Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study

Wenhua Yu, Tingting Ye, Yiwen Zhang, Rongbin Xu, Yadong Lei, Zhuying Chen, Zhengyu Yang, Yuxi Zhang, Jiangning Song, Xu Yue, Shanshan Li, Yuming Guo

Summary

Background Short-term exposure to ambient $PM_{2·5}$ is a leading contributor to the global burden of diseases and mortality. However, few studies have provided the global spatiotemporal variations of daily $PM_{2·5}$ concentrations over recent decades.

Methods In this modelling study, we implemented deep ensemble machine learning (DEML) to estimate global daily ambient $PM_{2·5}$ concentrations at 0·1°×0·1° spatial resolution between Jan 1, 2000, and Dec 31, 2019. In the DEML framework, ground-based $PM_{2·5}$ measurements from 5446 monitoring stations in 65 countries worldwide were combined with GEOS-Chem chemical transport model simulations of $PM_{2·5}$ concentration, meteorological data, and geographical features. At the global and regional levels, we investigated annual population-weighted $PM_{2·5}$ concentrations and annual population-weighted exposed days to $PM_{2·5}$ concentrations higher than 15 μg/m³ (2021 WHO daily limit) to assess spatiotemporal exposure in 2000, 2010, and 2019. Land area and population exposures to $PM_{2·5}$ above 5 μg/m³ (2021 WHO annual limit) were also assessed for the year 2019. $PM_{2·5}$ concentrations for each calendar month were averaged across the 20-year period to investigate global seasonal patterns.

Findings Our DEML model showed good performance in capturing the global variability in ground-measured daily $PM_{2·5}$, with a cross-validation R^2 of 0·91 and root mean square error of 7·86 μg/m³. Globally, across 175 countries, the mean annual population-weighted $PM_{2·5}$ concentration for the period 2000–19 was estimated at 32·8 μg/m³ (SD 0·6). During the two decades, population-weighted $PM_{2·5}$ concentration and annual population-weighted exposed days ($PM_{2·5}$ >15 μg/m³) decreased in Europe and northern America, whereas exposures increased in southern Asia, Australia and New Zealand, and Latin America and the Caribbean. In 2019, only 0·18% of the global land area and 0·001% of the global population had an annual exposure to $PM_{2·5}$ at concentrations lower than 5 μg/m³, with more than 70% of days having daily $PM_{2·5}$ concentrations higher than 15 μg/m³. Distinct seasonal patterns were indicated in many regions of the world.

Interpretation The high-resolution estimates of daily $PM_{2·5}$ provide the first global view of the unequal spatiotemporal distribution of $PM_{2·5}$ exposure for a recent 20-year period, which is of value for assessing short-term and long-term health effects of $PM_{2·5}$, especially for areas where monitoring station data are not available.

Funding Australian Research Council, Australian Medical Research Future Fund, and the Australian National Health and Medical Research Council.

Lancet Planet Health 2023; 7: e209–18

Climate, Air Quality Research Unit, School of Public Health and Preventive Medicine (W Yu MPH, T Ye MSc, Yi Zhang MSc, R Xu PhD, Z Yang MPH, Yu Zhang PhD, S Li PhD, Prof Y Guo PhD)**, Turner Institute for Brain and Mental Health, School of Psychological Sciences** (Z Chen PhD)**, and Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology** (J Song PhD)**, Monash University, Melbourne, VIC, Australia; State Key Laboratory of Severe Weather and Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing, China** (Y Lei PhD)**; Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China** (X Yue PhD)

Correspondence to:
Prof Yuming Guo, Climate, Air Quality Research Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia
yuming.guo@monash.edu
or
Dr Shanshan Li, Climate, Air Quality Research Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia
shanshan.li@monash.edu

For the **WHO global ambient air pollution database** see https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ambient-air-pollution

Introduction

Fine particulate matter ($PM_{2·5}$) is a leading risk factor for premature mortality and morbidity worldwide. According to a Global Burden of Disease Study, outdoor air pollution (including ambient particulate matter) was estimated to cause 6·67 million premature deaths in 2019.[1] In addition, no safe threshold for $PM_{2·5}$ has been identified below which no damage to health is observed.[2,3] An abundance of evidence has supported the adverse effects of short-term and long-term ambient $PM_{2·5}$ exposure on human health, even at low $PM_{2·5}$ concentrations.[4] Therefore, the latest version of the WHO global air quality guidelines published in 2021 has adjusted the recommended limit for outdoor $PM_{2·5}$ exposure from 10 μg/m³ to 5 μg/m³ for the annual mean exposure and from 25 μg/m³ to 15 μg/m³ for 24-h mean exposure.[5]

Estimates of global $PM_{2·5}$ concentrations are a prerequisite for health risk assessments of global air pollution. Despite global expansion of the ground-based $PM_{2·5}$ monitoring network (an example of which is the WHO ambient air pollution database), the frequently sparse spread and poor homogeneity in the distribution of monitoring stations make accurate measurement of global $PM_{2·5}$ exposure challenging. Fortunately, with improvement of statistical and machine learning methods, the combination of emerging satellite products, chemical transport model simulations, and ground monitor

Research in context

Evidence before this study

Assessing the disparities in global exposure to $PM_{2\cdot5}$ is crucial for population health risk assessment. However, little is known about the global short-term (daily) $PM_{2\cdot5}$ exposure and its spatiotemporal variations. We searched PubMed and Google Scholar for articles in English published between Jan 1, 2011, and Dec 31, 2021, with the terms “fine particulate matter”, “$PM_{2\cdot5}$”, “global estimate”, and “short-term daily exposure”. We found that most previous studies estimated daily average $PM_{2\cdot5}$ distribution at the city or national levels. Several studies have estimated global long-term trends in $PM_{2\cdot5}$ concentrations, but few articles have reported daily $PM_{2\cdot5}$ distribution worldwide. One article investigated the global daily $PM_{2\cdot5}$ distribution from 1997 to 2014; however, it did not explore the inequality and spatiotemporal changes in daily $PM_{2\cdot5}$ population exposure over decades.

Added value of this study

Our study provides a global perspective on the spatial and temporal distribution of daily $PM_{2\cdot5}$ concentrations in 2000–19 and assesses the inequalities in global population exposure on the basis of the new $PM_{2\cdot5}$ limits in the 2021 WHO air quality guidelines. We used a deep ensemble machine learning (DEML) approach to estimate the daily mean $PM_{2\cdot5}$ concentrations in each global grid cell (0·1° × 0·1° spatial resolution) for 20 years with use of ground-based $PM_{2\cdot5}$ data from 5446 monitoring stations in 65 countries worldwide, combined with GEOS-Chem chemical transport model simulations, meteorological data, and geographical information on a global scale. Based on modelled estimates, only 0·001% of the global population was exposed to $PM_{2\cdot5}$ at concentrations lower than the WHO annual limit (5 μg/m³), and more than 70% of days of the year globally had a $PM_{2\cdot5}$ concentration exceeding the WHO daily limit (15 μg/m³) in 2019.

