Environmental Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Examining $PM_{2.5}$ concentrations and exposure using multiple models

James T. Kelly [a,*], Carey Jang [a], Brian Timin [a], Qian Di [b], Joel Schwartz [c], Yang Liu [d], Aaron van Donkelaar [e,f], Randall V. Martin [e,f,g], Veronica Berrocal [h], Michelle L. Bell [i]

[a] Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA
[b] Vanke School of Public Health, Tsinghua University, Beijing, China
[c] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[d] Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[e] Department of Energy, Environmental & Chemical Engineering, Washington University, St. Louis, MO, USA
[f] Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada
[g] Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA, USA
[h] Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA, USA
[i] School of the Environment, Yale University, New Haven, CT, USA

ARTICLE INFO

Keywords:
$PM_{2.5}$
Air quality modeling
Ensemble modeling
Exposure inequality

ABSTRACT

Epidemiologic studies have found associations between fine particulate matter ($PM_{2.5}$) exposure and adverse health effects using exposure models that incorporate monitoring data and other relevant information. Here, we use nine $PM_{2.5}$ concentration models (i.e., exposure models) that span a wide range of methods to investigate i) $PM_{2.5}$ concentrations in 2011, ii) potential changes in $PM_{2.5}$ concentrations between 2011 and 2028 due to on-the-books regulations, and iii) $PM_{2.5}$ exposure for the U.S. population and four racial/ethnic groups. The exposure models included two geophysical chemical transport models (CTMs), two interpolation methods, a satellite-derived aerosol optical depth-based method, a Bayesian statistical regression model, and three data-rich machine learning methods. We focused on annual predictions that were regridded to 12-km resolution over the conterminous U.S., but also considered 1-km predictions in sensitivity analyses. The exposure models predicted broadly consistent $PM_{2.5}$ concentrations, with relatively high concentrations on average over the eastern U.S. and greater variability in the western U.S. However, differences in national concentration distributions (median standard deviation: 1.00 μg m^{-3}) and spatial distributions over urban areas were evident. Further exploration of these differences and their implications for specific applications would be valuable. $PM_{2.5}$ concentrations were estimated to decrease by about 1 μg m^{-3} on average due to modeled emission changes between 2011 and 2028, with decreases of more than 3 μg m^{-3} in areas with relatively high 2011 concentrations that were projected to experience relatively large emission reductions. Agreement among models was closer for population-weighted than uniformly weighted averages across the domain. About 50% of the population was estimated to experience $PM_{2.5}$ concentrations less than 10 μg m^{-3} in 2011 and $PM_{2.5}$ improvements of about 2 μg m^{-3} due to modeled emission changes between 2011 and 2028. Two inequality metrics were used to characterize differences in exposure among the four racial/ethnic groups. The metrics generally yielded consistent information and suggest that the modeled emission reductions between 2011 and 2028 would reduce absolute exposure inequality on average.

1. Introduction

Epidemiologic studies have reported associations between concentrations of ambient fine particulate matter ($PM_{2.5}$) and health effects including mortality (e.g., Di et al., 2017; Pappin et al., 2019; Pope et al., 2020; USEPA, 2019). Globally, 4.2 million deaths have been attributed to $PM_{2.5}$ air pollution in 2015 (Cohen et al., 2017). Epidemiologic and health-impact studies rely on accurate characterizations of population exposure to provide understanding of the health effects of ambient $PM_{2.5}$. Due to the limited spatial and temporal coverage of routine

* Corresponding author. Office of Air Quality, Planning & Standards, US Environmental Protection Agency, 109 TW Alexander Drive, Research Triangle Park, NC, 27711, USA.
E-mail address: kelly.james@epa.gov (J.T. Kelly).

https://doi.org/10.1016/j.envres.2020.110432
Received 4 September 2020; Received in revised form 22 October 2020; Accepted 3 November 2020
Available online 7 November 2020
0013-9351/Published by Elsevier Inc.

monitoring networks, many $PM_{2.5}$ concentration models have been developed in recent years to provide complete coverage for exposure applications over large domains including the conterminous U.S.

$PM_{2.5}$ concentration models (hereafter exposure models) have been developed using various approaches including geophysical process modeling (Ostro et al., 2015; Wang et al., 2017), monitor-only methods (e.g., monitor averaging, nearest-monitor, interpolation) (Keller and Peng, 2019), hybrid land-use regression modeling (Beckerman et al., 2013; Jerrett et al., 2017), satellite-derived aerosol optical depth (AOD) methods (van Donkelaar et al., 2010), Bayesian statistical modeling (Berrocal et al., 2010), and data-rich machine learning models (Chen et al., 2019; Di et al., 2016, 2019; Hu et al., 2017). These exposure models often demonstrate good predictive capability based on statistical performance metrics from cross-validation tests with withheld observations. For instance, several machine learning methods have reported R^2 values of 0.80 or greater based on ten-fold cross validation of daily $PM_{2.5}$ predictions over the conterminous U.S. in 2011 (Di et al., 2016, 2019; Hu et al., 2017). However, model performance may vary with the spatial and temporal resolution of the application, and the suitability of a model could depend on the resolution and type of the epidemiologic analysis to be performed (e.g., cross-sectional analysis versus case-crossover design).

Epidemiologic studies have examined the influence of different exposure characterizations on $PM_{2.5}$-mortality associations and found some agreement in results. For instance, Jerrett et al. (2017) reported associations between $PM_{2.5}$ exposure and mortality risk using multiple exposure models, although methods that incorporated ground-based observations yield larger risks than those estimated based on a remote-sensing-only model. Di et al. (2017) reported associations between $PM_{2.5}$ exposure and mortality risks using the neural network exposure model of Di et al. (2016) and a nearest-monitor approach, with slightly larger risks based on the national exposure model. Health impact assessments have also examined the influence of differences in exposure characterization on outcomes. For instance, Jin et al. (2019) estimated a 28% uncertainty in the state-level mortality burden for New York associated with the choice of method used to estimate $PM_{2.5}$ exposure. Ford and Heald (2016) estimated similar variability in the mortality burden for the U.S. during 2004–2011 based on satellite-derived exposure estimates.

Comparisons of $PM_{2.5}$ concentrations based on different exposure models have also been performed. Jin et al. (2019) reported that seven publicly available $PM_{2.5}$ products captured the decrease in $PM_{2.5}$ concentrations over NY between 2002 and 2012, but that the models had limited ability to resolve the intra-urban spatial patterns of $PM_{2.5}$ evident in a high-resolution monitoring network. Studies have also reported increasing differences in concentration among methods and degradation in performance with increasing distance to the nearest monitor (Berrocal et al., 2020; Huang et al., 2018; Jin et al., 2019; Kelly et al., 2019a). These results indicate that performance for many methods degrades in very places where we need exposure estimates, locations far from monitors. Diao et al. (2019) recently reported inconsistencies in $PM_{2.5}$ estimates from several publicly available $PM_{2.5}$ products and encouraged future research efforts.

Investigations of $PM_{2.5}$ concentrations over the conterminous U.S. based on multiple methods of exposure are limited and have not included predictions of some of the most widely used methods. Also, previous studies have not examined the influence of exposure modeling methods on the future-year air quality projections commonly used in regulatory planning. Moreover, the implications of differences in modeling approaches for characterizing population exposure have received little consideration, despite the importance of exposure characterization for epidemiologic and health impact assessments. Finally, exposure models have recently been used to examine exposure inequality in Massachusetts (Rosofsky et al., 2018), but modeling of exposure inequality over larger domains is limited (Bravo et al., 2016) and has not considered possible future changes in $PM_{2.5}$ exposure inequality.

Here, we characterize annual average $PM_{2.5}$ concentrations in 2011 over the conterminous U.S. using nine exposure models that span a wide range of complexity and data sources. We then estimate the change in $PM_{2.5}$ concentrations between 2011 and 2028 due to modeled emission changes associated with a suite of on-the-books regulations. Finally, we characterize $PM_{2.5}$ exposure (i.e., population-weighted concentrations) and exposure inequality according to two metrics using modeled $PM_{2.5}$ concentrations for the 2011 and 2028 cases. The use of multiple models is intended to provide insights on current exposure modeling methods as well as a thorough characterization of $PM_{2.5}$ concentrations and exposure.

2. Methods

We characterized $PM_{2.5}$ concentrations for 2011 due to the greater availability of exposure model predictions for this period than more recent years. Also, chemical transport model (CTM) simulations are available from previous work to project 2011 $PM_{2.5}$ concentrations in response to emission changes estimated to occur between 2011 and 2028 (USEPA, 2017a). Nine exposure models were used that can loosely be classified into categories of geophysical process models, interpolation-based models, Bayesian statistical regression, a satellite-AOD-based model, and machine learning models. The models are based on a wide range of input data and algorithms and were developed independently by multiple research groups.

2.1. Exposure models

$PM_{2.5}$ concentration estimates from two geophysical CTMs were used. CTMs simulate concentrations using scientific parameterizations of the key atmospheric processes that govern pollutant concentrations including emissions, transport, chemical reaction, and deposition. CTMs are therefore valuable for predicting $PM_{2.5}$ concentrations under hypothetical scenarios where observations are not possible. For exposure applications, CTM predictions are often calibrated to observations using statistical models. Here, we consider the uncalibrated CTM predictions as a reference for the statistical model predictions, and because CTM predictions are sometimes directly used in health studies (Ostro et al., 2015; Wang et al., 2017). First, $PM_{2.5}$ exposure fields were developed from a simulation of 2011 with the Community Multiscale Air Quality model, version 5.0.2. CMAQ was configured with the Carbon Bond 05 (CB05) chemical mechanism, and anthropogenic emissions were based version 1 of the 2011 national emission inventory (NEI11). The model configuration, performance evaluation, and emission inventory have been described in detail previously (Kelly et al., 2019b; USEPA, 2015a). Second, $PM_{2.5}$ exposure fields were developed using a simulation of 2011 based on the Community Atmosphere Model with Extensions (CAMx), version 6.3.2. CAMx was configured with the Carbon Bond chemical mechanism (CB6r4), and anthropogenic emissions were based on version 2 of NEI11. The CAMx model configuration and performance evaluation have been described in detail by USEPA (2017a), and model emissions have been described by USEPA (2016) and USEPA (2017b). The modeling domain for the CTM simulations covered the conterminous U.S. with 12-km grid spacing. The national emission totals for the CMAQ and CAMx simulations were within 2% for NOx, SO_2, and volatile organic compounds (VOCs), and primary $PM_{2.5}$ emissions were 4% higher in the CMAQ simulation.

