

## Geocoding and spatiotemporal modeling of long-term PM<sub>2.5</sub> and NO<sub>2</sub> exposure in the Mexican Teachers' Cohort

Karla CERVANTES-MARTÍNEZ<sup>1</sup>, Horacio RIOJAS-RODRÍGUEZ<sup>1</sup>, Carlos DÍAZ-ÁVALOS<sup>2</sup>,  
Hortensia MORENO-MACÍAS<sup>3</sup>, Ruy LÓPEZ-RIDAURA<sup>4</sup>, Dalia STERN<sup>5</sup>,  
Jorge Octavio ACOSTA-MONTES<sup>6</sup> and José Luis TEXCALAC-SANGRADOR<sup>1\*</sup>

<sup>1</sup> Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública, Universidad 655, Col. Santa María Ahuacatlán, 62100 Cuernavaca, Morelos, México.

<sup>2</sup> Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México, Av. Universidad 3000, Ciudad Universitaria, 04510 Ciudad de México, México.

<sup>3</sup> Departamento de Economía, Universidad Autónoma Metropolitana, San Rafael Atlixco 186, Col. Vicentina, 09340 Ciudad de México, México.

<sup>4</sup> Centro Nacional de Programas Preventivos y Control de Enfermedades (CENAPRECE), Secretaría de Salud, Benjamín Franklin 132, Col. Escandón, 11800, Ciudad de México, México.

<sup>5</sup> Consejo Nacional de Ciencia y Tecnología (CONACyT)-Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública, Universidad 655, Col. Santa María Ahuacatlán, 62100 Cuernavaca, Morelos, México.

<sup>6</sup> Facultad de Enfermería y Nutriología, Universidad Autónoma de Chihuahua, C. Escorza 900, Col. Centro, 31000 Chihuahua, Chihuahua, México.

\*Corresponding author; email: jtexcalac@insp.mx

Received: August 3, 2021; accepted: November 16, 2021

### RESUMEN

Los estudios epidemiológicos sobre los efectos de la contaminación del aire en México a menudo utilizan las concentraciones ambientales de los monitores más cercanos al hogar como indicadores de exposición, sin embargo, este enfoque ignora los gradientes espaciales de los contaminantes y asume que las personas no tienen movilidad dentro de la ciudad. Nuestro objetivo fue desarrollar modelos espaciales y temporales de alta resolución para predecir la exposición a largo plazo a PM<sub>2.5</sub> y NO<sub>2</sub> en una población de ~16 500 participantes de un estudio de Cohorte de Maestras Mexicanas (ESMaestras). Geocodificamos las direcciones de la casa y el trabajo de las participantes. Utilizando información de fuentes secundarias sobre variables geográficas y meteorológicas, así como otros contaminantes, ajustamos dos modelos aditivos generalizados para predecir las concentraciones mensuales de PM<sub>2.5</sub> y NO<sub>2</sub> en el periodo 2004-2019. Los modelos se evaluaron mediante validación cruzada de 10 iteraciones. Ambos mostraron una alta precisión predictiva con datos fuera de la muestra y no sobreajuste (CV RMSE = 0.102 para PM<sub>2.5</sub> y CV RMSE = 4.497 para NO<sub>2</sub>). Los participantes estuvieron expuestos a un promedio mensual de 24.38 (6.78) µg m<sup>-3</sup> de PM<sub>2.5</sub> y 28.21 (8.00) ppb de NO<sub>2</sub> durante el periodo de estudio. Estos modelos ofrecen una alternativa prometedora para estimar la exposición a PM<sub>2.5</sub> y NO<sub>2</sub> con alta resolución espaciotemporal para estudios epidemiológicos en la región del Valle de México.

### ABSTRACT

Epidemiological studies on air pollution in Mexico often use the environmental concentrations of pollutants as measured by monitors closest to the home of participants as exposure proxies, yet this approach does not account for the space gradients of pollutants and ignores intra-city human mobility. This study aimed to develop high-resolution spatial and temporal models for predicting long-term exposure to PM<sub>2.5</sub> and NO<sub>2</sub> in

~16 500 participants from the Mexican Teachers' Cohort study. We geocoded the home and work addresses of participants, and used secondary source information on geographical and meteorological variables as well as other pollutants to fit two generalized additive models capable of predicting monthly  $\text{PM}_{2.5}$  and  $\text{NO}_2$  concentrations during the 2004-2019 period. Both models were evaluated through 10-fold cross-validation, and showed high predictive accuracy with out-of-sample data and no overfitting (CV-RMSE = 0.102 for  $\text{PM}_{2.5}$  and CV-RMSE = 4.497 for  $\text{NO}_2$ ). Participants were exposed to a monthly average of 24.38 (6.78)  $\mu\text{g m}^{-3}$  of  $\text{PM}_{2.5}$  and 28.21 (8.00) ppb of  $\text{NO}_2$  during the study period. These models offer a promising alternative for estimating  $\text{PM}_{2.5}$  and  $\text{NO}_2$  exposure with high spatiotemporal resolution for epidemiological studies in the Mexico City Metropolitan Area.

**Keywords:** air pollution, generalized additive models, exposure assessment.

## 1. Introduction

In 2015, the World Health Organization (WHO) recognized ambient air pollutant exposure and its effects on human health as a public health priority requiring further study (WHA68, 2015). The rapid expansion of urban areas worldwide has increased the number of individuals exposed to pollutant concentrations above the upper limits recommended by the WHO (Riojas-Rodríguez et al., 2016). For instance, the Mexico City Metropolitan Area (MCMA) is the third-largest metropolitan area of the member states of the Organization for Economic Cooperation and Development (OECD) and is considered one of three megacities in Latin America (OECD, 2016). Historically, this area has registered high rates of air pollution (INE-SEMARNAT, 2011), posing a constant risk to the health of its nearly 21 million inhabitants (INEGI, 2015).

Epidemiological studies have shown that chronic exposure to ambient  $\text{PM}_{2.5}$  and  $\text{NO}_2$  is associated with health outcomes including increased mortality (Beelen et al., 2014), cardiovascular disease (Dockery, 2001; Hoek et al., 2013), lung cancer (Pope et al., 2002; Chen et al., 2008; Hamra et al., 2014), cardiopulmonary conditions (Krewski et al., 2009) and diabetes (Li et al., 2014). These studies have used a variety of methods to estimate air pollutant exposure and have developed deterministic and/or probabilistic models as the cornerstone of their analyses. In Mexico, the majority of epidemiological studies on the health effects of air pollutants have estimated pollutant exposure using proximity methods, such as those based on data from the nearest air quality monitor (Rojas-Martínez et al., 2007; Barraza-Villarreal et al., 2008; Escamilla-Nuñez et al., 2008; Hernández-Cadena et al., 2009), averages by city or municipality

(Téllez-Rojo et al., 2000; Carbajal-Arroyo et al., 2011), or Inverse Distance Weighting (IDW) interpolation (Riojas-Rodríguez et al., 2014; Trejo-González et al., 2019). One disadvantage of these approaches is their failure to capture the spatial variability of exposures caused by local pollution sources, urban topography and local meteorological factors (Jerrett et al., 2005), which leads to considerable uncertainty regarding accurate exposure estimation. To the best of our knowledge, the available literature offers few epidemiological studies which have employed more sophisticated, high-resolution methods (e.g., based on satellite data) for estimating exposure to ambient air pollutants in the Mexican population (Rosa et al., 2017; Gutiérrez-Ávila et al., 2018; Téllez-Rojo et al., 2020). Furthermore, previous studies have noted that human exposure to air pollutants can differ between individuals due to variable factors such as time spent outdoors and mobility patterns, thereby introducing bias into individual-level exposure estimates in epidemiological analyses (Setton et al., 2011; Nyhan et al., 2019; Yu et al., 2020).

To offset these limitations, our study has geocoded the home and work addresses of the study population and fitted generalized additive models (GAMs), which predict  $\text{PM}_{2.5}$  and  $\text{NO}_2$  concentrations at both locations. The principal advantage of the GAMs is their use of semiparametric methods to model non-linear functions through penalized splines (Yanosky et al., 2008).

This study represents a refinement of exposure models previously developed and applied in the MCMA (Just et al., 2015; Rivera-González et al., 2015; Son et al., 2018). In order to craft a feasible model adaptable to the limited information available in developing countries, we used the highest-quality

and largest amount of secondary source data available on air pollutant predictors in Mexico. This study therefore provides a novel methodological tool to be used in the Mexican Teachers' Cohort (MTC) study, as well as for subsequent epidemiological studies which seek to assess the effects of PM<sub>2.5</sub> and NO<sub>2</sub> exposures on health. Our aim was to determine the level of monthly exposure to outdoor PM<sub>2.5</sub> and NO<sub>2</sub> of approximately 16 500 MTC participants living and working in the MCMA between 2004 and 2019.