Implications of all the available evidence

Our grid-based daily $PM_{2\cdot5}$ estimates could fill knowledge gaps with regard to the assessment of global $PM_{2\cdot5}$-attributable health burden and both short-term and long-term health effects, especially for areas where monitoring station data are not available. The study results offer a novel global perspective of daily spatiotemporal variations in exposure of the global population to ambient $PM_{2\cdot5}$. Additionally, the DEML approach achieved good performance in estimating global air pollution concentrations.

expansion offers novel opportunities to accurately assess $PM_{2\cdot5}$ concentrations globally.[6–8] Global $PM_{2\cdot5}$ estimation studies have typically focused on long-term (annual or monthly average) $PM_{2\cdot5}$ estimates,[7,8] whereas few studies have explored the short-term (from hours to days) exposure to $PM_{2\cdot5}$ concentrations at a global level.[9] Although increasing numbers of studies have estimated daily $PM_{2\cdot5}$ concentrations at national and regional levels, such as in China,[10] Europe,[11] and the USA,[12] few studies have assessed short-term $PM_{2\cdot5}$ exposure and its spatiotemporal variations at the global level. The absence of uniformity in global training data and inconsistency in estimation methods create difficulties in comparing the previous regional estimates and in providing a global view of the spatiotemporal distribution in $PM_{2\cdot5}$ exposure. Therefore, estimating global daily $PM_{2\cdot5}$ concentration and its spatiotemporal variations with unified study designs, modelling approaches, and data sources is warranted.

Machine learning algorithms, especially ensemble machine learning technologies, have proven to be promisingly efficient for geospatial air pollution prediction[13] and have been increasingly applied in the estimation of $PM_{2\cdot5}$ concentration, offering high accuracy and the ability to handle large numbers of features with nonlinear associations.[10] Ensemble learning algorithms use a collection of multiple machine learning algorithms to achieve an optimal combination of predictions.[14] Multiple studies have indicated that ensemble machine learning could provide better estimations in environmental exposure assessment than a single machine learning model alone.[15,16]

The deep ensemble machine learning (DEML) framework is a multilevel stacked ensemble machine learning algorithm that combines the advantages of several diverse base models and meta models to achieve an optimal prediction performance.[17] Our previous study indicated the advantages of DEML modelling in environmental exposure assessment.[17] In this study, we used a DEML model to assess the global daily $PM_{2\cdot5}$ distribution at high spatial resolution (0·1°×0·1°) from 2000 to 2019. We also investigated regional and national population-weighted $PM_{2\cdot5}$ concentrations and population-weighted exposed days at $PM_{2\cdot5}$ concentration higher than the 2021 WHO daily limit (15 μg/m³) to identify the spatiotemporal exposure distribution from 2000 to 2019.

Methods

Overview and data sources

In this modelling study, we utilised a DEML framework by integrating ground monitoring measurements of $PM_{2\cdot5}$, chemical transport model simulations of all-source $PM_{2\cdot5}$ concentrations, meteorological conditions, and geographical features. All data used in the modelling were for the period between Jan 1, 2000, and Dec 31, 2019.

Daily mean (24-h) concentrations of $PM_{2\cdot5}$ recorded by government air quality monitoring stations in 2000–19 were obtained from multiple sources, including the national environmental protection agencies of the USA and China, the European Environmental Agency, and Australian state and territory government agencies.[18] Details on the air quality station data sources are provided

For the **US Environmental Protection Agency air data** see https://aqs.epa.gov/aqsweb/airdata/download_files.html

For the **China National Environmental Monitoring Centre** see http://www.cnemc.cn/en/

For the **European Environmental Agency air quality database** see https://www.eea.europa.eu/data-and-maps/data/aqereporting-9

For the **Australian National Air Pollution Monitor Database** see https://osf.io/jxd98/

in the appendix (p 4). Figure 1 shows the spatial distribution of the included 5446 monitoring stations in 65 countries. These countries covered 73% of the global population in 2019 according to WorldPop data, and 56% of the global land area in 2019 according to land cover data from the MCD12Q1.061 product. A summary of the sampled monitoring stations by country is presented in the appendix (pp 6–7).

We extracted simulations of global daily all-source $PM_{2\cdot5}$ concentrations for the study period by using a three-dimensional (3D) GEOS-Chem chemical transport model (version 12.0.0) at a spatial resolution of 2·0°×2·5° as described by Yue and colleagues.[19] Further details on the data sources for the GEOS-Chem model are provided in the appendix (p 1).

Satellite-based meteorological data with a spatial resolution of 0·1°×0·1° were collected from the ERA5 dataset (the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis set)[20] through the Google Earth Engine platform. For the study period, we collected ambient temperature, ambient dew point temperature (at 2 m above the land surface), wind speed at 10 m height above sea level, surface pressure, surface solar radiation, total precipitation, and total evaporation. All were collected as hourly measurements for the period under study. Daily mean data were calculated as an average of 24 h of observations, starting each day from 12 noon. Daily relative humidity was calculated with an algorithm in the humidity package[21] in R (version 0.1.5) on the basis of daily mean ambient temperature and dew point temperature.

Satellite-based annual land cover data with a spatial resolution of 500 m were obtained from the MCD12Q1.061 product through the Google Earth Engine platform. Satellite-based annual digital elevation data with a spatial resolution of 90 m were extracted from the Shuttle Radar Topography Mission project. Global annual population counts at 1 km spatial resolution were obtained from the WorldPop platform.

All data variables and sources are summarised in the appendix (pp 4–5). Bilinear interpolation was used to upscale the GEOS-Chem simulations, land cover, ground elevation, and population density to 0·1°×0·1° spatial resolution for further grid cell estimation.[22] Missing values in the collected datasets were excluded in the analysis to avoid the potential bias introduced by imputation. For ground measurements of $PM_{2\cdot5}$, we removed extreme outliers higher than the 99·9th quantiles of the total dataset, and removed unsuitable data obtained from six stations in central eastern Asia (Kazakhstan, n=1; Uzbekistan, n=1; Tajikistan, n=2; and Afghanistan, n=2) due to data being constantly or extremely high.

Statistical analysis

We estimated global daily $PM_{2\cdot5}$ concentration using the DEML framework,[17] which has a three-stage structure. At the first stage, three base models (random forest, light gradient boosting machine [LightGBM], and extreme gradient boosting [XGboost])[23] were used to generate new $PM_{2\cdot5}$ concentration data. At the second stage, two meta-models (random forest and generalised linear model)[23] were used to estimate daily $PM_{2\cdot5}$ concentrations with the predictions from the first-stage models. At the third stage, a non-negative least squares algorithm was used to obtain optimal weights for $PM_{2\cdot5}$ estimations. Details of the DEML framework are presented in the appendix (pp 1–2).

