Air Quality, Atmosphere & Health (2023) 16:2225–2238
https://doi.org/10.1007/s11869-023-01404-2

Assessing air pollution exposure misclassification using high-resolution $PM_{2.5}$ concentration model and human mobility data

Yougeng Lu[1]

Received: 10 April 2023 / Accepted: 1 August 2023 / Published online: 8 August 2023

Abstract

Due to a paucity of human movement data, the traditional method for estimating pollution exposure is static: Exposure is based on place of residence. However, local air quality varies over both time and space. This study explores exposure measurement errors associated with ignoring human mobility and its impact on exposure-health effect estimates. Using a random forest classification model, this study examines the impact of a variety of factors on potential measurement errors in personal exposure to outdoor $PM_{2.5}$. Mobility data at the individual level was combined with hourly $PM_{2.5}$ surfaces at the neighborhood level to estimate and compare residence-based and mobility-based exposures for 100,784 Los Angeles County residents. The results show that exposure measurement errors increase for individuals with high mobility levels. Significant sociodemographic disparities are observed across different exposure classification groups. Exposures of low-income people who have high mobility and reside in polluted neighborhoods tend to be overestimated. In contrast, exposures of high-income people living in neighborhoods with cleaner air are likely to be underestimated. The result on the exposure-health effect suggests that health risks of the socially disadvantaged after exposure to $PM_{2.5}$ is likely to be underestimated due to the exposure measurement error introduced by ignoring human mobility.

Keywords Exposure measurement error · $PM_{2.5}$ · Mobility · Exposure-health effect

Introduction

Personal air pollution exposures are very challenging to quantify accurately. Traditional approaches to quantifying exposure to outdoor air pollution assume that concentrations at the residential address are adequate surrogates of personal exposure to air pollution of outdoor origin (Bae et al. 2007; Bell and Ebisu 2012; Elliott and Smiley 2019; Houston et al. 2004; Rowangould 2013). The underlying assumption is that individuals spend most of their time indoors at the residence (Klepeis et al. 2001) and that outdoor air pollution infiltrates into the indoor environment where exposure occurs. However, given that people are mobile and their exposure to air pollution can occur in various locations, this static residential approach will inevitably introduce exposure measurement error and potential bias in air pollution and health assessments, which may lead to ineffective public health policy interventions.

Measurement errors in air pollution exposure can come from several sources. Recent studies report that exposure measurement may be substantially biased low if not considering human mobility (Gurram et al. 2019; Lu 2021; Park and Kwan 2017; Tayarani and Rowangould 2020). Park and Kwan (2017) argue that the individual's time-activity pattern determines their exposure levels as personal exposure to air pollution occurs through dynamic spatiotemporal interactions between individuals and air pollutant distribution. Although people generally spend more time at home, the majority of their exposure occurs in other places (Park 2020). For example, workers generally spend more time exposed in traffic during commuting. Overlooking human mobility will lead to inaccurate air pollution exposure measurements.

Beyond ignoring human mobility, spatiotemporal resolution of air pollution surfaces or prediction models used is another important factor that may cause exposure measurement error. Coarse resolution cannot reflect important spatial gradients (Clark et al. 2022; Korhonen et al. 2019; Li

✉ Yougeng Lu
yougengl@stanford.edu

1 Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA

et al. 2016). The spatial resolution of assigned air pollution concentrations in recent studies varied substantially, ranging from 0.25 to 10 km^2 (Dewulf et al. 2016; Gurram et al. 2019; Park and Kwan 2017; Yu et al. 2020). The temporal resolution of outdoor $PM_{2.5}$ surfaces is another important factor, with several studies relying on daily, monthly, or annual averaged pollution models to estimate residential exposure (M. Nyhan et al. 2016; Pennington et al. 2017; Setton et al. 2008). Greater temporal aggregation does not capture how concentrations vary over time. Evidence shows that personal exposure to outdoor air pollution tended to be overestimated at the residence and underestimated at daily activity spaces when daily, monthly, or annual average air pollution concentrations are used to estimate personal exposure (Dhondt et al. 2012).

Moreover, exposure measurement error will further distort the air pollution-health effect estimates (Basagaña et al. 2013; S. Y. Kim et al. 2009; Samoli and Butland 2017; Sellier et al. 2014), leading to biased estimates between exposure and health outcome. Prior studies have documented the impact of potential exposure measurement errors on the air pollution-health relationship. Jerrett et al. (2005) show that inaccurate personal exposure measurements resulting from poorly spatially resolved air pollution prediction models may significantly impact the relationship between air pollution exposure and mortality. Inconsistent results in effect estimates of NO_2 on newborns' birthweight were obtained when different spatially resolved air pollution prediction models were used (Sellier et al. 2014). Another study presents that exposure measurement error might lead to bias of regression coefficients and to inflation of their variance when personal exposures assessed through air pollution prediction models with different spatial and temporal resolutions are used as explanatory variables in models for exposure-health estimates (Basagaña et al. 2013).

Substantial effort has been invested in air pollution epidemiology research to develop statistical models to predict personal exposures at subjects' locations in situations where measurements at the desired locations are not available. However, most existing air pollution epidemiology studies focus on the impact of air pollution prediction models on the exposure-health relationship (Basagaña et al. 2013; S. Y. Kim et al. 2009; Samoli and Butland 2017; Sellier et al. 2014). Few studies have examined the effects of human mobility in exposure measurements on the association between air pollution exposure and health outcomes. Yu et al. (2020) find that people with higher mobility levels tend to have larger exposure measurement errors. Shareck et al. (2014) argue that unequally distributed features and resources across spaces may induce air pollution exposure disparities by constraining sites where people perform their everyday activities. For example, due to accessibility limitations, low-income groups usually travel shorter distances from their homes than their high-income counterparts (Morency et al. 2011; Vallée et al. 2010). Compared to whites, blacks and Latinos usually have lower mobility levels (Hu et al. 2020). Full-time employees tend to travel longer daily distances (Järv et al. 2015; Morency et al. 2011; Páez et al. 2010), whereas part-time employees and unemployed people are more place-bound (Lu 2021; Vallée et al. 2010). Little is known about how distinct travel behaviors and mobility patterns of different sociodemographic groups may influence their air pollution exposure levels. The impact of human mobility on the exposure-health effect is also less studied in previous research.

This study aims to examine the impact of human mobility in exposure measurement errors and health effects associated with air pollution and disentangle the complex relationship between sociodemographic variables and exposure measurement errors. In this study, residence-based and mobility-based $PM_{2.5}$ exposures for a sample of Los Angeles County residents on a typical weekday were estimated by coupling hourly 500 × 500m $PM_{2.5}$ surfaces at the neighborhood level and simulated daily mobility data at the individual level. The study samples were classified into three exposure groups based on differences between their residence- and mobility-based $PM_{2.5}$ exposures: individuals with similar residence and mobility exposure, individuals whose exposures were overestimated, and individuals whose exposures were underestimated. Random forest classification models were used to examine the impacts of a series of mobility and sociodemographic variables on exposure classification results. Last, sensitivity analysis was conducted to examine the impact of human mobility on exposure-health effects across exposure classification groups and sociodemographic groups.

The remainder of the paper is organized as follows. The "Method" section presents the data and methods used in this study. The "Results" section summarizes the results. The "Discussion" section discusses the findings. Conclusions are presented in the "Conclusion" section.

Method

Study area

Los Angeles is well recognized for its notoriously severe air pollution problem as one of the US metropolitan regions with the highest level of particulate matter pollution (American Lung Association 2020). Los Angeles has the most developed highway system and the busiest traffic in the US $PM_{2.5}$ pollution, a primary air pollutant created by vehicles, which has been a serious public health problem in Los Angeles for decades. In Los Angeles, $PM_{2.5}$ concentrations vary spatially and temporally, with the highest pollution observed during peak hours and within core urban areas (Lu et al. 2021). Los Angeles is, therefore, a good case study to examine variation in exposure levels across time and space taking daily travel patterns into account and test whether exposure patterns are related to sociodemographic population characteristics.

