

Article

# Satellite-Derived PM<sub>2.5</sub> Composition and Its Differential Effect on Children's Lung Function

Khang Chau <sup>\*,†</sup> , Meredith Franklin <sup>†</sup>  and W. James Gauderman

Keck School of Medicine, University of Southern California, Los Angeles, CA 90032, USA; meredith.franklin@usc.edu (M.F.); jimg@usc.edu (W.J.G.)

\* Correspondence: khang.chau@usc.edu

† These authors contributed equally to this work.

Received: 26 February 2020; Accepted: 19 March 2020; Published: 23 March 2020

**Abstract:** Studies of the association between air pollution and children's health typically rely on fixed-site monitors to determine exposures, which have spatial and temporal limitations. Satellite observations of aerosols provide the coverage that fixed-site monitors lack, enabling more refined exposure assessments. Using aerosol optical depth (AOD) data from the Multiangle Imaging SpectroRadiometer (MISR) instrument, we predicted fine particulate matter, PM<sub>2.5</sub>, and PM<sub>2.5</sub> speciation concentrations and linked them to the residential locations of 1206 children enrolled in the Southern California Children's Health Study. We fitted mixed-effects models to examine the relationship between the MISR-derived exposure estimates and lung function, measured as forced expiratory volume in 1 second (FEV<sub>1</sub>) and forced vital capacity (FVC), adjusting for study community and biological factors. Gradient Boosting and Support Vector Machines showed excellent predictive performance for PM<sub>2.5</sub> (test  $R^2 = 0.68$ ) and its chemical components (test  $R^2 = 0.53$ – $0.71$ ). In single-pollutant models, FEV<sub>1</sub> decreased by 131 mL (95% CI:  $-232, -35$ ) per 10.7- $\mu\text{g}/\text{m}^3$  increase in PM<sub>2.5</sub>, by 158 mL (95% CI:  $-273, -43$ ) per 1.2- $\mu\text{g}/\text{m}^3$  increase in sulfates (SO<sub>4</sub><sup>2-</sup>), and by 177 mL (95% CI:  $-306, -56$ ) per 1.6- $\mu\text{g}/\text{m}^3$  increase in dust; FVC decreased by 175 mL (95% CI:  $-310, -29$ ) per 1.2- $\mu\text{g}/\text{m}^3$  increase in SO<sub>4</sub><sup>2-</sup> and by 212 mL (95% CI:  $-391, -28$ ) per 2.5- $\mu\text{g}/\text{m}^3$  increase in nitrates (NO<sub>3</sub><sup>-</sup>). These results demonstrate that satellite observations can strengthen epidemiological studies investigating air pollution health effects by providing spatially and temporally resolved exposure estimates.

**Keywords:** Aerosol optical depth; particulate matter; particulate matter speciation; machine learning; exposure estimation; children's health; children lung function

## 1. Introduction

Numerous studies have examined the association between exposure to traffic-related air pollution and children's respiratory health [1–4]. Long-term exposure to air pollution, indicated by concentrations of nitric oxides (NO, NO<sub>2</sub>, NO<sub>x</sub>), ozone (O<sub>3</sub>), particulate matter with aerodynamic diameter less than 2.5  $\mu\text{m}$  (PM<sub>2.5</sub>), and particulate matter with aerodynamic diameter less than 10  $\mu\text{m}$  (PM<sub>10</sub>), has been shown to lead to reduced lung development in children [5]. Fortunately, decreases in air pollution in Southern California over the past 17 years have led to significant reductions in these detrimental effects [6].

The air pollution exposures used in the aforementioned studies were estimated from concentrations measured at national- and state-operated fixed-site monitors [7]. For example, in a longitudinal assessment of air quality and lung development, Gauderman et al. [6] used concentrations of nitrogen dioxide (NO<sub>2</sub>), O<sub>3</sub>, PM<sub>2.5</sub>, PM<sub>10</sub> from one monitor within each community to determine exposures. With this approach, the 120 to 300 study subjects residing in each community were assigned

the same exposure. Spatial statistical techniques such as kriging, smoothing, and land use regression have been used to incorporate additional information (e.g., traffic, population density, elevation, land cover, and other geographic data) to characterize the spatial relationships in fixed-site monitoring data and interpolate concentrations to the unmonitored locations where there are health data [8,9]. While these approaches are valuable for generating exposures with greater spatial coverage, it has been shown that if their prediction performance is poor, subsequent epidemiological studies can yield severe biases and underestimation of standard errors in the health effects estimates [10].

Advances in using satellite observations of aerosol optical depth (AOD) to estimate ground-level concentrations of particulate matter (PM) air pollution have been extremely valuable in improving the spatial and temporal coverage of exposure estimates [11–15]. Among the satellite instruments most commonly used for PM estimation are the Multi-angle Implementation for Aerosols (MAIAC) of the Moderate Resolution Imaging Spectroradiometer (MODIS) on-board the NASA Earth Observing System (EOS) Terra and Aqua satellites [16], and the Multiangle Imaging SpectroRadiometer (MISR) on-board the Terra satellite [17]. Recent algorithms applied to observations from these instruments provide global, near-daily AOD at a spatially resolved grid resolutions (1 km, 4.4 km) [16,18].

Some studies have combined AOD from MODIS and MISR to derive PM<sub>2.5</sub> concentrations [19] and more recently PM<sub>2.5</sub> speciation concentrations [20]. MISR, given its configuration of nine cameras and four spectral bands, has the capability of differentiating aerosol size and type resulting in fractionated AOD [21]. In a recent study over Southern California, we reliably estimated PM<sub>2.5</sub> with MISR 4.4-km resolution AOD small+medium, and PM<sub>10</sub> with AOD large using generalized additive models (GAMs) [22]. The MISR AOD-derived PM concentrations were well correlated (confirmed by leave-one-site-out cross validation, CV) with EPA monitoring site data (PM<sub>2.5</sub> CV  $r = 0.71$ , PM<sub>10</sub> CV  $r = 0.66$ ). In the same region, speciated PM<sub>2.5</sub> (sulfate, SO<sub>4</sub><sup>2-</sup>; nitrate, NO<sub>3</sub><sup>-</sup>; organic carbon, OC; and elemental carbon, EC) were estimated using GAMs from 8 MISR component fractions combined with meteorology and geographic characteristics [23].

In a simulation study, high-resolution exposure estimates derived from satellite AOD were found to produce less biased acute and chronic health effects estimates with smaller standard errors than did exposure estimates derived from kriging PM<sub>2.5</sub> concentrations from fixed-site monitors [10]. Satellite-derived PM<sub>2.5</sub> concentrations have been instrumental in studies of the global burden of disease [24,25]. A few epidemiological studies of smaller cohorts have used satellite-derived PM<sub>2.5</sub> to estimate residential exposures in longitudinal children's health effects [26–28].

In this study, we derived daily PM<sub>2.5</sub> and PM<sub>2.5</sub> speciation (SO<sub>4</sub><sup>2-</sup>, NO<sub>3</sub><sup>-</sup>, EC, dust) exposures from 2000–2018 over the state of California by applying machine learning approaches to ground-level air quality measurements linked with the high-dimensional 4.4-km MISR AOD products and mixtures. Estimated annual average concentrations were then assigned to the residences of children in 8 Southern California communities to examine the chronic effects of exposure to PM<sub>2.5</sub> and the aforementioned PM<sub>2.5</sub> components on lung function. This study is unique in that it is the first of its kind to examine the differential effects of satellite-derived PM<sub>2.5</sub> speciation on children's respiratory health.

## 2. Materials and Methods

### 2.1. Particulate Matter Measurements

The United States Environmental Protection Agency (EPA) provides data on PM<sub>2.5</sub> concentrations from outdoor monitoring sites across the US through their Air Quality System (AQS) [29], and PM<sub>2.5</sub> speciation concentrations including ions, carbons, and metals through their Chemical Speciation Network (CSN) [30] (Figure A1). The AQS PM<sub>2.5</sub> data we used include daily averages of hourly measurements or 24-h integrated measurements from Federal Equivalent and Reference Method (FEM/FRM) instruments. The CSN sites measure 24-h chemical composition on a 1-in-3- or 1-in-6-day sampling schedule. In this study, we used all available AQS and CSN data in California beginning in March 2000, the earliest date for which MISR AOD data are available, and ending in July 2018 for PM<sub>2.5</sub>

speciation and in December 2018 for PM<sub>2.5</sub> mass. Among PM<sub>2.5</sub> species, we focused on predicting SO<sub>4</sub><sup>2-</sup> ion, NO<sub>3</sub><sup>-</sup> ion, elemental carbon (EC), and dust, which was calculated as a linear combination of aluminum (Al), calcium (Ca), iron (Fe), silicon (Si), and titanium (Ti) [31]:

$$Dust = 2.2 \times Al + 2.49 \times Si + 1.63 \times Ca + 1.94 \times Ti + 2.42 \times Fe.$$

This definition of dust pertains to fugitive geological materials, and has been shown to have a temporally stable compositional source profile over California [32].