## 2. Material and methods

We developed spatiotemporal prediction models for PM<sub>2.5</sub> and NO<sub>2</sub> using the pollutant concentrations at monitoring locations as the dependent variable, and meteorological and geographic variables, as well as other pollutants, as predictors. For statistical modeling, geostatistical modeling, geoprocessing and spatial analysis, we used R software version 3.5.3 (R Core Team, 2019).

### 2.1 Study population and geocoding

This study is nested within a larger study known as the MTC, an ongoing prospective observational study established in 2006-2008 that includes 115 314 female public-school teachers from 12 Mexican states, with a median age at enrollment of 44 years. A detailed description of the full cohort has been published elsewhere (Lajous et al., 2015). In brief, recruitment was divided into two phases: the first (2006) took place in two states (Jalisco and Veracruz,  $n = 27\,979$ ), and the second (2008) incorporated 10 additional states (Baja California, Chiapas, Mexico City [CdMx], Durango, Guanajuato, Hidalgo, Mexico State [EdoMex], Nuevo León, Sonora and Yucatán,  $n = 87\,335$ ). Participants completed a reference questionnaire with sociodemographic characteristics (including home address [HA] and work address [WA]), reproductive history, diet, lifestyle and health status. Follow-up takes place every 3-4 years to obtain information on any disease diagnoses and update participant risk factor profiles. The first follow-up cycle (2011-2013) obtained a response rate of 83%, and the second cycle (2014-2020, ongoing) of 61%.

For the present study, we selected a subsample of the MTC which excluded participants with the following characteristics: (1) from states other than

CdMx or EdoMex ( $n = 92\,571$ ), (2) without a valid HA or WA on file ( $n = 6071$ ), or (3) with an HA or WA unable to be geocoded inside the area of influence of the MCMA environmental monitoring network ( $n = 265$ , methods described below). The analysis included a total study population of  $n = 16\,407$  MTC participants.

The original MCMA is composed of 16 municipalities of CdMx, 59 of EdoMex and one of Hidalgo. We did not include the entire MCMA due to the low/under-representation of pollutant measurements at locations far from monitoring stations, which do not achieve full coverage of the area (Fig. 1). Based on this determination, we drew the boundaries of the analysis area using the extreme coordinate position of a 5 km radius around each of the most peripheral monitoring stations. Within these boundaries, we defined a grid composed of  $1 \times 1$  km cells (in gray, Fig. 1).

We geocoded the HA and WA of each participant within the study population. For HA, we geocoded zip codes (ZC) as reported at baseline. Based on the Mexican Postal Service (SEPOMEX) criteria for valid ZCs, we selected only those containing five digits and verified that the first two corresponded to the codes of home municipalities also reported by teachers. Finally, we used the geographical layer of SEPOMEX's ZCs (Gobierno de México, n.d.) to assign the coordinates of the ZC's centroid to each HA. Since information on HA changes was unavailable for before and after the baseline, we assumed that HA were constant throughout the follow-up period. To geocode WAs at baseline and follow-ups, we used the workplace codes provided by the National Teaching Career Program. The geographic coordinates (longitude and latitude) were matched from the National School Information System (SEP, n.d.).

### 2.2 Meteorological and air pollutant data

We obtained databases of hourly pollutant measurements (PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub> and O<sub>3</sub>) and relevant meteorological variables (relative humidity, temperature and wind speed) for all monitoring sites within the area of interest from 1:00 LT on January 1, 2004 to 12:00 LT on July 31, 2019. Datasets were provided by the monitoring network of Mexico City's Environmental Ministry and remain permanently available for public consultation on the official websites (Gobierno de la Ciudad de México, n.d.).

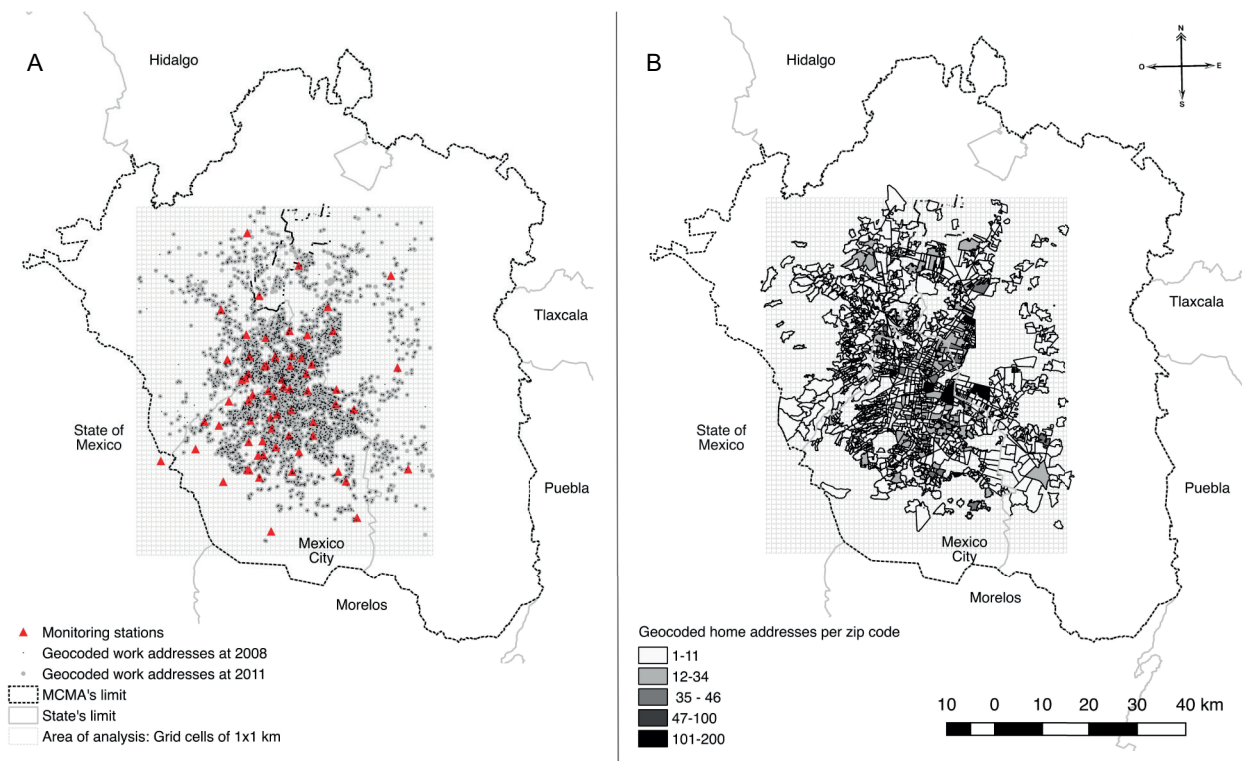


Fig. 1. Area of analysis and geocoded addresses. (a) Location of work addresses at baseline (2008) and first follow-up cycle (2011–2013). (b) Location of home addresses and number of teachers per zip code.

We calculated daily and monthly estimators (averages) for each pollutant and meteorological variable at each monitoring station where a minimum of 75% of the required observations were available (18 h and 23 days, respectively). Otherwise, we recorded data as missing and those observations were not part of the complete case analysis (more details in the statistical analysis section). Seasons, used as a categorical variable, were defined as follows: the cold-dry season from November to February, the hot-dry season from March to May, and the rainy season from June to October.

### 2.3 Geographic data and other covariables

Geographic profiles of the monitoring sites were characterized using a geographic information system (GIS). We created a geographical layer of the air quality monitoring stations in place from 2004 to 2019 in the study area, based on coordinates provided by the National Air Quality Information System (SINAICA) (INECC, n.d.). The altitude variable was provided by SINAICA in meters above sea level (masl) and cate-

gorized into tertiles as follows: low as  $\leq 2300$  masl, medium as  $\geq 2301$  and  $\leq 2500$  masl, and high as  $\geq 2501$  masl. The vehicle motorization index (number of motorized vehicles per 1000 inhabitants) was calculated annually at the municipal level, based on population projections from the National Population Council (Gobierno de México, n.d.) and historical databases of registered motor vehicles in circulation (automobiles, passenger buses, cargo trucks and motorcycles), according to the Mexican National Institute of Statistics, Geography and Informatics (INEGI, n.d.). This variable was categorized into tertiles as follows: low as  $\leq 222$  vehicles per 1000 inhabitants, medium as  $\geq 223$  and  $\leq 487$  vehicles per 1000 inhabitants, and high as  $\geq 488$  vehicles per 1000 inhabitants.

