Dear Editor,

Thank you for giving us the opportunity to submit the revised draft of our manuscript titled "Global $PM_{2.5}$ prediction and associated mortality to 2100 under different climate change scenarios" to *Environmental Science & Technology*. We thank the insightful comments from the reviewers. Most of the suggestions made by reviewers have been incorporated and the changes made within the manuscript have been highlighted. Please see below, in blue, for a point-by-point response to the reviewers' comments and concerns. The page numbers and line numbers refer to the revised manuscript file with tracked changes.

Reviewers' comments:

Reviewer: 1

Comments:

The authors conducted a comprehensive study on evaluating the impact of ambient fine particulate matter (i.e., $PM_{2.5}$) concentrations on public health or premature mortality burden by the end of the 21st century under different climate change scenarios. This is an interdisciplinary research, representing a great concern to research communities, general public, and policy makers.

Through this study, the authors developed a deep machine learning model and applied it to project the future ambient $PM_{2.5}$ concentrations at a high resolution ($0.1^{o} \times 0.1^{o}$) on a global scale. With such a dataset, the authors completed a rigorous assessment of the impact of climate change-constrained air quality on human health. The results have important implications for policymakers to formulate feasible shared socioeconomic pathways to support sustainable development. Overall, the study is carefully-designed, the manuscript is well written and organized. However, several aspects need to be clarified before it is accepted to publish in EST.

Author's Response: We thank the constructive comments and suggestions from the reviewer. Based on your review comments, we have added more descriptions of the methodology and analysis of the results into the manuscript. The point-to-point responses are provided as follows.

Specific comments

1. Reviewer #1 comment: Details about the emission: Emissions inputs are critical for generating accurate $PM_{2.5}$ forecasts or projections. However, it is not clear what global emission inventories were used in the projection or predictions of $PM_{2.5}$ concentrations from 2021 to 2100. As indicated on Line 91, PKU-FUEL could be one of them. Are there any other global emission inventories included in the study? In addition, it will be helpful if some additional detailed information about the global emission

inventories used in this study can be provided. For instance, what are the base year(s) of emission inventories and what is the spatial resolution of original emission inventories?

Author's response: Thanks for your suggestion. We have added more detailed information about the global emission inventories used in this study in the revised manuscript (Lines 104–109).

"*CMIP6 dataset contains historical emissions (1750 to 2014) and future emission data for SSP scenarios (2015-2100). CMIP6 emissions data were utilized in both the training and prediction processes. CMIP6 historical emissions data (1998–2014, the historical emissions data are available till 2014) were used to build the deep learning model, and the emission data of SSP scenarios for 2015 to 2019 were used in the deep learning model verification process. Future emission data for 2021–2100 were input into the trained deep learning model for prediction. The detailed narratives of emission inventory used in this study are summarized in Table S3.*"

Table S3. Narratives of CMIP6 emission inventory used in this study.

Role in this study	Emission inventory in this study	Time period	Original Resolution	Data citation
Training	Historical emissions	1998-2014	0.5 º x 0.5 º	Feng, et al. (2020)[1]
Verification	Emissions for SSP scenarios	2015-2019	0.5 º x 0.5 º	Feng, et al. (2020)[1]
Prediction	Emissions for SSP scenarios	2021-2100	0.5 º x 0.5 º	Feng, et al. (2020)[1]

<u>2. Reviewer #1 comment:</u> L88: Here the word "Primary" is a little bit confused since I assume the emissions of all $PM_{2.5}$ compositions are primary. When you use five pollutants as emission inputs to drive the deep learning model, what are the percentages of individual species?

Author's response: Thanks for your comment. We agreed with the reviewer, the word "primary" can be misleading. We have revised the below sentence in Lines 89-92: *"Since the deficiency in the emissions of primary $PM_{2.5}$ components (except organic carbon (BC) and black carbon (OC)) in the CMIP6 datasets, future $PM_{2.5}$ concentrations are driven by changes to precursor emissions (ammonia (NH_3), nitrogen oxides (NOx), and sulfur dioxide (SO_2)), BC, OC and climate in this study.*"

Global emission amounts and percentages of the five species have been included in the supplementary information (SI), as shown in Table S1 and Table S2.

Table S1. Global emission amount of BC, NOx, NH_3, OC and SO_2.