A $PM_{2.5}$ exposure field was also developed by interpolating monitor concentrations to the 12-km CTM grid. We used the Voronoi neighbor averaging (VNA) interpolation approach, which calculates the $PM_{2.5}$ concentration at a prediction location as the inverse-distance-squared weighted average of monitored concentrations in neighboring Voronoi polygons (Abt, 2012; Fann et al., 2012). Additionally, $PM_{2.5}$ concentrations were estimated using the extended VNA approach that weights the VNA interpolation using the ratio of CTM predictions between the

monitor location and the prediction point (Abt, 2012; Fann et al., 2012). This weighting effectively fuses the spatial gradients from the CTM field with the interpolated field of observations from VNA. The CAMx simulation described above was used in the eVNA calculations. VNA and eVNA predictions were made quarterly on the 12-km CTM grid using the Speciated Modeled Attainment Test-Community Addition (SMAT--CEv1.3) software. VNA and eVNA calculations are computationally efficient and can be made using publicly available software.

A $PM_{2.5}$ concentration field was also developed using a Bayesian statistical downscaling model (Berrocal et al., 2010) that incorporated $PM_{2.5}$ monitoring data and the CMAQ simulation described above. The downscaler model is a linear regression model that relates observed concentrations to CMAQ output using spatially and temporally varying coefficients. The downscaler is fit within a Bayesian framework and provides estimates of prediction uncertainty (Berrocal et al., 2020). The downscaler has been found to achieve relatively good model performance based on a limited amount of input data by explicitly accounting for the spatial dependence in $PM_{2.5}$ concentrations (Berrocal et al., 2020). Daily $PM_{2.5}$ predictions were made on the national 12-km CTM grid.

Also, $PM_{2.5}$ predictions were used from a model that first relates satellite-derived AOD retrievals to surface $PM_{2.5}$ concentrations using the geophysical relationship predicted by a global CTM (GEOS-Chem) with a nested simulation over North America. The initial $PM_{2.5}$ estimate (based on AOD and GEOS-Chem) is then calibrated to ground-based observations using geographically weighted regression as described by van Donkelaar et al. (2019; V4.NA.02). Here we characterize $PM_{2.5}$ concentrations over the conterminous U.S. using the V4.NA.02.MAPLE product, which modifies the original V4.NA.02 product with additional developments as part of the MAPLE (Mortality–Air Pollution Associations in Low-Exposure Environments) project. Annual predictions at 1-km resolution were available from this method (hereafter VD2019), although fine-scale features at 1-km resolution are not expected to be fully resolved due to the influence of some coarser resolution inputs on the final dataset. Previous studies have reported that earlier versions of the VD2019 approach perform relatively well in remote areas due to the use of remote sensing information (Jin et al., 2019; Lee et al., 2012).

$PM_{2.5}$ exposure fields from three data-rich machine learning methods were also compared. Machine learning methods can model the complex nonlinear and interactive relationships among predictor variables across wide range of datasets relevant to $PM_{2.5}$ formation. $PM_{2.5}$ predictions were used from a neural network model that incorporated >50 predictor variables including predictions of a global CTM and satellite-derived AOD (Di et al., 2016). Daily $PM_{2.5}$ predictions over the conterminous U.S. were available from this method at 1-km resolution (hereafter DI2016). Predictions based on a random forest model with about 40 predictors variables were also used (Hu et al., 2017). Important predictor variables in the random forest included satellite-derived AOD and a convolutional layer for nearby $PM_{2.5}$ measurements. Daily $PM_{2.5}$ concentration predictions were developed at 12-km resolution with this method (hereafter: HU2017). Finally, $PM_{2.5}$ concentrations based on an ensemble of predictions from three machine learning methods (gradient boosting, random forest, and neural network) were used (Di et al., 2019). $PM_{2.5}$ estimates from the individual learners are optimally combined using a generalized additive model that accounts for geographic variations with thin-plate spline functions. This model (hereafter DI2019) incorporated a wide range of predictors including the CMAQ simulation described above.

The nine exposure models are summarized in Table 1. Due to differences in the spatial and temporal resolution of the methods and domain coverage, we averaged the fields to the coarsest spatial (12 km) and temporal (annual) resolution and included only grid cells where all models had coverage. Spatial averaging was done by calculating the mean of model prediction points located within a 12-km grid cell. The study therefore focuses on 12-km resolution, but we also consider 1-km fields in some cases to understand the effects of spatial averaging. The

Table 1
Summary of exposure modeling methods.

Case	Method Description	Original Resolution[a]	Performance Statistics[b]	Reference
CMAQ	Geophysical process model based on unconstrained bottom-up methods	Hourly, 12-km	[c]Total R^2: 0.24 RMSE: 7.36 MB: 0.57 Slope: 0.62	USEPA (2015a); Kelly et al. (2019b)
CAMx	Geophysical process model based on unconstrained bottom-up methods	Hourly, 12-km	[c]Total R^2: 0.27 RMSE: 6.35 MB: −0.29 Slope: 0.53	USEPA (2017a)
VNA	Inverse-distance-squared weighted interpolation of $PM_{2.5}$ observations in neighboring Voronoi polygons	Quarterly, 12-km	[d]Total R^2: 0.68 RMSE: 3.5 MB: 0.20 Slope: 0.74	Abt (2012); Kelly et al. (2019a)
eVNA	VNA Interpolation of monitor data but with interpolation weighted by the ratio of CTM predictions in the monitor cell to the prediction cell.	Quarterly, 12-km	[d]Total R^2: 0.48 RMSE: 5.60 MB: 0.40 Slope: 0.88	Abt (2012); Kelly et al. (2019a)
Downscaler	Bayesian statistical regression of CTM predictions and observations with spatially varying coefficients	Daily, 12-km	[d]Total R^2: 0.66 RMSE: 3.7 Bias: 0.8 Slope: 0.78	Berrocal et al. (2010); Kelly et al. (2019a)
VD2019	CTM scaling of satellite AOD to surface $PM_{2.5}$ with geographic weighted regression of residuals.	Monthly, 1-km	[e]R^2: 0.73 (0.55,0.86) RMSE: 1.8 (1.5,2.4) Bias: 0.05 (−0.09,0.20) Slope: 0.88 (0.72,0.95).	van Donkelaar et al. (2019), following Table 1, modified as per V4. NA.02. MAPLE update
DI2016	Neural network model based on wide range of predictors including satellite AOD and CTM modeling	Daily, 1-km	[f]Total R^2: 0.81 RMSE: 2.83 Bias: 0.38 Slope: 0.99	Di et al. (2016)
HU2017	Random forest model based on ~40 predictors including satellite AOD and CTM modeling	Daily, 12-km	[f]Total R^2: 0.80 RMSE: 2.83 Slope: 1.00	Hu et al. (2017)
DI2019	Ensemble model based on random forest, gradient boosting, and neural network learners with a wide range of predictors	Daily, 1-km	[f]Total R^2: 0.832 RMSE: 2.670 Bias: 0.742 Slope: 0.940	Di et al. (2019)

[a] Fields were averaged here to the annual period and a common 12-km grid.
[b] RMSE: root-mean-square error (μg m^{-3}); MB: mean bias (μg m^{-3}).
[c] Statistics are based on daily $PM_{2.5}$ predictions for 2011 that were not constrained with observations.

[d] Statistics are based on 10-fold cross-validation of daily $PM_{2.5}$ predictions for 2015.

[e] Statistics are based on 10-fold cross-validation of annual $PM_{2.5}$ for 2000–2016 in the U.S. (mean and range for individual years).

[f] Statistics are based on 10-fold cross-validation of daily $PM_{2.5}$ predictions for 2011.

performance of the methods against withheld monitoring data has been reported previously and is consistent with the state-of-the-science for the individual approaches. Model performance is considered further in the Supporting Information (Text S1 and Figure S1), and the methods are summarized in Table 1. Note that the $PM_{2.5}$ from the CTMs is calculated as the sum of simulated $PM_{2.5}$ components and is not identical to the measured $PM_{2.5}$, which is operationally defined by the Federal Reference Method (Noble et al., 2001).

2.2. Projecting $PM_{2.5}$ based on 2011 and 2028 emissions

We projected the 2011 $PM_{2.5}$ concentrations from the nine exposure models to 2028 using CAMx modeling based on emissions for 2011 and 2028. In addition to the 2011 simulation described above, a second CAMx simulation was available that accounted for emission changes expected to occur between 2011 and 2028 due to a suite of on-the-books regulations and other factors (e.g., planned plant shutdowns). Emission inputs for the 2028 CAMx simulation were developed by projecting emission sectors individually from the 2011 CAMx case described above using spatially and temporally resolved information. The emission totals for the 2011 and 2028 simulations are provided in Table 2, and a full description of the emission projections is available in USEPA (2017b). The largest contribution to the 63% decrease in SO_2 emissions over the conterminous U.S. between the 2011 and 2028 simulations was from electric generation units (76% decrease). The largest contribution to the 50% decrease in modeled NO_x ($NO + NO_2$) emissions was from the onroad mobile source sector (77% decrease) followed by electric generation units (61% decrease). Note that the meteorology and chemical boundary conditions in the 2028 CAMx simulation were fixed at their 2011 values, and so the 2028 CAMx modeling reflects changes compared to 2011 due to emission changes alone.

To help reduce the influence of model bias, $PM_{2.5}$ concentrations were projected using relative response factors (RRFs) developed from the ratios of modeled 2011 and 2028 $PM_{2.5}$ concentrations (Cohan and Chen, 2014; NRC, 2004; USEPA, 2018). RRFs are generally applied on a species-specific basis to help mitigate the potential influence of a large model bias for one species dominating the overall response of $PM_{2.5}$. RRFs for total $PM_{2.5}$ in each grid cell were developed from RRFs for individual particle components by taking the species-weighted average of the component RRFs:

$$RRF_{species} = \frac{C_{2028,species}}{C_{2011,\ species}} \quad (1)$$

$$RRF_{Tot,PM2.5} = \frac{\sum C_{Obs,species} RRF_{species}}{\sum C_{Obs,species}} \quad (2)$$

where $C_{2028,species}$ and $C_{2011,species}$ are the modeled $PM_{2.5}$ component concentrations in the 2028 and 2011 CAMx simulations, respectively; $RRF_{species}$ is the RRF for a given $PM_{2.5}$ particle component; $C_{Obs,\ species}$ is the $PM_{2.5}$ component concentration in 2011 based on VNA interpolation of observations that have been adjusted for consistency with total $PM_{2.5}$ measurements (Frank, 2006); and $RRF_{Tot,PM2.5}$ is the RRF for total $PM_{2.5}$. Calculation of $C_{Obs,\ species}$ and $RRF_{species}$ was done using SMAT-CEv1.3. Finally, $PM_{2.5}$ concentrations were projected from 2011 to 2028 using the RRFs for total $PM_{2.5}$ as follows:

$$PM2.5_{Mod,2028} = RRF_{Tot,PM2.5} PM2.5_{Mod,2011} \quad (3)$$

where $PM2.5_{Mod,2011}$ is the modeled $PM_{2.5}$ concentration in 2011 and $PM25_{Mod,2028}$ is the modeled $PM_{2.5}$ concentration in the 2028 emission case. Therefore the 2028 projected $PM_{2.5}$ concentrations are influenced both by the RRFs (from CAMx) and the 2011 modeled concentrations (from each of the nine exposure models).