## 2.2. MISR Aerosol Optical Depth Data

The MISR instrument has been collecting data from nine camera angles and four spectral bands since early 2000. Due to its narrower retrieval swaths, the instrument overpasses any given location every 3–5 days (between 10:00 and 13:00 local time) instead of daily as is the case with MODIS/MAIAC. The latest version (V23) of the retrieval algorithm re-processed the entire MISR mission at a 4.4-km spatial resolution (from 17.6-km resolution) [18] with improved retrievals over water [33]. In addition to total column AOD, MISR also characterizes the size, shape, and type of aerosol particles via these fractionated measures: small, medium, and large AOD (amount of particles of each size), nonspherical AOD (amount of nonspherical particles), and absorption AOD (amount of light-absorbing particles). These AOD features are hereafter referred to as AOD products.

In its auxiliary data products, MISR further provides 74 AOD mixtures, which are more fine-grained groupings of aerosol particles characteristics [21]. Specifically, AOD mixtures numbered 1–30 are made up of spherical, non-absorbing components, mixtures 31–50 contain spherical, absorbing components, and mixtures 51–74 contain spherical and nonspherical dust analogues. The MISR particle properties provide robust aerosol-type classification for distinguishing aerosol mass types including polluted, smoky, maritime, and dusty conditions [21,34]. Details of processing MISR data, particularly AOD mixtures, are documented in our previous paper [35].

## 2.3. Meteorological Data

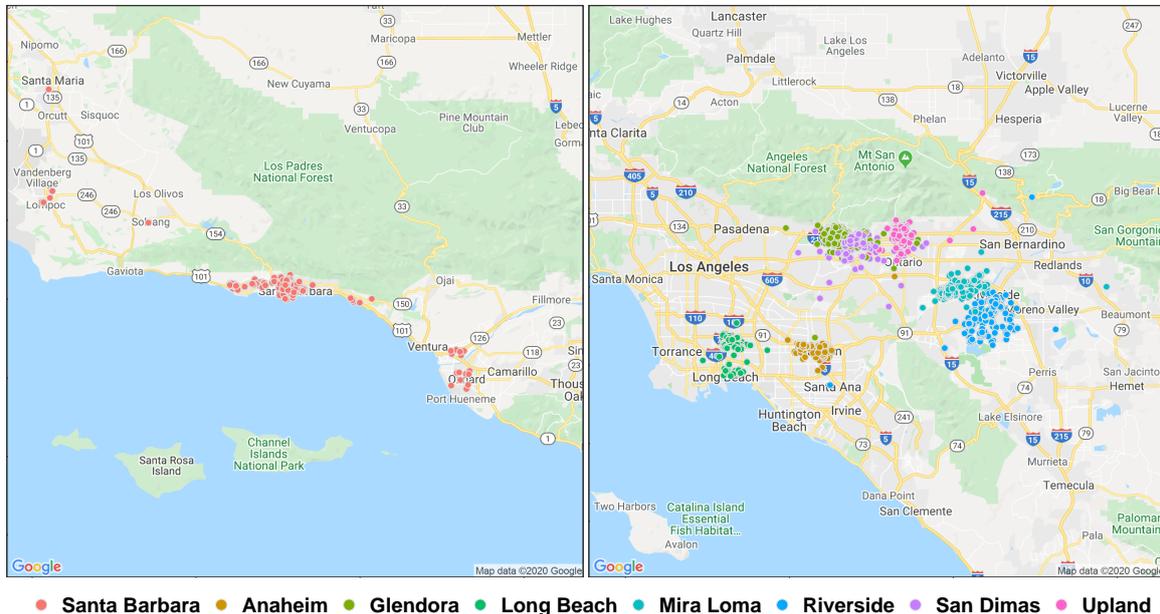
Similar to other studies [10,19,22,26,36], we used meteorological data including temperature, relative humidity, wind speed, and wind direction to better inform the association between ground-monitored air pollutants and satellite-observed AOD. The Climatology Lab at the University of Idaho provides daily meteorological data, called gridMET, in the contiguous United States from 1979–yesterday [37]. This dataset has been extensively validated [38] and, at 4-km resolution, is particularly useful for our study. The gridMET data also provide surface shortwave radiation (SSR—the amount of shortwave radiation that reaches the earth), whose negative correlation with aerosol emissions has been documented by Smith et al. [39] and Freychet et al. [40].

## 2.4. Health Data

Since its inception in the early 1990s, the Southern California Children's Health Study (CHS) has enrolled over 11,000 children in a series of five cohorts. In this study, we focused on the most recent cohort that began in 2003, enrolling approximately 3,000 children at age 6–7 years, and followed until 2012 when they were 15–16 years old. These children resided and went to school in eight communities in the greater Los Angeles, California area: Anaheim, Glendora, Long Beach, Mira Loma, Riverside, Santa Barbara, San Dimas, and Upland (Figure 1).

From 2007–2012, pulmonary function tests were conducted on each child by trained respiratory staff, measuring FEV<sub>1</sub> and FVC with pressure transducer-based spirometers (ScreenStar Spirometers, Morgan Scientific, Haverhill, Massachusetts, USA). A written questionnaire was also administered to obtain information including age, sex, self-identified race and ethnic background, parental education, occurrences of acute respiratory illness, exercise, tobacco-smoke exposure (personal smoking or environmental), and housing characteristics (air conditioning, age of house, presence of mildew,

pets in the home). Ethnic background in the CHS specifically relates to Hispanic ancestry, identifying Caucasian subjects with Hispanic and non-Hispanic ethnicity [41]. Study protocols were approved by the Institutional Review Board at the University of Southern California (USC), and additional details of CHS community and subject selection have been previously reported [42,43].



**Figure 1.** Map of the Children's Health Study participants who were recruited in Santa Barbara (left) and other communities (right).

### 2.5. Exposure Estimation Methods

We expand upon our previous work [22,35] to include MISR aerosol properties of absorption (absorbing or non-absorbing), shape (spherical or nonspherical), and type provided by 74 weighted aerosol optical depths (mixtures) [21] to predict  $PM_{2.5}$  and  $PM_{2.5} SO_4^{2-}$ ,  $NO_3^-$ , EC, and dust. We matched daily  $PM_{2.5}$  and  $PM_{2.5}$  speciation measurements from ground monitors to the nearest available MISR pixel within 4.4 km (Figure A2) and then further matched them to the nearest gridMET pixel within 4 km. As some  $PM_{2.5}$  monitors are close to each other, it is possible they share the same MISR or gridMET pixel. In these instances, we used an algorithm to pair the shared MISR/gridMET pixel to the nearest monitor, and the remaining monitors were paired with the next available MISR/gridMET pixel within 4.4/4 km. If there were no other available MISR/gridMET pixels within an appropriate distance, the  $PM_{2.5}$  measurement on that day for that monitor was removed in order to avoid duplicating AOD and meteorological data in the dataset.

We trained the  $PM_{2.5}$  models on 70% of the data using 5-fold CV and assessed performance of the best model on the remaining 30%. We also trained the  $PM_{2.5}$  speciation models on 70% of the data with 5-fold CV and tested on 30%, but repeated the process 20 times to assess model stability due to much smaller sample sizes. Five machine learning methods were considered: Ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), Gradient Boosting (GBM), Random Forests (RF), and Support Vector Machines (SVM), all within a regression setting. Inputs to the models were meteorology and either the MISR AOD products or the 74 MISR AOD mixtures. The optimal model for each pollutant was chosen based on its test  $R^2$  as the primary metric and its test RMSE as the secondary metric. We further supplemented the model with geospatial (coordinates of MISR pixels projected to UTM zone 11) and temporal (Julian date and month) predictors. The best predicting model for each pollutant was trained on the full dataset prior to estimating exposures for the epidemiological assessment.

## 2.6. Epidemiological Methods

To examine the association between air pollution exposure and lung function during the children's critical period of development, we focused on pulmonary function tests (FEV<sub>1</sub> and FVC) taken at ages 15–16 (2011–2012). During each assessment visit, the study participants reported their current addresses, which were geocoded. We identified MISR aerosol optical depth and gridMET meteorological data within 4.4 km of each study participant's residence for the 12 months prior to their pulmonary function test, and predicted PM<sub>2.5</sub>, NO<sub>3</sub><sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, EC, and dust concentrations spatially and temporally specific to each child. The 12-month means of these air pollutant estimates were assigned as exposures of interest for each child.