Species	1995	2000	2005	2010	2014	Unit
BC	8.12	7.46	8.84	9.66	9.74	Mt BC/yr
NOx	135.17	135.53	149.85	155.42	155.64	Mt NOx/yr
NH_3	54.03	54.51	58.88	62.45	65.04	Mt NH_3/yr
OC	32.08	27.98	32.72	34.61	36.15	Mt OC/yr
SO_2	121.40	111.15	124.98	116.32	105.00	Mt SO_2/yr

Table S2. Percentage (%) of historical BC, NOx, NH_3, OC and SO_2 emissions.

Species	1995	2000	2005	2010	2014
BC	2.3%	2.2%	2.4%	2.6%	2.6%
NOx	38.5%	40.3%	39.9%	41.1%	41.9%
NH_3	15.4%	16.2%	15.7%	16.5%	17.5%
OC	9.1%	8.3%	8.7%	9.1%	9.7%
SO_2	34.6%	33.0%	33.3%	30.7%	28.3%

3. Reviewer #1 comment: L117-118: There is a gap between historical simulations (1981-2010) and future projections (2021-2100). I'm curious as to the reasons for this.

Authors' response: Thank you for the comment. The classical period of climate, defined as the mean and variability of relevant quantities of certain variables over a period of time, is 30 years, which was recommended by the World Meteorological Organization (WMO).[2] WMO used a 30-year baseline for weather and climate, operating on the principle that 30 years of data provide enough information to even short-term variability and afford a reliable reference period for monitoring the general patterns of weather and climate. A widely used standard reference period for calculating climate normals is the 30-year period of 1981-2010.[3] In order to keep the same number of years (30-year average) as 1981-2010, we followed the previous downscaling study [4] and set the future scenarios as four 30-year periods for climate downscaling. The future periods are set as 2021–2050 (2030s), 2041–2070 (2050s), 2061–2090 (2070s), and 2071–2100 (2080s), respectively.

4. Reviewer #1 comment: As pointed out on L119-122, four SSP scenarios were classified by socioeconomic, land use, and environmental development assumptions. Can you add a table to summarize the key differences among them in the Supplementary even though they were described in other references?

Authors' response: Thank you for your suggestion. We have further added a summarized table of key differences among four SSP scenarios in Table S5 in the supplemental material.

Table S5. Descriptions of the critical elements for SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 relevant to air pollution

Scenario	Description	Land-use Change Regulation	Environmental Development	Pollution Control Policy	Energy Tech Change	Population Growth
SSP1-2.6	Sustainability	Strong regulation to avoid environmental tradeoffs	High environmental awareness; more inclusive development that respects perceived environmental boundaries	Strong	Directed away from fossil fuels, toward efficiency and renewables	Relatively Low
SSP2-4.5	Middle-of-the-road	Medium regulation; slow decline in the rate of deforestation	Medium environmental awareness; work toward but make slow progress in achieving sustainable development goals	Medium	Some investment in renewables but continued reliance on fossil fuels	Medium
SSP3-7.0	Regional rivalry	Limited regulation; continued deforestation	Low environmental awareness; strong environmental degradation in some regions	Weak	Slow tech change, directed toward domestic energy sources	High in developing countries; low in industrialized countries
SSP5-8.5	Fossil-fueled development	Medium regulation; slow decline in the rate of deforestation	High environmental awareness; local environmental problems are successfully managed	Strong	Directed toward fossil fuels; alternative sources not actively pursued	Relatively Low

5. Reviewer #1 comment: L122: Why do the authors highlight "radiative forcing" as an indicator to distinguish these four scenarios here? Is there any indication of "radiative forcing" to ambient levels of $PM_{2.5}$ in the future or to the shared socioeconomic pathways?

Authors' response: Thank you for your question. The term "radiative forcing" has been employed in the IPCC Assessments to denote an externally imposed perturbation in the radiative energy budget of the Earth's climate system. Conceptually, "radiative forcing" is the change in energy flux caused by both natural and anthropogenic factors that affect the global energy balance and force changes in the Earth's climate.[5] Changes in greenhouse gas concentrations in the atmosphere affect radiative forcing;

thus, radiative forcing corresponds to different greenhouse gas (GHG) emissions and the resulting climate change. The SSP scenarios translate the socioeconomic factors and mitigation goals into GHG emissions (and associated radiative forcing) in a standardized manner, indicating the social pathways associated with different levels of warming. For example, SSP 1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 correspond to the low, medium, high, and very high GHG emissions scenarios, resulting in subsequent radiative forcing of 2.6, 4.5, 7.0, 8.5 Watt/m^2 by 2100. Hence, the IPCC report and previous literature used "radiative forcing values" to represent four different future pathways. Under different future "radiative forcing" conditions, emissions and meteorological conditions are different and such variations can be further propagated to the $PM_{2.5}$ concentration. Previous studies have investigated the relationship between "radiative forcing" and the subsequent influence on ambient levels of $PM_{2.5}$. Colette et al. (2013)[6] estimated that, due to the air pollutant emission reduction, the decrease in $PM_{2.5}$ concentrations by 2050 in western Europe could reach up to 60% under radiative forcing of 8.5 Watt/m^2 and 75% under radiative forcing of 2.6 Watt/m^2.