Table 2
Modeled emission totals[a] for conterminous U.S. in 2011 and 2028 CAMx simulations.

Pollutant	2011 (t yr^{-1}-[a])	2028 (t yr^{-1})	Change 2011 to 2028
SO_2	6,403,986	2,351,163	−63%
NO_x	14,163,826	7,135,556	−50%
VOC	17,160,045	13,660,423	−20%
$PM_{2.5}$	4,599,665	4,413,668	−4%

[a] Emission totals were developed by projecting sectors individually with spatially and temporally resolved information. Projections for electric generation unit emissions include the Final Mercury and Air Toxics rule (December 21, 2011); the Cross-State Air Pollution Rule (CSAPR) (July 6, 2011); and the CSAPR Update Rule (October 26, 2016). Projections for onroad mobile sources include the Tier-3 Vehicle Emissions and Fuel Standards Program (March 2014); the California Low-Emission Vehicle (LEVIII) program; and Light-Duty Vehicle Greenhouse Gas Emissions and Corporate Average Fuel Economy Standards (October 2012); and Greenhouse Gas Emissions Standards and Fuel Efficiency Standards for Medium- and Heavy-Duty Engines and Vehicles (September 2011). See USEPA (2017b) for more details.

2.3. Population data and exposure metrics

To examine population exposure using the modeled concentration fields, we calculated population-weighted concentrations for the total population in the domain and for racial/ethnic subgroups as follows:

$$C_{PM2.5,Pop.weighted} = \frac{\sum Pop_i C_{PM2.5,i}}{\sum Pop_i} \quad (4)$$

where $C_{PM25,Pop.weighted}$ is the population-weighted concentration of $PM_{2.5}$ for a given population group, Pop_i is the population size in grid cell i for the given population group, and $C_{PM2.5,i}$ is the $PM_{2.5}$ concentration in grid cell i.

We used 2010 Census data based on the 1-km gridded product available from the Socioeconomic Data and Applications Center (CIESIN, 2017) to characterize population distributions. In addition to total U.S. population, we considered the following racial/ethnic groups: non-Hispanic Black (NH-Black), non-Hispanic white (NH-white), non-Hispanic other (NH-other), and Hispanic. For exposure calculations based on 12-km concentration fields, the 1-km population data were regridded to the 12-km domain used for the exposure models. For calculations with 1-km fields, modeled concentrations were matched with the nearest-neighbor grid point from the 1-km population data product. Population data were fixed at 2010 levels in all calculations, and so changes in exposure between 2011 and 2028 are due to concentration changes only.

Two metrics were used to characterize differences in exposure for population sub-groups. First, we calculated the difference in population-weighted $PM_{2.5}$ concentrations between the most and least exposed groups by state (i.e., the exposure gap), which is similar to the risk gap of Thind et al. (2019). The exposure gap is an absolute metric that uses the best-off group as the reference and does not include an explicit inequality aversion parameter (Harper et al., 2013; Levy et al., 2006). Second, we calculated the Between Group Atkinson index (AI_{BG}), which was recently applied by Rosofsky et al. (2018) to examine trends in $PM_{2.5}$ exposure inequality in Massachusetts. The AI_{BG} is a relative metric that uses average exposure as the reference and includes an explicit inequality parameter, ε, which reflects the level of societal concern for inequality (Harper et al., 2013). The AI_{BG} ranges from zero to one, with higher values indicating a less-equal distribution. Following Rosofsky et al. (2018), we calculate the AI_{BG} as follows:

$$AI_{BG} = 1 - \left(\sum_{j=1}^{n} f_j \left[\frac{\bar{y}_j}{\bar{y}} \right]^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}} \quad (5)$$

where n represents the number of subgroups in the population, f_j represents the fraction of the total population in each subgroup, $\bar{y}_j$ represents mean exposure of each subgroup, and $\bar{y}$ represents the mean exposure over the full population within a given geographic boundary (i.e., state).

Here we set ε to 0.75 consistent with previous studies (Fann et al., 2011, 2018; Levy et al., 2006; Rosofsky et al., 2018). We perform all AI_{BG} calculations using the inverse of the $PM_{2.5}$ concentrations, since the AI was originally developed for income distributions, where high income is considered desirable, whereas we are assessing $PM_{2.5}$ exposure, where high exposure is undesirable (Harper et al., 2013; Rosofsky et al., 2018).

3. Results and discussion

3.1. 2011 $PM_{2.5}$ concentrations

Annual average $PM_{2.5}$ concentrations predicted by the nine models are shown in Fig. 1. In general, $PM_{2.5}$ concentrations are higher in the eastern than western U.S., with some exceptions such as the high concentrations in the San Joaquin Valley and South Coast Air Basin of California. By U.S. climate region (Figure S2), the highest concentrations on average were predicted in the Ohio Valley region, where mean $PM_{2.5}$ ranged from 9.2 to 10.7 μg m^{-3} across all models (see Table S1 for the mean and range of model concentrations by region). The lowest concentrations were predicted in the Southwest region, where mean concentrations ranged from 2.6 to 5.3 μg m^{-3}. The spatial variation in concentrations in the western U.S., where some of the cleanest and most polluted areas are located, tended to be greater than in the eastern U.S. (Fig. 1).

The model predictions are in general agreement and are consistent with previous observation-based characterizations of $PM_{2.5}$ spatial distributions in the U.S. (Hand et al., 2012; USEPA, 2019). However, differences in the methodologies and input datasets lead to a range of model predictions. For instance, the CTMs predicted lower domain-wide average concentrations than the non-CTM models (i.e. 5.7 μg m^{-3} vs. 6.3–7.6 μg m^{-3}) but relatively high concentrations in urban areas in the eastern U.S. (Fig. 1). The standard deviation (sd) across the models (excluding the CTMs) is shown in Fig. 2a to illustrate the spatial distribution of the spread in model predictions (see Figure S3 for the maximum difference across models). The median (25th, 75th percentile) sd over all U.S. grid cells is 1.00 (0.78,1.26) μg m^{-3}, with the highest values in the western U.S. The sd divided by the mean (i.e., the coefficient of variation, cv) is shown in percentage units in Fig. 2b. The median (25th, 75th percentile) cv over all U.S. grid cells is 15.49 (9.34, 25.19) %. The cv is lower in the east than the west—e.g., the median value is 9.60% for grid cells east of 100W and 25.05% for cells west of 100W. This difference is due in part to the generally higher $PM_{2.5}$ concentrations in the eastern than western U.S. Modeling $PM_{2.5}$ concentrations in the west is more challenging than in the east due to factors such as complex terrain, prevalent wildfire, relatively sparse monitoring, and areas with low $PM_{2.5}$ concentration and sharp $PM_{2.5}$ gradients (Di et al., 2016; Geng et al., 2018; Kelly et al., 2019b).

The sd has limited variation as a function of $PM_{2.5}$ concentration, with slightly higher values in the 5–7 μg m^{-3} range (Figure S4a). In contrast, the cv increases with decreasing $PM_{2.5}$ concentration (Figure S4b). Previous studies have reported some degradation in the R^2 performance statistic with decreasing concentration, whereas bias and RMSE were relatively insensitive to concentration (e.g., Just et al., 2020; van Donkelaar et al., 2019). This behavior could be due to increases in