We fitted single-pollutant models to examine the effects of each predicted air pollutant exposure on lung function. Multi-pollutant models were also fitted to assess whether multiple predicted exposures better informed the associations of interest. Using mixed-effects models, we adjusted for study community with a random intercept and for biological characteristics such as age, gender, height, BMI, and race/ethnicity as fixed effects. Based on previous studies of the CHS, log transformation of FEV<sub>1</sub> and FVC as well as quadratic terms for height and BMI were considered [1,3,4,6]. We separately fitted similar models with central-site PM<sub>2.5</sub> to compare with MISR-derived PM<sub>2.5</sub> models. Akaike information criterion (AIC) was used as the main metric for model comparison and Bayesian information criterion (BIC), which favors more parsimonious models, as the secondary metric. Generalized variance inflation factor was calculated to assess potential collinearity among the predictors (using GVIF ≤ 10 as the cutoff).

## 3. Results

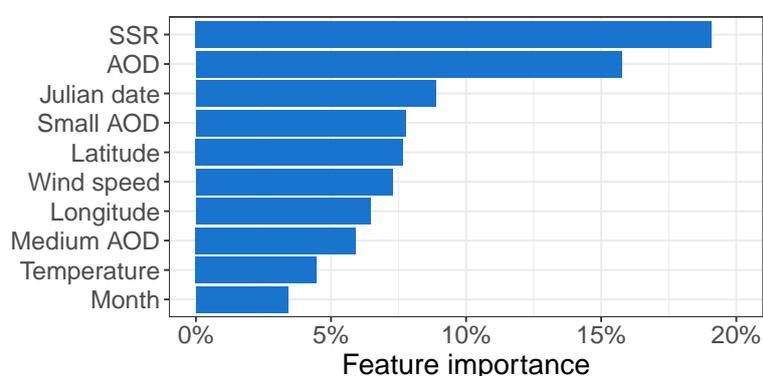
### 3.1. Exposure Estimation

From March 2000 through December 2018, there were 2828 days where MISR had complete retrievals of AOD products that could be matched to EPA-PM<sub>2.5</sub> stations, resulting in a dataset of  $N \approx 56,000$ . In the same period, there were 2864 days where MISR had complete retrievals of AOD mixtures that could be matched to EPA-PM<sub>2.5</sub> stations, resulting in a dataset of  $N \approx 35,000$ . While there were more days with successful AOD mixtures retrieved, the AOD-products datasets was larger because AOD-products were successfully retrieved in more locations per day compared to AOD mixtures. In a similar time frame (ending in July 2018), MISR had 698 days with AOD-products data and 544 days with AOD-mixtures data that were matched to PM<sub>2.5</sub> speciation concentrations from CSN sites in California, resulting in datasets of sizes  $N = 1965$  for models using AOD products and  $N = 1090$  for models using AOD mixtures.

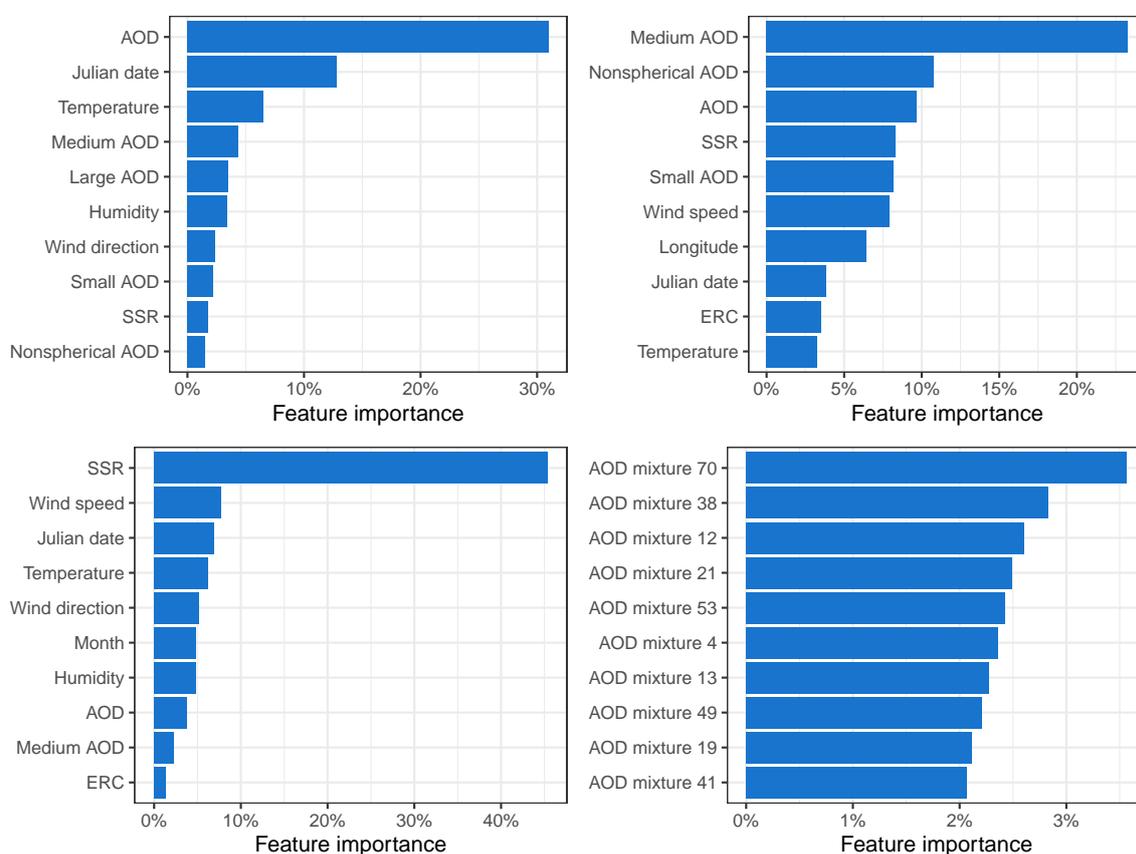
Non-linear machine learning methods (GBM, RF, and SVM) generally predicted better than linear methods (ridge and LASSO) for all exposures of interest and for both AOD-products and AOD-mixtures models. Although the AOD-mixtures models for PM<sub>2.5</sub> had a smaller training dataset, the increase in model complexity with 74 AOD mixtures versus 6 AOD products led to much more time-expensive model fitting. We experimented with a smaller sample size ( $N = 8000$ ) but the AOD-mixtures models did not perform better than the AOD-products models. The GBM model using AOD products had the best performance for PM<sub>2.5</sub> ( $R^2 = 0.68$ ) (Table 1).

The boxplots of test  $R^2$  (Figure A3) for PM<sub>2.5</sub> speciation models show that AOD-products models outperformed AOD-mixtures models and did so consistently (narrower boxplots) with the exception of dust, where AOD-mixtures models performed better. The best test  $R^2$  for SO<sub>4</sub><sup>2-</sup> and NO<sub>3</sub><sup>-</sup> was 0.71, EC 0.63, and dust 0.53 (Table 1). The most important variables for PM<sub>2.5</sub> include an interpretable mix of AOD, small and medium AOD, as well as meteorological variables (surface shortwave radiation, wind speed, and temperature) (Figure 2). Similar variables were important for SO<sub>4</sub><sup>2-</sup> and NO<sub>3</sub><sup>-</sup>, but nonspherical AOD played a larger role in both (Figure 3). Interestingly, AODs only ranked 8th and 9th for EC, with meteorology and temporal indicators playing a larger role in its prediction. Finally,

most important for dust were AOD mixtures relating primarily to dust (mixtures 70 and 53) and non-absorbing (mixtures 4, 12, 13, 19, 21) particles.

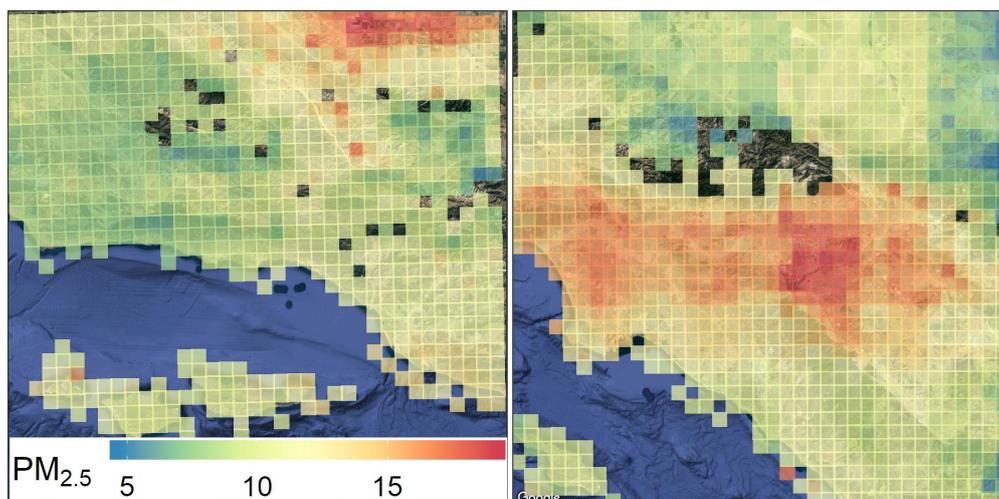


**Figure 2.** Relative importance of the ten most important features in model predicting  $PM_{2.5}$  using gradient boosting.