To explain the "radiative forcing" as an indicator to climate change and its relationship with $PM_{2.5}$ concentration, we have added the following description from Line 148 to Line 158.

"Changes in greenhouse gas concentrations in the atmosphere affect radiative forcing; thus, 'radiative forcing' mentioned in this work corresponds to different greenhouse gas (GHG) emission scenarios and the resulting climate change. Both SSP1-2.6 and SSP5-8.5 represent strong climate mitigation scenarios,[7, 8] with the distinction that the anthropogenic radiative forcings by 2100 are 2.6 Watt/m^2 and 8.5 Watt/m^2, respectively.[9, 10] SSP2-4.5 represents a moderate mitigation scenario and the radiative forcing is stabilized at 4.5 Watt/m^2 until 2100 by implementing moderately restrictive emission reduction measures and strategies.[11] SSP3-7.0 is the weakest climate mitigation scenario with an anthropogenic radiative forcing of 7.0 Watt/m^2 by 2100.[12]"

<u>6. Reviewer #1 comment:</u> Some details about the 80 years of global $PM_{2.5}$ projection (why not use prediction?): It requires significant computer resources to complete long-term global high-resolution ($0.1^o \times 0.1^o$) $PM_{2.5}$ predictions by using a 3D numerical climate-chemistry or air quality model. How much computer resources does the deep learning model need to complete the 80 years of simulations in this study?

Authors' response: Thanks for your comment. "Prediction" and "projection" are used to describe two different kinds of information that both significantly represent estimations about future conditions. "Prediction" is a probabilistic statement, i.e., a predictive relationship based on modeling some phenomenon from available data. In other words, "prediction" describes the forecasted climate for the coming months or years, which are strongly influenced by natural cycles of variability. While "projection" is a probabilistic statement that indicates about what happens in the future is possible if certain conditions develop. Projection relies on prescribed scenario information which illuminates how scenarios of gas emissions and associated socioeconomic pathways will influence climate and air quality in the long term. In summary, the critical difference between “prediction” and “projection” is that projections generally rely on scenarios or assumptions regarding the future.

Regarding the computation resources cost, detailed information has been added to Text S2 in the SI (Page 4).

“In this study, all experiments were performed on a high-performance computing server with an Intel i7-10700 CPU and an NVIDIA Quadro RTX 4000 GPU. It takes about 20 seconds to execute each training epoch, and the total training time is about 5 hours. For the projection, it takes only 10 seconds on average to simulate every decade.”

<u>7. Reviewer #1 comment:</u> As shown in Fig. 2, S2-4, the CNN model predictions show very good agreement with satellite-retrieved data. I know this is beyond of this study scope, but I am still interested in whether this model can be used to generating regional operational $PM_{2.5}$ forecasts. Any comments on that?

Authors’ response: Thank you for your comment. Yes, the CNN model can also be used to predict regional $PM_{2.5}$ concentration. $PM_{2.5}$ is a cross-boundary pollution issue; thus, it is influenced not only by local but also by regional situations (e.g., meteorological conditions and emission intensities). At the same time, CNN is adept at obtaining spatial information and considering the regional influence, due to the fact that this type of deep learning model typically consists of a series of convolutional layers, pooling layers, and fully-connected layers.[13, 14]

The CNN technique was found to have good performance in estimating the regional $PM_{2.5}$ concentration. A recent study developed a CNN model and used it for regional air quality forecast in South Korea.[15] Results show that the R^2 values of $PM_{2.5}$ and PM_{10} predictions reached 0.975 and 0.976 respectively. For another $PM_{2.5}$ forecast research in Beijing, China, results revealed that CNN has a more stable improvement in regional $PM_{2.5}$ concentration prediction for the future 24 hours.[16] Therefore, the CNN approach has the potential to perform reliable regional $PM_{2.5}$ forecast.

<u>8. Reviewer #1 comment:</u> L152: Even though some explanations are provided here. I am still not very clear the reason(s) of implementing 8-fold cross-validation and the way of calculation. Are there any references on this?

