectation Maximization) is a comprehensive statistical software for

78   analyzing and mapping of environmental data that was introduced in 2013, which is programmed and

79   run in Matlab environment (The MathWorks). The underlying D-STEM general model accounts for

80   spatial and temporal autocorrelations and measurement errors, spatial and temporal and

81   spatiotemporal predictors, fixed and random effect coefficients, concurrent missing imputation and

82   model fitting using Expectation Maximization (EM) algorithm. In addition, D-STEM supports multivariate

83   modeling even in heterogonous stations (for example, simultaneous modeling of meteorological

84   variables and air pollutants measured in the same or different stations), and data fusion capability for

85   joint modeling (and calibration) of remote sensing data measured by a satellite along with air pollution

86   concentrations measured in the ground stations (Finazzi, 2013; Finazzi and Fasso, 2014).

87   D-STEM software has been previously used in several air pollution studies in Europe (Calculli et al., 2015;

88   Fassò, 2013; Fassò et al., 2016; Finazzi, 2013; Finazzi et al., 2013), but primary focus of those studies

89   have been on statistical theory. Moreover, the resolution of their output maps has been about 1 km,

90   which might be course for epidemiological studies that need individual-level exposure estimates. In

91   addition, none of those studies used a pool of GIS generated PPVs for extensive independent variables

92   selection. Overall, D-STEM has not been used so far for LUR modeling and for exposure estimation with

93   very fine spatial resolution.

94   To date, several LUR studies have been published in Tehran to predict air pollution. These have been for

95   particulate matter with an aerodynamic diameter smaller than 10 μm (PM_{10}) and sulfur dioxide (SO_2)

96   (Amini et al., 2014), nitrogen oxides (NO, NO_2, and NOx) (Amini et al., 2016), $PM_{2.5}$, carbon monoxide

97   (CO), and nitrogen dioxide (NO_2) (Hassanpour Matikolaei et al., 2017), and alkylbenzenes (benzene,

98   toluene, ethylbenzene, *m*-xylene, *p*-xylene, and *o*-xylene (BTEX), and total BTEX) (Amini et al., 2017b). In

99   fact, multiple regression has been used in majority of these studies for modeling of annual or seasonal

100   averages of air pollutants. However, the study of $PM_{2.5}$, CO, and NO_2 (Hassanpour Matikolaei et al.,

101   2017), used panel regression to predict hourly concentrations with limited capability in accounting for

102   spatial and temporal autocorrelations, and not supporting spatiotemporal missing imputation

103   (Hassanpour Matikolaei et al., 2017).

5

104   In the present study, we aimed to utilize D-STEM with a relatively large pool of GIS generated PPVs for

105   producing high resolution daily estimations of airborne PM$_{2.5}$ in the Middle Eastern megacity of Tehran,

106   Iran. We further aimed to assess some characteristics of the recently introduced D-STEM software. In

107   addition, we aimed to compare performance of air pollution missing data imputation under two

108   scenarios: (1) using mean substitution in each monitoring station, and (2) using D-STEM concurrent

109   spatiotemporal missing imputation and LUR modeling.

110

2. Materials and Methods

2.1 Study area and data

113   The study area of this research is the megacity of Tehran, Iran which has 9 million urban residents and,

114   due to diurnal migration, a daytime population of more than 10 million. Alborz Mountains are located in

115   north of the city and there is a desert in the south. The populated area covers approximately 613 km^2

116   and its elevation ranges from 1,000 to 1800 meters above sea level from south to north. The prevailing

117   winds blow from west and north. The mean daily temperature ranges from about −15 °C in January to

118   43 °C in July with an annual average of 18.5 °C. The precipitation ranges from 0 millimeters (mm) in

119   September to 40 mm in January, with an annual average of 150 mm. The weather of most days is sunny,

120   and the mean cloud cover is about 30%. Geographic location of Tehran is in zone 39N of Universal

121   Transverse Mercator (UTM) (Amini et al., 2016; Amini et al., 2014). Figure S1 of Supporting information

122   shows the study area and location of monitoring stations.