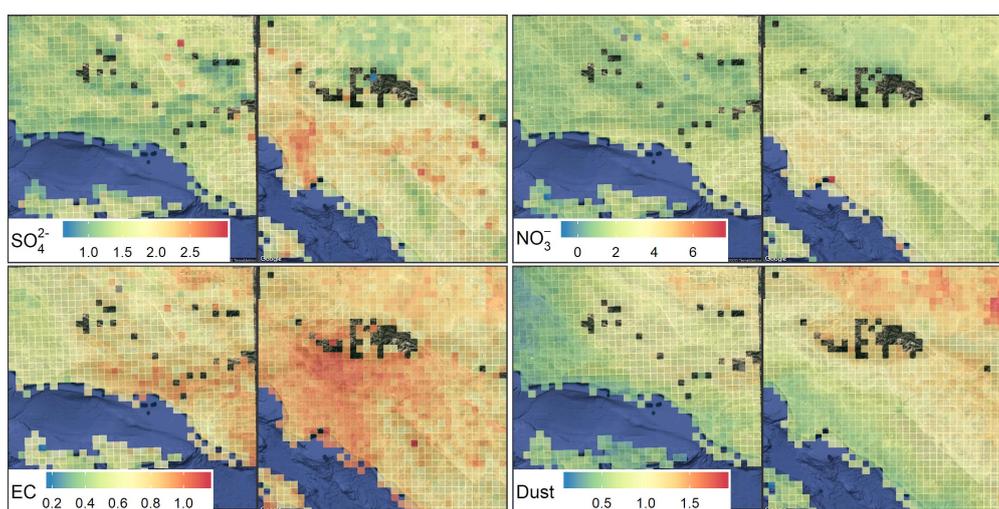


**Figure 3.** Relative importance of the ten most important features in models predicting  $SO_4^{2-}$  (top left),  $NO_3^-$  (top right), EC (bottom left), and dust (bottom right). Feature importance were calculated using relative influences in GBM models ( $PM_{2.5}$ ,  $SO_4^{2-}$ , and EC) [44] and global sensitivity analysis in SVM models ( $NO_3^-$  and dust) [45]. SSR: surface shortwave radiation; ERC: energy release component.

MISR AOD-products were used to predict all exposures except for dust, which was predicted using MISR AOD mixtures. Figures 4 and 5 show 6-year (2007–2012) means of the predicted air pollution exposures over greater Los Angeles and Santa Barbara. Over this period the correlation between annual mean MISR-derived  $PM_{2.5}$  and  $PM_{2.5}$  speciation with its central-site counterpart were positive and statistically significant, with Spearman  $r = 0.68$  and  $p < 0.001$  (Table A1).



**Figure 4.** Maps of mean MISR-derived  $PM_{2.5}$  ( $\mu\text{g}/\text{m}^3$ ) from 2007 through 2012 over Santa Barbara (left) and other communities in Southern California (right).



**Figure 5.** Maps of mean MISR-derived  $SO_4^{2-}$  (top left),  $NO_3^-$  (top right), EC (bottom left), and dust (bottom right) from 2007 through 2012, all measured in  $\mu\text{g}/\text{m}^3$  over Santa Barbara (left in each) and other communities in Southern California (right in each).

**Table 1.** Summary of machine learning methods, MISR predictors, and best test  $R^2$  for predicting each exposure of interest.

Exposure	ML Method	MISR Predictors	Best Test $R^2$	Best Test RMSE ( $\mu\text{g}/\text{m}^3$ )
$PM_{2.5}$	GBM	AOD Products	0.68	4.75
$SO_4^{2-}$	GBM	AOD Products	0.71	0.69
$NO_3^-$	SVM	AOD Products	0.71	2.13
EC	GBM	AOD Products	0.63	0.37
Dust	SVM	AOD Mixtures	0.53	0.60

### 3.2. Health Outcomes

There were 1206 children assessed in 2011–2012 with mean age 15.2 (SD = 0.6) years, whose geocoded addresses were linked with MISR-derived exposure estimates. Summary statistics of the study population, including gender, height, weight, race/ethnicity, and study community, are shown

in Table 2. Both FEV<sub>1</sub> and FVC were significantly higher among boys compared to girls; mean FEV<sub>1</sub> was 4111 mL among boys and 3304 mL among girls (*t*-test  $p < 0.001$ ), mean FVC was 4801 mL among boys and 3752 mL among girls (*t*-test  $p < 0.001$ ). Among race/ethnicity groups, there were significant differences in mean FEV<sub>1</sub> and FVC with one-way ANOVA test  $p = 0.003$  and  $p < 0.001$ , respectively (Table A2). The associations between biological factors and lung function remained significant when evaluated in fully adjusted models (Table A3).

**Table 2.** Characteristics of the study population

	Mean (SD)	N (%)
N	1206	
<b>Lung function</b>		
Forced Vital Capacity (mL)	4265 (856)	Gender
Forced Expiratory Volume (mL)	3700 (703)	Boys
<b>Exposures (µg/m<sup>3</sup>)</b>		Girls
<i>Central site</i>		<i>Race/Ethnicity</i>
PM <sub>2.5</sub>	11.99 (2.9)	White
<i>MISR-derived</i>		Asian
PM <sub>2.5</sub>	14.22 (2.6)	Black
Sulfate	1.37 (0.3)	Hispanic
Nitrate	2.36 (0.4)	Others
EC	0.82 (0.1)	<i>Community</i>
Dust	1.12 (0.3)	Anaheim
<b>Demographics</b>		Glendora
Age	15.24 (0.6)	Long Beach
Height (cm)	166.20 (8.6)	Mira Loma
Weight (kg)	64.41 (15.8)	Riverside
BMI (kg/m <sup>2</sup> )	23.23 (5.0)	San Dimas
		Santa Barbara
		Upland

**Table 3.** Air pollution single-pollutant effect estimates by outcome adjusted for demographic factors.

Source	Exposure	FEV <sub>1</sub>			FVC		
		Estimate *	95% CI	<i>p</i>	Estimate *	95% CI	<i>p</i>
<i>Central site</i>	PM <sub>2.5</sub>	−41	(−161, 80)	0.521	−21	(−162, 119)	0.775
<i>MISR-derived</i>	PM <sub>2.5</sub>	−131	(−232, −35)	<b>0.013</b>	−122	(−260, 25)	0.103
	SO <sub>4</sub> <sup>2−</sup>	−158	(−273, −43)	<b>0.008</b>	−175	(−310, −29)	<b>0.015</b>
	NO <sub>3</sub> <sup>−</sup>	−75	(−265, 124)	0.447	−212	(−391, −28)	<b>0.026</b>
	EC	−161	(−446, 128)	0.289	−218	(−547, 106)	0.206
	Dust	−177	(−306, −56)	<b>0.011</b>	−106	(−305, 95)	0.316

\* Effect estimates are the difference in FEV<sub>1</sub> and FVC from the highest to the lowest concentration of each air pollutant in µg/m<sup>3</sup>: [*Central sites*] PM<sub>2.5</sub>: 8.9; [*MISR-derived*] PM<sub>2.5</sub>: 10.7, SO<sub>4</sub><sup>2−</sup>: 1.2, NO<sub>3</sub><sup>−</sup>: 2.5, EC: 0.6, dust: 1.6.

In single-pollutant mixed models adjusted for biological characteristics, MISR-derived PM<sub>2.5</sub>, SO<sub>4</sub><sup>2−</sup>, and dust were significantly associated with decreases in FEV<sub>1</sub>, and SO<sub>4</sub><sup>2−</sup> and NO<sub>3</sub><sup>−</sup> were significantly associated with decreases in FVC. Notably, central-site PM<sub>2.5</sub> was not associated with either lung function measure (Table 3). Among multi-pollutant models, we did not find any grouping of two or more pollutants to be significantly associated with FEV<sub>1</sub> or FVC. Furthermore, while certain multi-pollutant models achieved modest reductions in AIC, none outperformed single-pollutant models in BIC. Log transformation of the lung function measures, with slight improvements in normality on Q-Q plots, did not improve model fit for either outcomes. Collinearity was not an issue in any of the models (all GVIF ≤ 3).

Effect estimates in Table 3 represent the difference in mean FEV<sub>1</sub> and FVC between the highest and the lowest exposure level for each pollutant in single-pollutant models. After accounting for

biological and community effects in single-pollutant mixed models, FEV<sub>1</sub> on average decreased by 131 mL (95% CI: −232, −35) per 10.7- $\mu\text{g}/\text{m}^3$  increase in PM<sub>2.5</sub>, by 158 mL (95% CI: −273, −43) per 1.2- $\mu\text{g}/\text{m}^3$  in SO<sub>4</sub><sup>2−</sup>, and by 177 mL (95% CI: −306, −56) per 1.6- $\mu\text{g}/\text{m}^3$  increase in dust. Meanwhile, FVC on average decreased by 175 mL (95% CI: −310, −29) per 1.2- $\mu\text{g}/\text{m}^3$  increase in SO<sub>4</sub><sup>2−</sup> and by 212 mL (95% CI: −391, −28) per 2.5- $\mu\text{g}/\text{m}^3$  increase in NO<sub>3</sub><sup>−</sup>.