Authors' response: Thanks for your comment. We agree with the reviewer that more descriptions and references are required to clarify the 8-fold cross-validation.

The core idea of cross-validation lies in dividing the dataset into several groups and averaging the results of multiple evaluations, thus thoroughly evaluating the proposed model and fully implementing all data for validation.[17, 18] With cross-validation, all data can have the opportunity to become both training and validation datasets. Therefore, cross-validation can better evaluate the generalization ability and model performance.

We have added the calculation of 8-fold cross-validation in Text S7 in the SI (Page 7).

"*Cross-validation is a resampling process used to evaluate machine/deep learning models on a limited sample of data,*[19, 20] *which has also been widely used in the verification of machine learning models for air pollution prediction.* [21, 22] *The core idea of cross-validation lies in dividing the dataset into several groups and averaging the results of multiple evaluations, thus thoroughly evaluating the proposed model and fully implementing all data for validation.*[17]

The procedure of 8-fold cross-validation is as follows.

1. *Full set of the input data is divided into 8 parts.*
2. *In the training process, each time, one part of the input data is withheld, and the remaining 7 parts are input to the CNN framework to train the model.*
3. *The correlation coefficient ($CORR_i$) and RMSEi are calculated by validation with the withheld one-part data.*
4. *Steps 2 and 3 are performed 8 times, and then 8 times of $RMSE_i$ and $CORR_i$ can be calculated.*
5. *The final RMSE and CORR are calculated by averaging the 8 different $RMSE_i$ and $CORR_i$.*"

9. Reviewer #1 comment: $PM_{2.5}$ exposure concentrations are widely used in health impact studies while ambient levels of $PM_{2.5}$ are more commonly used in air quality or atmospheric environment studies. Is(are) there any difference(s) between both in terms of values?

Authors' response: Thanks for your comment. Yes, there exist differences between the $PM_{2.5}$ exposure concentrations and ambient levels in terms of values.

The "exposure concentration" can be used to indicate the people's actual exposure to environmental $PM_{2.5}$ concentrations.[23] In real life, due to the influence of the natural environment and urban layout, residents usually gather in the city center, and the urban population is unevenly distributed. Directly using the ambient $PM_{2.5}$ concentration to assess the population exposure, without considering the disproportionate spatial-temporal distributions of the pollution and the population, might not reflect the true impact of air pollution on health. The criticism that "the ambient levels of $PM_{2.5}$ may not properly reflect individual exposures" has been supported by studies since the 1980s. A series of studies have found that personal exposures to particulate matter were much higher than the average ambient concentrations.[24-26]

To clarify, we have added the description and calculation results of "exposure" in Text S5 in the SI (Page 6).

"Population-weighted concentration can better reflect the impact of $PM_{2.5}$ pollution on the exposed population, and thus build a more reliable theoretical basis for public health assessment. Eq. (S5) and Eq. (S6) were used to calculate the exposure concentration ($\boldsymbol{C_{PM_{2.5}}^{i-pop}}$) and spatial ambient levels ($\boldsymbol{C_{PM_{2.5}}^{i-space}}$) of $PM_{2.5}$, respectively.

$$\boldsymbol{C_{PM_{2.5}}^{i-pop}} = \left(\sum_{k=1}^{N} \boldsymbol{C_k^i} * \boldsymbol{P_k^i}\right) / \sum_{k=1}^{N} \boldsymbol{P_k^i} \qquad \text{Eq. (S5)}$$

$$\boldsymbol{C_{PM_{2.5}}^{i-space}} = \frac{1}{N}\left(\sum_{k=1}^{N} \boldsymbol{C_k^i}\right) \qquad \text{Eq. (S6)}$$

Here, different countries and the total grid number of the country are represented by i and N, respectively. C_k^i and P_k^i are the $PM_{2.5}$ concentration and corresponding population of grid k in country i, respectively."

A clear difference between the exposure and spatial ambient level can be found in Table A1, with exposure at the national level being generally higher in values than the corresponding spatial ambient level of $PM_{2.5}$.

Table A1. Country-level population-weighted exposure and spatial-weighted $PM_{2.5}$ concentrations in 2015.