123   Hourly data of fine particulate matter (PM$_{2.5}$), which have been measured using fixed beta-gauge

124   monitors, were obtained from Tehran Air Quality Control Company (AQCC) and Iranian Department of

125   Environment (DOE) for year 2015. Daily PM$_{2.5}$ was calculated by 24-hours averaging of hourly measured

126 PM$_{2.5}$ concentrations when days had complete records in all hours. Daily mean of PM$_{2.5}$ for a set of 19

127 fixed measuring sites and for 365 days of year 2015 was used as response variable in this study.

128 A pool of 212 potentially predictive variables (PPVs) in six classes and 73 sub-classes were used for

129 selecting model independent variables. These variables were in raster format for whole study area with

130 resolution of 5×5 meters. The six GIS generated classes were Distance (60 variables), Product terms (52

131 variables), Land Use (50 variables), Traffic Surrogates (26 variables), Population Density (22 variables),

132 and Geographic Location (2 variables). The full list of our PPVs could be found elsewhere (Amini et al.,

133 2016; Amini et al., 2014).

134 2.2 Model Specification

135 D-STEM spatiotemporal LUR model for PM$_{2.5}$ is given by the following equation where Y is response

136 variable, which is spatiotemporally indexed at location s and time t, provided X's the GIS generated

137 spatial predictors that are assumed temporally constant over time and spatially indexed at location s:

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

139 While the left-hand side of the above equation defines the response variable, the right-hand side of this

140 equation consists of four parts. First, sum of p summand: β_i multiplied by $X_i(s)$, which are a fixed

141 effect coefficient and the corresponding GIS generated spatial predictor. Second, sum of q summand:

142 α_j multiplied by $\omega_j(s,t)$ and multiplied by $X_j(s)$, which are a scale changing coefficient and the

143 corresponding spatiotemporal random effect coefficient, and the corresponding GIS generated spatial

144 predictor, respectively. Third part is $z(t)$ which is a latent temporal state variable that concurrently

145 estimated using a sub-model. Forth part is $\varepsilon(s,t)$ which denotes measurement error that assumed

146 normally distributed with zero mean and uncorrelated over space and time.

7

147   Latent spatiotemporal random effect coefficients $\omega_j(s,t)$ assumed to be independent of each other

148   and normally distributed with zero mean and unit variance that uncorrelated over time but correlated

149   over space with exponential spatial auto-covariance function $C_{\theta_j}(d) = \exp(\frac{-d}{\theta_j})$ where d is the

150   distance between two generic spatial points and θ_j is auto-covariance parameter.

151   Since the latent spatiotemporal random effect coefficients $\omega_j(s,t)$ assumed to have zero means, we

152   are not allowed to incorporate any GIS generated spatial predictor in random effect part, except if we

153   already used it in fixed effect part. In the current representation, β could be interpreted as global

154   effect of GIS generated predictor, and α could be used for testing spatiotemporal variation in the effect

155   of corresponding predictor.

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

157   model:

158
$$z(t) = G \times z(t-1) + \eta(t)$$

159   Where G is coefficient of temporal state transition and η is temporal state innovation that assumed

160   normally distributed with zero mean.

161   The estimated parameters of the D-STEM spatiotemporal LUR model are as follows: fixed effect

162   coefficients, scaling changing coefficients of the corresponding spatiotemporal random effect

163   coefficients, parameters of the underlying auto-covariance functions of the corresponding

164   spatiotemporal random effect coefficients, temporal state transition coefficient, variance of the

165   temporal state innovation, and variance of the measurement error.

8

166 ## 2.3 Model fitting, assessing, and mapping

167 A forward algorithm, based on maximization of spatial Leave-One-Out Cross-Validation (LOOCV) R-

168 squared and consistency with prior knowledge, was used to select suitable predictors from pool of PPVs.

169 This algorithm considers sign consistency of fixed effect coefficients with a priori assumed direction of

170 effects. Effect directions defined as positive, negative, or arbitrary (unknown) based on knowledge

171 about emission of air pollutions.

172 The final daily LUR model was assessed in three forms: a) spatiotemporal, b) spatial (temporally

173 averaged to assess spatial variations), and c) temporal (spatially averaged to assess temporal variations).

174 Also, Normalized Mean Bias Factor (NMBF) was used for checking whether models underestimate or

175 overestimate observed data (Yu et al., 2006).