#### 4. Discussion

In this study, we used satellite observations of AOD, characterized by size, shape, and absorption properties as well as fractionated into 74 mixtures, to estimate PM<sub>2.5</sub> and select PM<sub>2.5</sub> chemical components. We then incorporated these estimates into an epidemiological assessment of their association with children's lung function. In terms of exposure estimation, MISR AOD products resulted in better and more robust estimates than did AOD mixtures, except for dust. Non-linear models (GBM, RF, and SVM) performed better than linear models (Ridge and LASSO), which was consistent with previous studies where linear models were inadequate in explaining the relationship between AOD and ground-monitored PM [22,35,46]. Although MISR aerosol data have coarser temporal and spatial resolution compared to MAIAC (every 3–5 days vs. daily and 4.4 km vs. 1 km, respectively), our model achieved high prediction performance (Table 1) using MISR-specific data products on size, shape, and absorption, which proved vital in the prediction models. At least two MISR AOD products were among the five most important features for predicting PM<sub>2.5</sub>, SO<sub>4</sub><sup>2−</sup>, and NO<sub>3</sub><sup>−</sup>, and the ten most important features for predicting dust were all AOD mixtures (Figures 2 and 3).

Our PM<sub>2.5</sub> prediction performance was similar to those by Sorek-Hamer et al. [46], who modeled PM<sub>2.5</sub> using AOD data from MODIS (Collection 5 Level 2) and the Ozone Monitoring Instrument (OMI) at 10-km resolution over the Central Valley in California. In Southern California, we improved upon previous work that predicted PM<sub>2.5</sub>, SO<sub>4</sub><sup>2−</sup>, and NO<sub>3</sub><sup>−</sup> by Franklin et al. [22], whose meteorological data from NOAA weather stations did not provide the spatial coverage of gridMET data. Surface shortwave radiation provided by gridMET was also among the most important predictors for PM<sub>2.5</sub>, NO<sub>3</sub><sup>−</sup>, and EC (Figures 2 and 3). Our PM<sub>2.5</sub>, SO<sub>4</sub><sup>2−</sup>, NO<sub>3</sub><sup>−</sup>, and EC models performed comparably to those by Meng et al. [23], who reconstructed fractional AOD using the AOD mixtures (V22) while we relied on MISR AOD products (V23).

One limitation in our PM<sub>2.5</sub> speciation prediction models is the scarcity of data. As the number of CSN sites in California increased from 3 in 2000 to 19 in 2013 and decreased to 16 in 2018 (California PM<sub>2.5</sub> mass sites increased from 95 in 2000 to 157 in 2018), spatial coverage was certainly restricted (Figure A1). Furthermore, the locations of these sparsely available monitors are not necessarily representative of the population density of Southern California. We used the coordinates of MISR pixels, which lacked a fixed grid, instead of monitoring sites as geospatial predictors to help mitigate this problem by introducing additional spatial variability. We also attempted prediction models for PM<sub>10</sub> and SO<sub>2</sub> (detailed results not reported here) to compare with our previous work over Mongolia [35]. While PM<sub>10</sub> models performed about the same, i.e., average test  $R^2 < 0.10$ , SO<sub>2</sub> models over California performed much worse, with average test  $R^2 < 0.10$  (vs. test  $R^2 > 0.45$  over Ulaanbaatar). Poor prediction performance for SO<sub>2</sub> was likely due to much lower concentrations of SO<sub>2</sub> in California (mean SO<sub>2</sub> = 2.4 ppb in 2000–2018) compared to Ulaanbaatar (9.7 ppb in 2008–2017), where SO<sub>2</sub> is the more dominant source of PM. For the epidemiological purposes of the current study, we focused on models predicting PM<sub>2.5</sub> and its chemical components.

This study is unique in estimating air pollution exposure specifically to the residence and follow-up period of each subject. Previously, exposures were assigned using annual means from one central air pollution monitoring site for each study community [5,6,42], even if the children might have lived far away from these sites. Furthermore, the follow-up period spanned about 211 days, yet the annual means of central-site air pollutants for each community were calculated using a fixed time window. Leveraging MISR aerosol and gridMET meteorological data, we improved upon these limitations by assigning exposures that were spatially within 4.4 km of where each child lived and

temporally specific to the 12 months prior to each child's assessment visit. Nevertheless, our exposure prediction models are not without unexplained residual variance; our best models had CV  $R^2$  from 0.53 (dust) to 0.71 ( $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ ). As noted by Alexeeff et al. [10], there can be 1–5% upward bias in subsequent health effects estimates when exposure predictions have performance statistics in the range we observed, and their standard errors may be underestimated. It is difficult to mitigate these issues due to imperfect exposure models, but it is worth keeping in mind while interpreting our epidemiological results.

While this is not the first study using AOD-derived  $\text{PM}_{2.5}$  concentration in an epidemiological context, it is the first examining satellite-derived  $\text{PM}_{2.5}$  speciation. Previous studies of AOD-derived  $\text{PM}_{2.5}$  include Rice et al. [27], who found that each  $2 \mu\text{g}/\text{m}^3$  increase in AOD-derived  $\text{PM}_{2.5}$  was associated with a 28 mL (−43.9 to 0.2 mL) decrease in forced vital capacity (FVC) and higher odds of forced expiratory volume in 1 second ( $\text{FEV}_1$ ) being less than 80% predicted (OR = 1.41, 1.03 to 1.93). In another study, AOD-derived  $\text{PM}_{2.5}$  concentrations were associated with an increased rate of asthma onset (HR = 1.31, 1.28 to 1.33) in Quebec [28].

Similar to previous evaluations of the CHS [3,5,6], biological characteristics (age, gender, race/ethnicity, height, height squared, BMI, and BMI squared) were significantly associated with both measurements of lung function (Table A3). With these adjustments, several MISR-derived estimates of air pollutants were able to explain the residual differences in lung function measurements among the children. Another strength of this study is identifying the effect of specific  $\text{PM}_{2.5}$  chemical components on lung function. While MISR-derived  $\text{PM}_{2.5}$  was significantly associated with decreases in  $\text{FEV}_1$ , its effect, measured as the difference in  $\text{FEV}_1$  between the highest and lowest exposure level for each pollutant, is smaller than those of  $\text{SO}_4^{2-}$  and dust (Table 3). Similarly, although FVC was only marginally statistically significantly associated with MISR-derived  $\text{PM}_{2.5}$ , its associations with  $\text{SO}_4^{2-}$  and  $\text{NO}_3^-$  were clinically significant. In California, secondary aerosols including nitrate and sulfate have been shown to be the most abundant contributors to ambient  $\text{PM}_{2.5}$ , with nitrate accounting for as much as 55% of the total mass [47]. Geologic dust can also contribute up to 20% of the mass in summer in more arid regions of Southern California. Importantly, we were able to distinguish that these  $\text{PM}_{2.5}$  species had differentially stronger associations with children's  $\text{FEV}_1$  and FVC.

Urman et al. [3], who examined the cross-sectional effect of central-site air pollution on lung function in the same cohort but at an earlier visit when the children were at ages 11–12, found central-site  $\text{PM}_{2.5}$  to be significant with both log-transformed  $\text{FEV}_1$  and FVC. In our study, we did not find central-site  $\text{PM}_{2.5}$  to be significant with either outcomes. We did find MISR-derived  $\text{PM}_{2.5}$  to be significantly associated with log-transformed  $\text{FEV}_1$ , with a similar effect size, but not with log-transformed FVC. Children exposed to the highest level of MISR-derived  $\text{PM}_{2.5}$  on average were −3.54% (95% CI: −6.24%, −0.49%) lower in  $\text{FEV}_1$  compared to those exposed to the lowest level. In the same CHS cohort and during the same follow-up period, Franklin and Fruin [4] found a significant association between  $\text{NO}_x$  on FVC when adjusted for traffic-related noise exposure. We found a similarly significant relationship between  $\text{NO}_3^-$  and FVC where an IQR increase in  $\text{NO}_3^-$  ( $0.64 \mu\text{g}/\text{m}^3$ ) is associated with 53 mL decrease in FVC (95% CI: −98, −7).

## 5. Conclusions

We have shown in this study that MISR AOD observations distinguishing size, shape, absorption and mixture properties can aid in predicting  $\text{PM}_{2.5}$  and its chemical speciation including  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ , EC, and dust, particularly when supplemented with spatio-temporal information and high-resolution meteorological data. Machine learning methods such as Gradient Boosting and Support Vector Machines were more suitable for characterizing the non-linear relationships between air pollutants and AOD. We further showed that different MISR-derived  $\text{PM}_{2.5}$  composition, estimated specifically to the residence and follow-up period of CHS study participants, were able to explain clinically significant differences in lung function measurements  $\text{FEV}_1$  and FVC. This demonstrates that satellite-observed aerosol data products can be incorporated to strengthen epidemiological studies investigating the

health effects of environmental pollution. Epidemiological assessments will only be made more viable, particularly as the quality of remote sensing data and estimation models continue to improve and exposure measurement error decreases.