Data

$PM_{2.5}$ pollution modeling

In this study, ground-level $PM_{2.5}$ concentrations were estimated from our recently developed $PM_{2.5}$ model (Lu et al. 2021). We have created an hourly, 0.25-km gridded $PM_{2.5}$ model for Los Angeles County that incorporates low-cost air sensor data (i.e., PurpleAir) and machine learning techniques. Ambient air pollution has been traditionally monitored at regulatory stations at high instrumentation and maintenance costs. Sparse and uneven regulatory monitoring has a limited ability to reflect pollution details, especially in unmonitored areas. Dense deployment allows low-cost PurpleAir sensors to capture spatiotemporal variations of localized $PM_{2.5}$ concentrations at finer resolution than regulatory air quality stations (Bi et al. 2020; Lu et al. 2021; Mousavi and Wu 2021). A number of recent studies have used PurpleAir sensors to develop $PM_{2.5}$ modeling at fine spatiotemporal resolution (Bi et al. 2020; Lu et al. 2021; Mousavi and Wu 2021).

In this study, twenty-four hourly $PM_{2.5}$ concentration surfaces over the course of a typical weekday in 2019 were generated at a 500 m × 500 m grid level for Los Angeles County. A suite of spatiotemporal variables, including meteorological conditions, land use variables, and traffic counts, was integrated with the random forest method to estimate $PM_{2.5}$ concentration at the sub-daily and neighborhood level. Estimated $PM_{2.5}$ concentrations were then validated against measured $PM_{2.5}$ concentrations by the 10-fold cross-validation method. The results showed that the PurpleAir-based $PM_{2.5}$ prediction model could capture more than 90% of variations. A comprehensive description and validation results can be found in Lu et al. (2021).

Activity-based travel demand modeling

Daily travel trajectories for 100,784 Los Angeles County residents were simulated using an activity-based travel demand model developed by the Southern California Association of Governments (SCAG) for an average weekday in 2019 (Pendyala et al. 2012; Ziemke et al. 2015). American Community Survey (ACS) 2003 and Census 2000 have been used to validate this SCAG simulated travel trajectory data. Validation results show that the SCAG activity-based travel demand model has a good performance in predicting "activity purpose-number" and mimicking corresponding population features at the individual level. According to the validation results, the majority of the synthetic population deviated less than 5% from the reference group in terms of demographic and socioeconomic characteristics (Pendyala et al. 2012).

The SCAG travel trajectory dataset contains 387,398 trip records for 100,784 Los Angeles County residents (approximately 10% of the total Los Angeles County population). Each trip record includes a personal ID, origin-destination pair of the trip, trip purpose, trip departure and arrival timestamps, trip duration, and travel mode. The personal ID is unique for each individual and is used to connect with synthetic demographic features provided by the SCAG (Bhat et al. 2013; Lu 2021). The origin and destination of each trip are allocated to the geographic unit of the traffic analysis zone (TAZ), whose size is similar to the census tract. The centroid of TAZ is assumed to be each trip's origin or destination point.

However, the travel trajectory dataset lacks information on travel paths between the activity sites. This study used the OSMnx python package to estimate probable travel paths between two activity TAZs in the shortest path distance (Boeing 2017; Lu 2021).

Individual exposure assessment

Static and dynamic exposure assessment

The study region was subdivided into 0.25 km^2 hexagon grids. As noted earlier, hourly $PM_{2.5}$ concentrations of each grid (twenty-four hours in total) were generated by utilizing the $PM_{2.5}$ model developed by Lu et al. (2021) for Wednesday, September 18, 2019. Due to limitations in computing resources, $PM_{2.5}$ concentrations are assumed to be constant during an hour within each hexagon grid. The hourly $PM_{2.5}$ concentrations were spatially matched to each TAZ in compliance with the travel trajectory data and averaged within each TAZ if multiple $PM_{2.5}$ hexagon grids locating in the same TAZ. Two types of individual $PM_{2.5}$ exposures were then assessed: (1) static $PM_{2.5}$ exposure at residence and (2) dynamic $PM_{2.5}$ exposure that considers individuals' daily mobility patterns.

The individual static and dynamic exposures are estimated as in Eqs. (1) and (2):

$$\text{Static}_i = \frac{\sum_{t=1}^{T} PM_{h,t}}{T} \tag{1}$$

$$\text{Dynamic}_i = \frac{\sum_{t=1}^{T} \sum_{n=1}^{N} PM_{n,t} \cdot P_n}{T} \tag{2}$$

where $PM_{h,\,t}$ is $PM_{2.5}$ concentration in hour t at TAZ h, where individual i's home is located. T denotes 24 h of a day. $PM_{n,t}$ is $PM_{2.5}$ concentration in hour t at TAZ n, where individual i is located within hour t. N represents the total number of TAZs (microenvironments) individual i has stayed during hour t ($N \geq 1$). P_n denotes the percentage of time during hour t that individual i stays in TAZ n.

Exposure classification based on exposure measurement error

Prior research has documented the occurrence of exposure misclassification when human mobility is not taken into account in exposure assessment (Guo et al. 2020; Yu et al. 2020). Exposure of individuals who have high residence-based exposures is likely to be reduced by their mobility, while exposure of individuals who have relatively low residence-based exposures is likely to be increased (J. Kim and Kwan 2021; Kwan 2018). All study subjects were subdivided into three groups according to their exposure measurement errors shown in Eq. (3): (1) individuals with similar dynamic and static exposures, which is referred to as the "Accurate" group; (2) individuals with higher static exposures than their dynamic exposures, which is referred to as the "Overestimated" group; and (3) individuals with higher dynamic exposures than their static exposures, which is referred to as the "Underestimated" group.

The magnitude and direction of two statistical indicators were employed to categorize exposure classification groups: (1) exposure measurement error and (2) mean absolute percentage error (MAPE). The exposure measurement error was calculated by subtracting an individual's static exposure from their dynamic exposure (i.e., $\text{Dynamic}_i - \text{Static}_i$). A positive exposure measurement error indicates an individual's exposure is underestimated, while a negative measurement error indicates overestimated exposure. MAPE was adopted as an additional criterion to evaluate the degree of agreement between an individual's static and dynamic exposures: $\left|\frac{\text{Dynamic}_i - \text{Static}_i}{\text{Static}_i}\right| \times 100\%$. Higher MAPE values indicate differences between static and dynamic exposures as a result of overestimated or underestimated exposures. The thresholds for exposure measurement error and MAPE were set to $\pm 0.5\ \mu g/m^3$ and 10%, respectively, to determine an individual's exposure classification group. A comprehensive description of the classification method is shown in Eq. (3).

$$E_i = \begin{cases} \text{Overestimated} & \text{if } \text{Error}_i < -0.5 \text{ and } MAPE_i > 10\% \\ \text{Accurate} & \textit{if} - 0.5 \leq \text{Error}_i \leq 0.5 \text{ and } MAPE_i \leq 10\% \\ \text{Underestimated} & \text{if } \text{Error}_i > 0.5 \text{ and } MAPE_i > 10\% \end{cases} \tag{3}$$

where E_i denotes the exposure classification group that individual i belongs to; Error_i is the exposure measurement error for individual i.

Random forest classification model

This study utilized the random forest classification model to examine associations of a variety of mobility and sociodemographic variables with exposure classification results. In contrast to traditional linear regression, the random forest model can capture nonlinear relationships between response variables and predictors and provide a flexible and automated process for predicting target variables (Breiman 2001). The random forest model generates a number of decision trees and trains each decision tree independently using a random sample of the data. This randomness contributes to the model being more robust than a single decision tree and less prone to overfitting the training data. Furthermore, the random forest model avoids the probable multicollinearity across sociodemographic variables, which violates the underlying premise of independence in many regression models.

In this study, 90% of samples were randomly subsampled as training set and the remaining 10% as a testing set to evaluate the model performance. Since the classes were unbalanced (79% of the study sample was classified as the Accurate group, 9% as the Overestimated group, and 12% as the Underestimated group), a combination of the Synthetic Minority Over-sampling Technique (SMOTE) and random under-sampling methods was utilized to resample the dataset until balanced training classes were achieved (Chawla et al. 2002; He and Garcia 2009). The class-wise sensitivity and specificity, as well as the mean classification accuracy, were calculated to evaluate the random forest model performance. Confusion matrices were utilized to calculate the specificity and sensitivity of candidate models.

The optimal number of randomly sampled features at each node (m) and decision trees (k) were determined by minimizing the out-of-bag (OOB) error rate through iterative cross-validation (Lu et al. 2021). The relative importance of each predictor variable was determined using the mean decrease in accuracy based on OOB error. Partial dependence plots were produced to depict the correlations between predictor variables and the probability of being classified into a given class. A partial dependence plot demonstrates the marginal effect of a predictor variable on the predicted response while controlling for all other variables in the model (Friedman 2001). Figure 1 shows the framework and process of assessing exposure classification error and factors affecting it.