176 The power of missing data imputation was compared between mean substitution in each monitoring

177 station, and final spatiotemporal LUR model by 10 independent simulations. In each simulation 10% of

178 spatiotemporal observed data was randomly selected and was replaced by missing data. Afterwards,

179 both models were independently employed for missing imputation. The predicted values of the

180 randomly missed data and original observed values (that randomly missed), was used to compute Mean

181 Absolute Error (MAE) in missing imputation of each model. Finally, MAEs that computed in these

182 independent simulations and percent of MAEs change between two models, was averaged and

183 reported.

184 Mapping was done using D-STEM for each day, by providing GIS generated maps that correspond to

185 predictors of the final daily LUR model. Seasonal maps were generated by averaging of daily output

186 maps in the corresponding time periods. Finally, we applied a quantification limit (QL) that its lower

187 bound was the minimum of within monitoring station averages of observed concentration in the

188 corresponding time period divided by square root of two, and its upper bounds was 120% of the

9

189   maximum of within monitoring station averages of observed concentration in the same time period.

190   Grid cells in the final output maps that had values outside QL interval were set to nearest QL bound and

191   frequency of these corrections was calculated and reported (Amini et al., 2016; Amini et al., 2014;

192   Henderson et al., 2007)

193

194   ## 3. Results and Discussion

195   ### 3.1 Description of data

196   Number of $PM_{2.5}$ daily observations at 19 monitoring stations was 5016. Frequency of missing data

197   varies from 12% to 47% in monitoring stations with an average of 28% over all stations. Frequency of

198   available data for each day varied from 2 stations to 19 stations with an average of 14 stations. Data

199   availability pattern of daily $PM_{2.5}$ in monitoring stations of Tehran in 2015 is presented in Figure S2 of

200   Supporting information.

201   Annual mean of observed $PM_{2.5}$ concentrations (spatially and temporally average of monitoring stations)

202   was 33 $\mu g/m^3$ that is more than three times of World Health Organization (WHO) guideline (10 $\mu g/m^3$)

203   and about three times of US Environmental Protection Agency (EPA) guideline (12 $\mu g/m^3$) (US EPA, 2013;

204   WHO, 2006), with a range of 21 $\mu g/m^3$ to 48 $\mu g/m^3$ in different monitoring stations. This is in line with

205   the findings of Brauer et al. (2016) and Shaddick et al. (2018) where they reported very high population-

206   weighted exposure estimates for $PM_{2.5}$ in the Middle East (Brauer et al., 2016; Shaddick et al., 2018).

207   Faridi et al. (2018) also reported very similar annual mean for $PM_{2.5}$ in Tehran (Faridi et al., 2018). Daily

208   mean of Tehran $PM_{2.5}$ concentrations in 2015 (spatially average of monitoring stations) was greater than

209   WHO daily standard (25 $\mu g/m^3$) in 246 days (67% of year), and was greater than US EPA daily standard

210   (35 $\mu g/m^3$) in 121 days (33% of year), and eventually was greater than 55 $\mu g/m^3$ in 31 days. As a result,

211   the air quality index (AQI) in Tehran in 2015 was unhealthy for at least sensitive groups in one third of

10

212 year, and was unhealthy for all groups in about one month (8.6% of year that exceeded 55 µg/m^3),

213 based on US EPA breakpoints (US EPA, 2013; WHO, 2006). The very high values of air pollution have

214 been confirmed in large Tehran Study of Exposure Prediction for Environmental Health Research

215 (SEPEHR) measurement campaigns (Amini et al., 2017a).

216 3.2 Description of the LUR model

217 Natural logarithm of PM$_{2.5}$ was standardized (using mean and standard deviation of its all

218 spatiotemporal values) and utilized in spatiotemporal LUR model as response variable. Also,

219 standardized version of GIS generated PPVs was employed in the model. Summary statistics for these

220 variables are presented in Table S1 of Supporting information.

221 Table 1 presents estimated parameters of spatiotemporal daily LUR model. This model accounts for both

222 temporal and spatial autocorrelations, and moreover supports inclusion of predictors for concurrent

223 missing imputation and LUR model fitting. Among 212 GIS generated predictors that produced a huge

224 possible combination of predictors, eventually 8 predictors were selected by our described algorithm to

225 include in the spatiotemporal final daily LUR model.