**Author Contributions:** Conceptualization, M.F.; methodology, K.C., M.F., W.J.G.; formal analysis, K.C.; data curation, K.C., M.F.; writing—original draft preparation, K.C., M.F.; writing—review and editing, K.C., M.F., W.J.G.; supervision, M.F.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Aeronautics and Space Administration grant number NNN12AA01C and under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No. CR-83590201) and certain motor vehicle and engine manufacturers.

**Acknowledgments:** The authors would like to thank Olga Kalashnikova, Michael Garay, and Dave Diner from the Jet Propulsion Laboratory’s MISR team for their guidance and support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
AOD	Aerosol Optical Depth
CHS	Children’s Health Study
CSN	Chemical Speciation Network
FEV <sub>1</sub>	Forced expiratory volume in 1 second
FVC	Forced vital capacity
MAIAC	MultiAngle Implementation of Atmospheric Correction
MISR	Multiangl e Imaging SpectroRadiometer
MODIS	MoDerate resolution Imaging Spectroradiometer
NO <sub>3</sub> <sup>-</sup>	Nitrate
NO <sub>2</sub>	Nitrogen Dioxide
NO <sub>x</sub>	Nitrogen Oxides
PM	Particulate Matter
PM <sub>2.5</sub>	Particulate Matter with aerodynamic diameter ≤ 2.5 μm
PM <sub>10</sub>	Particulate Matter with aerodynamic diameter ≤ 10 μm
PM	Particulate Matter
SO <sub>4</sub> <sup>2-</sup>	Sulfate

## Appendix A

**Table A1.** Spearman correlation coefficients for annual means of air pollution exposures during the three CHS follow-up periods (2007–2012).

	PM <sub>2.5</sub> <sup>*</sup>	PM <sub>2.5</sub> <sup>†</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	EC	Dust
PM <sub>2.5</sub> <sup>*</sup>	1	0.68	0.55	0.34	0.29	0.68
PM <sub>2.5</sub> <sup>†</sup>		1	0.74	0.59	0.15	0.38
SO <sub>4</sub> <sup>2-</sup>			1	0.92	0.29	0.14
NO <sub>3</sub> <sup>-</sup>				1	0.37	-0.12
EC					1	0.32
Dust						1

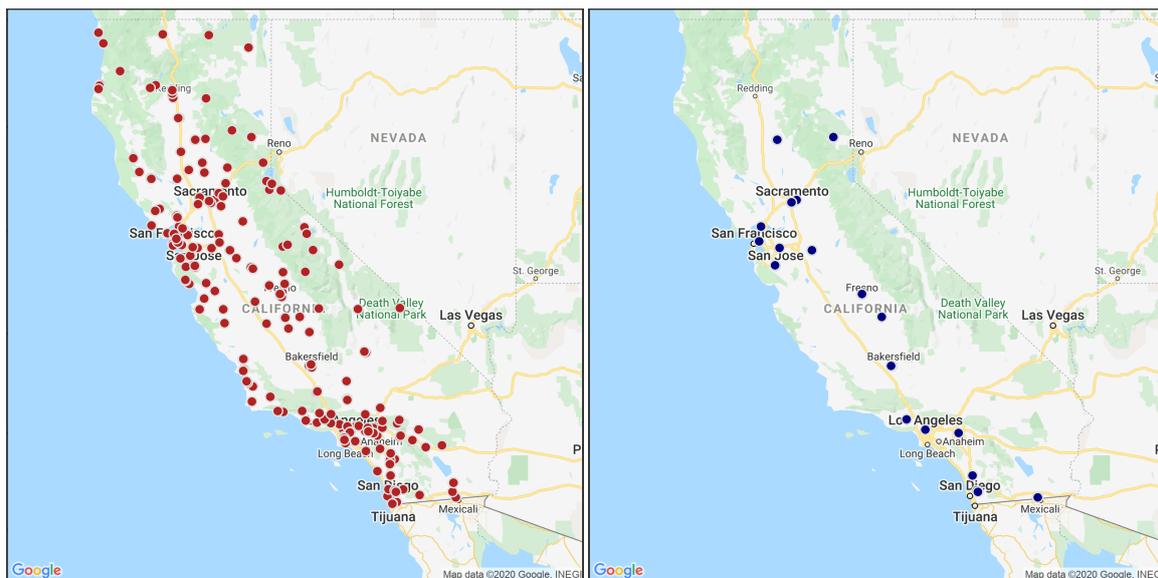
\*Central-site PM<sub>2.5</sub>; †MISR-derived PM<sub>2.5</sub>.

**Table A2.** Mean (SD) of FEV<sub>1</sub> and FVC by gender and race/ethnicity.

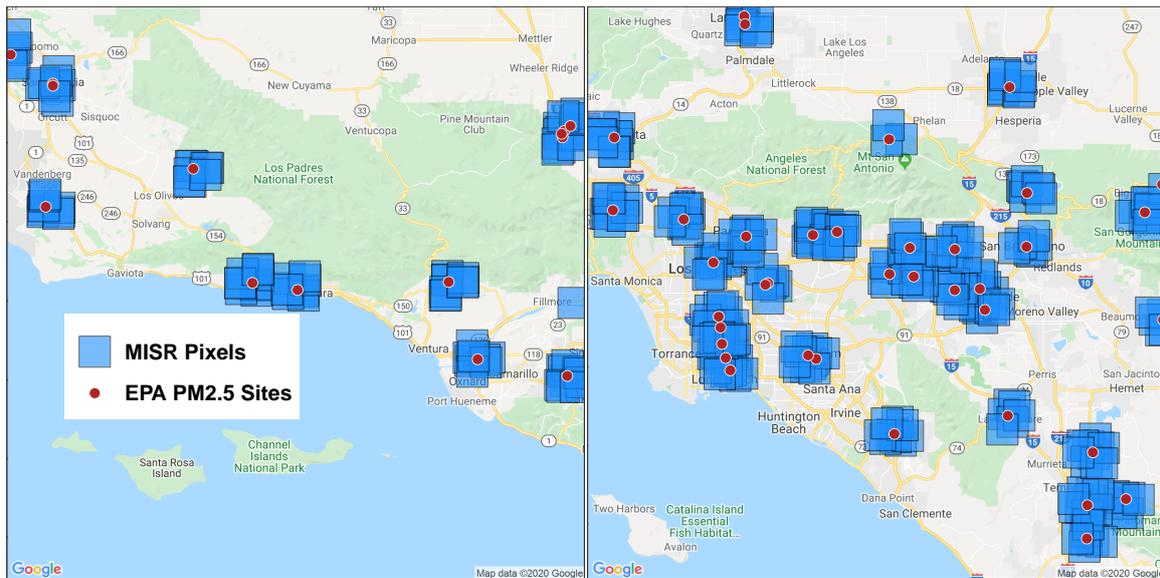
	FEV <sub>1</sub>	FVC
<i>Gender</i>		
Female	3304 (453)	3752 (537)
Male	4111 (680)	4801 (797)
<i>Race/Ethnicity</i>		
White	3746 (665)	4377 (831)
Asian	3492 (635)	3892 (746)
Black	3271 (548)	3777 (683)
Hispanic	3715 (733)	4258 (884)
Others	3634 (688)	4162 (774)

**Table A3.** Effect estimates of biological characteristics on lung function in mixed models without air pollution exposure.