Exposure and health effect across exposure classification groups

Ordinary least squares (OLS) regression models were run as multivariate models to assess the association between personal exposures and health outcomes for static and dynamic exposures, respectively. The exposure-health effects were

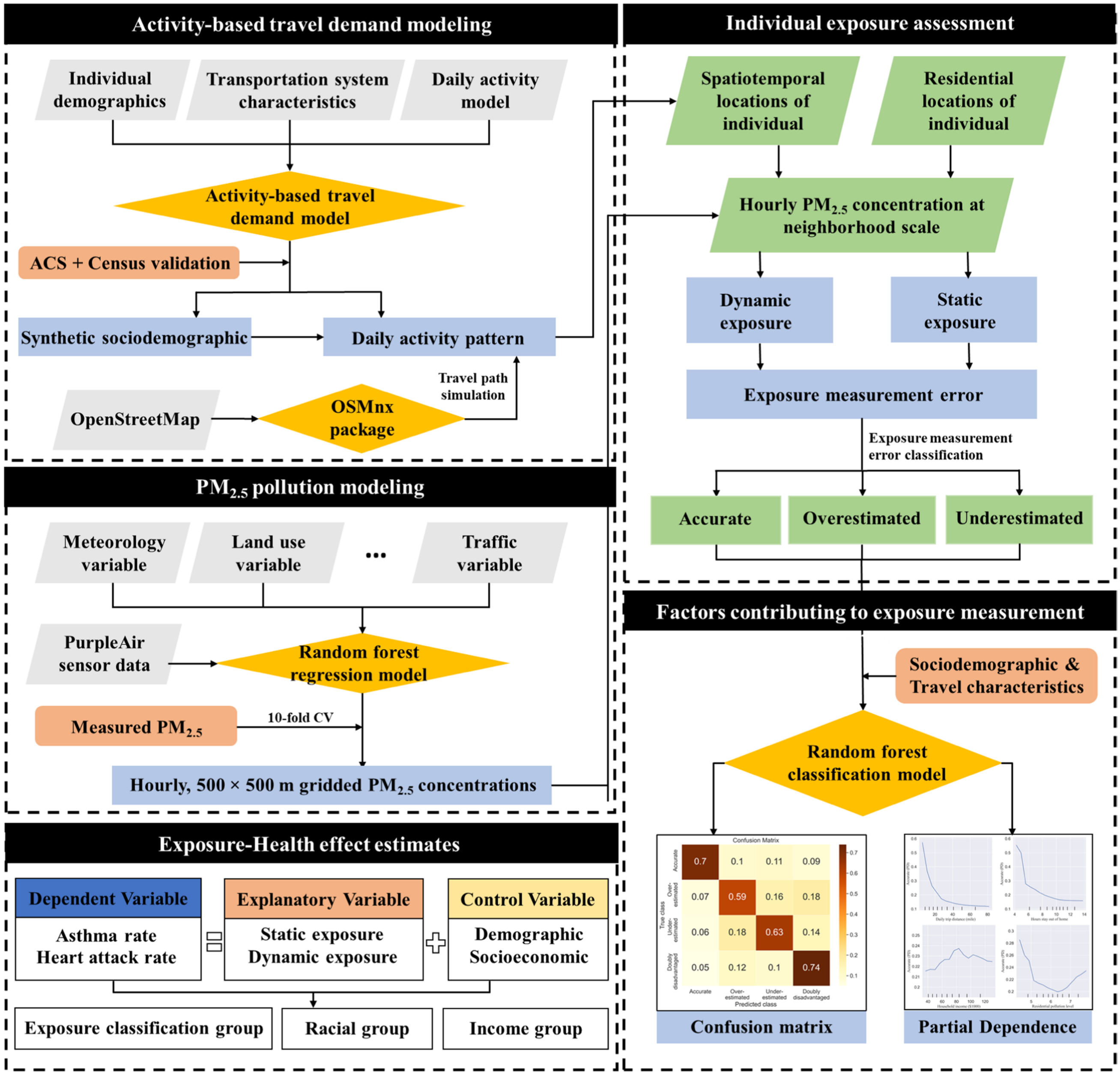

Fig. 1 Research conceptual framework

further assessed across different exposure classification groups, racial groups, and income groups. Recent studies have shown that exposure to $PM_{2.5}$ is linked to acute respiratory symptoms (Bose et al. 2015) and cardiovascular disease (Madrigano et al. 2013; Neophytou et al. 2014). Thus, two health outcomes were adopted as the dependent variables for the OLS models: the rate of emergency department visits for asthma and the rate of emergency department visits for heart attacks per 10,000 persons.

The health outcome data were obtained from CalEnviroScreen at the census tract level (California Office of Environmental Health Hazard Assessment 2023). Since the individual-level health outcome data were not available, the rate of asthma and rate of heart attack was assigned to each study subject based on their residential locations. That is, study subjects who live in the same TAZ are assumed to have the same health outcomes. The relationship between exposure and the two health indicators was estimated by OLS models adjusted several confounding variables including demographic variables (gender, age, race) and socioeconomic status (income, employment status, education). Table 2 presents a summary of these variables.

In summary, this study first measures the residence-based static and mobility-based dynamic exposures for study subjects, respectively. The study subjects are then classified to three groups according to the magnitude and direction of their exposure measurement errors. The random forest classification model is used to examine the impact of human mobility and sociodemographic characteristics on exposure measurement error. Last, this study explores the exposure-health effect by using static and dynamic exposure measurements through OLS regression models among different cohorts (Fig. 1).

Results

$PM_{2.5}$ and population activity distribution

Figure 2a–d presents the estimated hourly $PM_{2.5}$ concentrations across Los Angeles County at the neighborhood level on a typical weekday in 2019. A heterogeneous spatiotemporal distribution pattern of $PM_{2.5}$ pollution can be observed. In general, $PM_{2.5}$ concentrations are higher in daytime than evening and night, especially in the morning peak hours (Fig. 2b), and most pollution concentrates in urban cores and along highways. These patterns are in line with the findings of previous studies (Lu et al. 2021).

Figure 2e–h presents the simulated activity patterns of Los Angeles residents at four different hours on a typical weekday in 2019. These figures reflect the distribution pattern of people's residences and workplaces. Overall, Los Angeles residents generally travel from their sparsely distributed places of residence (Fig. 2e) to urban cores (Downtown Los Angeles, Wilshire-Santa Monica corridor, Long Beach) in the early morning (Fig. 2f) and stay all day (Fig. 2g) until they return to their residences again in the evening (Fig. 2h).

Exposure classification error analysis

To investigate potential exposure measurement errors resulting from ignoring human mobility, flows between different quartiles of study subject's static exposure and dynamic exposure were plotted in Fig. 3a. A high percentage of the population was misclassified into other quartiles, especially for study subjects with static exposures in middle quartiles (Q2 and Q3). About one-third of populations in the middle quartiles was classified into other quartiles when human mobility was omitted in exposure measurement. Three exposure classification groups were identified by quantifying the difference between individuals' static and dynamic exposures based on Eq. (3). Table 1 gives summary statistics. Figure 3b–d shows the distributions of static and dynamic exposures for each group.

Table 1 shows that the Accurate group is the largest. For about 80% of the observations, there is no difference between static and dynamic exposures (the difference is statistically significant but not meaningful). Figure 3b shows how close the two distributions are. The Overestimated group is the smallest (9% of the study sample). For individuals in the Overestimated group, mean static exposure was 0.96 μg/m^3 higher than their dynamic exposure. This difference is large (about 10%) and significant. This group has the highest static exposure level of all groups (Fig. 3c). The Underestimated group accounts for the remaining 12% of observations. The mean difference between static and dynamic estimates is 1.15 μg/m^3 or about 17%, even larger than the difference for the Overestimated group (Fig. 3d).