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

Category	Term	Estimate	Std. Error	p-value
Fixed effect	LNDIST to AMB	-0.170	0.009	< 0.001
Coefficients	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

11

	GRS.500	-0.151	0.013	< 0.001
	BGD.500	+0.082	0.009	< 0.001
Scale changing Coefficients of Random effect	LNDIST to AMB	0.089	0.018	< 0.001
	SEN.400	0.037	0.019	0.05
	GRS.500	0.145	0.025	< 0.001
Spatial covariance Parameters	LNDIST to AMB	101.30	284.15	-
	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

The model response was standardized natural logarithm of $PM_{2.5}$ concentrations

Also, Predictors were standardized (see Table S1 of Supporting information).

Abbreviations:

LN $PM_{2.5}$: natural logarithm of particulate matter with diameter<=2.5 micrometer

LNDIST to AMB: natural logarithm of distance to nearest ambulance service

PD.1000: total population density in buffer 1000

LNDIST to HZRFAC: natural logarithm of distance to nearest hazardous facility

ELEV: elevation

SEN.400: sensitive areas in buffer 400

DIST to SCSC: distance to nearest secondary or high school

GRS.500: green space area in buffer 500

BGD.500: bridge length in buffer 500 divided by distance to the bridge

227

228   GIS generated predictors that had a positive coefficient in the final spatiotemporal daily LUR model

229   were: total population density in buffer of 1000 m, natural logarithm of distance to nearest hazardous

230   facility (with an arbitrary assumed direction of the effect), distance to the nearest secondary or high

231   school, and bridge length in buffer of 500 m divided by distance to the bridges.

232   On the other hand, the predictors that had a negative coefficient in the final spatiotemporal daily LUR

233   model were: natural logarithm of distance to nearest ambulance service (assumed sign was arbitrary),

234   elevation (assumed sign was arbitrary), sensitive areas in buffer of 400 m (assumed sign was arbitrary),

235   and green space areas in buffer of 500 m.

236   Random effect predictors in the final spatiotemporal daily LUR model were as follow: natural logarithm

237   of distance to nearest ambulance service, sensitive areas in buffer of 400 m, and green space areas in

238   buffer of 500 m. The corresponding scale changing coefficients were significant or borderline significant.

239   That means coefficient of these three predictors vary in time and space, and observed data are spatially

240   auto-correlated.

241   Temporal state transition coefficient was significant ($p < 0.001$) and positive that indicates a positive

242   temporal autocorrelation existed between successive observations. Moreover, variance of the temporal

243   state innovation was significant ($p < 0.001$) that means variations existed in temporal trend of air

244   pollution.

3.3 Assessing fit of the LUR model

246   Figure 1 shows temporal pattern (spatially averaged) of daily $PM_{2.5}$ for observed and predicted data of

247   the final daily LUR model. Figure 1 confirms that D-STEM succeeds in capturing natural turbulence of