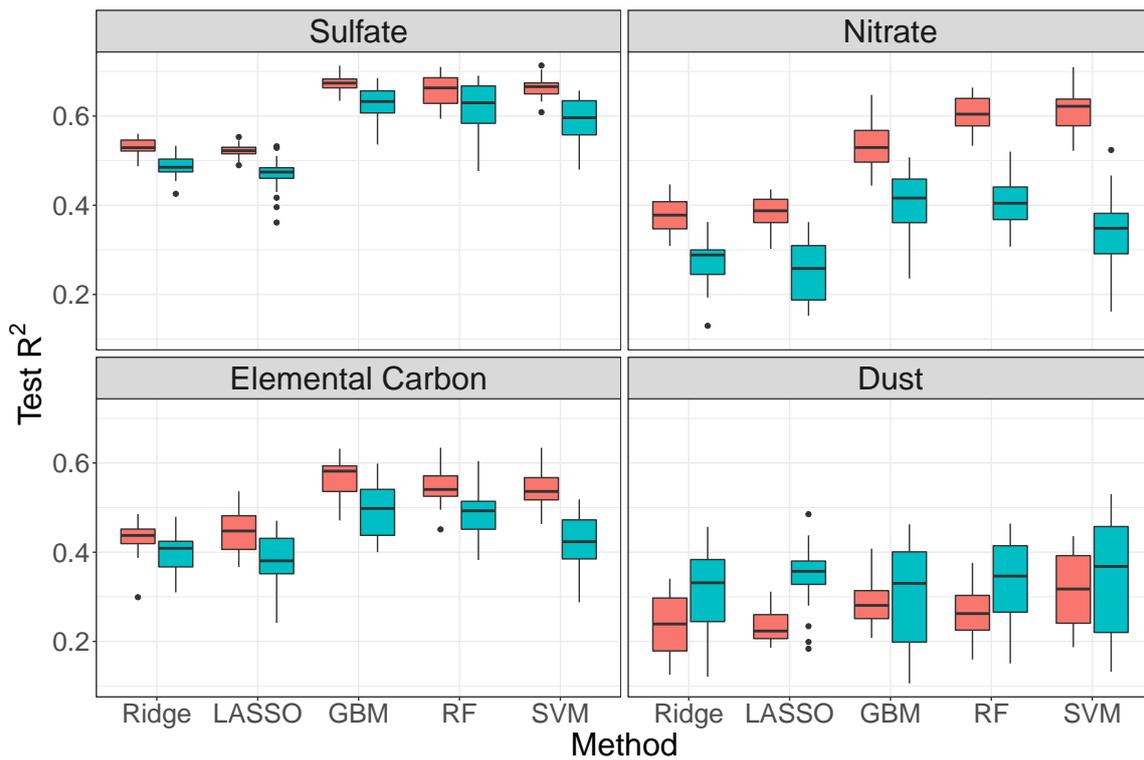
Characteristic	FEV <sub>1</sub>		FVC	
	Estimate	95% CI	Estimate	95% CI
Age (year)	121	(−81, 162)	112	(67, 156)
<i>Gender</i>				
Girls	(ref.)	−	(ref.)	−
Boys	317	(−256, 378)	440	(372, 506)
<i>Race/ethnicity</i>				
White	(ref.)	−	(ref.)	−
Asian	−21	(−145, 102)	−188	(−323, −54)
Black	−331	(−513, −151)	−454	(−653, −254)
Hispanic	96	(−38, 154)	40	(−24, 103)
Others	52	(−42, 146)	−12	(−115, 91)
Height (cm)	47	(44, 51)	59	(55, 63)
Height <sup>2</sup>	0.3	(0.0, 0.5)	0.5	(0.3, 0.8)
BMI (kg/m <sup>2</sup> )	35.7	(29, 42)	49	(42, 56)
BMI <sup>2</sup>	−2.4	(−3.0, −1.8)	−2.8	(−3.5, −2.2)



**Figure A1.** Maps of EPA PM<sub>2.5</sub> monitoring sites (left) and CSN monitoring sites (right) from 2000 to 2018.



**Figure A2.** Maps of EPA PM<sub>2.5</sub> monitoring sites and the MISR pixels within 4.4 km of each site, in Santa Barbara (left) and other communities (right).



**Figure A3.** Boxplots of test  $R^2$  from 20 iterations of each machine learning method applied to predicting PM<sub>2.5</sub> speciation, in AOD-products (coral) and AOD-mixtures (teal) models.

**References**

- Gauderman, W.J.; Vora, H.; McConnell, R.; Berhane, K.; Gilliland, F.; Thomas, D.; Lurmann, F.; Avol, E.; Kunzli, N.; Jerrett, M.; et al. Effect of exposure to traffic on lung development from 10 to 18 years of age: A cohort study. *Lancet* **2007**, *369*, 571–577. doi:10.1016/S0140-6736(07)60037-3. [CrossRef]

2. McConnell, R.; Islam, T.; Shankardass, K.; Jerrett, M.; Lurmann, F.; Gilliland, F.; Gauderman, J.; Avol, E.; Künzli, N.; Yao, L.; et al. Childhood incident asthma and traffic-related air pollution at home and school. *Environ. Health Perspect.* **2010**, *118*, 1021–1026. doi:10.1289/ehp.0901232. [[CrossRef](#)] [[PubMed](#)]
3. Urman, R.; McConnell, R.; Islam, T.; Avol, E.L.; Lurmann, F.W.; Vora, H.; Linn, W.S.; Rappaport, E.B.; Gilliland, F.D.; Gauderman, W.J. Associations of children’s lung function with ambient air pollution: Joint effects of regional and near-roadway pollutants. *Thorax* **2014**, *69*, 540–547. doi:10.1136/thoraxjnl-2012-203159. [[CrossRef](#)]
4. Franklin, M.; Fruin, S. The role of traffic noise on the association between air pollution and children’s lung function. *Environ. Res.* **2017**, *157*, 153–159. doi:10.1016/j.envres.2017.05.024. [[CrossRef](#)] [[PubMed](#)]
5. Gauderman, W.J.; Avol, E.; Gilliland, F.; Vora, H.; Thomas, D.; Berhane, K.; McConnell, R.; Kuenzli, N.; Lurmann, F.; Rappaport, E.; et al. The Effect of Air Pollution on Lung Development from 10 to 18 Years of Age. *N. Engl. J. Med.* **2004**, *351*, 1057–1067. doi:10.1056/NEJMoa040610. [[CrossRef](#)]
6. Gauderman, W.J.; Urman, R.; Avol, E.; Berhane, K.; McConnell, R.; Rappaport, E.; Chang, R.; Lurmann, F.; Gilliland, F. Association of Improved Air Quality with Lung Development in Children. *N. Engl. J. Med.* **2015**, *372*, 905–913. doi:10.1056/NEJMoa1414123. [[CrossRef](#)]
7. AQS (Air Quality System) User Guide. Available online: <https://www.epa.gov/aqs/aqs-user-guide> (accessed on 16 January 2020).
8. Ryan, P.H.; Lemasters, G.K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007**, *19*, 127–133. doi:10.1080/08958370701495998. [[CrossRef](#)]
9. Franklin, M.; Vora, H.; Avol, E.; McConnell, R.; Lurmann, F.; Liu, F.; Penfold, B.; Berhane, K.; Gilliland, F.; Gauderman, W.J. Predictors of intra-community variation in air quality. *J. Expo. Sci. Environ. Epidemiol.* **2012**, *22*, 135–147. doi:10.1038/jes.2011.45. [[CrossRef](#)]
10. Alexeeff, S.E.; Schwartz, J.; Kloog, I.; Chudnovsky, A.; Koutrakis, P.; Coull, B.A. Consequences of kriging and land use regression for PM<sub>2.5</sub> predictions in epidemiologic analyses: Insights into spatial variability using high-resolution satellite data. *J. Expo. Sci. Environ. Epidemiol.* **2015**, *25*, 138–144. doi:10.1038/jes.2014.40. [[CrossRef](#)]
11. Brauer, M.; Amann, M.; Burnett, R.T.; Cohen, A.; Dentener, F.; Ezzati, M.; Henderson, S.B.; Krzyzanowski, M.; Martin, R.V.; Van Dingenen, R.; et al. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environ. Sci. Technol.* **2012**, *46*, 652–660. doi:10.1021/es2025752. [[CrossRef](#)]
12. De Sherbinin, A.; Levy, M.A.; Zell, E.; Weber, S.; Jaiteh, M. Using satellite data to develop environmental indicators. *Environ. Res. Lett.* **2014**, *9*, 084013. doi:10.1088/1748-9326/9/8/084013. [[CrossRef](#)]
13. Van Donkelaar, A.; Martin, R.V.; Spurr, R.J.; Drury, E.; Remer, L.A.; Levy, R.C.; Wang, J. Optimal estimation for global ground-level fine particulate matter concentrations. *J. Geophys. Res. Atmos.* **2013**, *118*, 5621–5636. doi:10.1002/jgrd.50479. [[CrossRef](#)]
14. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Boys, B.L. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* **2015**, *123*, 135–143. doi:10.1289/ehp.1408646. [[CrossRef](#)] [[PubMed](#)]
15. Shaddick, G.; Thomas, M.L.; Green, A.; Brauer, M.; van Donkelaar, A.; Burnett, R.; Chang, H.H.; Cohen, A.; Dingenen, R.V.; Dora, C.; et al. Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution. *J. R. Stat. Soc. Ser. Appl. Stat.* **2018**, *67*, 231–253. doi:10.1111/rssc.12227. [[CrossRef](#)]
16. Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741–5765. doi:10.5194/amt-11-5741-2018. [[CrossRef](#)]
17. Diner, D.J.; Beckert, J.C.; Bothwell, G.W.; Rodriguez, J.I. Performance of the MISR instrument during its first 20 months in earth orbit. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1449–1466. doi:10.1109/TGRS.2002.801584. [[CrossRef](#)]
18. Garay, M.J.; Kalashnikova, O.V.; Bull, M.A. Development and assessment of a higher-spatial-resolution (4.4 km) MISR aerosol optical depth product using AERONET-DRAGON data. *Atmos. Chem. Phys.* **2017**, *17*, 5095–5106. doi:10.5194/acp-17-5095-2017. [[CrossRef](#)]
19. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. doi:10.1289/ehp.0901623. [[CrossRef](#)]