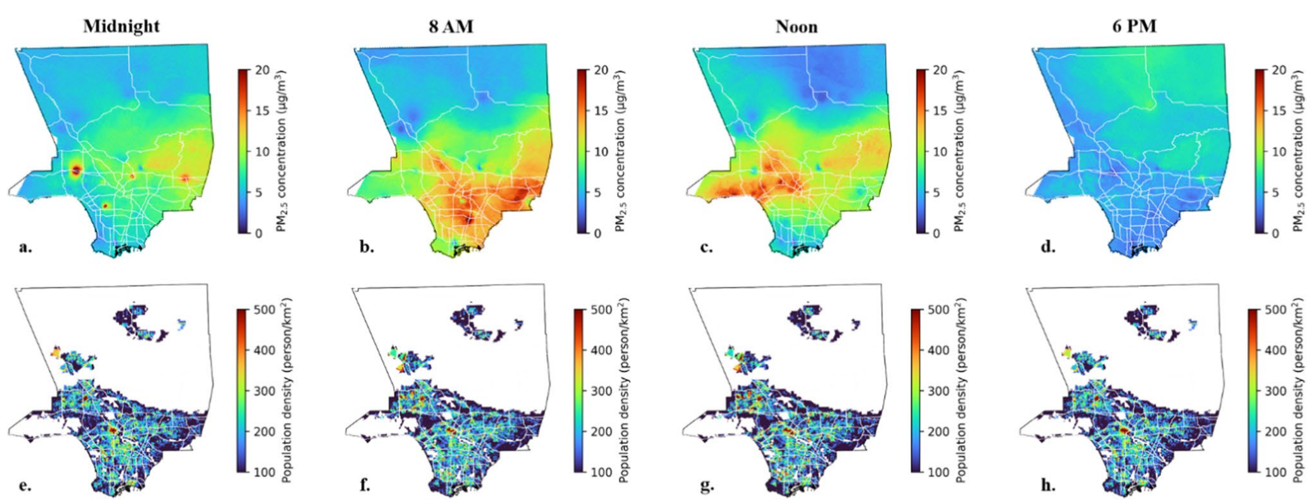

Fig. 2 Estimated hourly $PM_{2.5}$ concentrations at (**a**) midnight, (**b**) 8 AM, (**c**) noon, and (**d**) 6 PM and distribution of population activity at (**e**) midnight, (**f**) 8 AM, (**g**) noon, and (**h**) 6 PM on a typical weekday in Los Angeles County

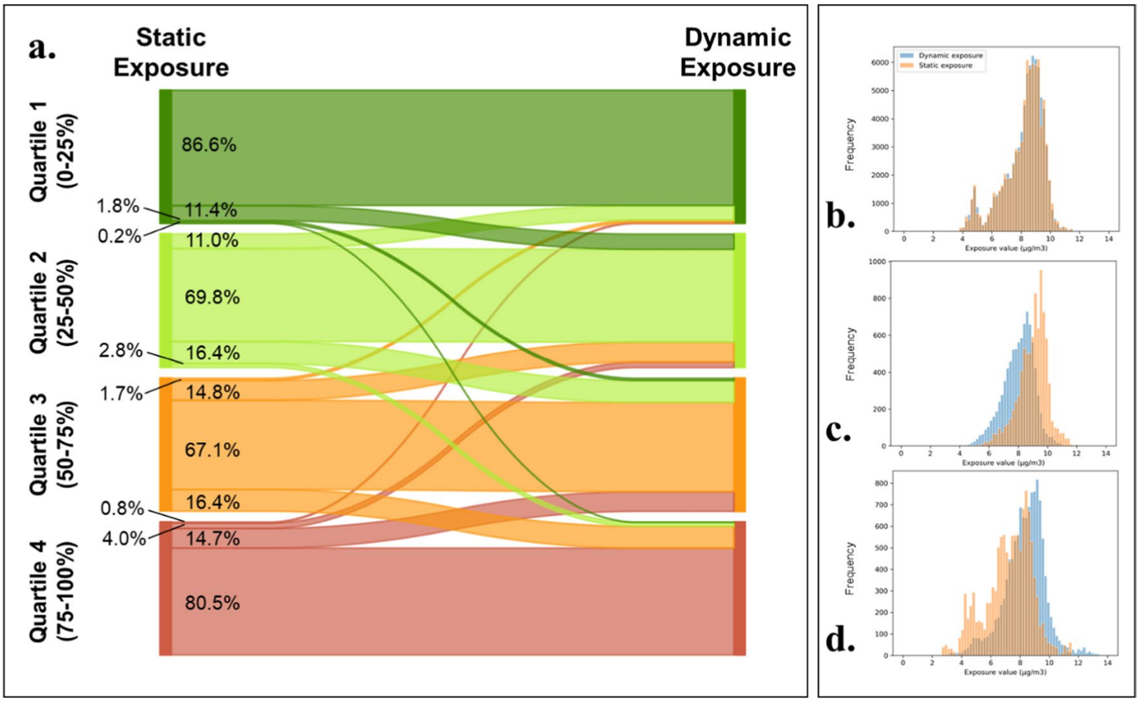

Fig. 3 The distribution of exposure measurement error: (**a**) direction of potential $PM_{2.5}$ exposure misclassifications between static exposure and dynamic exposure; (**b**) distribution of static and dynamic exposures for the Accurate group; (**c**) distribution of static and dynamic exposures for the Overestimated group; (**d**) distribution of static and dynamic exposures for the Underestimated group

Table 1 Summary statistics of static and dynamic $PM_{2.5}$ exposures (µg/m^3) across exposure classification groups (***$p < 0.001$)

		Mean	Std. dev	Median	Minimum	Maximum
Accurate (N = 80,411)	Dynamic exposure	8.17	1.38	8.51	2.71	12.27
	Static exposure	8.16	1.39	8.48	2.65	11.97
	Difference	0.01***				
Overestimated (N = 8655)	Dynamic exposure	8.03	1.06	8.16	3.35	11.46
	Static exposure	8.99	1.01	9.15	4.20	11.97
	Difference	−0.96***				
Underestimated (N = 11,718)	Dynamic exposure	7.84	1.16	8.06	3.06	13.86
	Static exposure	6.69	1.33	6.93	2.65	11.43
	Difference	1.15***				

Random forest results

Model performance and variable importance

The descriptive analysis has revealed mobility and sociodemographic differences across exposure classification groups. Random forest models were further trained using the same set of mobility and sociodemographic variables to examine their correlation with exposure classification errors. Table 2 lists all mobility and sociodemographic variables and their summary statistics used for the random forest model. As noted earlier, three exposure classification groups were defined, and the random forest classification algorithm was used to develop a predictive classification model based on individual's mobility patterns, residential pollution level, and sociodemographic characteristics. The hyperparameters for the random forest model were set to 1500 decision trees with a minimum sample leaf of 50.

The random forest model yielded a mean classification accuracy (adjusted across all classes) of 71%. Figure 4a presents the confusion matrix for evaluating the random forest model's performance. The sensitivity values for the Accurate, Overestimated, and Underestimated groups are 73%, 71%, and 70%, respectively, implying good agreement between actual and predicted classifications.

The relative contribution value of predictor variables to the random forest classification results is shown in Fig. 4b, sorted in order of importance. The variable importance

Table 2 Descriptive statistics of the mobility and sociodemographic variables used in the analysis of exposure classification errors

	Variables	Description	Mean	Std. dev.	Minimum	Maximum
Travel behavior	Trip number	The number of daily trips made by an individual	3.84	2.05	2	18
	Trip distance	The daily trip distance in miles an individual travels	24.48	21.44	0.19	234.28
	Hours stay out-of-home	The hours an individual spends out-of-home per day	8.07	3.49	0	22.43
	%Driving trip	Proportion of daily trips of an individual made by driving	84.3%	32.2%	0%	100%
	%Public transit trip	Proportion of daily trips of an individual made by public transit	2.5%	14.1%	0%	100%
	%Walk/bike trip	Proportion of daily trips of an individual made by walking or bicycling	10.0%	25.8%	0%	100%
Residential pollution	Residence pollution level	The standard index showing overall air pollution level at places of residence (between 1 and 10): higher value means higher pollution	6.10	1.16	2.32	9.62
Sociodemographic	Age	Age	33.99	21.43	0	94
	Income	Household income ($1000)	72.22	30.12	12.27	230.90
	Male	Dummy variable: 1 = male; 0 = female			0	1
	Non-Hispanic White	Dummy variable: 1 = non-Hispanic White; 0 = otherwise			0	1
	Black	Dummy variable: 1 = Black; 0 = otherwise			0	1
	Hispanic	Dummy variable: 1 = Hispanic/Latino; 0 = otherwise			0	1
	Asian	Dummy variable: 1 = Asian; 0 = otherwise			0	1
	Worker	Dummy variable: 1 = employed; 0 = otherwise (unemployed, homemaker, student, or retired)			0	1
	College	Dummy variable: 1 = college or higher degree; 0 = otherwise			0	1