### 2.4 Statistical analysis

#### 2.4.1 Generalized additive models (GAMs) and validation

To determine the contribution of each predictor to  $PM_{2.5}$  and  $NO_2$  variability at monitoring stations

between 2004 and 2019, we fitted two independent GAMs (Eq. [1])

$$E(Y|X_1, X_2, \dots, X_p) = \alpha_0 + \sum_{j=1}^p f_j(X_j) \quad (1)$$

where  $f_j(\cdot)$  are smooth functions,  $Y$  is the vector of outcomes and the  $X_j$  are the vectors of covariates. We performed a complete case analysis for each model, using monthly PM<sub>2.5</sub> or NO<sub>2</sub> concentrations per monitoring site as the dependent variable. Due to its skewed distribution, the vector PM<sub>2.5</sub> was log-transformed. Models only included covariables expected a priori to exert a physical influence on the PM<sub>2.5</sub> and NO<sub>2</sub> levels. Predictors introduced in the adjusted models were the group of other pollutants, the meteorological variables and the geographical variables. To account for the correlation between pollutants, we included them with interaction terms. To ensure a parsimonious model specification, we tested the contribution of each term by eliminating it in the model and retained only those which improved predictive accuracy and showed statistical significance ( $p < 0.05$ ). Model comparison and selection were performed according to deviance explained (DE), Akaike information criterion and root mean square error (RMSE). To test the goodness of fit, we performed a residual diagnosis allowing us to verify normality and the absence of any overdispersion patterns. Finally, we performed stepwise selection to test and select the number of base functions needed to achieve a smooth curve which would maximize our data fit and reduce RMSE.

To test the GAMs for possible overfitting and evaluate their predictive accuracy, we carried out a 10-fold cross-validation and determined the predictive accuracy of each model based on its cross-validation root mean square error (CV-RMSE).

#### 2.4.2 Model predictions and geostatistical conditional simulation

We used the GAM coefficients to predict the monthly concentrations of PM<sub>2.5</sub> and NO<sub>2</sub> in each  $1 \times 1$  km grid cell from January 2004 to July 2019. For each grid cell and for each month of the study period, we carried out 3000 simulation replicates using the Geostatistical Conditional Simulation (GCS) method (ESRI, n.d.) with the aim to obtain the mean value of the predictor variables: PM<sub>10</sub>, O<sub>3</sub>, SO<sub>2</sub>, NO<sub>2</sub>, wind

speed, relative humidity and temperature in each cell. As part of the GCS, we fitted Gaussian and exponential variograms with and without the nugget effect to empirical semivariograms in order to characterize the spatial structures at unmeasured locations (Eq. [2]). To this end, we used the Cholesky decomposition of covariance matrix (Chilès and Delfiner, 2012):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(X_i) - Z(X_i + h)]^2 \quad (2)$$

where the degree of spatial dependence among the values of attribute  $Z$  in two different locations or points in space was semivariance  $\gamma(h)$ , and the distance between the points was lag  $h$ .

To account for error  $\epsilon_{ij}$  in the GAMs caused by the spatial correlation of the observations, we carried out the same simulation procedure with the residuals before adding them to the predicted values for each grid cell.

#### 2.5 Assigning of cohort participant exposures

To maximize the precision of exposure estimates, we accounted for the intra-city mobility of study participants by assigning predicted monthly values for PM<sub>2.5</sub> and NO<sub>2</sub> for each month during the study period to the grid cell where the HA and WA of each teacher were located. With the assumption that mobility-based estimates ( $HWA_{weig}$ ) were closer to actual exposure (individual monitoring), HA and WA exposures were weighted according to the number of hours spent at each address, assuming the average length of the Mexican workday (8 h) (Eq. [3]).

$$HWA_{weig} = 2HA/3 + WA/3 \quad (3)$$

To assess the relevance of mobility to our study, we compared HA and WA exposures by calculating the difference between  $HA - HWA_{weig}$ . The null hypothesis (a difference of zero) was proved using a mixed-effects model, where the teacher was the grouping variable.

## 3. Results

### 3.1 Study population and geocoding

The national-level MTC included 115 314 total participants, of which 22 743 lived either in the CdMx or EdoMex. From this subsample, we were able to

geocode HA and WA within the MCMA of 73% ( $n = 16672$ ). In terms of HAs, the median of teachers per ZC was 23, with a range of one to 175 (Fig. 1). The areas of these ZCs ranged from 0.005 to 17.003 km<sup>2</sup>. Therefore, the final study population included 16407 MTC participants who lived and worked in the area of analysis (Fig. 2), covering (totally or partially) 72 municipalities in the CdMx or EdoMex.

### 3.2 Descriptive statistics

Monthly averages of PM<sub>2.5</sub>, NO<sub>2</sub> and their predictor variables at the monitoring site level are summarized in Table I. During the 2004-2019 period, a maximum of 11 and 17 monitoring stations for PM<sub>2.5</sub> and NO<sub>2</sub>, respectively, provided data on all variables required to adjust the GAMs. The geometric mean and SD of monthly PM<sub>2.5</sub> concentrations were 3.18 and 0.27  $\mu\text{g m}^{-3}$ , respectively, and the mean and SD of monthly NO<sub>2</sub> concentrations were 28.34 and 8.59 ppb, respectively. Over the study period, the mean ambient temperature was around 17 °C (range 5, 23), wind speed averaged 1.85 m/s (range 0.49, 4.95) and average humidity was 55% (range 28, 88). The lowest concentrations of both pollutants were registered during the rainy season, and the highest during the cold-dry season. PM<sub>2.5</sub> concentrations were greatest at the lowest altitudes.

### 3.3 Generalized additive model for PM<sub>2.5</sub> and predicted values

The PM<sub>2.5</sub> model predictors were altitude, relative humidity, wind speed, season and year, in addition to the pollutants PM<sub>10</sub> and NO<sub>2</sub>. Only PM<sub>10</sub>, relative humidity and wind speed were included as continuous variables with smoothing functions, using penalized splines of up to seven degrees of freedom (Fig. 3).

The maps in Figure 4 illustrate the average spatial distribution of the monthly PM<sub>2.5</sub> levels predicted by the model per year on the surface of the area of analysis (the grid). In general, predicted levels were higher further north within the MCMA and trended downwards until 2014. The 2019 map considers only measurements until July; therefore, the rainy season, which yielded the lowest PM<sub>2.5</sub> levels (see Table I), was underrepresented. The percentage of deviance explained by the model was high (DE = 87.3%) for the 2004-2019 period (Table I). The RMSE in the sample indicated high predictive accuracy with a low level of error (RMSE = 0.100), and showed no overfitting of data, as determined by its proximity to the cross-validation error (CV-RMSE = 0.102).

### 3.4 Generalized additive model for NO<sub>2</sub> and predicted values

The NO<sub>2</sub> model predictors were the vehicle motorization index, wind speed, temperature, season and

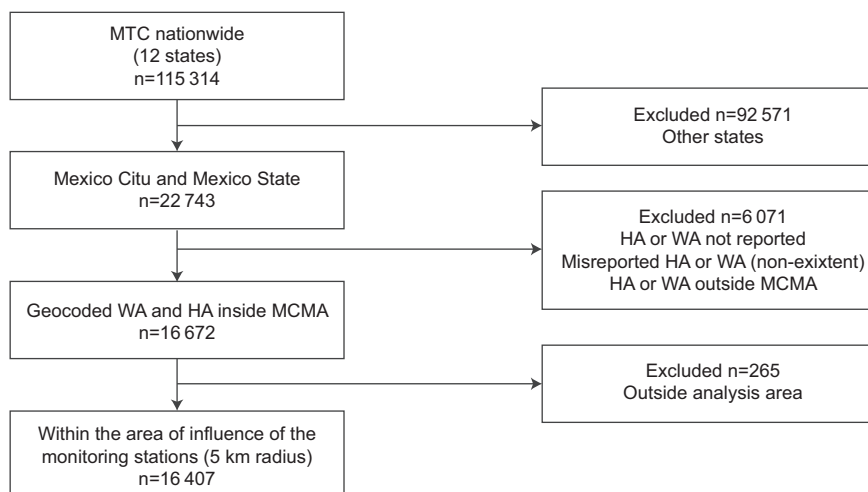


Fig. 2. Flow chart of geocoding process and identification of study population. (MTC: Mexican Teachers' Cohort; MCMA: Mexico City Metropolitan Area; SE-POMEX: Mexican Postal Service; HA: home address; WA: work address; WC: workplace code).

Table I. Descriptive statistics for final GAM predictor variables, and bias and precision statistics from cross-validation.