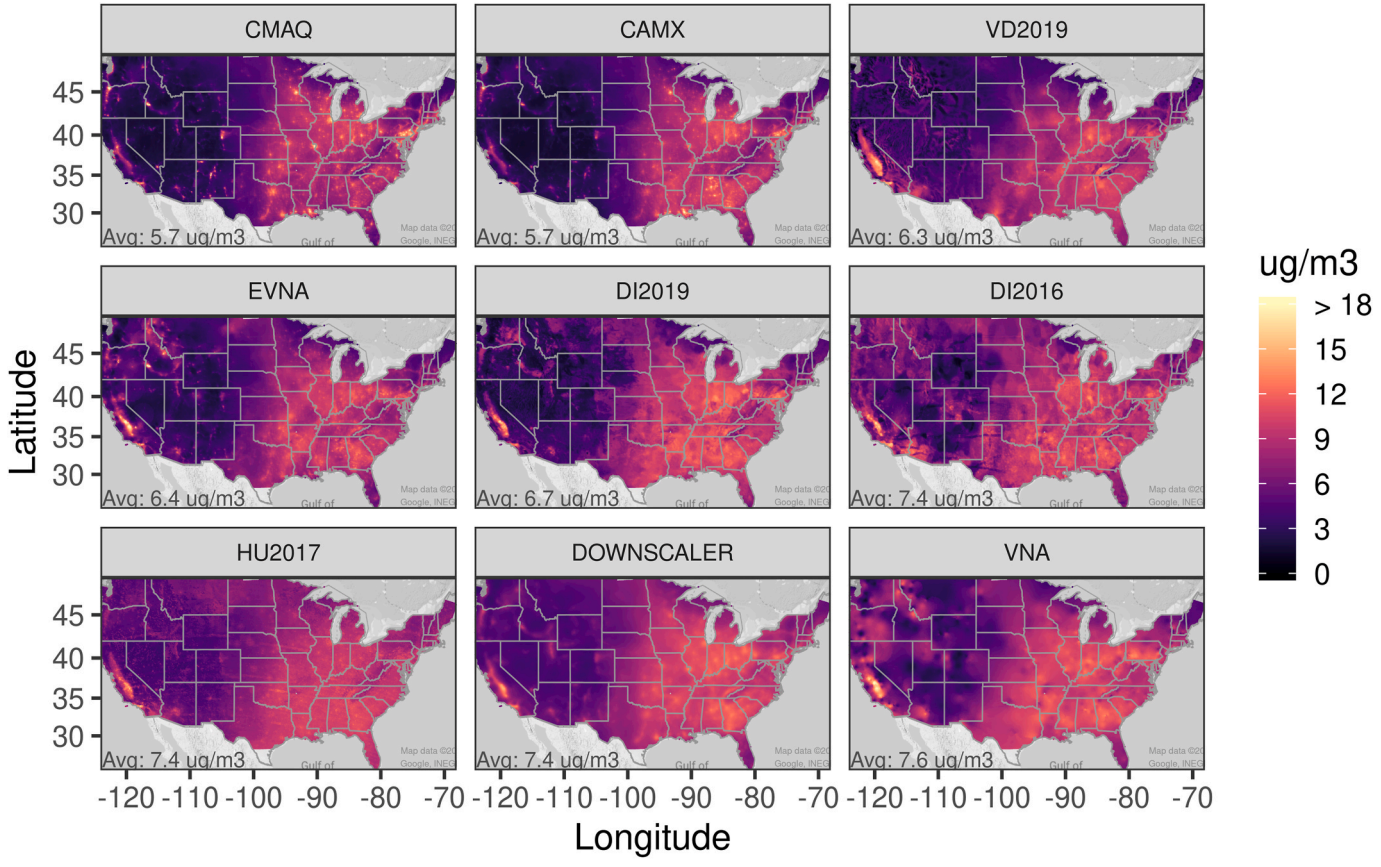

Fig. 1. Annual average $PM_{2.5}$ concentrations (μg m^{-3}) in 2011 based on nine exposure models averaged to a common 12-km grid.

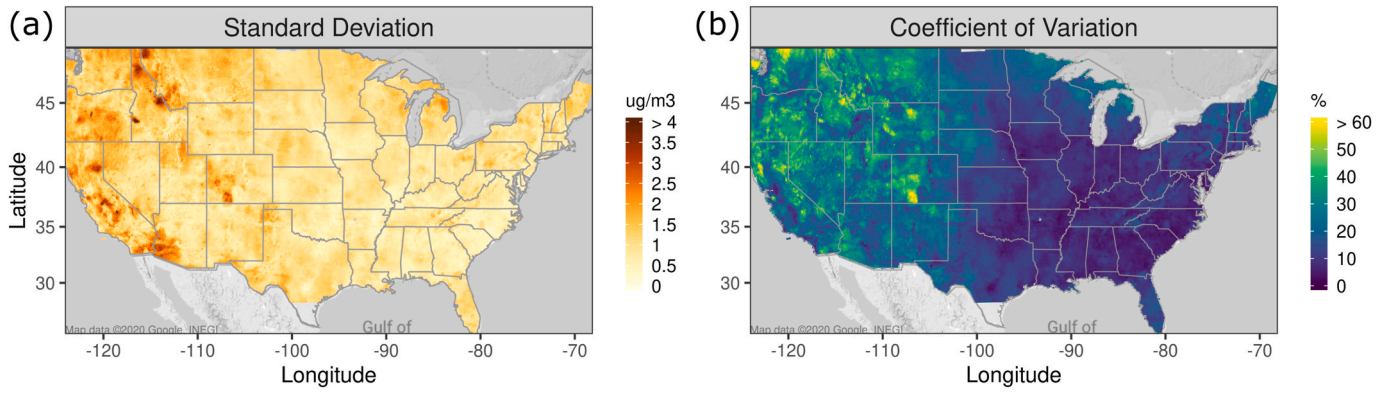

Fig. 2. Standard deviation and coefficient of variation for $PM_{2.5}$ concentrations from the exposure models in Fig. 1, excluding the CTMs (CMAQ and CAMx).

the proportion of the $PM_{2.5}$ distribution that is explained by less predictable stochastic variation as concentrations decrease (Just et al., 2020). The implications for specific applications of differences in performance according to different metrics is an area of current research and relates to what types of performance are most relevant for the research question of interest (e.g., daily variation or absolute pollutant levels).

In Fig. 3a, $PM_{2.5}$ concentrations over Los Angeles in 2011 are shown for the exposure fields averaged to a common 12-km grid. The spatial distributions in $PM_{2.5}$ concentrations vary across methods, but realistic features are generally evident in the fields. These features include relatively high $PM_{2.5}$ concentrations near downtown LA, the Long Beach ports, and to the east of downtown LA (i.e., downwind of LA on the seabreeze). Also, the methods generally capture the relatively low $PM_{2.5}$ over the mountains to the north of LA, although this feature is muted in the relatively smooth VNA and Downscaler fields. Since VNA predictions are based solely on monitor data, the limited basin-to-mountain gradient in $PM_{2.5}$ concentrations may be attributed to limited monitoring in the elevated terrain. The higher $PM_{2.5}$ concentrations for CMAQ than CAMx could be related to differences in the emission inventories and potentially less atmospheric mixing in CMAQ (updates were made to enhance mixing in CMAQ version 5.1 (USEPA, 2015b) compared with version 5.0.2 used here).

In Fig. 3b, $PM_{2.5}$ concentrations over Los Angeles are shown for the three 1-km fields. Terrain features and fine-scale spatial variability in $PM_{2.5}$ over the area are better represented in the 1-km than 12-km fields. However, differences in fine-scale spatial features exist among the three 1-km exposure fields, and some of the features are counterintuitive. For instance, $PM_{2.5}$ concentrations are slightly lower over the major roadways than over neighboring areas in the DI2019 method (the opposite pattern exists in some other urban areas such as Denver, Figure S5), and roadway features are indistinct in the DI2016 and VD2019 fields. Variability in the roadway features across cities in the DI2019 method may be due to differences in the relative values of the predictor variables (e.g., elevation) or influence of base learners across cities. Differences in roadway features between DI2016 and DI2019 could be related to the use of random forest and gradient boosting (in addition to neural network) in DI2019 but neural network alone in DI2016, since both models were trained using road density variables. Traffic influence was considered indirectly in the VD2019 method (via AOD, GEOS-Chem, and monitoring), which may not fully resolve the fine-scale features associated with roadways. Considering that the models were optimized over large domains (i.e., conterminous U.S. for DI2016 and DI2019, and North America for VD2019) using different data and algorithms, differences in fine-scale features among the models are not surprising. For city-specific applications, recent studies have demonstrated the promise of urban-scale models that incorporate non-regulatory data (Huang et al., 2019) and low-cost monitors (Bi et al., 2020; Eilenberg, 2020; Li et al., 2020).

In Fig. 3c, the difference in $PM_{2.5}$ concentrations between the 12-km and 1-km grids is shown. The largest differences occur for grid cells that span a wide range of terrain. For instance, a band of cells to the north of Pasadena had lower concentrations at 12-km than 1-km resolution just south of the mountains, and higher concentrations over the mountains. Similarly, relatively large differences between 12-km and 1-km fields are evident for cells to the east of LA that include a relatively wide range of terrain elevation (e.g., see VD2019 field). Spatial maps for additional urban areas are provided in Figures S5-S10 to further illustrate features of the exposure fields.

3.2. $\Delta PM_{2.5}$ concentrations: 2028–2011

The modeled change in $PM_{2.5}$ concentrations between the simulations for 2028 and 2011 is shown in Fig. 4. Nationally, the average $\Delta PM_{2.5}$ ranged from −0.85 to −1.02 μg m^{-3} across the models. Regional air quality improvements were greater than the national average for the Ohio Valley region (−2.1 to −2.4 μg m^{-3}) and lower for the Southwest region (−0.1 to −0.2 μg m^{-3}). These patterns are consistent with the locations of the modeled emission reductions (e.g., large SO_2 emission reductions from electric generation units were projected for the Ohio Valley region).

Since the same RRFs (i.e., from CAMx) were used to project all exposure fields, the differences in $\Delta PM_{2.5}$ among models are associated with differences in the 2011 $PM_{2.5}$ concentrations alone (i.e., $PM2.5_{Mod,2011}$ in Eqn. (3)). The spatial patterns of $\Delta PM_{2.5}$ in Fig. 4 therefore resemble the concentration fields in Fig. 1 (e.g., Downscaler predicts relatively large, smooth values in the Ohio Valley and CMAQ predicts relatively narrow peaks in urban areas). The spatial distribution of the maximum difference in $\Delta PM_{2.5}$ across the non-CTM models is illustrated in Figure S11. Developing RRFs from multiple CTM simulations would introduce additional variability into results, although such modeling is not available.

In Fig. 5, distributions of $\Delta PM_{2.5}$ are shown as a function of different strata of the 2011 $PM_{2.5}$ concentrations for the different models. The air quality improvements from 2011 to 2028 generally increase with increasing 2011 concentration. This behavior reflects the influence of emission reductions from sources that contributed to the relatively high $PM_{2.5}$ concentrations in 2011. There was general agreement in $\Delta PM_{2.5}$ estimates among models as a function of 2011 $PM_{2.5}$ concentration, especially for the non-CTM models. For instance, the median $\Delta PM_{2.5}$ ranged from −1.77 to −2.01 μg m^{-3} across models for 2011 $PM_{2.5}$ concentrations between 9 and 11 μg m^{-3}. The range in the median $\Delta PM_{2.5}$ across methods is slightly greater for 2011 concentrations between 7 and 9 μg m^{-3} (−0.95 to −1.54 μg m^{-3}) and 11 and 20 μg m^{-3} (−1.92 to −2.67 μg m^{-3}). Overall, the impacts of the emission changes on $PM_{2.5}$ concentrations are broadly consistent among methods, although the CTMs tended to have slightly greater impacts, which may be due to the larger uncertainty in the CTM simulations than the observationally constrained