248    $PM_{2.5}$ time series, even without employing any temporal predictors.
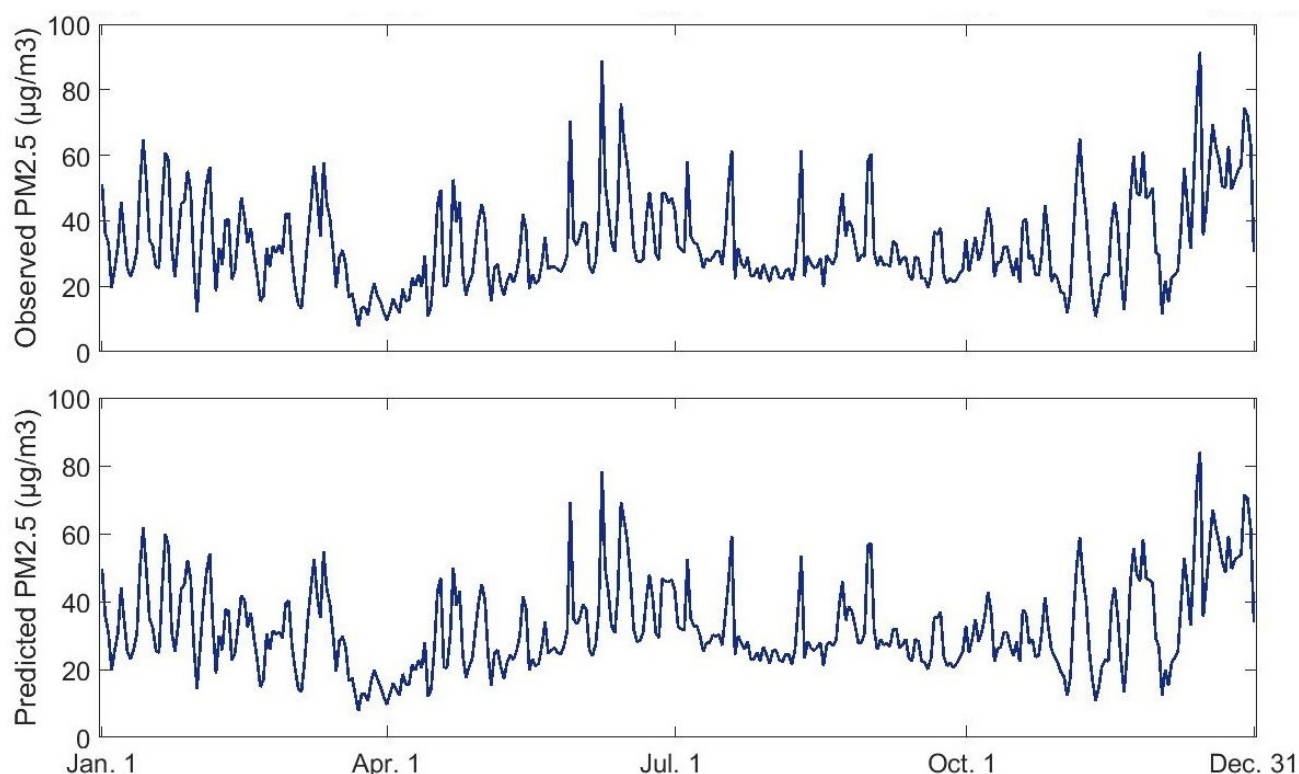
13

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 (Calculli et al., 2015; Fassò, 2013; Fassò et al., 2016; Finazzi, 2013; Finazzi et al., 2013), so it is not clear

262   that this slight deficiency is belonging to D-STEM software or it is an accidental observation that

263   happened in our study.

264   However, the temporal prediction of D-STEM LUR model was very well. In fact, temporal R-squared and

265   temporal LOOCV R-squared were greater than 99%. In a study by Kloog et al. (Kloog et al., 2011) where

266   they employed land use and satellite-derived Aerosol Optical Depth (AOD) and multiple stations

267   (spatiotemporal) meteorological predictors in a mixed model, the temporal cross-validated R-squared

268   was 84%. Note that we did not use any spatiotemporal or pure temporal predictors in the D-STEM

269   software. A similar study in Tehran that modelled $PM_{2.5}$ with a panel regression used pure temporal

270   predictors, i.e., one station meteorological variables, just in order to partly capture temporal innovations

271   of air pollution data (Hassanpour Matikolaei et al., 2017). Hence, the use of pure temporal predictors in

272   D-STEM for spatiotemporal modeling of air pollution is completely unnecessary.

273   The spatial R-squared of our D-STEM final LUR model (and even spatial LOOCV R-squared) was greater

274   than 65% that reported in a recently published study on LUR modelling of $PM_{2.5}$ in Tehran. That study,

275   that used a panel regression for spatiotemporal modeling of fixed monitoring air pollution data,

276   provided GIS generated spatial data, meteorological data of one station, and real time google traffic

277   data as predictors (Hassanpour Matikolaei et al., 2017). Therefore, despite of the fact that predictors of

278   our D-STEM spatiotemporal LUR model were limited to GIS generated spatial data, D-STEM performed

279   better.

280   3.4 Assessing missing data imputation

281   Mean Absolute error (MAE) in 10 independent simulations for spatiotemporal missing imputation, was

282   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,

283   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

284   station.

285   Changes between MAEs showed a decreased range from 24.3% to 29.9% with an average of 25.9% in

286   using D-STEM final spatiotemporal daily LUR model instead of Mean substitution in each monitoring

287   station. Details of MAEs and changes between MAEs are presented in Table S3 of Supporting

288   information.

289   This result is important because missing data exists in almost any air pollution study of fixed air

290   monitoring stations, such as our two previous LUR studies of PM_{10}, SO_2, and nitrogen oxides in Tehran.