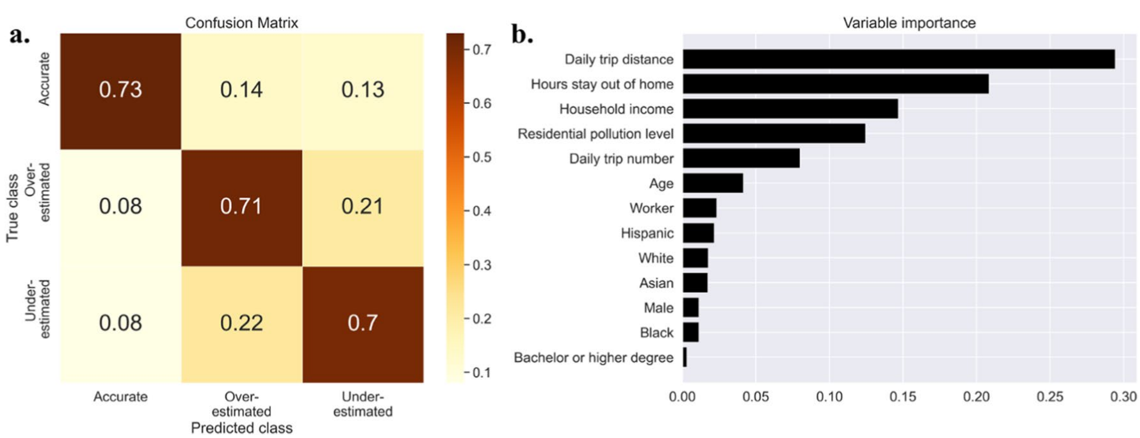

Fig. 4 Random Forest model performance: (a) confusion matrix of predicted exposure classification groups; (b) variable importance rank

rank shows that daily trip distance, hours stay out of home, household income, and residential pollution level are among the most important features. By contrast, ethnicity, employment status, and education play weaker roles in affecting exposure classification errors. These findings suggest that individual's exposure measurement error is mainly affected by their mobility levels, income, and pollution levels at residence.

Partial dependence analysis

The partial dependence plots illustrate the marginal effect of a single variable on the predicted classification outcome. According to variable importance results, the partial dependence of daily trip distance, hours stay out of home, household income, and residential pollution levels on the probability of classification results were examined. Figure 5 plots

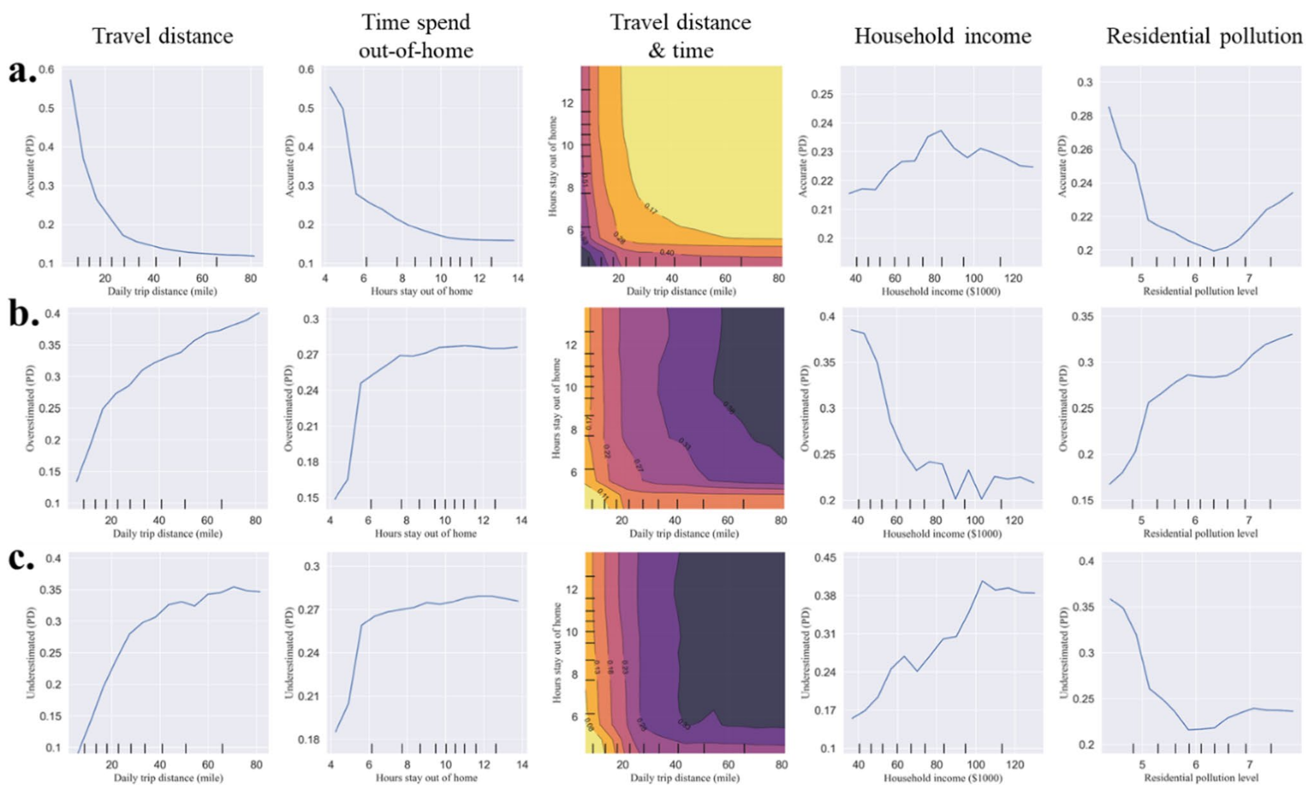

Fig. 5 Partial dependence (PD) plots for the most important variables in the random forest classification model for (**a**) the Accurate group, (**b**) the Overestimated group, and (**c**) the Underestimated group

the partial dependence of the abovementioned variables for all groups.

As shown in Fig. 5a, increasing probabilities of an individual belonging to the Accurate group were associated with shorter daily trip distance, fewer hours spent out of home, and higher residential pollution levels. Household income displays a nonlinear relationship with probabilities of the Accurate group. The most significant marginal influence was depicted at around $80,000. The middle column of Fig. 5a shows a two-dimensional partial dependence plot of daily trip distance and hours stay out of home to explore the effects of combining two mobility variables on probabilities of the Accurate group. The color scheme represents different probability levels. Yellow tones indicate a lower probability, and purple tones denote a higher probability. The two-dimensional plot shows that individuals who travel longer distances and time away from home are least likely to be categorized to the Accurate group.

A similar effect of mobility variables on classification probability can be observed in Fig. 5b,c. Both probabilities of the Overestimated group and the Underestimated group grow with the daily trip distance and hours stay out of home. The two-dimensional plots indicate that exposure is more likely to be overestimated or underestimated for people with high mobility levels. Although the effects of mobility variables on the magnitude of exposure classification error were similar for the Overestimated group and the Underestimated group, different household income and residential pollution levels for the two groups resulted in completely opposite directions of exposure classification errors. Increasing probabilities of the Overestimated group were associated with lower household income and higher residential pollution levels (Fig. 5b). By contrast, lower household income and higher residential pollution levels were associated with reduced probabilities of the Underestimated group (Fig. 5c). The opposite associations of household income and residential pollution level with the probabilities of the Overestimated and Underestimated groups suggest that as mobility levels increased, exposures were more likely to be overestimated for low-income residents living in highly polluted areas, while exposures were more likely to be underestimated for high-income residents living in areas with cleaner air.