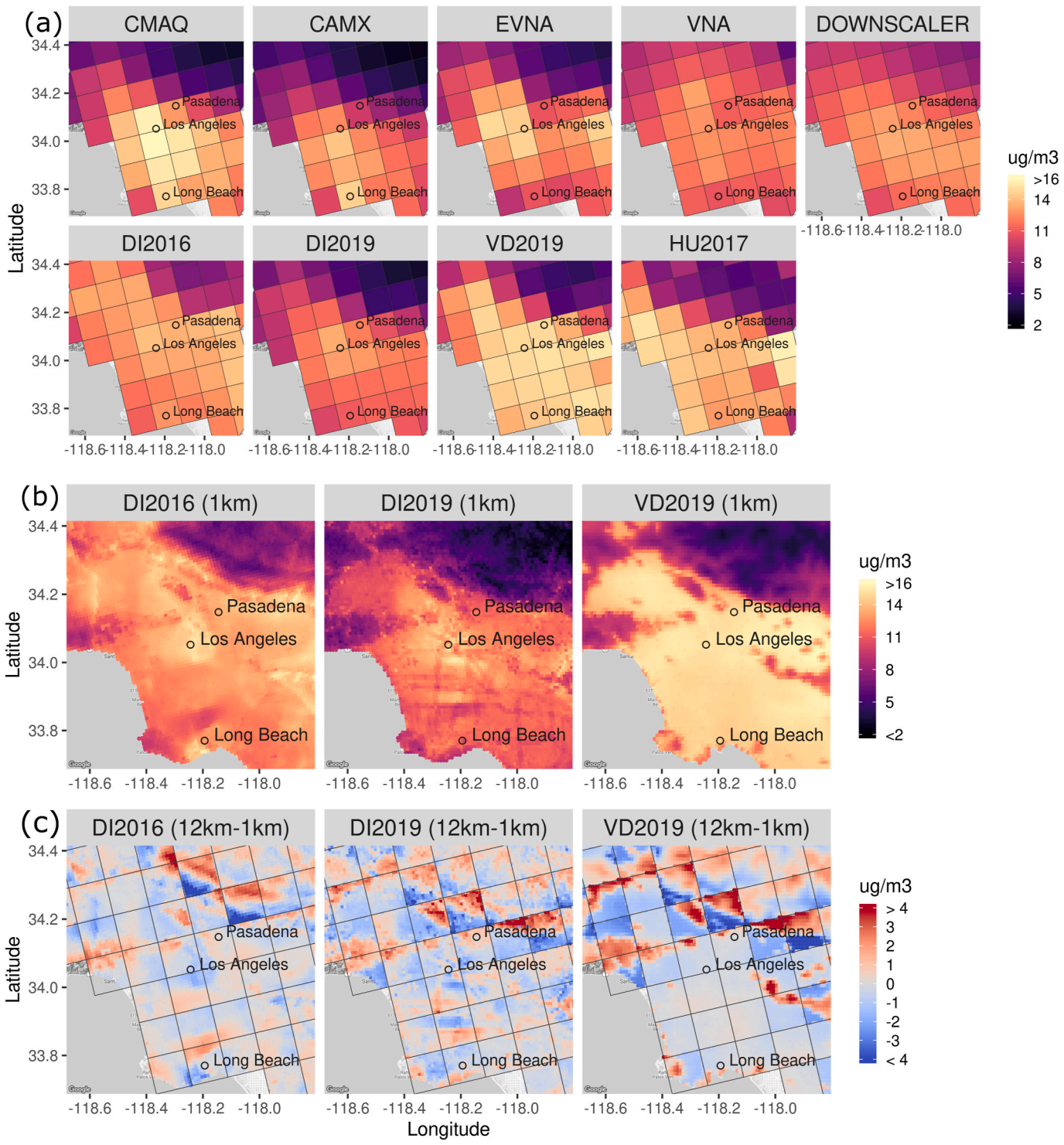

Fig. 3. Annual average $PM_{2.5}$ concentrations over Los Angeles in 2011 based on (a) 12-km and (b) 1-km gridded exposure fields, and (c) difference between 12-km and 1-km fields (12 km–1 km).

methods. Note also that the agreement in $\Delta PM_{2.5}$ among models is due in part to the use of the same RRFs for projecting the 2011 $PM_{2.5}$ fields. Uncertainty in the RRFs, which was not examined here, likely contributes more to the uncertainty in $\Delta PM_{2.5}$ than does uncertainty in 2011 $PM_{2.5}$ field.

3.3. Population exposure

In this section, we combine the modeled $PM_{2.5}$ concentrations with population data to examine $PM_{2.5}$ exposure levels in an approach often called population-weighting. In Fig. 6a, the percentage of U.S. grid cells and population exposed to 2011 $PM_{2.5}$ levels are shown for the exposure models. For instance, a negligible percentage of grid cells have concentrations less than about 2 μg m^{-3} across all models, and almost all grid cells have concentrations less than about 12.5 μg m^{-3} (Fig. 6a, left panel). A roughly 2.8 μg m^{-3} spread exists among models (1.7 μg m^{-3} excluding the CTMs) in the concentration predicted at the 50% exposure level across U.S. grid cells. Agreement is closer for population-weighted

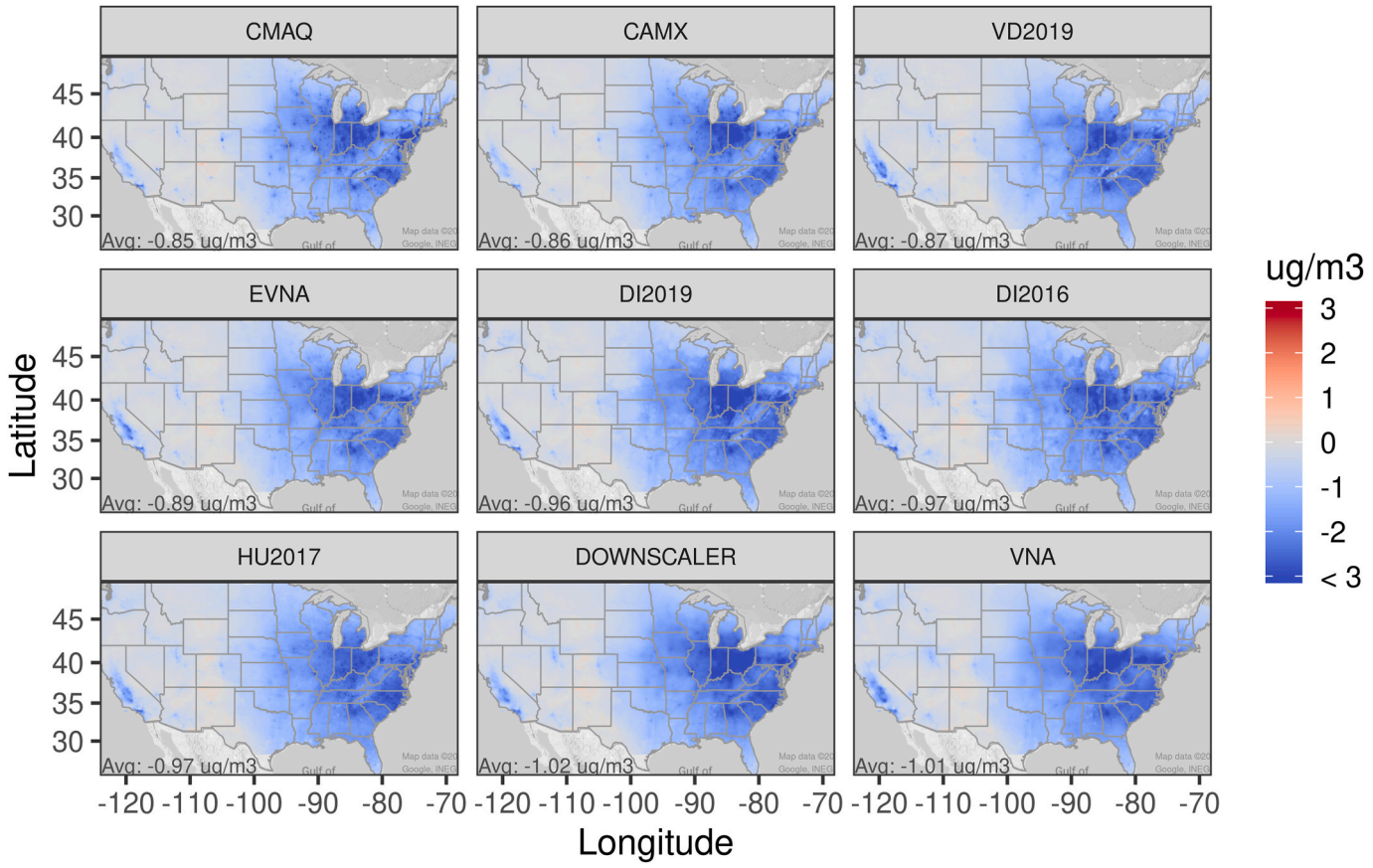

Fig. 4. Change in annual average $PM_{2.5}$ concentrations (2028–2011; μg m^{-3}) based on nine exposure models averaged to a common 12-km grid.

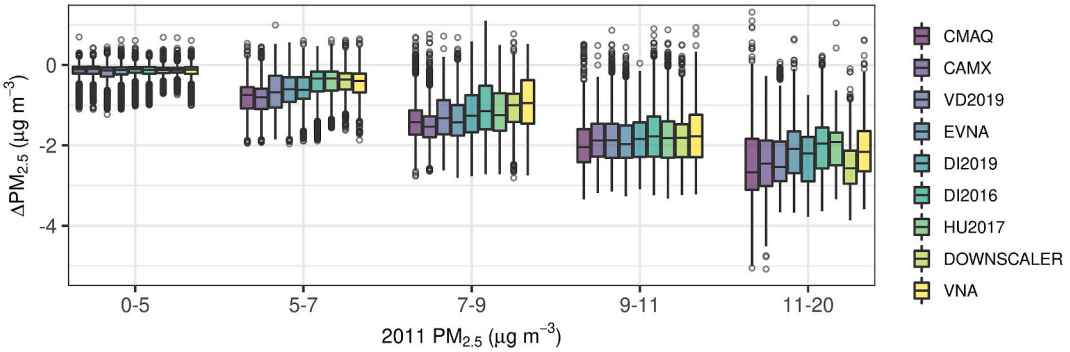

Fig. 5. Box and whisker plot depicting distributions of modeled $\Delta PM_{2.5}$ (2028–2011) as a function of 2011 $PM_{2.5}$ concentration. The boxes bracket the interquartile range (IQR; 25th to 75th percentiles), the horizontal line within the box represents the median, whiskers represent 1.5 times the IQR from either end of the box, and circles represent individual values less than and greater than the range of the whiskers, respectively.

averages (Fig. 6a, right panel) than uniformly weighted averages across the domain. A roughly 1 μg m^{-3} spread exists among models (0.8 μg m^{-3} excluding the CTMs) in the concentration predicted at the 50% exposure level for the U.S. population. The closer agreement in terms of population weighting than for equal weighting of all grid cells is because relatively large differences among models occur in sparsely populated parts of the western U.S. (e.g., Fig. 2a).