291   In those studies we imputed air pollution missing data using EM algorithm in a multivariate way but

292   without considering spatial and temporal autocorrelations (Amini et al., 2016; Amini et al., 2014).

293   Although, some other LUR studies such as one earlier mentioned LUR study of $PM_{2.5}$ in Tehran that

294   modeled using a panel regression, did not describe the missing data issue (Hassanpour Matikolaei et al.,

295   2017).

296   This study proved that D-STEM missing imputation is better than ignoring of missing data, which is

297   common at least when the amount of missing data is small (Meng et al., 2016). We believe D-STEM is

298   worthy to be applied in missing imputation of (fixed monitoring stations) air pollution data, because it

299   imputes missing data concurrently with modeling, and we see that the underlying temporal model of D-

300   STEM perform very well. Moreover, D-STEM can capture spatial variation of air pollution data by

301   incorporating suitable land use and GIS generated predictors. Furthermore, D-STEM spatiotemporal

302   model that concurrently impute missing data has ability to employ spatiotemporal meteorological

303   predictors, and even remote sensing data.

304   3.5 Generating output maps

305   We produced daily $PM_{2.5}$ concentrations maps with resolution of 20×20 meters for whole study area that

306   cover 613 km^2. Cooler and warmer season maps of $PM_{2.5}$ were produced by averaging of D-STEM

307   generated daily maps in the corresponding periods.

16

308   Quantification limits (QL) that calculated based on averaged observed values of air quality monitoring

309   stations in the corresponding periods, had a lower bound of 13.6 μg/m^3 in cooler season, and 15.2

310   μg/m^3 in warmer season; and had upper bound of 68.1 μg/m^3 in cooler season, and 50.1 μg/m^3 in

311   warmer season.

312   Values beyond these quantification limits were corrected by setting to the nearest upper or lower limits.

313   Percent of these corrections for lower and upper bounds of the corresponding QL limits were 0.03% and

314   1.9% in the cooler season map, and were 0.1% and 3.4% in warmer season map, respectively. Hence,

315   the under- or over-estimation of PM$_{2.5}$ in the D-STEM LUR generated maps were relatively small when

316   compared to other LUR studies (Amini et al., 2016; Amini et al., 2014).

317   The seasonal estimation maps of PM$_{2.5}$ concentrations in Tehran in 2015, with resolution of 20×20

318   meters, are presented in Figure 2. The scatter plot of predicted versus observed values are presented in