Country	Exposure [μg/m³]	Ambient level [μg/m³]	Population [in million]
Global	44	14	7366
Australia	6	1.5	24
Brazil	11	6.1	207
Canada	7	2.9	36
China	58	18.7	1383
France	12	8.7	65
India	72	39.4	1311
Japan	13	10.8	128
Philippines	23	4.6	100
Russia	17	6.7	148
Singapore	19	27.6	4

South Africa	30	13.1	53
South Korea	29	24.7	50
South Sudan	32	12.1	12
Spain	10	6.3	48
Turkey	36	12.9	78
United States	8	6.0	323
Vietnam	28	18.0	93

10. Reviewer #1 comment: L298-301: The impact of driving factors could be quite complex. The authors highlighted the importance of the planetary layer height for the SSP3-7.0 scenario in Figure. S7 and on lines 298-301. However, if you look closely at Figs. S5, S6, and S8, the delta changes of surface temperature, specific humidity, and sea level pressure are more significant than the change in PBLH for Scenarios SSP 3-7.0 and 5-8.5. Any comments on that?

Authors' response: Thank you very much for your comment. We agree with the reviewer that meteorological conditions affect the accumulation and diffusion of $PM_{2.5}$ through multiple mechanisms. Influence of PBLH on $PM_{2.5}$ concentrations has been reported in several previous studies. $PM_{2.5}$ concentrations are highly sensitive to PBLH, which decides the vertical space to which $PM_{2.5}$ can disperse.[27] With a given amount of emitted pollutants, lower PBL height can generally cause the pollution episode of $PM_{2.5}$.[28, 29] That's the main reason why we want to emphasize the role of PBLH on $PM_{2.5}$ concentration in the manuscript. Although the differences in pressure, temperature, and specific humidity are also significant as pointed out by the reviewer, their impacts on $PM_{2.5}$ pollution are not linear and straightforward. Based on your comment, we have modified our description of the influence of meteorological factors (Lines 425–436).

"*Under the SSP3-7.0 scenario, the planetary boundary layer height (PBLH) exerts strong influence on $PM_{2.5}$ dispersion, and thus its decreases in East Asia, South Asia, and eastern Africa (Figure S10) will increase the $PM_{2.5}$ concentrations. Besides PBLH, other meteorological conditions, such as higher temperature,*[30] *are also favorable for $PM_{2.5}$ accumulation in these regions and therefore exacerbate the $PM_{2.5}$-associated mortality burden until the 2050s.*"

11. Reviewer #1 comment: Differences in key meteorological fields (e.g., PBLH, surface temperature, etc.) are presented in Figures S5-S9. It will be helpful to include similar changes in emissions for four scenarios if they are different.

Authors' response: Thank you for your suggestion. We have added below figures to reveal the changes of emissions in the supplementary material. (SI Page 23-25)

"*Emissions (SO_2, NH_3, OC, BC, NOx) for the four SSP scenarios up to 2100 are shown in Figures S13-S17.*"