Cumulative exposure curves for the U.S. population to $\Delta PM_{2.5}$ concentrations (2028–2011) are shown in Fig. 6b. Good agreement exists among model estimates for population-weighted $\Delta PM_{2.5}$ concentrations. For instance, a 0.2 μg m^{-3} spread exists among models in the $\Delta PM_{2.5}$ concentration predicted at the 50% exposure level. The agreement among models is due in part to the use of the same RRFs in projecting all exposure fields from 2011 to 2028. Overall, the results in Fig. 6 indicate that agreement among models is closer for population exposure than for average domain levels, and the change in population exposure is relatively insensitive to differences in the baseline exposure field. However, CTM results deviate by a relatively large amount from the other models and demonstrate the value of calibrating CTMs with monitoring data in exposure applications.

Modeled 2011 $PM_{2.5}$ exposure concentrations are shown for the total population and the four racial/ethnic groups in Fig. 7a. The spread among the non-CTM models in the average population-weighted concentration ranges from to 0.5–0.9 μg m^{-3} for the five population categories. The relative rank in exposure concentration for the racial/ethnic groups is similar among models, with all models estimating the lowest exposure concentrations for the NH-white population and highest for the NH-Black population. This result is consistent with previous studies (Bell

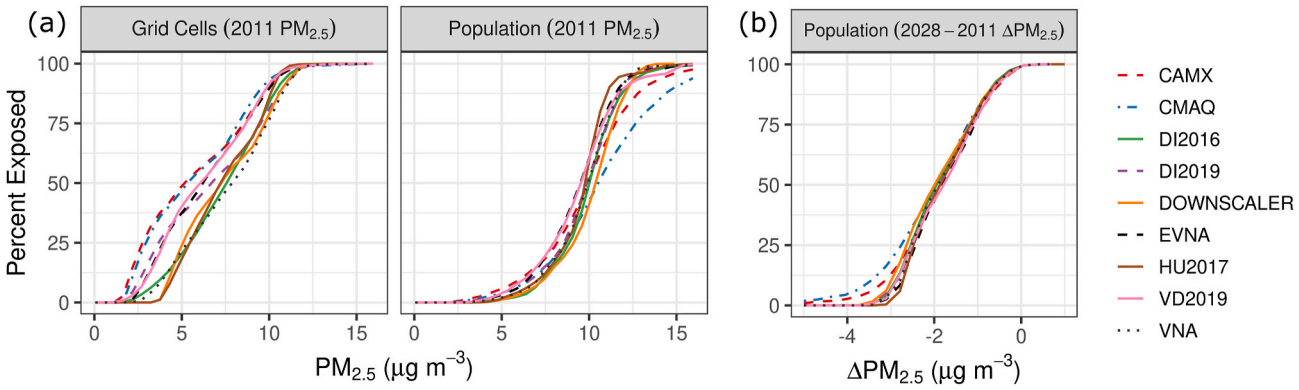

Fig. 6. The percentage of (a) U.S. grid cells and population exposed to $PM_{2.5}$ concentrations less than the values depicted by the curves and (b) U.S. population exposed to $\Delta PM_{2.5}$ concentrations less than the values depicted. Population data is based on the 2010 Census.

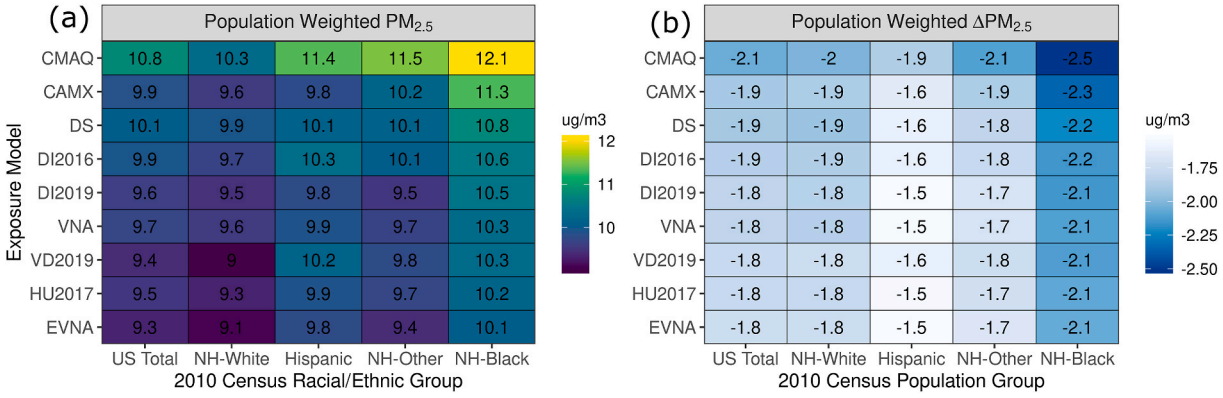

Fig. 7. Population-weighted average (a) 2011 $PM_{2.5}$ and (b) $\Delta PM_{2.5}$ (2028–2011) concentrations for the total population and four racial/ethnic groups based on 2010 Census data. DS: Downscaler.

and Ebisu, 2012; Miranda et al., 2011). Population-weighted average concentrations for the 1-km fields (i.e., DI2016, DI2019, and VD2019) are generally similar to those for the respective 12-km fields, but with slightly higher values for the 1-km field in the DI2019 (up to 0.3 μg m^{-3}) and VD2019 (up to 0.5 μg m^{-3}) cases (Figure S12).

Population-weighted average $\Delta PM_{2.5}$ concentrations (2028–2011) are shown for the total population and four racial/ethnic groups in Fig. 7b. The values are similar among models for the five population groups. The NH-Black population, which experienced the highest exposure in 2011, is projected to experience the greatest reduction in exposure among racial/ethnic groups by 2028. The smallest air quality improvement is projected for the Hispanic group, possibly due to the relatively large percentage of the Hispanic population in the western U. S. and parts of the domain that experienced fewer $PM_{2.5}$ reductions (c.f., Figure S13 and Fig. 4). On a percentage basis, the reductions in population-weighted $PM_{2.5}$ between 2011 and 2028 were similar for the NH-white (−19%) and NH-Black (−20%) populations and slightly lower for the Hispanic population (−15%).

In Fig. 8a, the difference in population-weighted $PM_{2.5}$ concentrations between the highest and lowest exposed racial/ethnic group (i.e., the exposure gap) is shown by state, with labels for the highest exposed group when differences exceed 0.5 μg m^{-3} (this value was selected considering the sd among model predictions, discussed below). Results in Fig. 8a are based on the mean concentration across the non-CTM models. The exposure gap in 2011 is less than 1 μg m^{-3} in most states but approaches 2 μg m^{-3} in Arizona and South Dakota. The relatively large gaps in these states appear to be driven by the distribution of the NH-Black population toward the the cities (e.g., Soux Falls, SD and Pheonix, AZ), where $PM_{2.5}$ concentrations are relatively high, compared with the relatively broad distribution of the NH-other population throughout the state, where $PM_{2.5}$ concentrations are lower on average (c.f., Fig. 1 and Figure S13). The distributions of the NH-other population in these states appear to be strongly influenced by those of the American Indian and Alaskan Native population (Figure S14). The exposure gap is projected to decrease from 2011 to 2028, with most states having values less than 0.5 μg m^{-3} and all states having values less than 1.5 μg m^{-3} in the 2028 case. This projected decrease continues a trend of decreasing absolute exposure disparity associated with roughly proportional reductions in $PM_{2.5}$ concentrations in recent decades (Colmer et al., 2020).

Although the general patterns in Fig. 8a are consistent with results for the individual models (Figure S15), some variability in the magnitude of the exposure gap and identity of the most-exposed group is evident. In Fig. 8b, the sd of the exposure gap across the non-CTM models is shown along with the number of models that identified the same most-exposed group as for the mean concentration field used in Fig. 8a. The sd for the fields is less than 0.5 μg m^{-3} in all states except Wisconsin, where a relatively large difference in concentration between the urban and northwestern areas led to a greater gap in the VD2019 field (c.f. Fig. 1 and S15). Also, there is general agreement in the most exposed group among models (Fig. 8b). This agreement builds confidence in the use of the models for large-scale exposure applications, although distinguishing small differences in exposure among groups in a given area may require development of an area-specific model that incoporates local information.

In Fig. 9, the AI_{BG} for 2011 and 2028 is shown by state based on the mean $PM_{2.5}$ concentrations across the seven non-CTM exposure models. The spatial pattern of AI_{BG} resembles that of the exposure gap metric in