319   Figure S3 of Supporting information for cooler and warmer seasons maps of PM$_{2.5}$ concentrations.

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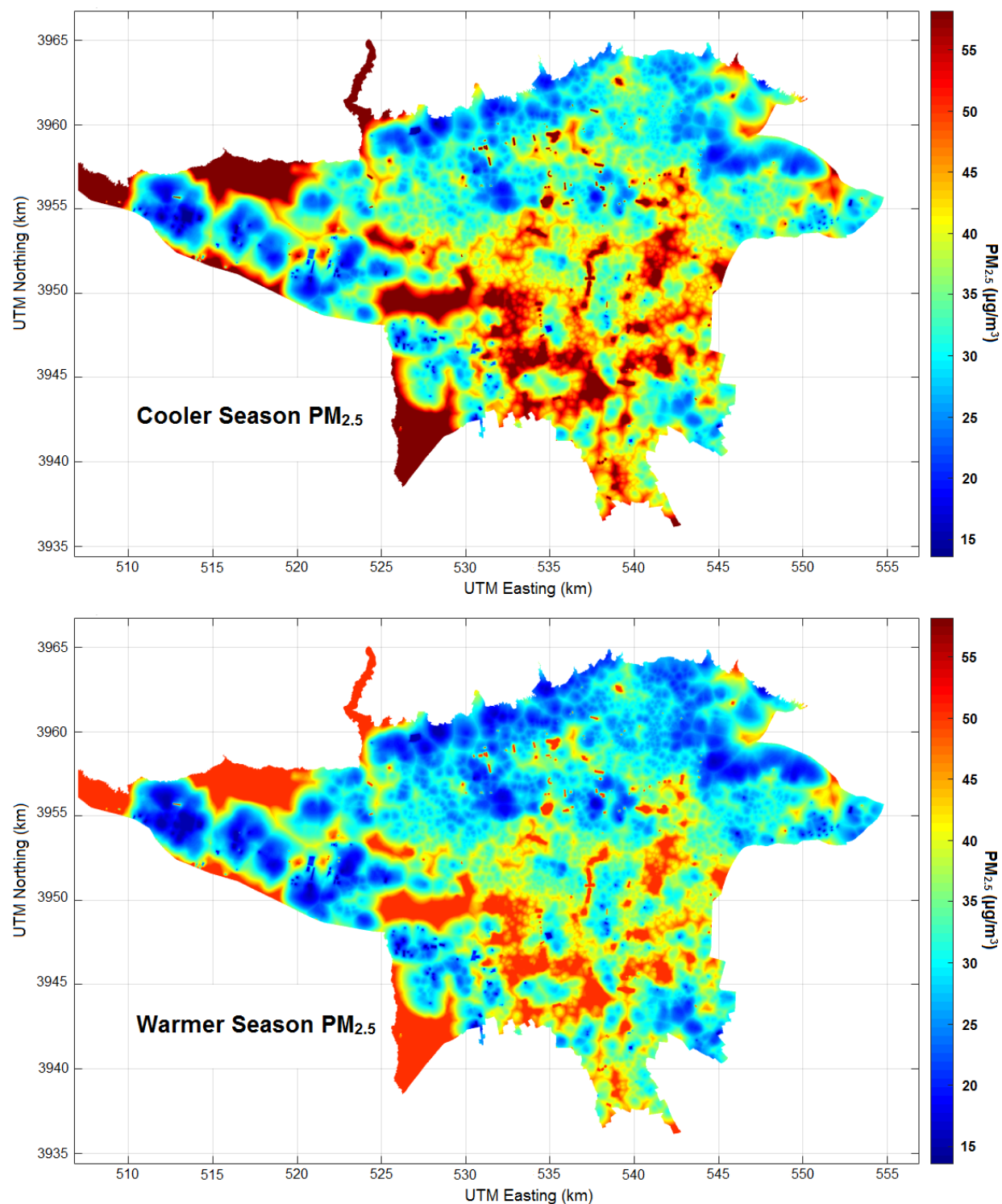
Figure 2. Estimated cooler and warmer seasons maps of $PM_{2.5}$ in Tehran 2015 with resolution of 20×20 meters that produced by averaging of daily LUR maps over corresponding periods.

323

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 km^2), 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

19

Acknowledgements

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.

References

Amini, H., Hosseini, V., Schindler, C., Hassankhany, H., Yunesian, M., Henderson, S.B., Künzli, N., 2017a. Spatiotemporal description of BTEX volatile organic compounds in a Middle Eastern megacity: Tehran Study of Exposure Prediction for Environmental Health Research (Tehran SEPEHR). Environmental Pollution 226, 219-229.

Amini, H., Schindler, C., Hosseini, V., Yunesian, M., Künzli, N., 2017b. Land Use Regression Models for Alkylbenzenes in a Middle Eastern Megacity: Tehran Study of Exposure Prediction for Environmental Health Research (Tehran SEPEHR). Environmental Science & Technology 51, 8481-8490.

Amini, H., Taghavi-Shahri, S.-M., Henderson, S.B., Hosseini, V., Hassankhany, H., Naderi, M., Ahadi, S., Schindler, C., Künzli, N., Yunesian, M., 2016. Annual and seasonal spatial models for nitrogen oxides in Tehran, Iran. Scientific Reports 6, 32970.

Amini, H., Taghavi-Shahri, S.M., Henderson, S.B., Naddafi, K., Nabizadeh, R., Yunesian, M., 2014. Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran. Science of the Total Environment 488, 343-353.

Amini, H., Yunesian, M., Hosseini, V., Schindler, C., Henderson, S.B., Künzli, N., 2017c. A systematic review of land use regression models for volatile organic compounds. Atmospheric Environment 171, 1-16.

Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R.V., Dentener, F., Dingenen, R.v., Estep, K., Amini, H., Apte, J.S., 2016. Ambient air pollution exposure estimation for the global burden of disease 2013. Environmental Science & Technology 50, 79-88.

Calculli, C., Fassò, A., Finazzi, F., Pollice, A., Turnone, A., 2015. Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy. Environmetrics.