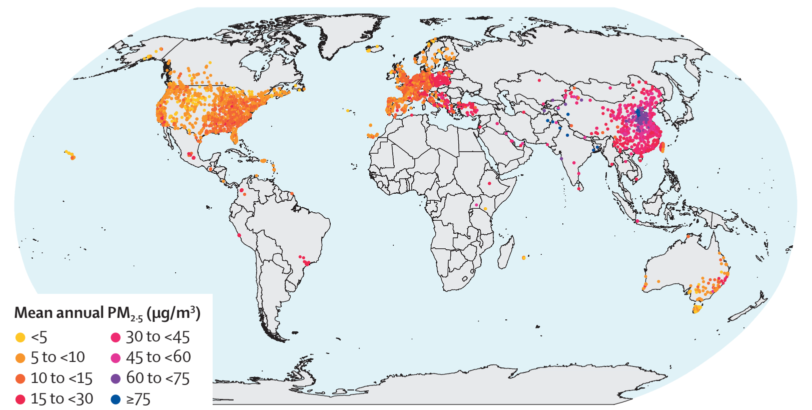

Figure 1: **Global monitor station distribution and mean annual $PM_{2\cdot5}$ concentration over two decades (2000–19)**

Spatiotemporal generalisation and reliability of the DEML framework were evaluated by a series of cross-validations, including a general 10-fold cross-validation, spatial cross-validations (based on monitor station sampling and grid cell sampling), a continent-stratified cluster cross-validation, and a temporal cross-validation by year. Details on the cross-validation strategies are provided in the appendix (p 3). Root mean square error (RMSE), mean absolute error, and coefficients of determination (R^2) were used to assess the prediction performance of the DEML model. Benchmark machine learning models (random forest, LightGBM, and XGboost) were used to compare the performance of DEML.

For observed and estimated daily values, we present both mean (SD), median (IQR), and range. We calculated the monthly and annual mean values (from daily mean values in each calendar month or year) for observed and predicted $PM_{2\cdot5}$ concentration and used density scatter plots to assess DEML performance for the daily, monthly, and annual mean $PM_{2\cdot5}$ estimations in the study period. Spearman's correlation analysis was used to compare consistency between the DEML-estimated and observed daily mean $PM_{2\cdot5}$ concentrations at the global and regional levels. We also compared the variability in temporal trends of observed and predicted daily mean $PM_{2\cdot5}$ in eight metropolises (Beijing, China; Deli, India; Ho Chi Minh, Viet Nam; Milan, Italy; New York, NY, USA; Sao Paolo, Brazil; Sydney, NSW, Australia; and Toronto, ON, Canada), which were selected as being regionally representative and for comparison with

See **Online** for appendix

For **WorldPop** see https://hub.worldpop.org/

For the **MCD12Q1.061 product** see https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD12Q1#description

For the **GEOS-Chem chemical transport model** see https://geos-chem.seas.harvard.edu/

For the **ERA5 dataset** see https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_HOURLY

For the **Shuttle Radar Topography Mission project** see https://bigdata.cgiar.org/srtm-90m-digital-elevation-database/

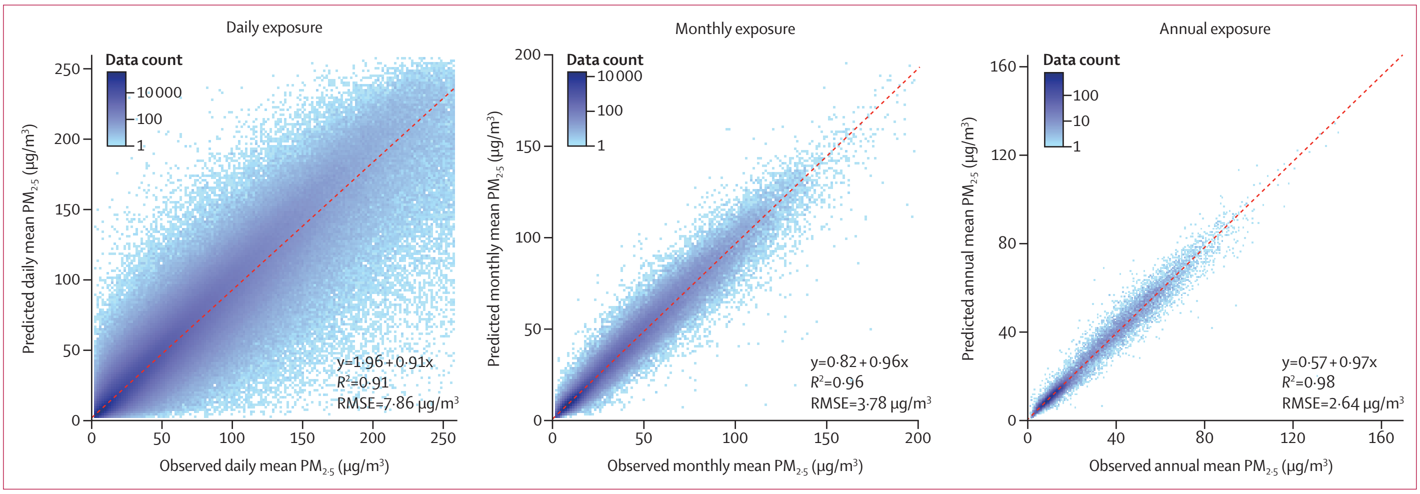

Figure 2: **Comparison between observed and model-estimated $PM_{2·5}$ concentrations in 2000–19**
The x-axis indicates observed mean $PM_{2·5}$ in the monitor stations; the y-axis indicates estimated mean $PM_{2·5}$ by the deep ensemble machine learning model in 10-fold cross-validation analysis; the points represent the corresponding $PM_{2·5}$ for both observed and predicted values. There are 9 289 613 datapoints for daily $PM_{2·5}$, 434 122 for monthly $PM_{2·5}$, and 38 488 for annual $PM_{2·5}$; the red line represents a regression line (simple linear regression) for the observed and predicted $PM_{2·5}$. R^2=coefficient of determination for the unseen independent data in 10-fold cross-validation analysis. RMSE=root mean square error.

previous studies. Overall averages for the 20-year study period were also calculated.

For the **country and regional classifications of the UN Statistics Division** see https://unstats.un.org/unsd/methodology/m49/

We assessed the global spatial distribution of $PM_{2·5}$ and the spatiotemporal changes per decade. We presented the global distribution of annual mean $PM_{2·5}$ in 2000, 2010, and 2019, and measured the changes in annual mean $PM_{2·5}$ per decade by multiplying 10 years with the coefficients of a linear regression model established from 20 annual mean $PM_{2·5}$ values for each global grid cell (0·1°×0·1° spatial resolution). Additionally, the proportion of global and regional land area exposure and population exposure at the previous 2005 WHO limit (10 μg/m³) and updated 2021 WHO limit (5 μg/m³) for annual mean $PM_{2·5}$ were assessed for the year 2019.