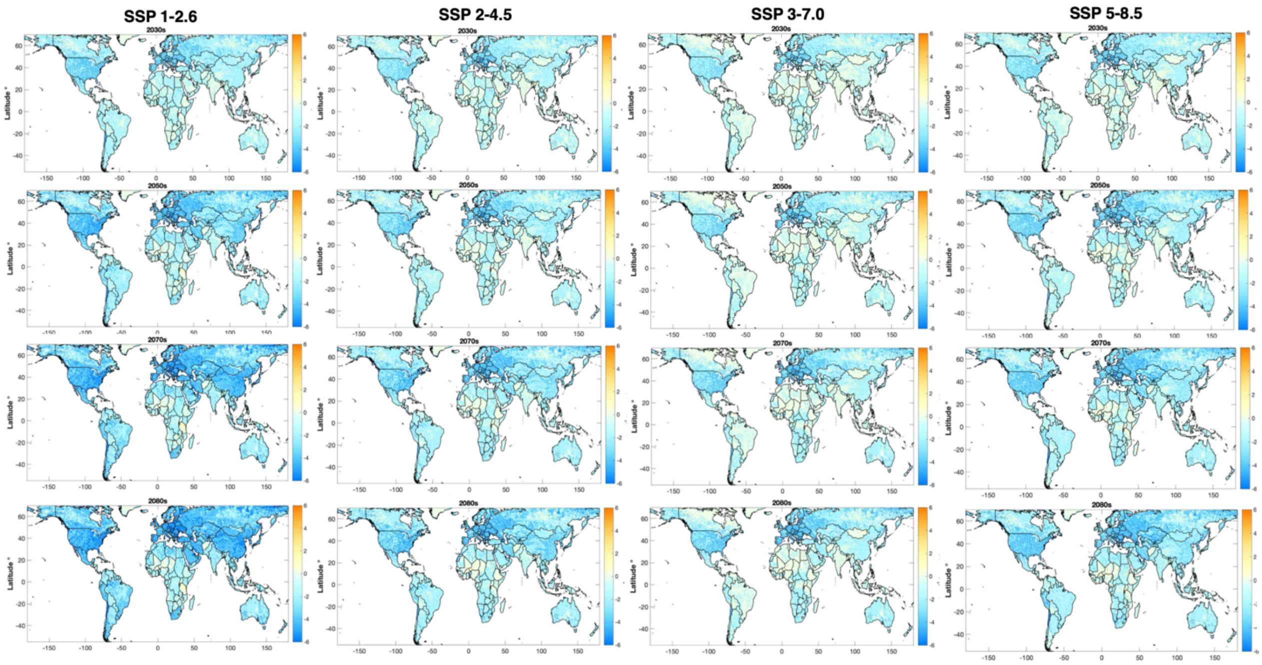


Figure S13. Changes in SO_2 emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

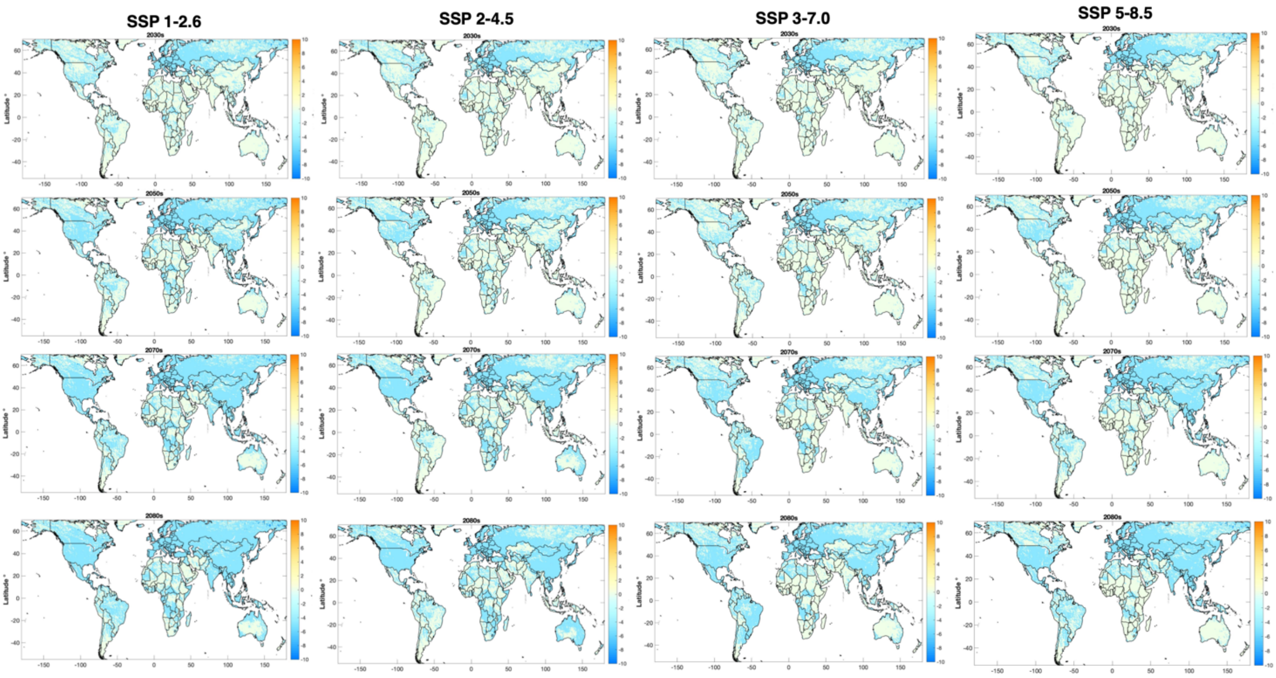


Figure S14. Changes in NH_3 emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