Variable	Time-varying	Model	
		PM <sub>2.5</sub> mean (SD) n = 779	NO <sub>2</sub> mean (SD) n = 1298
Monitoring stations included 2004-2019 period (min-max)		3-11	6-17
PM <sub>2.5</sub> (µg m <sup>-3</sup> )	Yes	25.05 (6.90)	---
NO <sub>2</sub> (ppb)	Yes	29.83 (8.02)	28.34 (8.59)
O <sub>3</sub> (ppb)	Yes	---	27.95 (7.14)
PM <sub>10</sub> (µg m <sup>-3</sup> )	Yes	49.67 (17.15)	50.41 (18.66)
SO <sub>2</sub> (ppb)	Yes	---	6.93 (4.23)
Temperature (°C)	Yes	---	16.71 (2.18)
Wind speed (m/sec)	Yes	1.90 (0.50)	1.81 (0.47)
Relative humidity (%)	Yes	52.54 (11.40)	---
Season <sup>a</sup>	No		
Rainy (ref)		19.73 (3.90)	23.42 (6.35)
Cold-dry		29.12 (6.30)	33.13 (8.62)
Hot-dry		27.90 (5.88)	28.83 (7.39)
Motorization index (no. vehicles/1,000 ha) <sup>a</sup>	Yes		
High (ref)		---	27.60 (7.47)
Medium		---	29.35 (8.77)
Low		---	27.52 (9.68)
Altitude (m) <sup>a</sup>	Yes		
High (ref)		19.80 (4.53)	---
Medium		24.96 (6.10)	---
Low		26.36 (7.20)	---
Year <sup>a</sup>			
2004 (ref)		28.32 (5.75)	33.97 (7.15)
2005		29.26 (9.94)	34.02 (8.86)
2006		29.09 (6.41)	33.14 (7.22)
2007		28.16 (4.92)	33.04 (7.64)
2008		26.80 (7.10)	31.73 (7.17)
2009		24.07 (5.88)	30.70 (7.25)
2010		23.63 (8.01)	32.14 (8.49)
2011		24.87 (7.95)	28.06 (8.88)
2012		23.39 (4.49)	27.78 (8.77)
2013		25.33 (7.47)	25.98 (7.96)
2014		23.22 (6.14)	26.06 (7.47)
2015		24.96 (5.75)	23.92 (7.21)
2016		24.39 (6.97)	24.80 (7.66)
2017		24.66 (7.75)	25.74 (8.25)
2018		24.05 (5.91)	25.72 (7.30)
2019		24.15 (6.37)	22.20 (6.67)
Statistics		Model	
		PM <sub>2.5</sub>	NO <sub>2</sub>
DE		87.3%	74.4%
RMSE		0.100	4.406
CV-RMSE		0.102	4.497

<sup>a</sup>Variable included is categorical; however, for descriptive purposes, mean (SD) is presented for the dependent variable in each category.

DE: deviance explained; RMSE: root mean square error; CV: cross-validation.

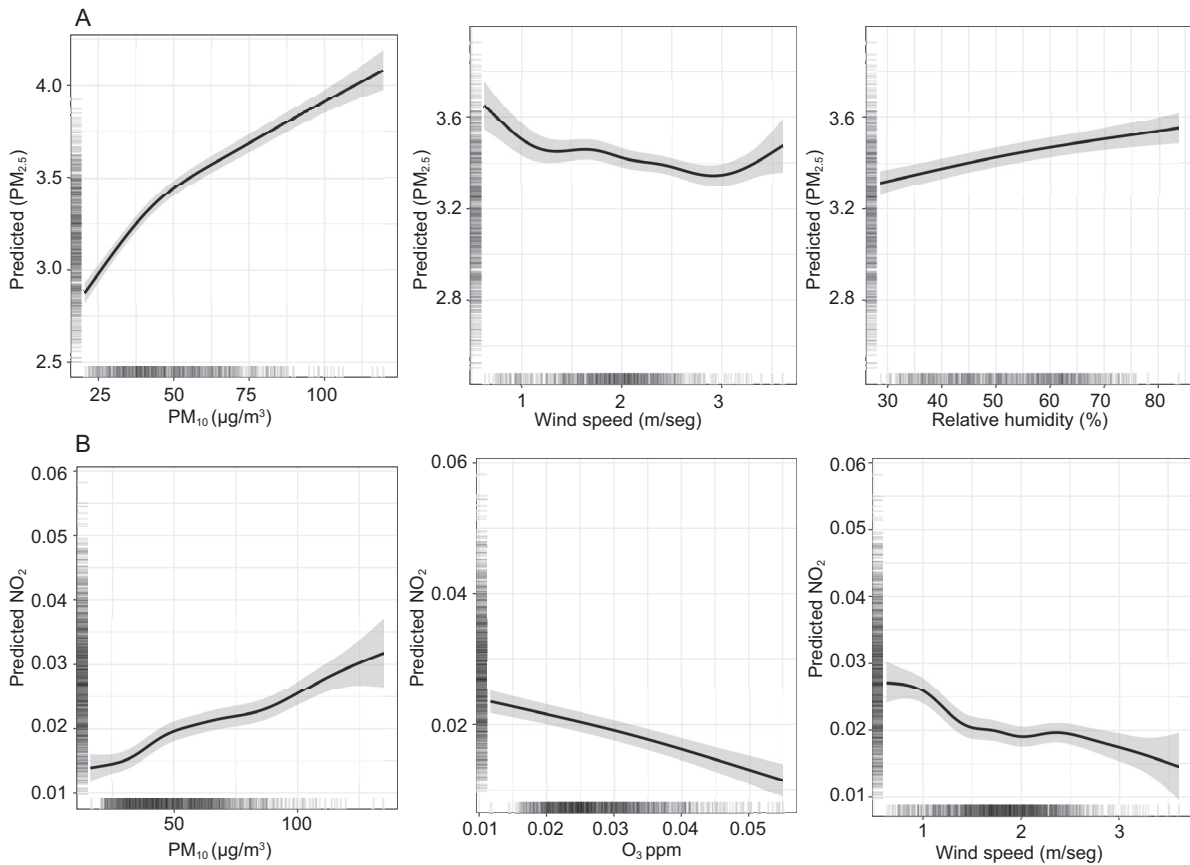


Fig. 3. Smooth plots from the final models showing non-linear effects of selected predictors and associated predicted  $PM_{2.5}$  (uptrend) and  $NO_2$  (downtrend) concentrations. CI is set at two standard errors above and below the mean.

year, in addition to the pollutants  $PM_{10}$ ,  $O_3$  and  $SO_2$ . Only  $PM_{10}$ ,  $O_3$  and wind speed were included as continuous variables with smoothing functions using penalized splines of up to seven degrees of freedom, since the other continuous variables showed a linear relationship with the dependent variable (Fig. 3). The maps in Figure 4 show the model predictions of average spatial distribution of monthly levels of  $NO_2$  within the study area per year. In general, the predicted levels of  $NO_2$  were highest in the center of the MCMA and trended downwards until 2016, with a mean of 16.87 ppb (SD = 7.77) within the study area. As with  $PM_{2.5}$ , only measurements until July 2019 were considered. The percentage of deviance explained by the model was intermediate (DE = 74.4%) for the 2004-2019 period (Table I). The RMSE in the sample indicated high predictive accuracy with a low level of error (RMSE = 4.406) and showed no

overfitting of data, as determined by its proximity to the cross-validation error (CV-RMSE = 4.497).

### 3.5. Exposure assignment adjusted for mobility

All teachers reported a HA different from their WA, which strengthens the case for considering mobility when estimating air pollutant exposure. For the time period and population analyzed, the ranges of differences in estimated exposure at the homes and workplaces of participants were 0-6.01  $\mu\text{g m}^{-3}$  for  $PM_{2.5}$  and 0-14.76 ppb for  $NO_2$ . Only 0.11% ( $n = 1799$ ) of teachers changed WAs between the first and second follow-up cycles. On average, they traveled 6.285 km (SD = 6.216) between their homes and their workplaces (min: 0.0098 km, Q1: 1.846 km, Q2: 4.282 km, Q3: 8.746 km, max: 56.155 km).

We obtained 187 weighted monthly averages of  $PM_{2.5}$  and 187 of  $NO_2$  assigned to each teacher as