For **The World Bank income classification** see https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Population-weighted $PM_{2·5}$ and population-weighted exposed days, when daily $PM_{2·5}$ concentration was higher than the 2021 WHO-recommended limit (15 μg/m³), were identified to indicate both the levels of air pollution and the size of the affected population in each region. Regional mean population-weighted $PM_{2·5}$ concentrations were identified as:

$$\text{Population}-\text{weighted concentration} = \sum\left(\frac{pi}{P}\times C_i\right)$$

Where C_i denotes the daily mean $PM_{2·5}$ concentration and pi the annual mean population in a specific grid cell, i; and, $P=\sum pi$, which is the total population of grid cells in a specified region. Regional population-weighted exposed days ($PM_{2·5}$ >15 μg/m³) were identified as:

$$\text{Population}-\text{weighted exposed days} = \sum_{j=1}^{365}\sum\left(\frac{pi}{P}\times D_{ij}\right)$$

Where D_{ij} is a Boolean value, indicating whether daily mean $PM_{2·5}$ in a specific j day of a year in a grid cell i was higher than 15 μg/m³ (D_{ij}=1) or not (D_{ij}=0). This indicator can be used to quantify the average exposed days to concentrations higher than the WHO guideline limit for the population in a specified region. Study regions were based on the country and regional groupings of the UN Statistics Division, and we provide estimates for 175 countries. The annual mean population-weighted $PM_{2·5}$ concentrations and annual accumulation of population-weighted exposed days at the global and regional levels were used to assess global and regional $PM_{2·5}$ exposure in 2000, 2010, and 2019. Countries with the highest exposures were ranked and labelled by income group (2019 World Bank classification). To investigate seasonal patterns, $PM_{2·5}$ concentrations for each calendar month were averaged across the 20-year period for each global grid cell. We also assessed mean population-weighted $PM_{2·5}$ values for each calendar month over the 20 years at the global population level and in eight specific countries (Australia, Brazil, China, India, South Africa, Thailand, the UK, and the USA), which were selected as regionally representative countries and for comparison with previous studies.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Globally, across 175 countries, the mean annual population-weighted $PM_{2·5}$ concentration for the period 2000–19 was estimated at 32·8 μg/m³ (SD 0·6). When estimating mean annual exposure across regions, the highest population-weighted $PM_{2·5}$ concentrations were distributed in eastern Asia (50·0 μg/m³ [SD 2·2]) and southern Asia (37·2 μg/m³ [1·2]), followed by northern

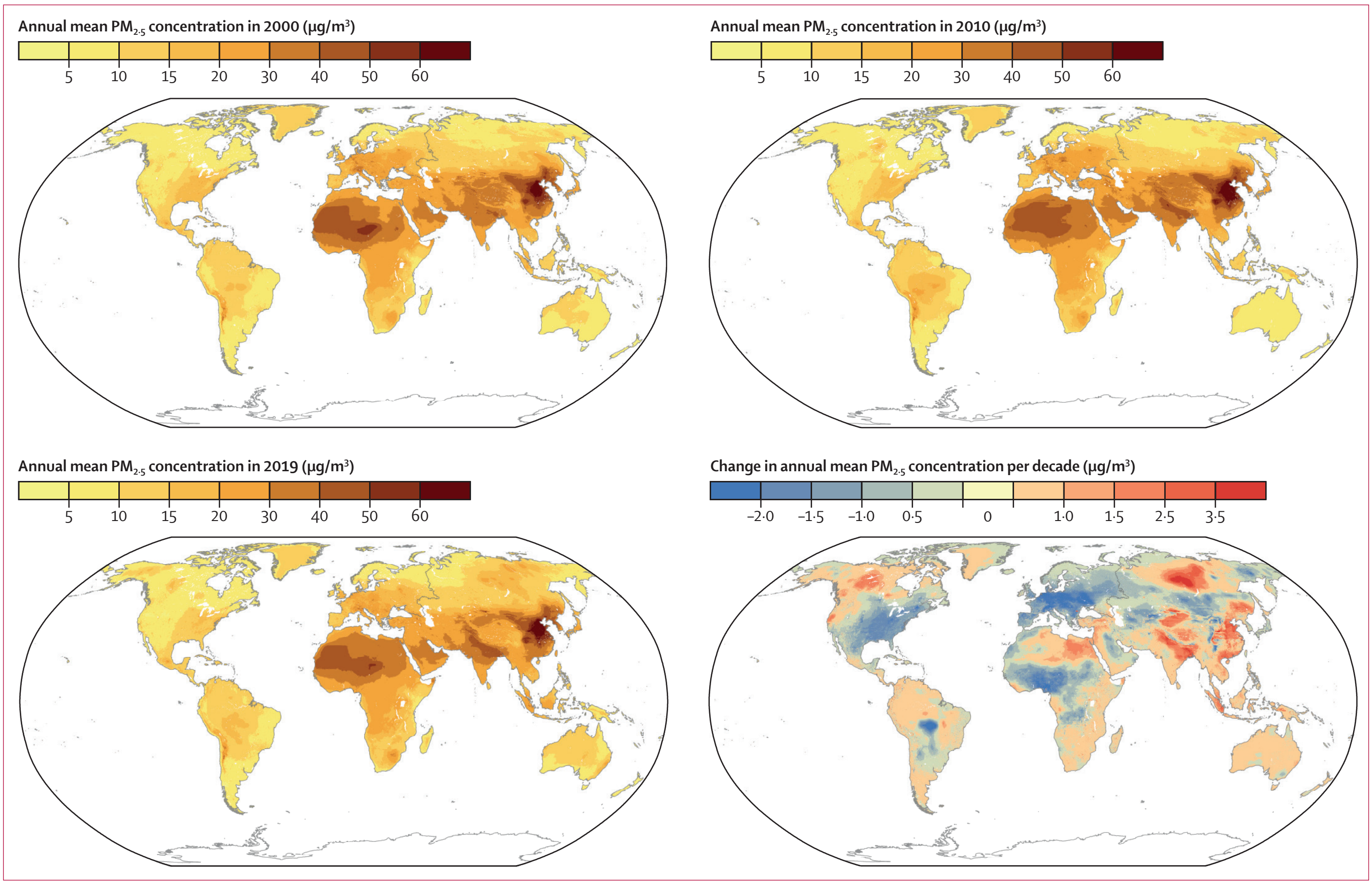

***Figure 3:* Annual mean $PM_{2·5}$ in 2000, 2010, and 2019 and changes in annual mean $PM_{2·5}$ per decade at 0·1° × 0·1° spatial resolution**
The change in annual $PM_{2·5}$ per decade was calculated by multiplying 10 years with the coefficients of a linear regression model established from 20 annual mean $PM_{2·5}$ values for each global grid cell.

Africa (30·1 µg/m³ [0·7]). Australia and New Zealand (8·5 µg/m³ [0·9]), other regions in Oceania (12·6 µg/m³ [0·5]), and southern America (15·6 µg/m³ [0·3]) had the lowest annual population-weighted $PM_{2·5}$ concentrations (appendix p 8). We observed the highest correlation between estimated and observed daily mean $PM_{2·5}$ in eastern Asia (Spearman's r=0·87) and western Europe (r=0·82), and the lowest correlation in Australia and New Zealand (r=0·59). Estimated global daily mean $PM_{2·5}$ concentrations had high correlation with the observed global mean (r=0·91; appendix p 9). The temporal trends between observed and predicted daily $PM_{2·5}$ in eight metropolises are presented in the appendix (p 21).

Observed and DEML-predicted $PM_{2·5}$ concentrations were compared (figure 2). DEML could accurately capture the global variability in ground-measured daily mean $PM_{2·5}$ with a 10-fold cross-validation R^2 of 0·91 and RMSE of 7·86 µg/m³, and with a cross-validation R^2 of 0·96 (RMSE 3·78 µg/m³) for monthly mean $PM_{2·5}$ and 0·98 (2·64 µg/m³) for annual mean $PM_{2·5}$. Results of other spatiotemporal cross-validations also indicated robust model performances (appendix pp 10–13).