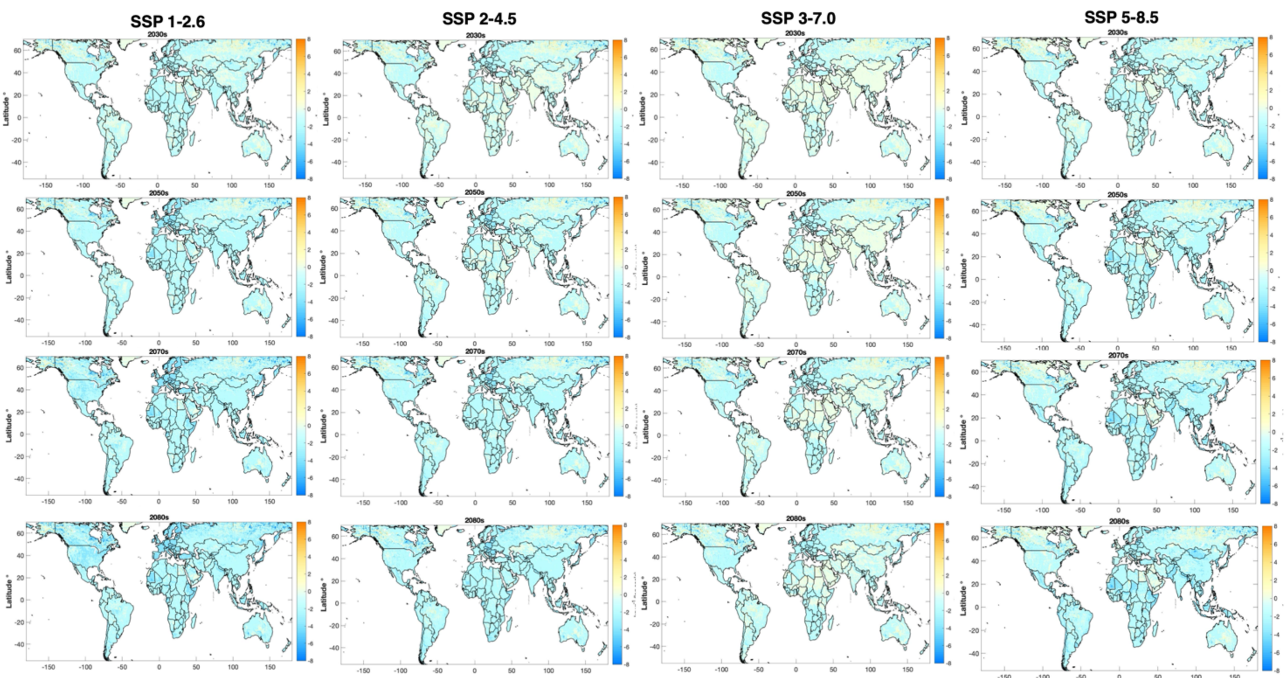


Figure S15. Changes in OC emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

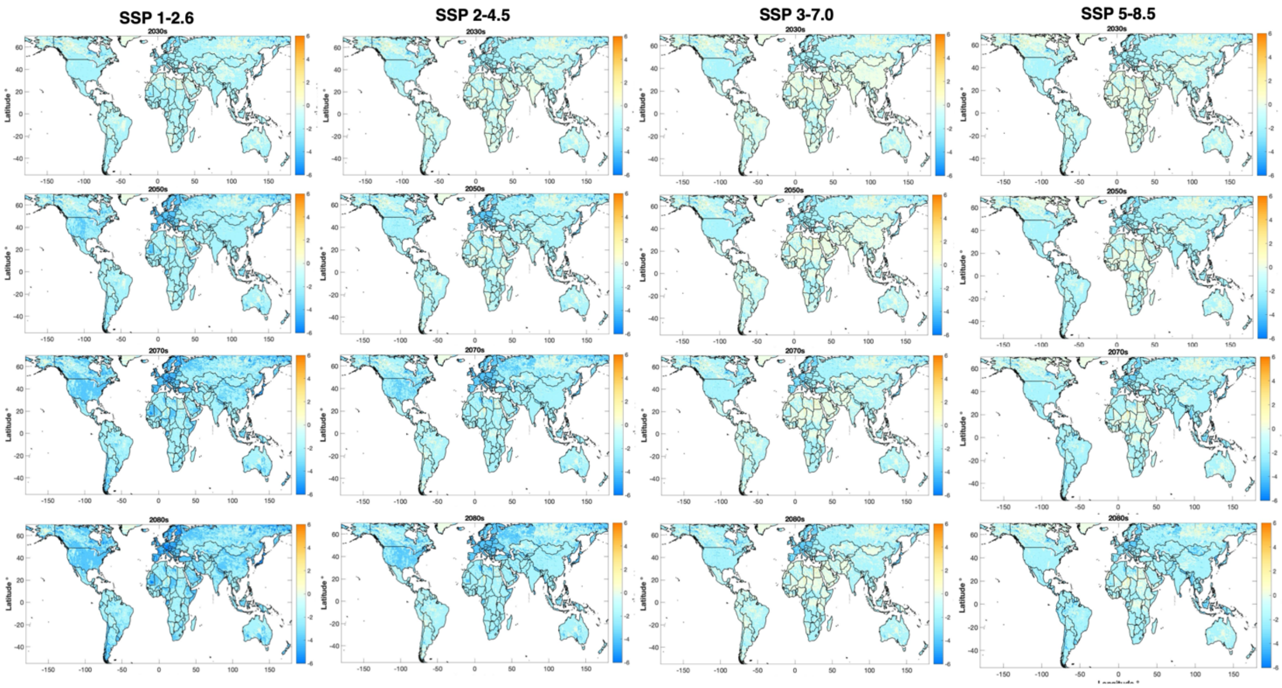


Figure S16. Changes in BC emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