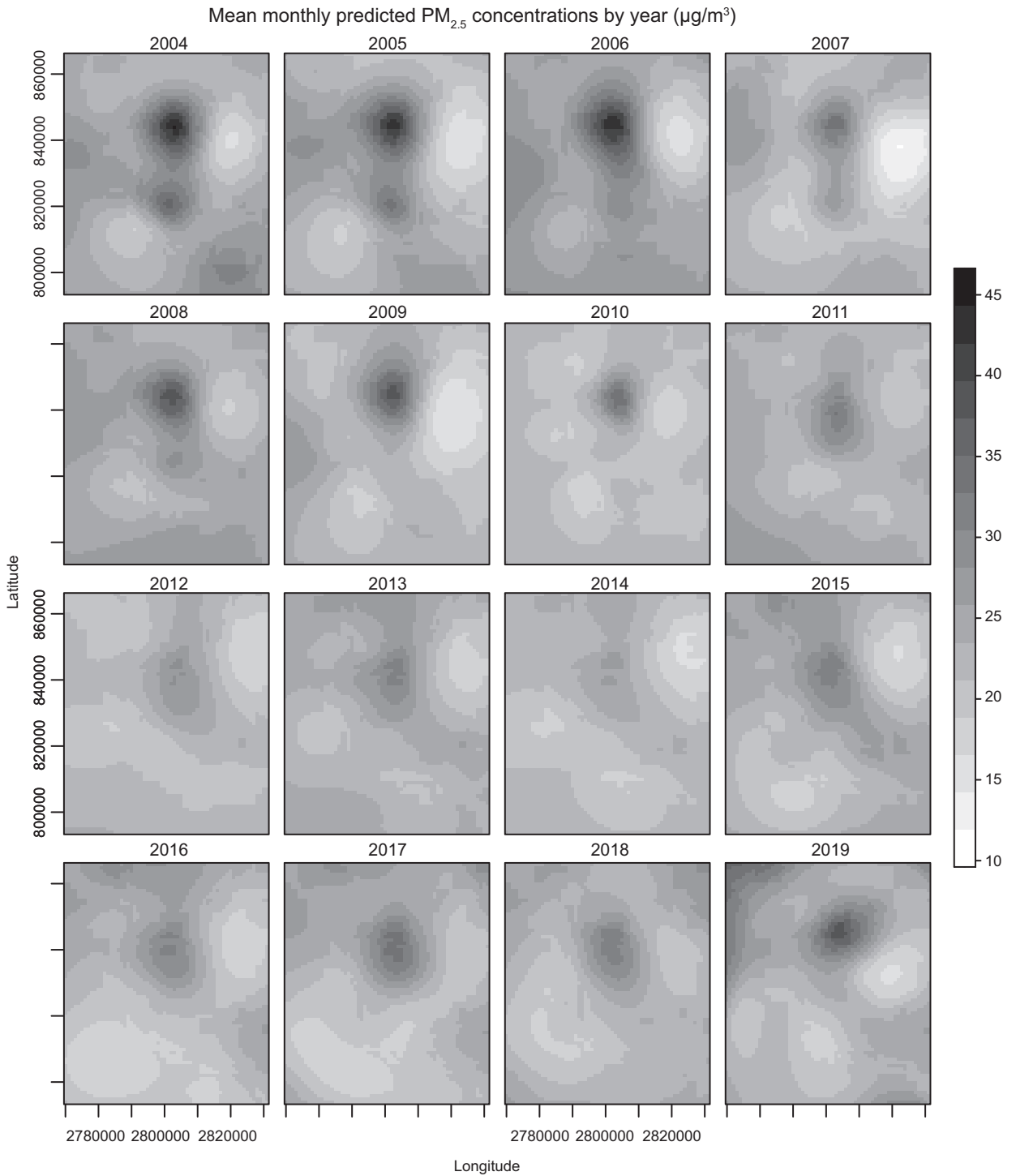


Fig. 4. Raster of the annual average of the monthly PM<sub>2.5</sub> and NO<sub>2</sub> predicted concentrations on the surface of the area of analysis (grid made up of 1 × 1 km cells) (2004-2019).

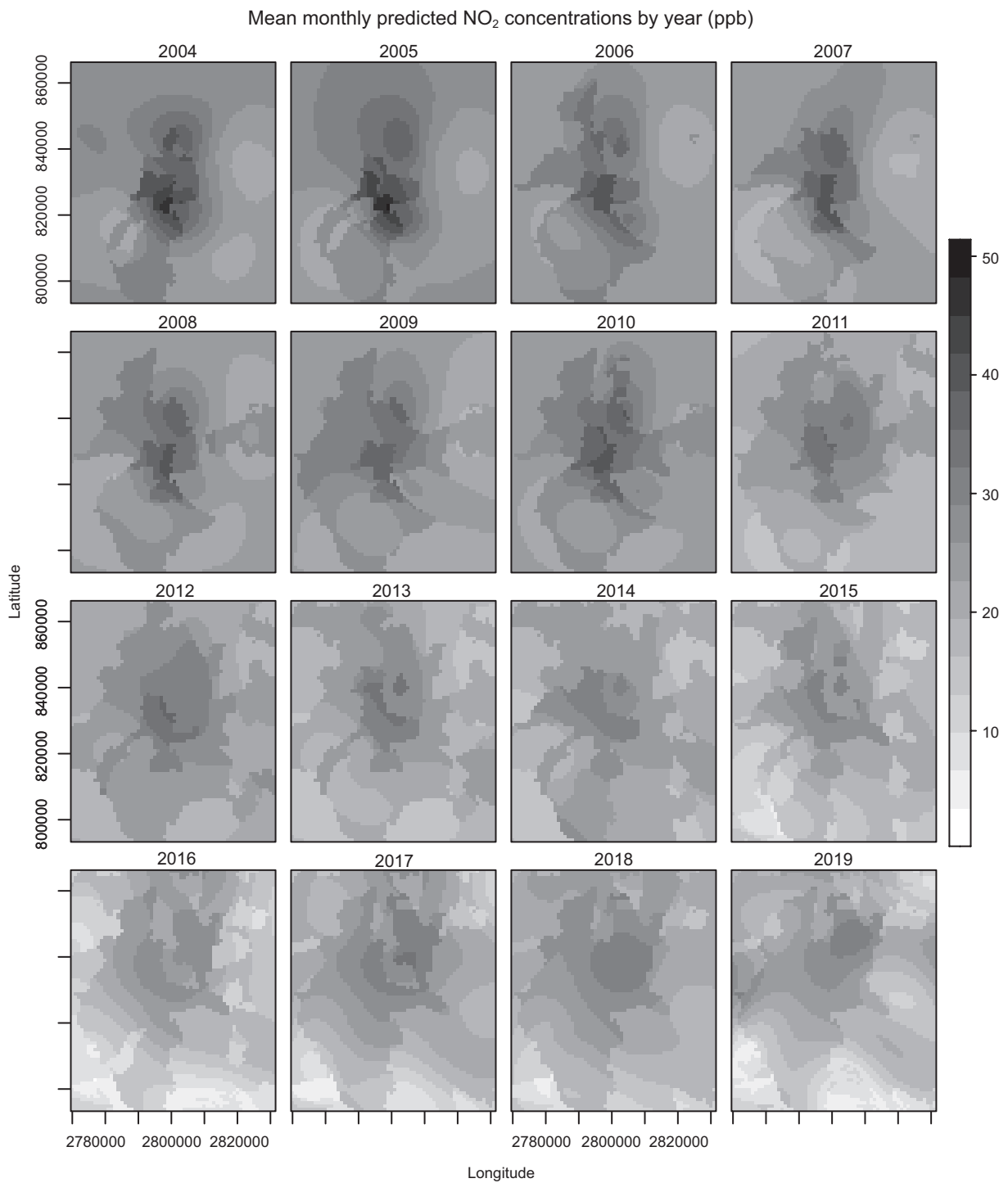


Fig. 4. Raster of the annual average of the monthly PM<sub>2.5</sub> and NO<sub>2</sub> predicted concentrations on the surface of the area of analysis (grid made up of 1 × 1 km cells) (2004-2019).

a proxy for long-term exposure. Figure 5 illustrates the annual distribution and principal statistics for the weighted monthly averages ( $HWA_{weig}$ ) assigned to the 16 407 MTC participants during the 2004–2019 period. For both pollutants, we determined the expected exposure variability attributable to the spatial distribution of pollutants and the location of HAs and WAs within the study area. On average, for the population and period analyzed, weighted exposures were  $24.38 \mu\text{g m}^{-3}$  (SD = 6.78) to PM<sub>2.5</sub> and 28.21 ppb (SD = 8.00) to NO<sub>2</sub>.

Average exposures to both pollutants were lower at HAs than at  $HWA_{weig}$  (PM<sub>2.5</sub>: HA  $24.37 \mu\text{g m}^{-3}$ ,  $HWA_{weig}$   $24.39 \mu\text{g m}^{-3}$ ; NO<sub>2</sub>: HA 28.15 ppb,  $HWA_{weig}$  28.21 ppb). Average differences were  $-0.016$  (range  $-2.00$ – $1.91$ ) for PM<sub>2.5</sub> and  $-0.063$  (range  $-4.92$ – $4.60$ ) for NO<sub>2</sub>. In both cases, the differences between HA and  $HWA_{weig}$  did not equal zero and were statistically significant ( $p < 0.0001$ ).

#### 4. Discussion

The models developed in our study provide a useful method for predicting monthly exposure to outdoor PM<sub>2.5</sub> and NO<sub>2</sub> at any location within large areas throughout the MCMA, with high spatial ( $1 \times 1 \text{ km}$ ) and temporal (16 years: 2004–2019) resolution. These models performed well and delivered high predictive accuracy, with cross-validation errors slightly higher than those of the models (CV-RMSE = 0.102 for PM<sub>2.5</sub> and CV-RMSE = 4.497 for NO<sub>2</sub>). A large difference between these values would have suggested that GAMs overfit the data, as performance would have been lower in data not included in the adjustment during cross-validation. Because of their high resolution, these models satisfy the permanent need of epidemiological studies to predict and assign exposures to ambient air pollutants in numerous MCMA locations with a high degree of precision. Furthermore, they cover a long time period (16 years) on a small temporal scale (monthly), allowing the assessment of effects of medium- and long-term exposures.

To our knowledge, only three published studies have proposed models for determining exposure to ambient air pollutants in Mexico (Just et al., 2015; Rivera-González et al., 2015; Son et al., 2018), of which all focus on Mexico City and/or the MCMA.