Exposure-health effect analysis

Figure 6 shows the correlation coefficient (95 percent confidence interval (CI)) between exposures (static and dynamic exposure) and health outcomes (rate of asthma and heart attack) across three different cohorts: exposure

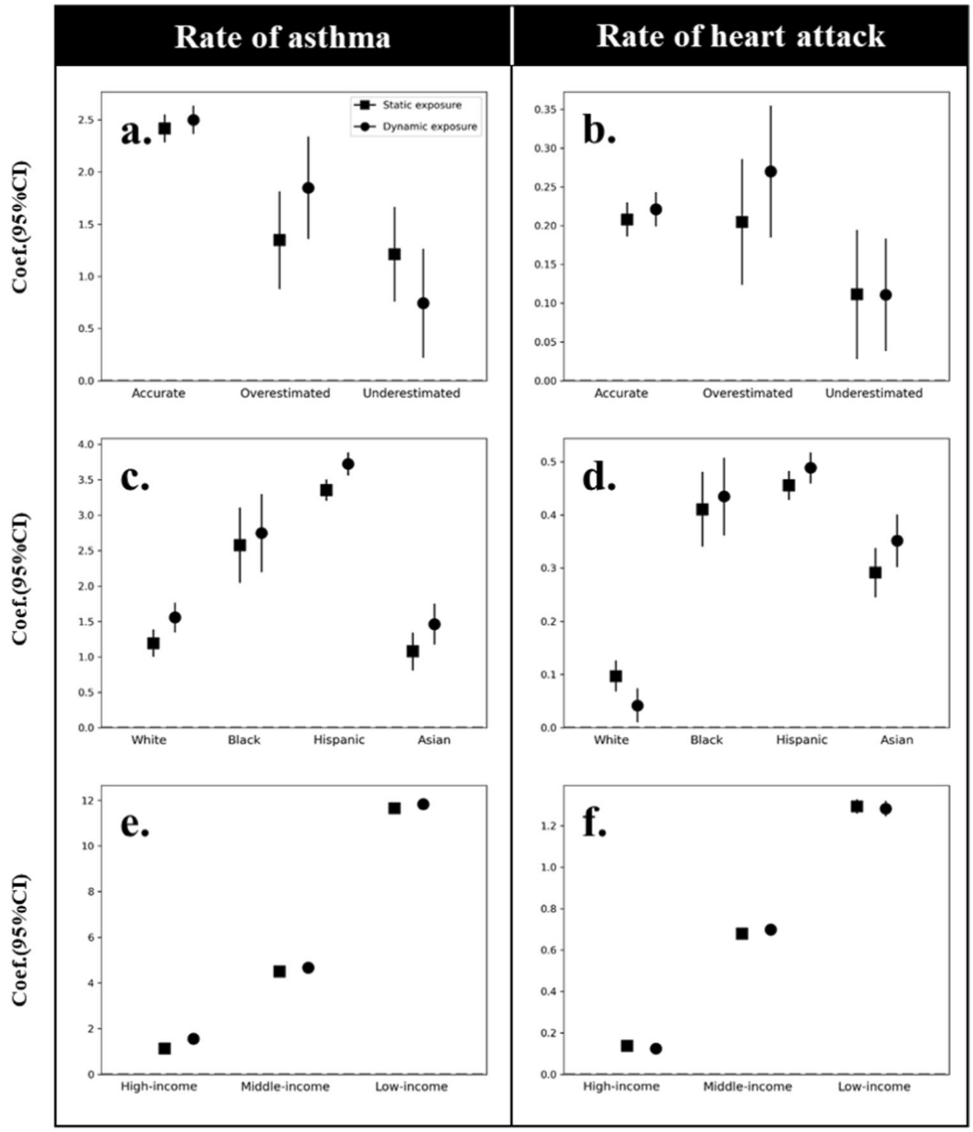

Fig. 6 Sensitivity analysis of exposure-health effect

classification groups, racial groups, and income groups. A positive association between exposure and adverse health outcomes was observed in Fig. 6. However, the coefficients of static and dynamic exposures were significantly different across different groups. Exposure measurement errors associated with omitting human mobility can result in bias in the correlation between exposures and health outcomes. For the Accurate group, the correlation coefficient of static exposure is similar to dynamic exposure (2.42 vs. 2.50 for asthma and 0.21 vs. 0.22 for heart attack). However, for people whose exposure is overestimated, the effect of $PM_{2.5}$ exposure on the risk of asthma (1.85) and heart attack (0.27) is greater than estimated by static exposure (1.35 and 0.20). Conversely, for people whose exposure is underestimated, their health risks related to exposure to $PM_{2.5}$ tended to be overstated. For the Underestimated group, a 1 μg/m^3 increase in static exposure to $PM_{2.5}$ would lead to a 1.21% increase in emergency department visits for asthma. At the same time, the rate decreases to 0.74% when considering human mobility in the exposure measurement (Fig. 6a).

The effect of $PM_{2.5}$ exposure on the risk of asthma and heart attack is found to be underestimated for most racial and income groups if ignoring human mobility in exposure measurement (Fig. 6c–f). Hispanics, blacks, and the low-income are found to be disproportionately burdened with health risks associated with air pollution, which is consistent with findings obtained from existing literature (Bae et al. 2007; Gilbert and Chakraborty 2011; Houston et al. 2004). The sensitivity analysis on the exposure-health effect suggests that health risks of the socially disadvantaged after exposure to $PM_{2.5}$ are likely to be underestimated due to the exposure mismeasurement introduced by ignoring human mobility.

Discussion

Overlooking human mobility may lead to incorrect exposure assessment and misleading conclusions and thus results in inefficient public health policy solutions (J. Kim and Kwan 2021; Park and Kwan 2017). A growing amount of research has highlighted the importance of human mobility in air pollution exposure assessment (Dewulf et al. 2016; Ma et al. 2020; M. M. Nyhan et al. 2019; Park 2020), but little is known about the impact of distinct mobility patterns on exposure measurement errors and how these errors influence exposure-health effect. This study offers important insights into the literature by investigating the underlying factors contributing to exposure measurement errors. This study indicates that the individual's mobility level is the most critical factor in determining exposure measurement errors. Exposure measurement errors increase with mobility. Individuals with high mobility have the most significant exposure measurement errors, especially those who travel long distances and spend more time out of the home.

There is also a significant correlation between individuals' sociodemographic characteristics and exposure measurement errors. Household income has the greatest effect on exposure measurement errors, likely due to the key role of wealth in determining where people live, their occupations, and places people often visit (Sampson 2019). According to the results, household income is more inclined to drive the direction of exposure measurement errors. Air pollution at residence is another critical factor influencing exposure measurement errors. On average, as mobility increases, exposure is likely overestimated for low-income residents of neighborhoods with poor air quality, while exposure is typically underestimated for high-income residents of neighborhoods with cleaner air. This finding is consistent with the conclusion obtained from prior empirical research: Exposures of individuals who are less exposed at residence are likely amplified by their mobility, and exposure of people with high residential exposure is usually attenuated (Dewulf et al. 2016; Picornell et al. 2019; Tayarani and Rowangould 2020; Yu et al. 2020). One probable explanation is that people from neighborhoods with cleaner air are more likely to carry out their daily activities in neighborhoods with poorer air quality (Boeing et al. 2023). In contrast, residents of neighborhoods with high air pollution tend to engage in daily activities in neighborhoods with less air pollution than their residential neighborhoods (J. Kim and Kwan 2021; Lu 2021).

The results show that the relative exposure measurement error is larger for wealthier people because they often reside in neighborhoods with cleaner air and their residence-based exposures start much lower than those with financial restrictions. However, the overall exposure and burden of health risks are much higher for the more disadvantaged populations as they are likely from more polluted neighborhoods. People tend to spend more time at home, even those with high mobility levels (Lu 2021; Park 2020). If the socially disadvantaged stay within their residential neighborhoods or vicinity most of the day and spend a lot more time in transit to move shorter distances, either or both of these mobility patterns can lead to much worse exposures but fewer exposure measurement errors. Given the significant contribution of residential air pollution to an individual's overall exposure, although the exposure of people who live in more polluted neighborhoods may be overestimated, they are still likely to have relatively higher exposures than those living in less polluted neighborhoods.

Moreover, exposure mismeasurement can result in bias in the correlation between air pollution exposure and health outcomes, which may further bias estimates of public health impact. The direction and magnitude of exposure measurement error can lead to incorrect estimates of the exposure-health effect. Our results show that the exposure-health effect may be underestimated for individuals with overestimated exposure. Conversely, for those whose exposure is underestimated, their health risks after exposure to $PM_{2.5}$ tend to be overstated. Ineffective public health and environmental interventions can be introduced by biased exposure-health effects as a result of exposure mismeasurement. Low-income and ethnic minorities have been burdened with more financial restrictions (Bae et al. 2007). They are also found to be exposed to high air pollution, which makes them doubly disadvantaged (Bae et al. 2007; Elliott and Smiley 2019; Gilbert and Chakraborty 2011; Houston et al. 2004; J. Kim and Kwan 2021). Low-income people generally spend more time exposed to traffic during commuting or live in areas with poorer air quality, thus increasing exposure. Also, higher incidence of co-morbidities, nutritional deficiency, and less access to information and education due to lack of economic resources impose an increased vulnerability for socially disadvantaged groups. It is important for policymakers to account for individual's exposure at not only places of residence but also all other activity locations. Accurate exposure measurements can help policymakers develop public health policies that reflect the interests of all people.