Faridi, S., Shamsipour, M., Krzyzanowski, M., Künzli, N., Amini, H., Azimi, F., Malkawi, M., Momeniha, F., Gholampour, A., Hassanvand, M.S., 2018. Long-term trends and health impact of PM 2.5 and O 3 in Tehran, Iran, 2006–2015. Environment International 114, 37-49.

Fassò, A., 2013. Statistical assessment of air quality interventions. Stoch Environ Res Risk Assess 27, 1651-1660.

Fassò, A., Finazzi, F., Ndongo, F., 2016. European population exposure to airborne pollutants based on a multivariate spatio-temporal model. Journal of Agricultural, Biological, and Environmental Statistics 21, 492-511.

Finazzi, F., 2013. D-STEM: a software for the analysis and mapping of environmental space-time variables. https://github.com/graspa-group/d-stem, GRASPA group, Italy.

Finazzi, F., Fasso, A., 2014. D-STEM: a software for the analysis and mapping of environmental space-time variables. Journal of Statistical Software 62, 1-29.

Finazzi, F., Scott, E.M., Fassò, A., 2013. A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 62, 287-308.

Forouzanfar, M.H., Afshin, A., Alexander, L.T., Anderson, H.R., Bhutta, Z.A., Biryukov, S., Brauer, M., Burnett, R., Cercy, K., Charlson, F.J., 2016. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 388, 1659-1724.

Hassanpour Matikolaei, S.A.H., Jamshidi, H., Samimi, A., 2017. Characterizing the effect of traffic density on ambient CO, NO2, and PM2. 5 in Tehran, Iran: an hourly land-use regression model. Transportation Letters, 1-11.

403   Henderson, S.B., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to
404   estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter.
405   Environmental science & technology 41, 2422-2428.
406   Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of
407   land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric
408   Environment 42, 7561-7578.
409   Huang, L., Zhang, C., Bi, J., 2017. Development of land use regression models for PM2. 5, SO2, NO2 and
410   O3 in Nanjing, China. Environmental research 158, 542-552.
411   Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Moody, E., Wang, Y.,
412   Lyapustin, A., Kloog, I., 2015. Using High-Resolution Satellite Aerosol Optical Depth To Estimate Daily
413   PM2. 5 Geographical Distribution in Mexico City. Environmental science & technology 49, 8576-8584.
414   Kashima, S., Yorifuji, T., Sawada, N., Nakaya, T., Eboshida, A., 2018. Comparison of land use regression
415   models for NO2 based on routine and campaign monitoring data from an urban area of Japan. Science
416   of The Total Environment 631-632, 1029-1037.
417   Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially
418   resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth
419   measurements. Atmospheric Environment 45, 6267-6275.
420   Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., Xue, W., Wang, J., Wang, D., Kan, H., Liu, Y., 2016.
421   Estimating ground-level PM10 in a Chinese city by combining satellite data, meteorological information
422   and a land use regression model. Environmental Pollution 208, 177-184.
423   Ryan, P.H., LeMasters, G.K., 2007. A review of land-use regression models for characterizing intraurban
424   air pollution exposure. Inhalation toxicology 19, 127-133.
425   Shaddick, G., Thomas, M.L., Amini, H., Broday, D., Cohen, A., Frostad, J., Green, A., Gumy, S., Liu, Y.,
426   Martin, R.V., Pruss-Ustun, A., Simpson, D., Donkelaar, A.v., Brauer, M., 2018. Data integration for the
427   assessment of population exposure to ambient air pollution for global burden of disease assessment.
428   Environmental Science & Technology.
429   The MathWorks, I., MATLAB, Statistics Toolbox, Optimization Toolbox and Mapping Toolbox, Natick,
430   Massachusetts, United States.
431   US EPA, 2013. National Ambient Air Quality Standards for Particulate Matter.
432   WHO, 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur
433   dioxide.
434   Yu, S., Eder, B., Dennis, R., Chu, S.H., Schwartz, S.E., 2006. New unbiased symmetric metrics for
435   evaluation of air quality models. Atmospheric Science Letters 7, 26-34.
436   Zou, B., Luo, Y., Wan, N., Zheng, Z., Sternberg, T., Liao, Y., 2015. Performance comparison of LUR and OK
437   in PM2. 5 concentration mapping: a multidimensional perspective. Scientific reports 5.

438

439

440