Rivera-González et al. (2015) provide a comparison of four basic methods based on proximity and interpolation for all criteria pollutants. The principal limitation of several of these approaches is that they do not capture the spatial variability of exposure (Jerrett et al., 2005). Our study overcomes this limitation by using high spatial resolution, which captures much greater variability. Son et al. (2018) proposed a mixed-effects regression model based on land use for all criteria pollutants. Although this study follows the traditional assumption of a linear relationship between predictors and the dependent variable, it is innovative in that it explores the statistical model-fitting techniques of least absolute shrinkage and selection operator (LASSO) method, in addition to analyzing various spatiotemporal scales (hourly, daily, monthly, semi-annual and annual). Instead of this model, our study proposes GAMs, which maximize the quality of dependent variable prediction by accounting for its non-linear relationship with the predictors, and incorporate smoothed (non-parametric) terms using penalized splines. However, we explored a monthly exposure scale for only two criteria pollutants. Finally, Just et al. (2015) developed a more sophisticated mixed-effects regression model, including satellite-based aerosol optical depth measurements as primary PM<sub>2.5</sub> predictor, over a period of 10 years. While innovative, this method is limited in that it cannot be applied where little or no PM<sub>2.5</sub> monitoring data exist for validation, and also because no satellite measurements can be taken on cloudy days (Paciorek and Liu, 2009; Bravo et al., 2012). We overcame this first limitation, an obstacle faced by many developing countries, by using information on predictors from secondary sources which are easily accessed and interpreted, and are permanently monitored. Furthermore, we covered a longer time span (16 years) and also modeled NO<sub>2</sub> exposure. Despite important distinctions from previous studies, we obtained largely similar results; predicted spatial patterns for PM<sub>2.5</sub> were comparable and, on average, predicted concentrations for both pollutants (PM<sub>2.5</sub> and NO<sub>2</sub>) were only slightly lower.

The results of our analyses demonstrate the relevance of using meteorological variables and other pollutants as primary PM<sub>2.5</sub> and NO<sub>2</sub> predictors. This could indicate the feasibility of generating highly predictive models using only secondary

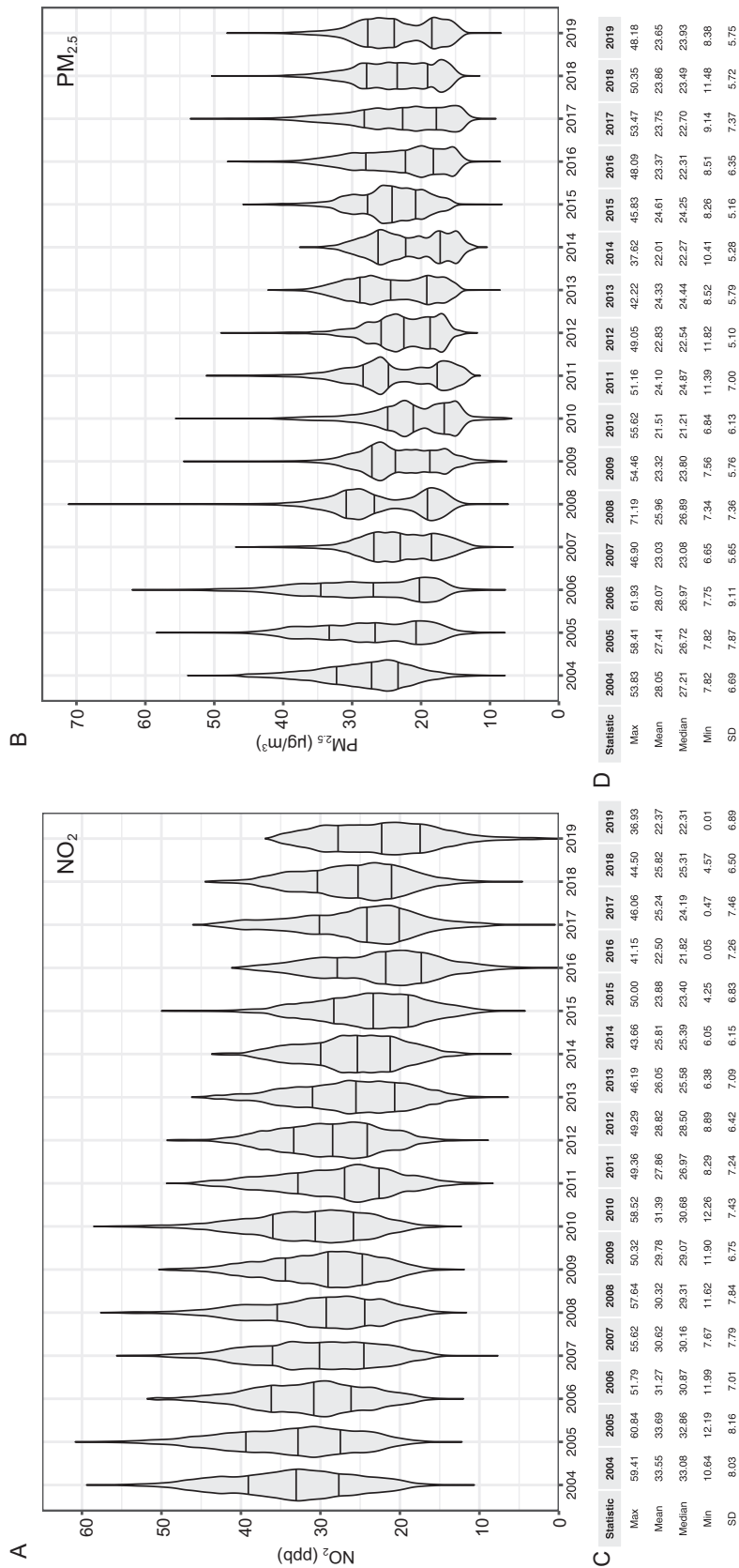


Fig. 5. (b, d) Violin plot and descriptive statistics for assignment of monthly  $PM_{2.5}$  (B and D) and  $NO_2$ . (a, c) weighted exposures () by year. First, second and third quartiles estimated for  $n = 16407$  teachers.

monitoring data from local networks. Our study therefore confirms that, despite limited availability of information, it is possible to design feasible methods for predicting outdoor exposure to air pollutants, including for other pollutants (such as PM<sub>10</sub>) when no data is available on PM<sub>2.5</sub>, as is the case in multiple cities within developing countries. In terms of spatiotemporal prediction of PM<sub>10</sub>, we expect to observe predictive accuracy as high as that of PM<sub>2.5</sub>, since in urban areas both fractions of PM have similar potential sources, though in different proportions (Karagulian et al., 2015). Furthermore, geographical predictors are a key consideration given the nature of the contaminants of interest and the geographical characteristics of the study area. In the two models we developed, the geographical predictors showed smaller coefficients than the other predictors. In addition to altitude and the vehicle motorization index, we generated and tested other geographical variables. The linear meters of different types of avenues in buffers of 100-, 300- and 500-m diameters around the monitoring stations did not prove to be significant and therefore were not included in the final models. We tested the UV-B radiation variable and found a significant relationship with NO<sub>2</sub>; however, it was excluded from the final model due to insufficient monitoring sites in the study area (only 0-6 stations during the 2004-2019 period) which did not allow a semivariographic analysis, causing an increase of 0.583 units of CV-RMSE in the NO<sub>2</sub> model (CV-RMSE including UV-B = 3.914).

A recent study which analyzed and compared the home-work mobility patterns of approximately 400 000 individuals, affirmed that ignoring mobility causes erroneous classification when estimating health effects (Nyhan et al., 2019). The principal strength of our study was in considering home-work mobility to estimate PM<sub>2.5</sub> and NO<sub>2</sub> exposures. Given the findings of Nyhan et al. (2019), we expect that this consideration will help reduce measurement errors that could lead to null association measurements in subsequent epidemiological studies. Another strength of our study was the method used to ascertain the spatial distribution of predictor variables at unmeasured sites. Although no existing method can offer an exact estimate of an unknown reality, GCS considers the dispersion of the phenomenon analyzed and allows

the obtention of one among the possible spatial of a random function (surface). The simulated variables thus present the same variability and correlation characteristics as the observed data (measured according to their mean, variance and semivariogram) (Goovaerts, 1997; Deutsch and Journel, 1998). Other potential methods considered included IDW interpolation or kriging; however, despite their extensive application in earth and environmental sciences, these provide only a smooth image of reality, whereas GCS assumptions include seasonality and a normal data distribution (Goovaerts, 1997; Deutsch and Journel, 1998).

We are aware that inherent precision and measurement errors exist in predictive models such as the one we developed, which depend not only on mobility but also on the residential history of participants and precise geocoding of addresses. Prior studies have suggested that the geocoding technique can influence health effect estimates when using fine-scale exposure models (Jacquemin et al., 2013). Therefore, we recognize three sources of error. First, we assumed that the HAs of cohort participants were permanent throughout follow-ups, supported by a previous study by Pimienta-Lastra and Toscana-Aparicio (2019) that registered low inter-municipal migration within the MCMA (4.8%) from 2010 to 2015 (Pimienta-Lastra and Toscana-Aparicio, 2019), thereby suggesting that our study population was similarly unlikely to change addresses. Second, we assessed HAs by geocoding the centroids of their corresponding ZCs as opposed to using exact HA coordinates, and ZC areas varied widely (0.005-17.003 km<sup>2</sup>) among participants. Previous studies have documented that estimating air pollution exposure by certain zones such as zip codes implies greater uncertainty, as compared to estimation by exact home addresses (Kinnee et al., 2020); however, since home addresses were not available, we used zip code centroids as a proxy of participant residence. Finally, we assumed that the long-term outdoor exposure levels determined were also a reasonable representation of interior and microenvironment exposure.