We presented the spatial distribution of estimated annual mean $PM_{2·5}$ in 2000, 2010, and 2019 and spatiotemporal changes per decade in 2000–19 (figure 3). Despite large differences between regions in $PM_{2·5}$ concentrations, most areas with high $PM_{2·5}$ were in eastern Asia, southern Asia, and northern Africa. Most areas in Asia, northern and sub-Saharan Africa, Oceania, and Latin America and the Caribbean had increases in $PM_{2·5}$ concentrations during the 20 years, while decreases were estimated in Europe, some regions of northern America, and some regions of Africa (figure 3). Based on the 2005 WHO guideline limit,[24] 29·4% of the global land area and 1·8% of the global population were exposed to an annual mean concentration of $PM_{2·5}$ lower than 10 µg/m³ in 2019. When restricting to the 2021 WHO guideline value of 5 µg/m³, only 0·18% of the global land area and 0·001% of the global population remained at an annual exposure lower than the guideline limit in 2019. Regional land area exposures and population exposures in 2019 are provided in the appendix (pp 14–15).

We assessed temporal changes in estimated annual mean population-weighted $PM_{2·5}$ concentration and the proportion of annual population-weighted exposed days

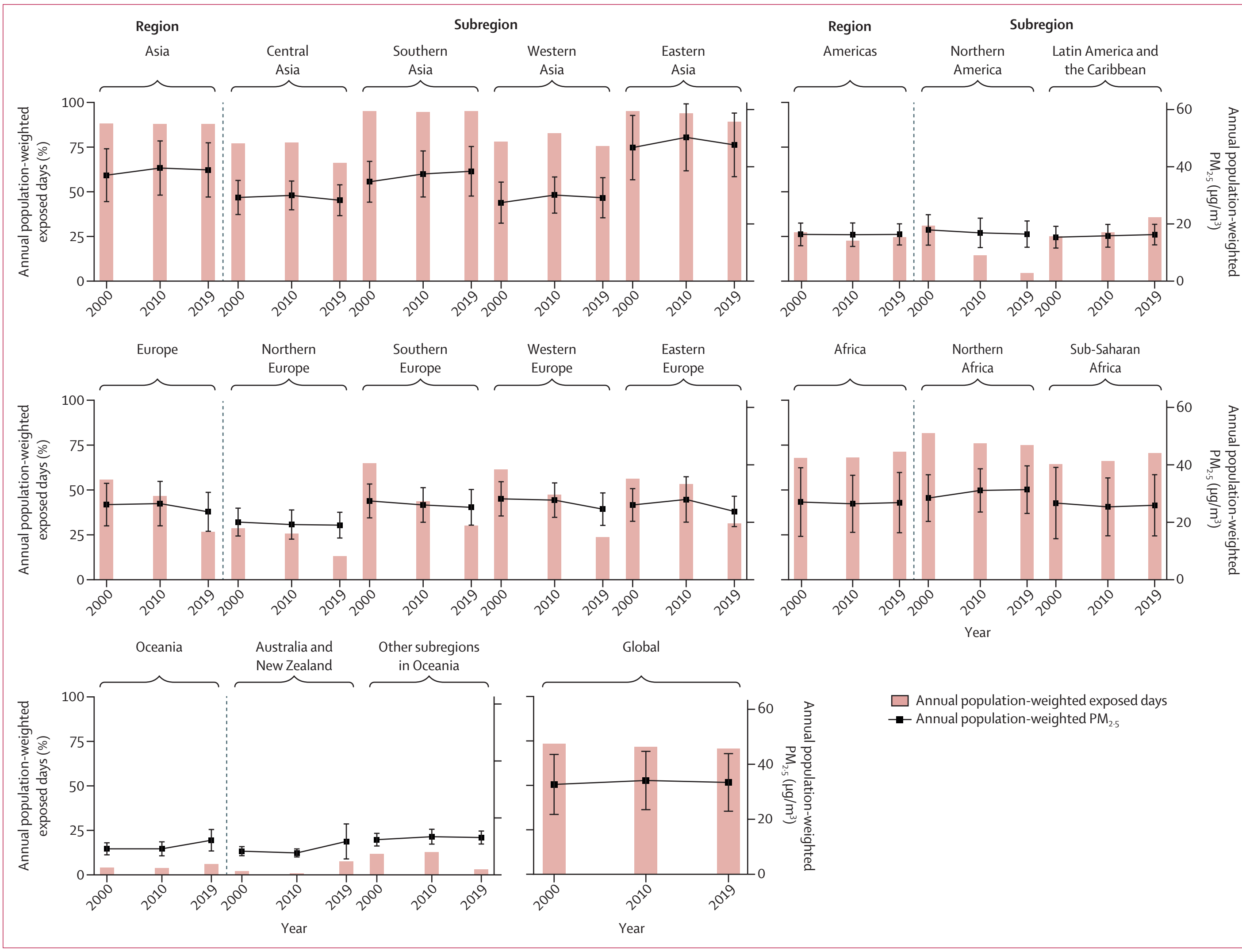

Figure 4: **Changes in annual mean population-weighted $PM_{2\cdot5}$ concentration and annual population-weighted exposed days by region in 2000, 2010, and 2019**
The bar chart shows the proportion of population-weighted exposed days with $PM_{2\cdot5}$ concentrations higher than the 2021 WHO air quality guideline limit (15 µg/m³) in a year. Datapoints and connecting lines show the annual mean population-weighted $PM_{2\cdot5}$ concentrations; error bars show standard deviation.

($PM_{2\cdot5}$ >15 µg/m³) in 2000, 2010, and 2019 (figure 4). Globally, we observed similar annual mean population-weighted $PM_{2\cdot5}$ at each timepoint, from 31·6 µg/m³ (SD 10·5) in 2000, to 33·0 µg/m³ (10·3) in 2010, to 32·3 µg/m³ (10·1) in 2019 (figure 4). Geographically, population-weighted $PM_{2\cdot5}$ concentration and annual population-weighted exposed days in Europe and northern America decreased over the two decades, whereas exposures increased in southern Asia, Australia and New Zealand, and Latin America and the Caribbean. Despite a slight decrease in population-weighted exposed days globally, by 2019 more than 70% of days still had $PM_{2\cdot5}$ concentrations higher than the 2021 WHO daily limit (15 µg/m³). In southern Asia and eastern Asia, more than 90% of days in 2000, 2010, and 2019 had daily $PM_{2\cdot5}$ concentrations higher than 15 µg/m³ (figure 4). Oceania, especially Australia and New Zealand, had a marked increase in the number of days with high $PM_{2\cdot5}$ concentrations in 2019. Trends in annual mean population-weighted concentrations by continent are presented in the appendix (p 22).