Several limitations in this study need to be addressed in future research. First, the human movement data used in this study were simulated from an activity-based travel demand model. Given this model simulated people's daily travel trajectories for a typical weekday in 2019, it was assumed that individuals have constant activity patterns throughout the year. However, people's daily mobility patterns are not consistent over time and may vary across weekdays, weekends, or seasons (Susilo and Kitamura 2005; Xianyu et al. 2017). It is debatable whether people's varied travel behaviors on a different day (e.g., weekend, holiday) can generate similar exposure classification patterns identified in this study. Thus, more effort should be placed into studying how different travel behaviors over time can affect exposure measurement error by collecting human movement data covering multiple time periods.

Second, in this study, only ambient $PM_{2.5}$ exposure was estimated as indoor $PM_{2.5}$ data were not available. Recent evidence shows that people spend more time indoors (e.g., at home, workplace, and school) during the day, especially those with less mobility (Lu 2021; Park 2020). Staying indoors may provide some protection from sources of ambient air pollution (e.g., traffic emission), leading to different results in air pollution exposure assessments. Future exposure research should consider indoor and outdoor $PM_{2.5}$ concentrations to measure individual exposure accurately.

Third, the unique characteristics of demographic composition and land use layout are recognized for Los Angeles. As a result, the spatiotemporal mobility patterns and ground-level $PM_{2.5}$ concentration distribution depicted in this study only represent the study area's distinct features. The population mobility pattern, spatiotemporal variabilities of air pollution concentrations, sociodemographic mix, and land use layout are expected to vary across different regions. Further research is needed to examine whether findings from this study can be applied to other areas.

Conclusion

Ignoring human mobility in exposure estimates can lead to erroneous exposure assessments and ineffective policy implications. Prior research has emphasized the importance of human mobility in estimating air pollution exposure, but little is known about how human mobility might lead to exposure measurement errors. To fill the literature gap, this study assesses residence-based and mobility-based $PM_{2.5}$ exposures for 100,784 Los Angeles County residents. It examines the impact of mobility and sociodemographic variables on potential exposure measurement errors. Detailed human mobility data was integrated with hourly $PM_{2.5}$ surfaces at the neighborhood level. The finding suggests that the magnitude of exposure measurement error is linked to people's mobility levels, and individuals' sociodemographic variables drive the direction of exposure measurement errors. Individuals with high mobility levels are likely to have increased exposure measurement errors. High income and low residential pollution are associated with exposure underestimation, and low income and high residential pollution levels are associated with exposure overestimation. The exposure measurement error introduced by the residence-based method can further lead to erroneous conclusions on the relationship between exposure and health risks. Policymakers should take into account human mobility and sociodemographic characteristics in exposure assessment and ensure that their policies reflect not only the preferences of socially advantaged populations but also the interests of disadvantaged populations.

Author contributions Yougeng Lu: conceptualization, formal analysis, methodology, software, data curation, writing – original draft preparation

Data availability The datasets of this study are available upon request.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The author declares no competing interests.

References

American Lung Association. (2020). Most polluted cities. https://www.stateoftheair.org/city-rankings/most-polluted-cities.html
Bae CHC, Sandlin G, Bassok A, Kim S (2007) The exposure of disadvantaged populations in freeway air-pollution sheds: a case study of the Seattle and Portland regions. Environ Plann B Plann Des 34(1):154–170. https://doi.org/10.1068/b32124
Basagaña X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J, Elosua R, Künzli N (2013) Measurement error in epidemiologic studies of air pollution based on land-use regression models. Am J Epidemiol 178(8):1342–1346. https://doi.org/10.1093/aje/kwt127
Bell ML, Ebisu K (2012) Environmental inequality in exposures to airborne particulate matter components in the United States. Environ Health Perspect 120(12):1699–1704. https://doi.org/10.1289/ehp.1205201
Bhat CR, Goulias KG, Pendyala RM, Paleti R, Sidharthan R, Schmitt L, Hu HH (2013) A household-level activity pattern generation model with an application for Southern California. Transportation 40(5):1063–1086. https://doi.org/10.1007/s11116-013-9452-y
Bi J, Wildani A, Chang HH, Liu Y (2020) Incorporating low-cost sensor measurements into high-resolution PM2.5 modeling at a large spatial scale. Environ Sci Technol 54(4):2152–2162. https://doi.org/10.1021/acs.est.9b06046
Boeing G (2017) OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput Environ Urban Syst 65:126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004
Boeing G, Lu Y, Pilgram C (2023) Local inequities in the relative production of and exposure to vehicular air pollution in Los Angeles. Urban Stud 0(0). https://doi.org/10.1177/00420980221145403
Bose S, Hansel NN, Tonorezos ES, Williams DL, Bilderback A, Breysse PN, Diette GB, Mccormack MC (2015) Indoor particulate matter associated with systemic inflammation in COPD. J Environ Prot 6:566–572
Breiman L (2001) Random forests. Machine Learning 45(1):5–32