20. Van Donkelaar, A.; Martin, R.V.; Li, C.; Burnett, R.T. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environ. Sci. Technol.* **2019**, *53*, 2595–2611. doi:10.1021/acs.est.8b06392. [[CrossRef](#)]
21. Kahn, R.A.; Gaitley, B.J. An analysis of global aerosol type as retrieved by MISR. *J. Geophys. Res. Atmos.* **2015**, *120*, 4248–4281. doi:10.1002/2015JD023322. [[CrossRef](#)]
22. Franklin, M.; Kalashnikova, O.V.; Garay, M.J. Size-resolved particulate matter concentrations derived from 4.4 km-resolution size-fractionated Multi-angle Imaging SpectroRadiometer (MISR) aerosol optical depth over Southern California. *Remote Sens. Environ.* **2017**, *196*, 312–323. doi:10.1016/j.rse.2017.05.002. [[CrossRef](#)]
23. Meng, X.; Garay, M.J.; Diner, D.J.; Kalashnikova, O.V.; Xu, J.; Liu, Y. Estimating PM<sub>2.5</sub> speciation concentrations using prototype 4.4 km-resolution MISR aerosol properties over Southern California. *Atmos. Environ.* **2018**, *181*, 70–81. doi:10.1016/j.atmosenv.2018.03.019. [[CrossRef](#)]
24. Burnett, R.T.; Pope, C.A.; Ezzati, M.; Olives, C.; Lim, S.S.; Mehta, S.; Shin, H.H.; Singh, G.; Hubbell, B.; Brauer, M.; et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **2014**, *122*, 397–403. doi:10.1289/ehp.1307049. [[CrossRef](#)] [[PubMed](#)]
25. Cohen, A.J.; Brauer, M.; Burnett, R.; Anderson, H.R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **2017**, *389*, 1907–1918. doi:10.1016/s0140-6736(17)30505-6. [[CrossRef](#)]
26. Kloog, I.; Nordio, F.; Coull, B.A.; Schwartz, J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM<sub>2.5</sub> exposures in the mid-atlantic states. *Environ. Sci. Technol.* **2012**, *46*, 11913–11921. doi:10.1021/es302673e. [[CrossRef](#)] [[PubMed](#)]
27. Rice, M.B.; Rifas-Shiman, S.L.; Litonjua, A.A.; Oken, E.; Gillman, M.W.; Kloog, I.; Luttmann-Gibson, H.; Zanobetti, A.; Coull, B.A.; Schwartz, J.; et al. Lifetime exposure to ambient pollution and lung function in children. *Am. J. Respir. Crit. Care Med.* **2016**, *193*, 881–888. doi:10.1164/rccm.201506-1058OC. [[CrossRef](#)]
28. Tétreault, L.F.; Doucet, M.; Gamache, P.; Fournier, M.; Brand, A.; Kosatsky, T.; Smargiassi, A. Childhood exposure to ambient air pollutants and the onset of asthma: An administrative cohort study in Québec. *Environ. Health Perspect.* **2016**, *124*, 1276–1282. doi:10.1289/ehp.1509838. [[CrossRef](#)]
29. Air Data: Air Quality Data Collected at Outdoor Monitors Across the US. Available online: <https://www.epa.gov/outdoor-air-quality-data> (accessed on 14 January 2020).
30. Chemical Speciation Network (CSN). Available online: <https://www.epa.gov/amtic/chemical-speciation-network-csn> (accessed on 14 January 2020).
31. Chow, J.C.; Lowenthal, D.H.; Chen, L.W.; Wang, X.; Watson, J.G. Mass reconstruction methods for PM<sub>2.5</sub>: A review. *Air Qual. Atmos. Health* **2015**, *8*, 243–263. [[CrossRef](#)]
32. Chow, J.C.; Watson, J.G.; Ashbaugh, L.L.; Magliano, K.L. Similarities and differences in PM<sub>10</sub> chemical source profiles for geological dust from the San Joaquin Valley, California. *Atmos. Environ.* **2003**, *37*, 1317–1340. doi:10.1016/s1352-2310(02)01021-x. [[CrossRef](#)]
33. Witek, M.L.; Garay, M.J.; Diner, D.J.; Bull, M.A.; Seidel, F.C. New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water. *Atmos. Meas. Tech.* **2018**, *11*, 429–439. doi:10.5194/amt-11-429-2018. [[CrossRef](#)]
34. Kahn, R.A.; Gaitley, B.J.; Garay, M.J.; Diner, D.J.; Eck, T.F.; Smirnov, A.; Holben, B.N. Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network. *J. Geophys. Res. Atmos.* **2010**, *115*, D23209. doi:10.1029/2010JD014601. [[CrossRef](#)]
35. Franklin, M.; Chau, K.; Kalashnikova, O.V.; Garay, M.J.; Enebish, T.; Sorek-Hamer, M. Using multi-angle imaging spectro radiometer aerosol mixture properties for air quality assessment in Mongolia. *Remote Sens.* **2018**, *10*, 1–14. doi:10.3390/RS10081317. [[CrossRef](#)]
36. Al-Saadi, J.; Szykman, J.; Pierce, R.B.; Kittaka, C.; Neil, D.; Chu, D.A.; Remer, L.; Gumley, L.; Prins, E.; Weinstock, L.; et al. Improving national air quality forecasts with satellite aerosol observations. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1249–1261. doi:10.1175/BAMS-86-9-1249. [[CrossRef](#)]
37. gridMET - Climatology Lab. Available online: <http://www.climatologylab.org/gridmet.html> (accessed on 14 January 2020).
38. Abatzoglou, J.T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **2013**, *33*, 121–131. doi:10.1002/joc.3413. [[CrossRef](#)]

39. Smith, D.M.; Booth, B.B.B.; Dunstone, N.J.; Eade, R.; Hermanson, L.; Jones, G.S.; Scaife, A.A.; Sheen, K.L.; Thompson, V. Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nat. Clim. Chang.* **2016**, *6*, 936–940. doi:10.1038/nclimate3058. [[CrossRef](#)]
40. Freychet, N.; Tett, S.F.B.; Bollasina, M.; Wang, K.C.; Hegerl, G.C. The Local Aerosol Emission Effect on Surface Shortwave Radiation and Temperatures. *J. Adv. Model. Earth Syst.* **2019**, *11*, 806–817. doi:10.1029/2018ms001530. [[CrossRef](#)]
41. Weaver, G.M.; James Gauderman, W. Traffic-Related Pollutants: Exposure and Health Effects among Hispanic Children. *Am. J. Epidemiol.* **2018**, *187*, 45–52. doi:10.1093/aje/kwx223. [[CrossRef](#)]
42. Peters, J.M.; Avol, E.; Navidi, W.; London, S.J.; Gauderman, W.J.; Lurmann, F.; Linn, W.S.; Margolis, H.; Rappaport, E.; Gong, H.; et al. A study of twelve Southern California communities with differing levels and types of air pollution: I. Prevalence of respiratory morbidity. *Am. J. Respir. Crit. Care Med.* **1999**, *159*, 760–767. doi:10.1164/ajrccm.159.3.9804143. [[CrossRef](#)]
43. McConnell, R.; Berhane, K.; Yao, L.; Jerrett, M.; Lurmann, F.; Gilliland, F.; Künzli, N.; Gauderman, J.; Avol, E.; Thomas, D.; et al. Traffic, susceptibility, and childhood asthma. *Environ. Health Perspect.* **2006**, *114*, 766–772. doi:10.1289/ehp.8594. [[CrossRef](#)]
44. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. doi:10.1214/aos/1013203451. [[CrossRef](#)]
45. Cortez, P.; Embrechts, M.J. Using sensitivity analysis and visualization techniques to open black box data mining models. *Inf. Sci.* **2013**, *225*, 1–17. doi:10.1016/j.ins.2012.10.039. [[CrossRef](#)]
46. Sorek-Hamer, M.; Strawa, A.W.; Chatfield, R.B.; Esswein, R.; Cohen, A.; Broday, D.M. Improved retrieval of PM<sub>2.5</sub> from satellite data products using non-linear methods. *Environ. Pollut.* **2013**, *182*, 417–423. doi:10.1016/j.envpol.2013.08.002. [[CrossRef](#)] [[PubMed](#)]
47. Hasheminassab, S.; Daher, N.; Saffari, A.; Wang, D.; Ostro, B.D.; Sioutas, C. Spatial and temporal variability of sources of ambient fine particulate matter (PM<sub>2.5</sub>) in California. *Atmos. Chem. Phys.* **2014**, *14*, 12085–12097. doi:10.5194/acp-14-12085-2014. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).