We ranked countries with the highest estimates of population-weighted $PM_{2\cdot5}$ exposure and exposed days above 15 µg/m³ $PM_{2\cdot5}$ in 2000, 2010, and 2019 (figure 5). Despite a decrease in estimated $PM_{2\cdot5}$ exposure between 2010 and 2019, China ranked first for annual mean population-weighted $PM_{2\cdot5}$ concentration at all three timepoints. Additionally, increased country rank in the past two decades according to annual mean population-weighted $PM_{2\cdot5}$ concentration was observed in southern Asian countries such as Bangladesh (from 11th in 2000 to 3rd in 2019), India (from 15th to 8th), and

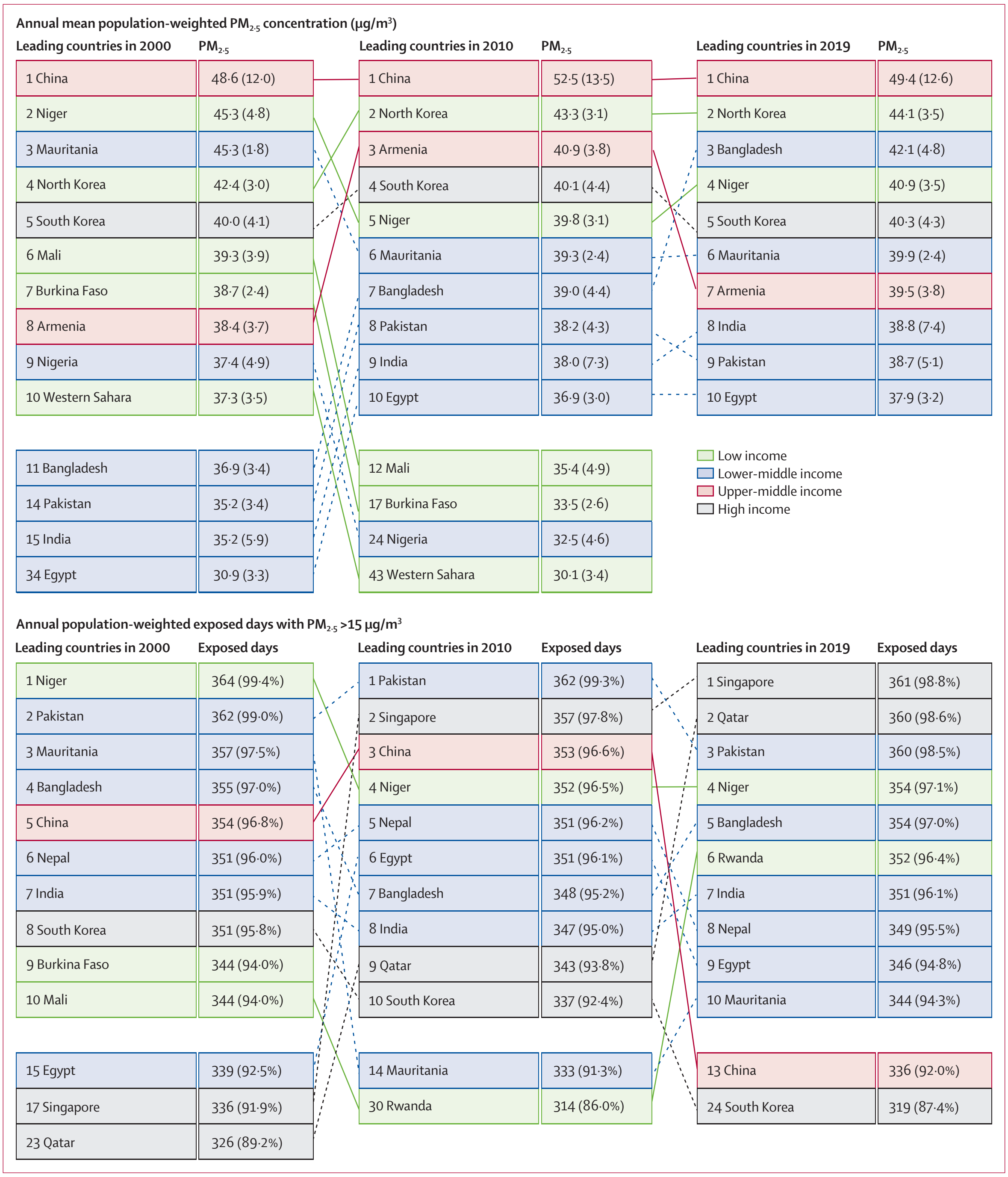

***Figure 5:* Leading countries for annual mean population-weighted $PM_{2.5}$ concentration and population-weighted exposed days in 2000, 2010, and 2019**
Values in parentheses are standard deviation (for $PM_{2.5}$ concentration) or % (for proportion of days out of 365). The 2021 WHO air quality guideline limit for daily $PM_{2.5}$ (15 μg/m³) was the threshold for population-weighted exposed days. Income categories are based on The World Bank classifications for the year 2019.

Pakistan (from 14th to 9th). For annual population-weighted exposed days, all countries ranked within the top 10 in 2000, 2010, and 2019 had more than 90% of days of the year exceeding 15 μg/m³, including several high-income countries such as Singapore and Qatar in 2010 and 2019, and South Korea in 2000 and 2010.

Estimates of annual mean population-weighted $PM_{2.5}$ and annual population-weighted exposed days in 2000, 2010, and 2019 by all countries and regions are listed in the appendix (pp 16–20).

Seasonal patterns in estimated $PM_{2.5}$ concentrations for the 20-year period were investigated (figure 6). The