The models developed were subject to various limitations. First, determination of the study area and population was restricted by the quality of secondary available data (the geographical layer of ZCs) to geocode the HAs of the MTC participants.

Although the cohort was extended across 12 Mexican states, we were able to partially include only two of them in this study. Second, given the suboptimal quality of available data, we did not include the characterization of pollution sources attributable to land use and vehicular traffic emissions. Nonetheless, as a proxy for the latter, we used a vehicle motorization index created with official data on registered vehicles at the municipal level. This data was used under the assumption that it would stay constant within each municipality, and inter-municipal mobility was not considered. The vehicle motorization index variable proved significant in the NO<sub>2</sub> model, but not the PM<sub>2.5</sub> model, which is consistent with our theoretical framework given that NO<sub>2</sub> is primarily generated by fuel combustion in vehicles. The result of including aggregated variables to the models may have been a misclassification of predictions' spatial variability, as a result of the dilution of all possible values within each aggregated area. Finally, another important limitation to consider was the lack of monitoring data, which limits the type of statistical and spatial analyses possible. Air pollution models require datasets with sufficient data for training models and validation, therefore less available data limits the statistical techniques than can be applied to achieve prime model performance or cross-validation.

## 5. Conclusions

In conclusion, it is crucial for environmental epidemiology to apply methods for the estimation of ambient air pollutant exposure that can overcome limited data availability, which is a particularly relevant factor in developing countries. Our results suggest that the model presented herein offers a promising alternative to be used in epidemiological studies for predicting PM<sub>2.5</sub> and NO<sub>2</sub> exposure with high spatiotemporal resolution within the MCMA in the 2004-2019 period. Beyond the MCMA context, these models can be replicated in other cities using similar information sources. Our results highlight the importance of developing pollutant exposure models with high predictive accuracy based on available secondary data from official sources. Further research should focus on expanding these models to other Mexican cities.

## Acknowledgments

This study was conducted with support from the national scholarship program of the CONACyT National Program for Quality Postgraduate Studies, granted to the first author during her doctoral studies. The authors wish to thank Karla Mariana Valdez Trejo, M.Sc. for her support in editing this paper. We are indebted to the participants in the Mexican Teachers' Cohort for their outstanding cooperation. We also thank the Social Security and Services Institute for Employees of the State (ISSSTE) and Víctor Sastre from Mexico's Ministry of Education for their support over the years.

## References

- Barraza-Villarreal A, Sunyer J, Hernández-Cadena L, Escamilla-Núñez MC, Sienna-Monge JJ, Ramírez-Agular M, Cortez-Lugo M, Holguín F, Díaz-Sánchez D, Olin AC, Romieu I. 2008. Air pollution, airway inflammation, and lung function in a cohort study of Mexico City schoolchildren. *Environmental Health Perspectives* 116: 832-838. <https://doi.org/10.1289/ehp.10926>
- Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, Wolf K, Samoli E, Fischer P, Nieuwenhuijsen M, Vineis P, Xun WW, Katsouyanni K, Dimakopoulou K, Oudin A, Forsberg B, Modig L, Havulinna AS, Lanki T, Turunen A, Oftedal B, Nystad W, Nafstad P, De Faire U, Pedersen NL, Östenson C-G, Fratiglioni L, Penell J, Korek M, Pershagen G, Eriksen KT, Overvad K, Ellermann T, Eeftens M, Peeters PH, Meliefste K, Wang M, Bueno-de-Mesquita B, Sugiri D, Krämer U, Heinrich J, de Hoogh K, Key T, Peters A, Hampel R, Concin H, Nagel G, Ineichen A, Schaffner E, Probst-Hensch N, Künzli N, Schindler C, Schikowski T, Adam M, Phuleria H, Vilier A, Clavel-Chapelon F, Declercq C, Gironi S, Krogh V, Tsai M-Y, Ricceri F, Sacerdote C, Galassi C, Migliore E, Ranzi A, Cesaroni G, Badaloni C, Forastiere F, Tamayo I, Amiano P, Dorronsoro M, Katsoulis M, Trichopoulos A, Brunekreef B, Hoek G. 2014. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 383: 785-795. [https://doi.org/10.1016/S0140-6736\(13\)62158-3](https://doi.org/10.1016/S0140-6736(13)62158-3)
- Bravo MA, Fuentes M, Zhang Y, Burr MJ, Bell ML. 2012. Comparison of exposure estimation methods for air

- pollutants: Ambient monitoring data and regional air quality simulation. *Environmental Research* 116: 1-10. <https://doi.org/10.1016/j.envres.2012.04.008>
- Carbajal-Arroyo L, Miranda-Soberanis V, Medina-Ramón M, Rojas-Bracho L, Tzintzun G, Solís-Gutiérrez P, Méndez-Ramírez I, Hurtado-Díaz M, Schwartz J, Romieu I. 2011. Effect of PM<sub>10</sub> and O<sub>3</sub> on infant mortality among residents in the Mexico City Metropolitan Area: a case-crossover analysis, 1997-2005. *Journal of Epidemiology and Community Health* 65: 715-721. <https://doi.org/10.1136/jech.2009.101212>
- Chen H, Goldberg MS, Villeneuve PJ. 2008. A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. *Reviews on Environmental Health* 23: 243-297. <https://doi.org/10.1515/reveh.2008.23.4.243>
- Chilès J-P, Delfiner P. 2012. *Geostatistics: Modeling spatial uncertainty*. 2nd ed. Wiley, Hoboken, NJ.
- Deutsch CV, Journel AG. 1998. *GSLIB: Geostatistical software library and user's guide*, Version 2.0. 2nd. ed. Oxford University Press, New York.
- Dockery DW. 2001. Epidemiologic evidence of cardiovascular effects of particulate air pollution. *Environmental Health Perspectives* 109 Suppl 4: 483-486. <https://doi.org/10.1289/ehp.01109s4483>
- Escamilla-Núñez MC, Barraza-Villarreal A, Hernández-Cadena L, Moreno-Macías H, Ramírez-Aguilar M, Sienra-Monge JJ, Cortez-Lugo M, Texcalac JL, del Río-Navarro B, Romieu I. 2008. Traffic-related air pollution and respiratory symptoms among asthmatic children, resident in Mexico City: The EVA cohort study. *Respiratory Research* 9: 74. <https://doi.org/10.1186/1465-9921-9-74>
- ESRI. n.d. Key concepts of geostatistical simulation. ArcGIS. Available at: <https://desktop.arcgis.com/es/arcmap/latest/extensions/geostatistical-analyst/key-concepts-of-geostatistical-simulation.htm> (accessed on January 7, 2016).
- Gobierno de la Ciudad de México. n.d. Dirección de Monitoreo Atmosférico. Available at: <http://www.aire.cdmx.gob.mx/default.php> (accessed on May 8, 2019).
- Gobierno de México. n.d. Servicio Postal Mexicano. Available at: <https://www.gob.mx/correosdemexico> (accessed on July 4, 2017).
- Gobierno de México. n.d. Consejo Nacional de Población. Available at: <https://www.gob.mx/conapo> (accessed on December 4, 2019).
- Goovaerts P. 1997. *Geostatistics for natural resources evaluation*. Applied Geostatistics Series. Oxford University Press, New York.
- Gutiérrez-Ávila I, Rojas-Bracho L, Riojas-Rodríguez H, Kloog I, Just AC, Rothenberg SJ. 2018. Cardiovascular and cerebrovascular mortality associated with acute exposure to PM<sub>2.5</sub> in Mexico City. *Stroke* 49: 1734-1736. <https://doi.org/10.1161/STROKEAHA.118.021034>
- Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, Vineis P, Forastiere F, Saldiva P, Yori-fuji T, Loomis D. 2014. Outdoor particulate matter exposure and lung cancer: A Systematic review and meta-analysis. *Environmental Health Perspectives* 122: 906-911. <https://doi.org/10.1289/ehp.1408092>
- Hernández-Cadena L, Holguín F, Barraza-Villarreal A, del Río-Navarro BE, Sienra-Monge JJ, Romieu I. 2009. Increased levels of outdoor air pollutants are associated with reduced bronchodilation in children with asthma. *Chest* 136: 1529-1536. <https://doi.org/10.1378/chest.08-1463>
- Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, Kaufman JD. 2013. Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environmental Health* 12: 43. <https://doi.org/10.1186/1476-069X-12-43>
- INECC. n.d. Sistema Nacional de Información de la Calidad del Aire (SINAICA). Instituto de Ecología y Cambio Climático. Available at: <http://sinaica.inecc.gob.mx> (accessed on June 11, 2013).
- INEGI. 2015. Encuesta Intercensal 2015. Available at: <https://www.inegi.org.mx/programas/intercensal/2015/default.html#Documentacion> (accessed on June 7, 2018).
- INEGI. n.d. Instituto Nacional de Estadística y Geografía. Available at: <https://www.inegi.org.mx/> (accessed on December 4, 2019).
- INE-SEMARNAT. 2011. Cuarto almanaque de datos y tendencias de la calidad del aire en 20 ciudades mexicanas (2000-2009). 1a ed. Instituto Nacional de Ecología, México.
- Jacquemin B, Lepeule J, Boudier A, Arnould C, Benmerad M, Chappaz C, Ferran J, Kauffmann F, Morelli X, Pin I, Pison C, Rios I, Temam S, Künzli N, Slama R, Siroux V. 2013. Impact of geocoding methods on associations between long-term exposure to urban air pollution and lung function. *Environmental Health Perspectives* 121: 1054-1060. <https://doi.org/10.1289/ehp.1206016>

- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, Morrison J, Giovis C. 2005. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology* 15: 185-204. <https://doi.org/10.1038/sj.jea.7500388>
- Just AC, Wright RO, Schwartz J, Coull BA, Baccarelli AA, Téllez-Rojo MM, Moody E, Wang Y, Lyapustin A, Kloog I. 2015. Using high-resolution satellite aerosol optical depth to estimate daily PM<sub>2.5</sub> geographical distribution in Mexico City. *Environmental Science & Technology* 49: 8576-8584. <https://doi.org/10.1021/acs.est.5b00859>
- Karagulian F, Belis CA, Dora CFC, Prüss-Ustün AM, Bonjour S, Adair-Rohani H, Amann M. 2015. Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric Environment* 120: 475-483. <https://doi.org/10.1016/j.atmosenv.2015.08.087>
- Kinnee EJ, Tripathy S, Schinasi L, Shmool JLC, Sheffield PE, Holguin F, Clougherty JE. 2020. Geocoding error, spatial uncertainty, and implications for exposure assessment and environmental epidemiology. *International Journal of Environmental Research and Public Health* 17: 5845. <https://doi.org/10.3390/ijerph17165845>
- Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA, Thurston G, Calle EE, Thun MJ, Beckerman B, DeLuca P, Finkelstein N, Ito K, Moore DK, Newbold KB, Ramsay T, Ross Z, Shin H, Tempalski B. 2009. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Research Report (Health Effects Institute)* 140: 5-114; discussion 115-136.
- Lajous M, Ortiz-Panozo E, Monge A, Santoyo-Vistrain R, García-Anaya A, Yunes-Díaz E, Rice MS, Blanco M, Hernández-Ávila M, Willett WC, Romieu I, López-Ridaura R. 2015. Cohort profile: The Mexican Teachers' Cohort (MTC). *International Journal of Epidemiology* 46: e10. <https://doi.org/10.1093/ije/dyv123>
- Li C, Fang D, Xu D, Wang B, Zhao S, Yan S, Wang Y. 2014. Main air pollutants and diabetes-associated mortality: a systematic review and meta-analysis. *European Journal of Endocrinology/European Federation of Endocrine Societies* 171: R183-190. <https://doi.org/10.1530/EJE-14-0287>
- Nyhan MM, Kloog I, Britter R, Ratti C, Koutrakis P. 2019. Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. *Journal of Exposure Science & Environmental Epidemiology* 29: 238-247. <https://doi.org/10.1038/s41370-018-0038-9>
- OECD. 2016. The economic consequences of outdoor air pollution. Organization for Economic Cooperation and Development, Paris.
- Paciorek CJ, Liu Y. 2009. Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. *Environmental Health Perspectives* 117: 904-909. <https://doi.org/10.1289/ehp.0800360>
- Pimienta-Lastra R, Toscana-Aparicio A. 2019. Migración intermunicipal permanente de la Zona Metropolitana del Valle de México 2010-2015. *Ciencia Ergo Sum* 26: 1-19. <https://doi.org/10.30878/ces.v26n3a2>
- Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* 287: 1132-1141. <https://doi.org/10.1001/jama.287.9.1132>
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Riojas-Rodríguez H, Álamo-Hernández U, Texcalac-Sangrador JL, Romieu I. 2014. Health impact assessment of decreases in PM<sub>10</sub> and ozone concentrations in the Mexico City Metropolitan Area: A basis for a new air quality management program. *Salud Pública de México* 56: 579-591. <https://doi.org/10.21149/spm.v56i6.7384>
- Riojas-Rodríguez H, da Silva AS, Texcalac-Sangrador JL, Moreno-Banda GL. 2016. Air pollution management and control in Latin America and the Caribbean: Implications for climate change. *Revista Panamericana de Salud Pública* 40: 150-159.
- Rivera-González LO, Zhang Z, Sánchez BN, Zhang K, Brown DG, Rojas-Bracho L, Osornio-Vargas A, Vadillo-Ortega F, O'Neill MS. 2015. An assessment of air pollutant exposure methods in Mexico City, Mexico. *Journal of the Air & Waste Management Association* (1995) 65: 581-591. <https://doi.org/10.1080/10962247.2015.1020974>
- Rojas-Martínez R, Pérez-Padilla R, Olaiz-Fernández G, Mendoza-Alvarado L, Moreno-Macías H, Fortoul T, McDonnell W, Loomis D, Romieu I. 2007. Lung function growth in children with long-term exposure to air pollutants in Mexico City. *American Journal of*



- Respiratory and Critical Care Medicine 176: 377-384. <https://doi.org/10.1164/rccm.200510-1678OC>
- Rosa MJ, Just AC, Guerra MS, Kloog I, Hsu H-HL, Brennan KJ, García AM, Coull B, Wright RJ, Téllez-Rojo MM, Baccarelli AA, Wright RO. 2017. Identifying sensitive windows for prenatal particulate air pollution exposure and mitochondrial DNA content in cord blood. *Environment International* 98: 198-203. <https://doi.org/10.1016/j.envint.2016.11.007>
- SEP. n.d. Sistema Nacional de Información de Escuelas. Secretaría de Educación Pública. Available at: <http://snie.sep.gob.mx/SNIESC/> (accessed on December 4, 2019).
- Setton E, Marshall JD, Brauer M, Lundquist KR, Hystad P, Keller P, Cloutier-Fisher D. 2011. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *Journal of Exposure Science & Environmental Epidemiology* 21: 42-48. <https://doi.org/10.1038/jes.2010.14>
- Son Y, Osornio-Vargas ÁR, O'Neill MS, Hystad P, Texcalac-Sangrador JL, Ohman-Strickland P, Meng Q, Schwander S. 2018. Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *Science of The Total Environment* 639: 40-48. <https://doi.org/10.1016/j.scitotenv.2018.05.144>
- Téllez-Rojo MM, Romieu I, Ruiz-Velasco S, Lezana MA, Hernández-Ávila MM. 2000. Daily respiratory mortality and PM<sub>10</sub> pollution in Mexico City: Importance of considering place of death. *The European Respiratory Journal* 16: 391-396. <https://doi.org/10.1034/j.1399-3003.2000.016003391.x>
- Téllez-Rojo MM, Rothenberg SJ, Texcalac-Sangrador JL, Just AC, Kloog I, Rojas-Saunero LP, Gutiérrez-Ávila I, Bautista-Arredondo LF, Tamayo-Ortiz M, Romero M, Hurtado-Díaz M, Schwartz JD, Wright R, Riojas-Rodríguez H. 2020. Children's acute respiratory symptoms associated with PM<sub>2.5</sub> estimates in two sequential representative surveys from the Mexico City Metropolitan Area. *Environmental Research* 180: 108868. <https://doi.org/10.1016/j.envres.2019.108868>
- Trejo-González AG, Riojas-Rodríguez H, Texcalac-Sangrador JL, Guerrero-López CM, Cervantes-Martínez K, Hurtado-Díaz M, de la Sierra-de la Vega LA, Zúñiga-Bello PE. 2019. Quantifying health impacts and economic costs of PM<sub>2.5</sub> exposure in Mexican cities of the National Urban System. *International Journal of Public Health* 64: 561-572. <https://doi.org/10.1007/s00038-019-01216-1>
- WHA68. 2015. Health and the environment: Addressing the health impact of air pollution. World Health Assembly, 68, World Health Organization, Geneva, Switzerland. Available at: <https://apps.who.int/iris/handle/10665/253237>
- Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH. 2008. Spatiotemporal modeling of chronic PM<sub>10</sub> exposure for the nurses? Health study. *Atmospheric Environment* 42: 4047-4062. <https://doi.org/10.1016/j.atmosenv.2008.01.044>
- Yu X, Ivey C, Huang Z, Gurram S, Sivaraman V, Shen H, Eluru N, Hasan S, Henneman L, Shi G, Zhang H, Yu H, Zheng J. 2020. Quantifying the impact of daily mobility on errors in air pollution exposure estimation using mobile phone location data. *Environment International* 141: 105772. <https://doi.org/10.1016/j.envint.2020.105772>