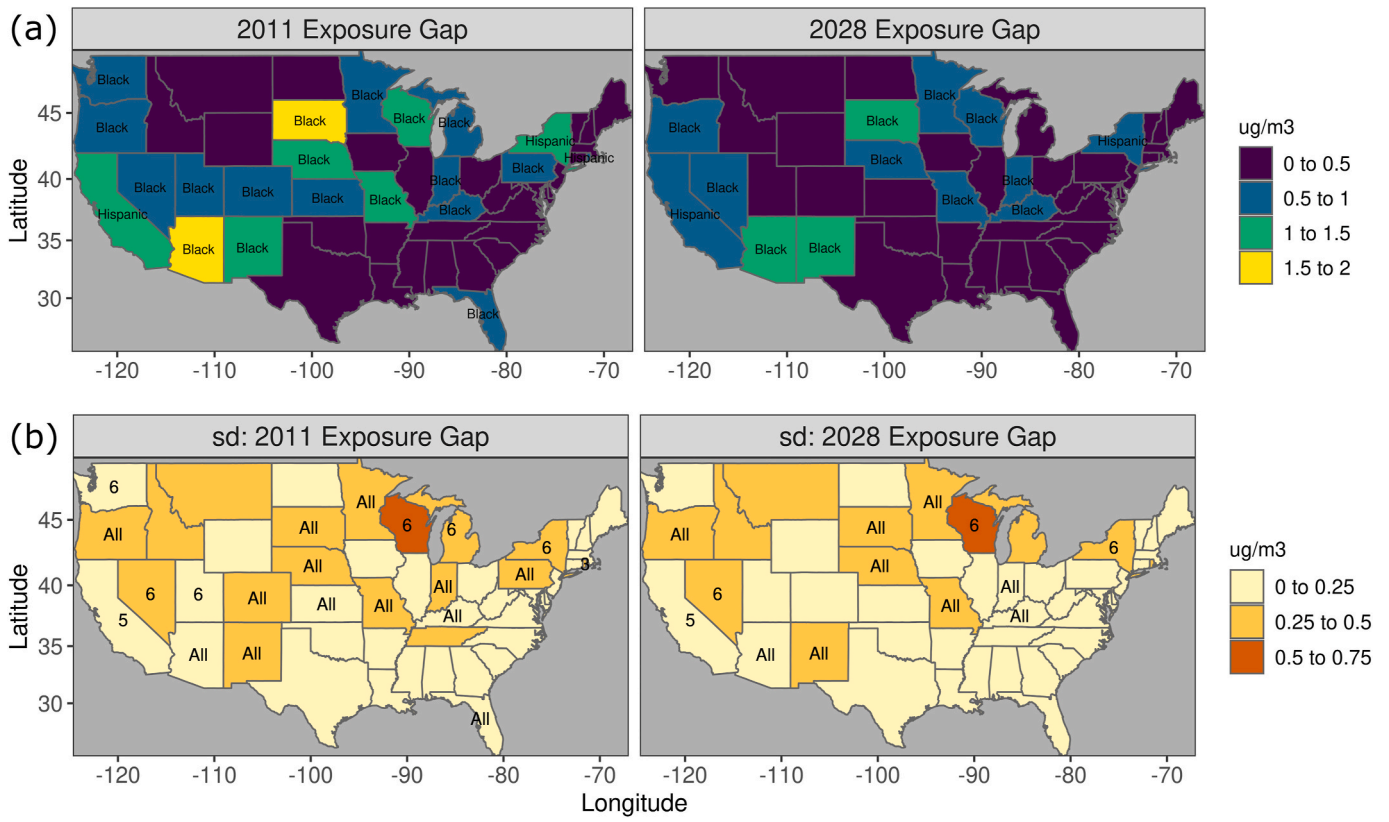

Fig. 8. (a) Difference in population-weighted $PM_{2.5}$ concentration based on the mean of the non-CTM models between the highest and lowest exposed racial/ethnic group (i.e., exposure gap) with labels for the highest exposed group when differences exceeded 0.5 μg m^{-3} and (b) the sd of the exposure gap calculated from the seven non-CTM models, with the number of models that identified the same most-exposed group as the median field.

Fig. 8a, with relatively high values in SD, AZ, NM, NY, and CA. The AI_{BG} also decreases from 2011 to 2028 due to reductions in the relatively high $PM_{2.5}$ concentrations (Fig. 5) that narrow the distribution of exposure across groups in 2028. However, the sd for the AI_{BG} appears to be relatively large in relation to the mean compared with the exposure gap metric. For example, the cv is 99% for the AI_{BG} in South Dakota in 2011, compared with 23% for the exposure gap. The relatively large cv for the AI_{BG} might be related to the use of the mean concentration as the reference (Eqn. (5)) compared with the best-off group for the exposure gap metric (i.e., deviations from the mean may be harder to distinguish compared with the full exposure gap). This information, as well as the general consistency in results for AI_{BG} and the exposure gap and other properties of the metrics (Harper et al., 2013; Levy et al., 2006), may be useful in guiding future study design.

4. Conclusions

At regional and national scales, the nine exposure models provided broadly consistent estimates of annual average 2011 $PM_{2.5}$ concentrations, with relatively high concentrations in the eastern U.S. and more variability in the west. The general agreement among the non-CTM models suggests that exposure studies based on the individual models may yield broadly similar conclusions in some cases. Yet differences in model performance for the methods are evident from previous reports, and some differences in $PM_{2.5}$ concentration predictions were found here due to the differences in model algorithms and input data. The median sd for the non-CTM models that had been averaged to a common 12-km grid was 1.00 μg m^{-3} nationally. The CTMs, which are not constrained by observations, predicted lower domain-average concentrations than the other models (i.e. 5.7 μg m^{-3} vs. 6.3–7.6 μg m^{-3}) but relatively high concentrations in urban areas in the eastern U.S.

All models estimated relatively high $PM_{2.5}$ concentrations over urban areas compared with surrounding areas, but differences in the spatial patterns within urban areas were evident. Fine-scale features associated with terrain and major roadways that were present in the 1-km fields were largely removed when concentrations were based on the 12-km grid. However, fine-scale features were inconsistent among the 1-km fields. Considering these differences and the similarity in results for the 12-km and 1-km fields in terms of national population-weighted concentrations and model-observation comparisons, the 12-km resolution may be justified for estimating exposure statistics over large domains. High-resolution exposure applications for individual cities may benefit in the future from the development of city-specific models that incorporate local information (e.g., low-cost sensors).

$PM_{2.5}$ concentrations were estimated to decrease by about 1 μg m^{-3} on average due to modeled emission changes between 2011 and 2028 associated with on-the-books regulations. Decreases of more than 3 μg m^{-3} were predicted in areas with relatively high 2011 $PM_{2.5}$ concentrations (e.g., the Ohio Valley region) due in part to SO_2 emission reductions from electric generation units. The projected pattern of greater $PM_{2.5}$ improvements for areas with relatively high 2011 $PM_{2.5}$ concentrations was generally consistent among the models.

Cumulative distributions of $PM_{2.5}$ concentrations over the U.S. illustrated the spread in predicted concentrations among models. For the non-CTM models, a roughly 1.7 μg m^{-3} spread exists in the 2011 $PM_{2.5}$ concentration estimated at the 50% exposure level for model grid cells, and a smaller spread (roughly 0.8 μg m^{-3}) exists in terms of population exposure. The closer agreement in terms of population-weighted averages than uniformly weighted averages is due to the relatively large differences among models in sparsely populated and monitored parts of the western U.S. About 50% of the population was estimated to experience $PM_{2.5}$ concentrations less than 10 μg m^{-3} in 2011 and $PM_{2.5}$ improvements of about 2 μg m^{-3} due to modeled emission changes between 2011 and 2028.

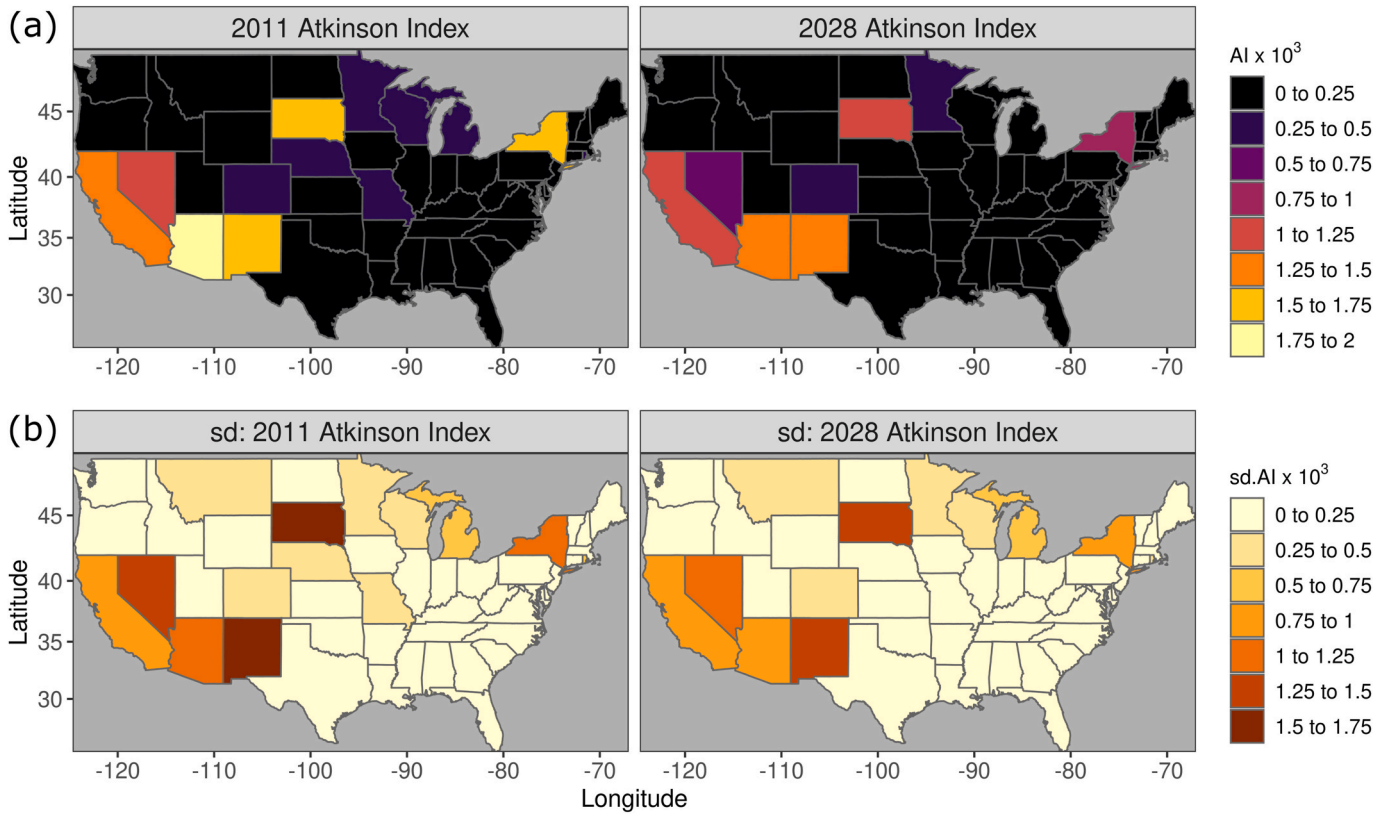

Fig. 9. (a) AI_{BG} for 2011 and 2028 based on the mean $PM_{2.5}$ concentrations from the non-CTM models and (b) the sd of the AI_{BG} calculated from the seven non-CTM models.

Differences in population-weighted concentrations among the non-CTM models ranged from 0.5 to 0.9 μg m^{-3} nationally for the total U. S. population and four racial/ethnic groups. The relative rank in the average exposure concentration for the racial/ethnic groups was similar among models. The NH-Black population was estimated by all models to have the highest $PM_{2.5}$ exposure in 2011 and experience the greatest reduction in $PM_{2.5}$ exposure between 2011 and 2028.