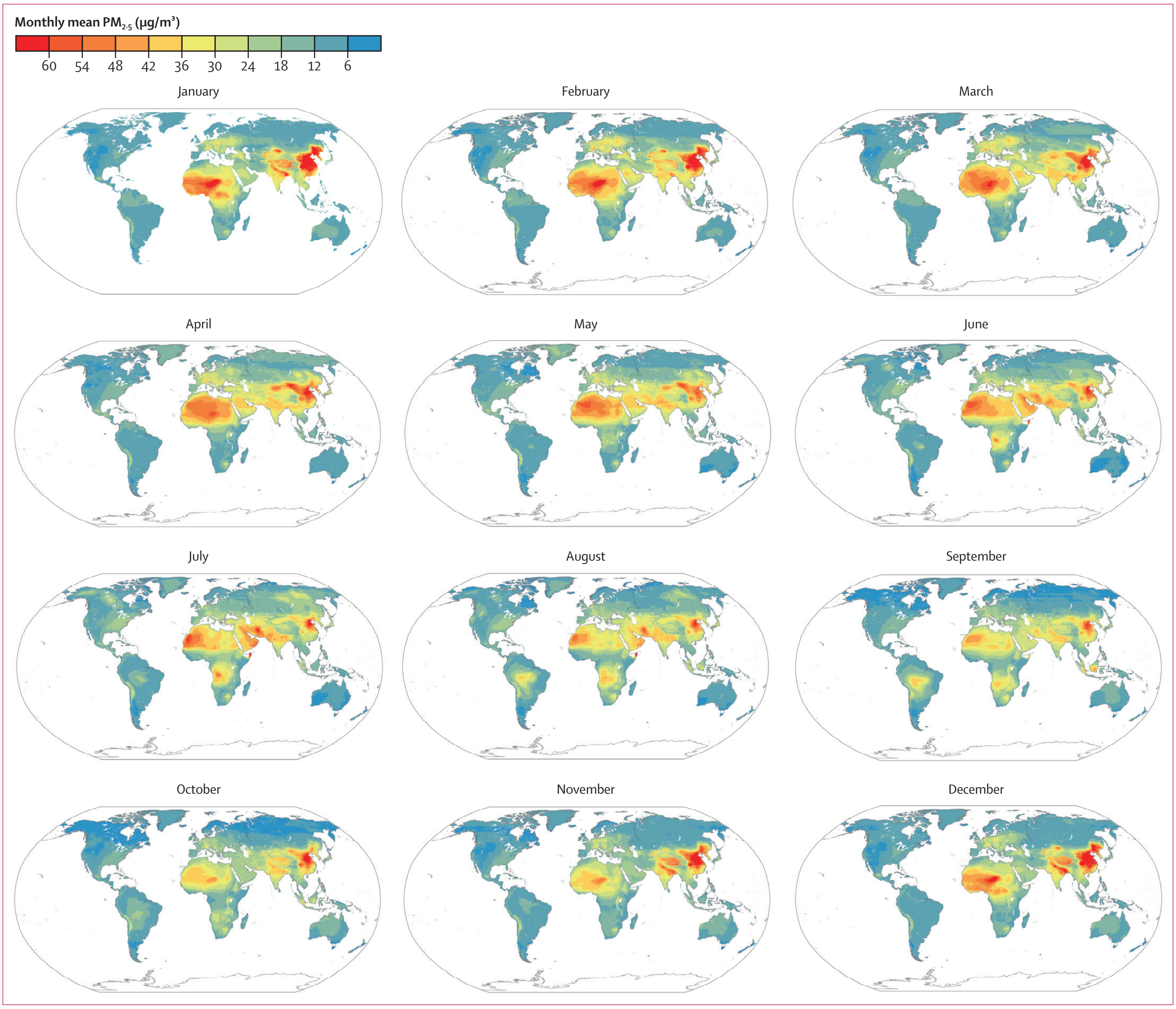

Figure 6: **Global mean $PM_{2·5}$ concentrations for each calendar month in 2000–19 at 0·1°×0·1° spatial resolution**
Estimated daily mean $PM_{2·5}$ values were used to calculate mean $PM_{2·5}$ for each calendar month, and the monthly mean values were averaged over the 20 years.

estimates of $PM_{2·5}$ showed marked seasonal variations across particular regions and countries. For example, northeast China and north India had high $PM_{2·5}$ concentrations during their winter months (December, January, and February), whereas eastern areas in northern America had high $PM_{2·5}$ in its summer months (June, July, and August). Relatively high $PM_{2·5}$ air pollution was observed in August and September in South America and from June to September in sub-Saharan Africa. Seasonal patterns for eight national distributions are presented in the appendix (p 23).

Discussion

Our grid-based daily $PM_{2·5}$ estimates provide the first global perspective on the spatial and temporal variations of population exposure in 2000–19. We implemented a validated DEML model on the basis of $PM_{2·5}$ data from 5446 monitoring stations in 65 countries worldwide, combined with chemical transport model simulations, meteorological conditions, and other geographical information, and achieved strong model performance. Consistent with previous studies,[6,8,10] our results for eastern and southern Asia indicated the highest $PM_{2·5}$ concentrations. Additionally, the estimates of $PM_{2.5}$ showed distinct seasonal patterns in many regions of the world.

In this study, the innovative DEML approach to estimate daily $PM_{2.5}$ concentrations attained high prediction accuracy. Global estimation of $PM_{2.5}$ concentrations has been described in many previous modelling studies.[6,9,25,26] An estimation of global daily $PM_{2.5}$ concentration in 1997–2014 used remote sensing and meteorological data, and ground-based observations of $PM_{2.5}$ in 55 countries, to train a machine learning algorithm, achieving a correlation coefficient for each independent validation dataset of 0·52–0·75.[9] Similarly, other studies have measured global monthly and annual mean $PM_{2.5}$ exposure. Hammer and colleagues[6] assessed global estimates of annual mean $PM_{2.5}$ concentrations in 1998–2018 with data from satellite observations, a chemical transport model, and ground-based monitoring, obtaining an R^2 of 0·90–0·92. Shaddick and colleagues[8] used a Bayesian hierarchical model with multiple data integration to estimate the global annual mean $PM_{2.5}$ exposures in 2014 with an R^2 of 0·91. However, comparisons with previous global $PM_{2.5}$ estimation studies are restricted given that few studies estimated the global daily (short-term) $PM_{2.5}$ exposure, and further comparison analyses are required.

Although ambient $PM_{2.5}$ concentrations vary substantially across the world, few global land areas and populations are exposed to $PM_{2.5}$ at concentrations lower than the 2021 WHO guideline limit.[5] Based on modelled estimates for 2019, we found that 0·18% of the global land area and 0·001% of the global population had an annual exposure to $PM_{2.5}$ lower than the WHO limit of 5 μg/m³, and more than 70% of days in the year, even in some high-income countries such as Singapore and Qatar, had a daily mean $PM_{2.5}$ exposure above 15 μg/m³. Additionally, the high spatiotemporal global $PM_{2.5}$ estimates provide valuable air pollution information for areas not covered by monitoring stations. For example, we observed a notable seasonal trend with regularly high regional $PM_{2.5}$ concentrations in August and September in the Amazon rainforest region, where monitoring stations are scarce.

The spatiotemporal variations in $PM_{2.5}$ concentrations might be the result of different types and components of anthropogenic fuel combustion emissions[27] and the changes in natural sources due to extreme weather events such as bushfires and windblown dust.[28] For example, northeast China had increased estimated $PM_{2.5}$ concentrations in winter, which might stem from conducive weather patterns[29] and winter heating-related fossil fuel combustion,[30] whereas southern American countries such as Brazil had increased estimated $PM_{2.5}$ concentrations in August and September, which might be associated with anthropogenic emissions such as slash-and-burn cultivation.[31] By contrast, the increasing frequency and scale of climate change-related air pollution events, such as windblown dust and bushfire events in 2019, might have contributed to the elevated $PM_{2.5}$ concentrations in south-eastern Australia in 2019.[32]

This study has several limitations. Despite conducting spatial and temporal cross-validations to test the robustness of estimations with stable results, biases might exist in the DEML model due to the sparse ground station distribution in most regions, limited availability of station data before 2014, and uncertainties from some predictors, such as the fixed anthropogenic emissions assumption in the chemical transport model data after 2013. In areas with few monitoring stations, the estimates should be interpreted with caution because our model predictions in these areas were only validated by a small number of monitoring stations, and the air quality records from these location-based ground stations are not representative of overall exposure variations across a wide region. In addition, despite a high spatial resolution, uncertainties exist in the aggregation of gridded estimates and calculation of population-weighted exposure and exposed days. Our study also cannot account for personal exposure, and assuming equivalent population exposures within each grid cell might lead to exposure misclassification due to population migration and different activity patterns and behaviours, such as time spent indoors and outdoors.

This study is the first to apply an innovative DEML approach to estimate high-resolution global daily $PM_{2.5}$ exposure over two decades. Our DEML model showed accurate performance in predicting global and regional short-term and long-term $PM_{2.5}$ concentrations, and provided a global perspective on the spatial and temporal variations of $PM_{2.5}$ concentrations between 2000 and 2019. The results highlight the inequality of global air pollution population exposure, particularly in southern and eastern Asia. Our findings are of importance for global air pollution mitigation strategies and for assessing the short-term and long-term health effects of global $PM_{2.5}$ exposure.