California Office of Environmental Health Hazard Assessment. (2023). CalEnviroScreen. https://oehha.ca.gov/calenviroscreen
Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357 https://www.jair.org/index.php/jair/article/view/10302
Clark LP, Harris MH, Apte JS, Marshall JD (2022) National and intraurban air pollution exposure disparity estimates in the United States: impact of data-aggregation spatial scale:9–14. https://doi.org/10.1021/acs.estlett.2c00403
Dewulf B, Neutens T, Lefebvre W, Seynaeve G, Vanpoucke C, Beckx C (2016) Dynamic assessment of exposure to air pollution using mobile phone data. Int J Health Geogr 15(1):1–14. https://doi.org/10.1186/s12942-016-0042-z
Dhondt S, Beckx C, Degraeuwe B, Lefebvre W, Kochan B, Bellemans T, Int Panis L, Macharis C, Putman K (2012) Health impact assessment of air pollution using a dynamic exposure profile: implications for exposure and health impact estimates. Environ Impact Assess Rev 36:42–51. https://doi.org/10.1016/j.eiar.2012.03.004
Elliott JR, Smiley KT (2019) Place, space, and racially unequal exposures to pollution at home and work. Soc Curr 6(1):32–50. https://doi.org/10.1177/2329496517704873
Friedman J (2001) Greedy function approximation : a gradient boosting machine. Ann Stat 29(5):1189–1232 https://www.jstor.org/stable/2699986
Gilbert A, Chakraborty J (2011) Using geographically weighted regression for environmental justice analysis : cumulative cancer risks from air toxics in Florida. Soc Sci Res 40(1):273–286. https://doi.org/10.1016/j.ssresearch.2010.08.006
Guo H, Zhan Q, Ho HC, Yao F, Zhou X, Wu J, Li W (2020) Coupling mobile phone data with machine learning: how misclassification errors in ambient PM2.5 exposure estimates are produced? Sci Total Environ 745:141034. https://doi.org/10.1016/j.scitotenv.2020.141034
Gurram S, Stuart AL, Pinjari AR (2019) Agent-based modeling to estimate exposures to urban air pollution from transportation: exposure disparities and impacts of high-resolution data. Comput Environ Urban Syst 75:22–34. https://doi.org/10.1016/j.compenvurbsys.2019.01.002
He H, Garcia EA (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21(9):1263–1284
Houston D, Wu J, Ong P, Winer A (2004) Structural disparities of urban traffic in Southern California: implications for vehicle-related air pollution exposure in minority and high-poverty neighborhoods. J Urban Aff 26(5):565–592. https://doi.org/10.1111/j.0735-2166.2004.00215.x
Hu L, Li Z, Ye X (2020) Delineating and modeling activity space using geotagged social media data. Cartogr Geogr Inf Sci 47(3):277–288. https://doi.org/10.1080/15230406.2019.1705187
Järv O, Müürisepp K, Ahas R, Derudder B, Witlox F (2015) Ethnic differences in activity spaces as a characteristic of segregation: a study based on mobile phone usage in Tallinn Estonia. Urban Stud 52(14):2680–2698. https://doi.org/10.1177/0042098014550459
Jerrett M, Burnett RT, Ma R, Arden Pope C, Krewski D, Newbold KB, Thurston G, Shi Y, Finkelstein N, Calle EE, Thun MJ (2005) Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16(6):727–736. https://doi.org/10.1097/01.ede.0000181630.15826.7d
Kim J, Kwan MP (2021) Assessment of sociodemographic disparities in environmental exposure might be erroneous due to neighborhood effect averaging: implications for environmental inequality research. Environ Res 195:110519. https://doi.org/10.1016/j.envres.2020.110519
Kim SY, Sheppard L, Kim H (2009) Health effects of long-term air pollution: influence of exposure prediction methods. Epidemiology 20(3):442–450. https://doi.org/10.1097/EDE.0b013e31819e4331
Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hern SC, Engelmann WH (2001) The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol 11(3):231–252. https://doi.org/10.1038/sj.jea.7500165
Korhonen A, Lehtomäki H, Rumrich I, Karvosenoja N, Paunu VV, Kupiainen K, Sofiev M, Palamarchuk Y, Kukkonen J, Kangas L, Karppinen A, Hänninen O (2019) Influence of spatial resolution on population PM2.5 exposure and health impacts. Air Qual Atmos Health 12(6):705–718. https://doi.org/10.1007/s11869-019-00690-z
Kwan MP (2018) The neighborhood effect averaging problem (NEAP): an elusive confounder of the neighborhood effect. Int J Environ Res Public Health 15(9). https://doi.org/10.3390/ijerph15091841
Li Y, Henze DK, Jack D, Kinney PL (2016) The influence of air quality model resolution on health impact assessment for fine particulate matter and its components:51–68. https://doi.org/10.1007/s11869-015-0321-z
Lu Y (2021) Beyond air pollution at home: assessment of personal exposure to PM2.5 using activity-based travel demand model and low-cost air sensor network data. Environ Res 201:111549. https://doi.org/10.1016/j.envres.2021.111549
Lu Y, Giuliano G, Habre R (2021) Estimating hourly PM2.5 concentrations at the neighborhood scale using a low-cost air sensor network: a Los Angeles case study. Environ Res 195:110653. https://doi.org/10.1016/j.envres.2020.110653
Ma J, Tao Y, Kwan MP, Chai Y (2020) Assessing mobility-based real-time air pollution exposure in space and time using smart sensors and GPS trajectories in Beijing. Ann Am Assoc Geogr 110(2):434–448. https://doi.org/10.1080/24694452.2019.1653752
Madrigano J, Kloog I, Goldberg R, Coull BA, Mittleman MA, Schwartz J (2013) Long-term exposure to PM2.5 and incidence of acute myocardial infarction. Environ Health Perspect 121(2):192–196. https://doi.org/10.1289/ehp.1205284
Morency C, Paez A, Roorda MJ, Mercado R, Farber S (2011) Distance traveled in three Canadian cities: spatial analysis from the perspective of vulnerable population segments. J Transp Geogr 19(1):39–50. https://doi.org/10.1016/j.jtrangeo.2009.09.013
Mousavi A, Wu J (2021) Indoor-generated PM2.5 during COVID-19 shutdowns across California: application of the PurpleAir indoor-outdoor low-cost sensor network. Environ Sci Technol 55(9):5648–5656. https://doi.org/10.1021/acs.est.0c06937
Neophytou AM, Costello S, Brown DM, Picciotto S, Noth EM, Hammond SK, Cullen MR, Eisen EA (2014) Marginal structural models in occupational epidemiology: application in a study of ischemic heart disease incidence and PM2.5 in the US aluminum industry. Am J Epidemiol 180(6):608–615. https://doi.org/10.1093/aje/kwu175
Nyhan M, Grauwin S, Britter R, Misstear B, Mcnabola A, Laden F, Barrett SRH, Ratti C (2016) " Exposure Track " - The Impact of Mobile-Device-Based Mobility Patterns on Quantifying Population Exposure to Air Pollution. https://doi.org/10.1021/acs.est.6b02385
Nyhan MM, Britter IKR, Koutrakis CRP (2019) Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. J Expo Sci Environ Epidemiol:238–247. https://doi.org/10.1038/s41370-018-0038-9
Páez A, Mercado RG, Farber S, Morency C, Roorda M (2010) Relative accessibility deprivation indicators for urban settings: definitions and application to food deserts in Montreal. Urban Stud 47(7):1415–1438. https://doi.org/10.1177/0042098009353626
Park YM (2020) Assessing personal exposure to traffic-related air pollution using individual travel-activity diary data and an on-road

source air dispersion model. Health Place 63:102351. https://doi.org/10.1016/j.healthplace.2020.102351

Park YM, Kwan MP (2017) Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health Place 43:85–94. https://doi.org/10.1016/j.healthplace.2016.10.002

Pendyala R, Bhat C, Goulias K, Paleti R, Konduri K, Sidharthan R, Hu HH, Huang G, Christian K (2012) Application of socioeconomic model system for activity-based modeling. Transp Res Rec 2303:71–80. https://doi.org/10.3141/2303-08

Pennington AF, Strickland MJ, Klein M, Zhai X, Russell AG, Hansen C, Darrow LA (2017) Measurement error in mobile source air pollution exposure estimates due to residential mobility during pregnancy. J Exposure Sci Environ Epidemiol 27(5):513–520. https://doi.org/10.1038/jes.2016.66

Picornell M, Ruiz T, Borge R, García-Albertos P, de la Paz D, Lumbreras J (2019) Population dynamics based on mobile phone data to improve air pollution exposure assessments. J Expo Sci Environ Epidemiol 29(2):278–291. https://doi.org/10.1038/s41370-018-0058-5

Rowangould GM (2013) A census of the US near-roadway population: public health and environmental justice considerations. Transp Res Part D: Transp Environ 25:59–67. https://doi.org/10.1016/j.trd.2013.08.003

Samoli E, Butland BK (2017) Incorporating measurement error from modeled air pollution exposures into epidemiological analyses. Curr Environ Health Rep 4(4):472–480. https://doi.org/10.1007/s40572-017-0160-1

Sampson RJ (2019) Neighbourhood effects and beyond: explaining the paradoxes of inequality in the changing American metropolis. Urban Stud 56(1):3–32. https://doi.org/10.1177/0042098018795363

Sellier Y, Galineau J, Hulin A, Caini F, Marquis N, Navel V, Bottagisi S, Giorgis-Allemand L, Jacquier C, Slama R, Lepeule J (2014) Health effects of ambient air pollution: do different methods for estimating exposure lead to different results? Environ Int 66:165–173. https://doi.org/10.1016/j.envint.2014.02.001

Setton EM, Peter CP, Cloutier-Fisher D, Hystad PW (2008) Spatial variations in estimated chronic exposure to traffic-related air pollution in working populations: a simulation. Int J Health Geogr 7(2):1–17. https://doi.org/10.1186/1476-072X-7-39

Shareck M, Frohlich KL, Kestens Y (2014) Considering daily mobility for a more comprehensive understanding of contextual effects on social inequalities in health: a conceptual proposal. Health Place 29:154–160. https://doi.org/10.1016/j.healthplace.2014.07.007

Susilo YO, Kitamura R (2005) Analysis of day-to-day variability in an individual's action space: exploration of 6-week mobidrive travel diary data. Transp Res Rec 1902:124–133. https://doi.org/10.3141/1902-15

Tayarani M, Rowangould G (2020) Estimating exposure to fine particulate matter emissions from vehicle traffic: exposure misclassification and daily activity patterns in a large, sprawling region. Environ Res 182:108999. https://doi.org/10.1016/j.envres.2019.108999

Vallée J, Cadot E, Grillo F, Parizot I, Chauvin P (2010) The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities: the case of cervical screening in the Paris metropolitan area (France). Health Place 16(5):838–852. https://doi.org/10.1016/j.healthplace.2010.04.009

Xianyu J, Rasouli S, Timmermans H (2017) Analysis of variability in multi-day GPS imputed activity-travel diaries using multidimensional sequence alignment and panel effects regression models. Transportation 44(3):533–553. https://doi.org/10.1007/s11116-015-9666-2

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. Environ Int 141:105772. https://doi.org/10.1016/j.envint.2020.105772

Ziemke D, Nagel K, Bhat C (2015) Integrating CEMDAP and MATSIM to increase the transferability of transport demand models. Transp Res Rec 2493:117–125. https://doi.org/10.3141/2493-13