Eight states were predicted to have exposure gaps among racial/ethnic groups greater than 1 μg m^{-3} in 2011, but this decreased to three states in 2028 due to the modeled emission reductions. The AI_{BG} metric also suggested improvements in exposure inequality between 2011 and 2028. The projected decrease in absolute exposure inequality continues a trend over recent decades (Colmer et al., 2020). The sd in the AI_{BG} metric among the non-CTMs was relatively large compared with the sd in the exposure gap metric in relation to the respective values calculated using the mean concentration field. These differences could be due to the use of mean exposure as a reference in the AI_{BG} metric and best-off group in the exposure gap.

By examining predictions of multiple models, our study provided insights on current exposure modeling methods as well as a thorough characterization of $PM_{2.5}$ concentrations and exposure. However, the performance of the individual models and their suitability for operational use varies, and so a multi-model approach may be less effective than a single-model approach in a specific application, and the choice of method may influence results depending on the exposure features that are of interest and model performance. Improved understanding of how model skill in terms of various performance metrics relates to outcomes in specific applications would help inform future study design. Relating the $PM_{2.5}$ exposure concentrations described above to health risks would also be valuable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments and Disclaimer

The authors thank Dr. Yawen Guan for helpful conversations related to the manuscript and Dr. Adam Reff for running the downscaler model. The views expressed in this manuscript are those of the authors alone and do not necessarily reflect the views and policies of the U.S. Environmental Protection Agency. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2020.110432.

Author contributions

James T. Kelly, conceptualized the study, performed the analysis and wrote the original draft. Carey Jang, conceptualized the study. Brian Timin, conceptualized the study. Qian Di, developed the exposure models. Joel Schwartz, developed the exposure models. Yang Liu, developed the exposure models. Aaron van Donkelaar, developed the exposure models. Randall V. Martin, developed the exposure models. Veronica Berrocal, developed the exposure models. Michelle L. Bell, conceptualized the study. All authors provided interpretation and reviewed and edited the draft.

References

Abt, 2012. Model Attainment Test Software, User's Manual. https://www3.epa.gov/scram001/guidance/guide/MATS-2-5-1_manual.pdf.

Beckerman, B.S., et al., 2013. A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States. Environ. Sci. Technol. 47, 7233–7241.

Bell, M.L., Ebisu, K., 2012. Environmental inequality in exposures to airborne particulate matter components in the United States, 120, 1699–1704.

Berrocal, V.J., et al., 2010. A spatio-temporal downscaler for output from numerical models. J. Agric. Biol. Environ. Stat. 15, 176–197.

Berrocal, V.J., et al., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM2.5 concentration. Atmos. Environ. 222, 117130.

Bi, J., et al., 2020. Contribution of low-cost sensor measurements to the prediction of PM2.5 levels: a case study in Imperial County, California, USA. Environ. Res. 180, 108810.

Bravo, M.A., et al., 2016. Racial isolation and exposure to airborne particulate matter and ozone in understudied US populations: Environmental justice applications of downscaled numerical model output. Environ. Int. 92–93, 247–255.

Chen, Z.-Y., et al., 2019. Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. Atmos. Environ. 202, 180–189.

Center for International Earth Science Information Network - CIESIN - Columbia University, 2017. U.S. Census Grids (Summary File 1), 2010. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H40Z716C. Accessed 20 April 2020.

Cohan, D.S., Chen, R., 2014. Modeled and observed fine particulate matter reductions from state attainment demonstrations. J. Air Waste Manag. Assoc. 64, 995–1002.

Cohen, A.J., et al., 2017. 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 389, 1907–1918.

Colmer, J., et al., 2020. Disparities in PM2.5 air pollution in the United States. Science 369, 575.

Di, Q., et al., 2016. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50, 4712–4721.

Di, Q., et al., 2017. Air Pollution and Mortality in the Medicare Population 376, 2513–2522.

Di, Q., et al., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909.

Diao, M., et al., 2019. Methods, availability, and applications of PM2.5 exposure estimates derived from ground measurements, satellite, and atmospheric models. J. Air Waste Manag. Assoc. 69, 1391–1414.

Eilenberg, R.S., et al., 2020. Using a network of lower-cost monitors to identify the influence of modifiable factors driving spatial patterns in fine particulate matter concentrations in an urban environment. J. Expo. Sci. Environ. Epidemiol. https://doi.org/10.1038/s41370-020-0255-x.

Fann, N., et al., 2011. Maximizing health benefits and minimizing inequality: incorporating local-scale data in the design and evaluation of air. Quality Policies 31, 908–922.

Fann, N., et al., 2012. Estimating the national public health burden associated with exposure to ambient PM2.5. and Ozone 5, 81–95.

Fann, N., et al., 2018. The estimated change in the level and distribution of PM2.5-attributable health impacts in the United States: 2005–2014. Environ. Res. 167, 506–514.

Ford, B., Heald, C.L., 2016. Exploring the uncertainty associated with satellite-based estimates of premature mortality due to exposure to fine particulate matter. Atmos. Chem. Phys. 16, 3499–3523.

Frank, N.H., 2006. Retained nitrate, hydrated sulfates, and carbonaceous mass in federal reference method fine particulate matter for six eastern U.S. Cities. J. Air Waste Manag. Assoc. 56, 500–511.

Geng, G., et al., 2018. Satellite-based daily PM2.5 estimates during fire seasons in Colorado. Journal of Geophysical Research-Atmospheres 123, 8159–8171.

Hand, J.L., et al., 2012. Seasonal composition of remote and urban fine particulate matter in the United States, 117.

Harper, S., et al., 2013. Using inequality measures to incorporate environmental justice into regulatory analyses. Int. J. Environ. Res. Publ. Health 10, 4039–4059.

Hu, X., et al., 2017. Estimating PM2.5 concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 51, 6936–6944.

Huang, R., et al., 2018. Air pollutant exposure field modeling using air quality model-data fusion methods and comparison with satellite AOD-derived fields: application over North Carolina, USA. Air Quality, Atmosphere & Health 11, 11–22.

Huang, K., et al., 2019. Estimating daily PM2.5 concentrations in New York City at the neighborhood-scale: implications for integrating non-regulatory measurements. Sci. Total Environ. 697, 134094.

Jerrett, M., et al., 2017. Comparing the health effects of ambient particulate matter estimated using ground-based versus. Remote Sensing Exposure Estimates 125, 552–559.

Jin, X., et al., 2019. Comparison of multiple PM 2.5 exposure products for estimating health benefits of emission controls over New York State, USA. Environ. Res. Lett. 14, 084023.

Just, A.C., et al., 2020. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2.5) using satellite data over large regions. Atmos. Environ. 239, 117649.

Keller, J.P., Peng, R.D., 2019. Error in estimating area-level air pollution exposures for epidemiology, 30, e2573.

Kelly, J.T., et al., 2019a. A system for developing and projecting PM2.5 spatial fields to correspond to just meeting. In: National Ambient Air Quality Standards. Atmospheric Environment: X. https://doi.org/10.1016/j.aeaoa.2019.100019.

Kelly, J.T., et al., 2019b. Assessing PM2.5 model performance for the conterminous U.S. with comparison to model performance statistics from 2007-2015. Atmos. Environ. 214, 116872.

Lee, S.-J., et al., 2012. Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM2.5 concentrations across the continental United States. Environ. Health Perspect. 120, 1727–1732.

Levy, J.I., et al., 2006. Incorporating concepts of inequality and inequity into health benefits analysis. Int. J. Equity Health 5, 2.

Li, J., et al., 2020. Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5. Atmos. Environ. 223, 117293.

Miranda, M.L., et al., 2011. Making the environmental justice grade: the relative burden of air pollution exposure in the United States. Int. J. Environ. Res. Publ. Health 8, 1755–1771.

Noble, C.A., et al., 2001. Federal reference and equivalent methods for measuring fine particulate matter. Aerosol. Sci. Technol. 34, 457–464.

NRC, 2004. Air Quality Management in the United States. The National Academies Press, Washington, DC. https://doi.org/10.17226/10728.

Ostro, B., et al., 2015. Associations of mortality with long-term exposures to fine and ultrafine particles, species and sources: results from the California teachers study cohort. Environ. Health Perspect. 123, 549–556.

Pappin, A.J., et al., 2019. Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census health and environment cohort, 127, 107008.

Pope, C.A., et al., 2020. Fine particulate air pollution and human mortality: 25+ years of cohort studies. Environ. Res. 183, 108924.

Rosofsky, A., et al., 2018. Temporal trends in air pollution exposure inequality in Massachusetts. Environ. Res. 161, 76–86.

Thind, M.P.S., et al., 2019. Fine particulate air pollution from electricity generation in the US: health impacts by race, income, and geography. Environ. Sci. Technol. 53, 14010–14019.

USEPA, 2015a. Bayesian Space-Time Downscaling Fusion Model (Downscaler) -Derived Estimates of Air Quality for 2011. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC. https://www.epa.gov/hesc/rsig-related-downloadable-data-files.

USEPA, 2015b. Release Notes for CMAQv5.1 - November 2015. https://www.airqualitymodeling.org/index.php/CMAQ_version_5.1_(November_2015_release)_Technical_Documentation#Asymmectric_Convective_Model_version_2_.28ACM2.29.

USEPA, 2016. Technical Support Document (TSD) Preparation of Emissions Inventories for the Version 6.3, 2011 Emissions Modeling Platform. U.S. Environmental Protection Agency.

USEPA, 2017a. Documentation for the EPA's Preliminary 2028 Regional Haze Modeling. Office of Air Quality Planning and Standards United States Environmental Protection Agency. Available: https://www3.epa.gov/ttn/scram/reports/2028_Regional_Haze_Modeling-TSD.pdf.

USEPA, 2017b. Technical Support Document (TSD) Updates to Emissions Inventories for the Version 6.3, 2011 Emissions Modeling Platform for the Year 2028. https://www.epa.gov/sites/production/files/2017-11/documents/2011v6.3_2028_update_emismod_tsd_oct2017.pdf.

USEPA, 2019. Integrated Science Assessment (ISA) for Particulate Matter (Final Report, 2019). U.S. Environmental Protection Agency, Washington, DC.

USEPA, U.S. EPA, 2018. Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze. Office of Air Quality Planning and Standards, Research Triangle Park, NC. https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling_Guidance-2018.pdf.

van Donkelaar, A., et al., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based. Aerosol Optical Depth: Development and Application 118, 847–855.

van Donkelaar, A., et al., 2019. 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. 53, 2595–2611.

Wang, J., et al., 2017. Historical trends in PM2.5-related premature mortality during 1990–2010 across the northern hemisphere. Environ. Health Perspect. 125, 400–408.