Contributors

YG and SL obtained funding. YG, JS, and SL are the study guarantors. YG and WY conceived the idea for the study. WY wrote the original draft and reviewed and edited the manuscript under the supervision of YG, JS, and SL. WY also performed the literature search, established the model, and analysed the data. TY was responsible for results visualisation and parts of the results presentation. RX, YL, WY, ZY, and YuZ contributed to data collection, cleaning, and preparation. RX and YL also verified the analytical methods. XY, YiZ, and ZC contributed to the interpretation of parts of the results and code improvement. YG and SL supervised the project. YG and WY accessed and verified the data. All authors had full access to all the data in the study. All authors discussed the results, provided critical feedback, helped revise the final manuscript, and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The global annual mean $PM_{2.5}$ data will be freely available in Monash Planetary Health Data Hub (https://deepearth.erc.monash.edu/), which is under development and may become available in 2023. The daily $PM_{2.5}$ datasets and related study documents including the monitoring station dataset are available upon request from the corresponding authors.

Acknowledgments

This study was supported by the Australian Research Council (grant number DP210102076), the Australian Medical Research Future Fund (grant number 2015916), and the Australian National Health and Medical Research Council (grant number APP2000581). YG was supported by the Career Development Fellowship (award number APP1163693) and Leader Fellowship (award number APP2008813) of the Australian National Health and Medical Research Council. SL was supported by an Emerging Leader Fellowship of the Australian National Health and Medical Research Council (award number APP2009866). WY was supported by a Monash Graduate Scholarship, Monash International Tuition Scholarship, and a Centre for Air Pollution, Energy, and Health Research (Australia) PhD top-up scholarship. TY was supported by the China Scholarship Council (award number 201906320051). We thank Fiona Hore-Lacy from the Monash Centre for Occupational and Environmental Health, School of Public Health and Preventive Medicine, Monash University, for assisting in proofreading the manuscript.

Editorial note: The Lancet Group takes a neutral position with respect to territorial claims in published figures and tables.

References

1 Murray CJ, Aravkin AY, Zheng P, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; **396:** 1223–49.

2 Yu W, Guo Y, Shi L, Li S. The association between long-term exposure to low-level $PM_{2.5}$ and mortality in the state of Queensland, Australia: a modelling study with the difference-in-differences approach. *PLoS Med* 2020; **17:** e1003141.

3 Chen J, Hoek G. Long-term exposure to PM and all-cause and cause-specific mortality: a systematic review and meta-analysis. *Environ Int* 2020; **143:** 105974.

4 Liu C, Chen R, Sera F, et al. Ambient particulate air pollution and daily mortality in 652 cities. *N Engl J Med* 2019; **381:** 705–15.

5 WHO. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Sept 22, 2021. https://www.who.int/publications/i/item/9789240034228 (accessed Nov 5, 2021).

6 Hammer MS, van Donkelaar A, Li C, et al. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ Sci Technol* 2020; **54:** 7879–90.

7 van Donkelaar A, Hammer MS, Bindle L, et al. Monthly global estimates of fine particulate matter and their uncertainty. *Environ Sci Technol* 2021; **55:** 15287–300.

8 Shaddick G, Thomas ML, Green A, 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 C Appl Stat* 2018; **67:** 231–53.

9 Lary DJ, Faruque FS, Malakar N, et al. Estimating the global abundance of ground level presence of particulate matter ($PM_{2.5}$). *Geospat Health* 2014; **8:** S611–30.

10 Chen G, Li S, Knibbs LD, et al. A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. *Sci Total Environ* 2018; **636:** 52–60.

11 Shtein A, Kloog I, Schwartz J, et al. Estimating daily $PM_{2.5}$ and PM_{10} over Italy using an ensemble model. *Environ Sci Technol* 2020; **54:** 120–28.

12 Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. *Environ Sci Technol* 2016; **50:** 4712–21.

13 Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 2018; **6:** e5518.

14 Zhou Z-H. Ensemble methods: foundations and algorithms. New York: Chapman and Hall/CRC, 2012.

15 Xiao Q, Chang HH, Geng G, Liu Y. An ensemble machine-learning model to predict historical $PM_{2.5}$ concentrations in China from satellite data. *Environ Sci Technol* 2018; **52:** 13260–69.

16 Di Q, Amini H, Shi L, et al. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int* 2019; **130:** 104909.

17 Yu W, Li S, Ye T, Xu R, Song J, Guo Y. Deep ensemble machine learning framework for the estimation of $PM_{2.5}$ concentrations. *Environ Health Perspect* 2022; **130:** 37004.

18 Riley M, Kirkwood J, Jiang N, Ross G, Scorgie Y. Air quality monitoring in NSW: from long term trend monitoring to integrated urban services. *Air Qual Clim Chang* 2020; **54:** 44–51.

19 Yue X, Unger N. Fire air pollution reduces global terrestrial productivity. *Nat Commun* 2018; **9:** 5413.

20 Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc* 2020; **146:** 1999–2049.

21 Cai J. Humidity: calculate water vapor measures from temperature and dew point. 2019. https://cran.r-project.org/web/packages/humidity/index.html (accessed Sept 15, 2020).

22 Stroh E, Harrie L, Gustafsson S. A study of spatial resolution in pollution exposure modelling. *Int J Health Geogr* 2007; **6:** 19.

23 Bishop CM. Pattern recognition and machine learning. New York, NY: Springer, 2006.

24 Krzyzanowski M, Cohen A. Update of WHO air quality guidelines. *Air Qual Atmos Health* 2008; **1:** 7–13.

25 van Donkelaar A, Martin RV, Brauer M, et al. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ Sci Technol* 2016; **50:** 3762–72.

26 Southerland VA, Brauer M, Mohegh A, et al. Global urban temporal trends in fine particulate matter ($PM_{2.5}$) and attributable health burdens: estimates from global datasets. *Lancet Planet Health* 2022; **6:** e139–46.

27 McDuffie EE, Martin RV, Spadaro JV, et al. Source sector and fuel contributions to ambient $PM_{2.5}$ and attributable mortality across multiple spatial scales. *Nat Commun* 2021; **12:** 3594.

28 Chen G, Guo Y, Yue X, et al. Mortality risk attributable to wildfire-related $PM_{2.5}$ pollution: a global time series study in 749 locations. *Lancet Planet Health* 2021; **5:** e579–87.

29 Li J, Hao X, Liao H, et al. Winter particulate pollution severity in north China driven by atmospheric teleconnections. *Nat Geosci* 2022; **15:** 349–55.

30 Xiao Q, Ma Z, Li S, Liu Y. The impact of winter heating on air pollution in China. *PLoS One* 2015; **10:** e0117311.

31 Barlow J, Berenguer E, Carmenta R, França F. Clarifying Amazonia's burning crisis. *Glob Change Biol* 2020; **26:** 319–21.

32 Yu P, Xu R, Abramson MJ, Li S, Guo Y. Bushfires in Australia: a serious health emergency under climate change. *Lancet Planet Health* 2020; **4:** e7–8.