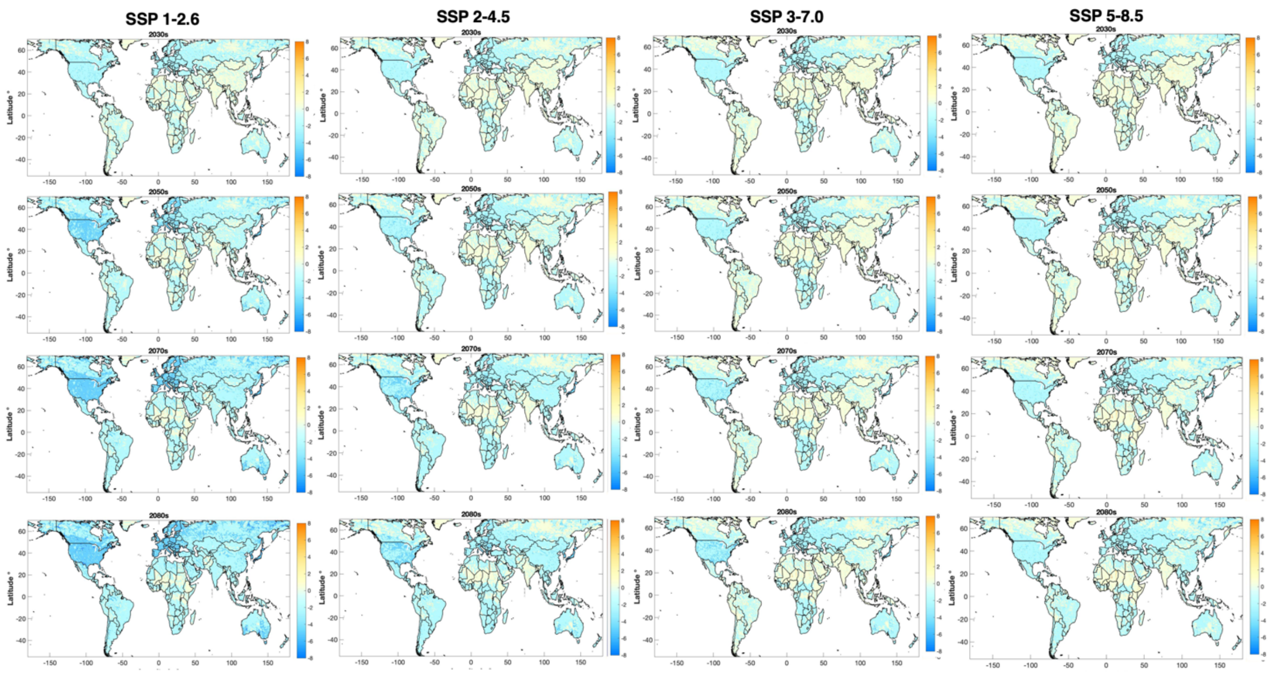


Figure S17. Changes in NOx emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

Minor comments:

<u>1. Reviewer #1 comment:</u> L25: Please spell out CI.

Authors' response: Thanks for your comment. We have spelled out the CI (Confidence Interval) in the manuscript (Line 377).

<u>2. Reviewer #1 comment:</u> L63: "and emissions" or "emissions"? Do you need a "and" before "emissions"?

Authors' response: We are sorry for the typo, and it has been corrected.

<u>3. Reviewer #1 comment:</u> L97: What are the "primary spatial resolutions"?

Authors' response: To avoid misunderstanding, we have revised the "primary spatial resolutions" to "original spatial resolutions". Also, we have added more information about the original spatial resolutions in the manuscript. (Line 113–115)

"Before training the deep learning model, all meteorological and emissions data were re-interpolated from their original spatial resolutions (meteorological data with 0.625° × 0.500° and emission data with 0.5° × 0.5°) into the same grid as the surface $PM_{2.5}$ data with a resolution of 0.1° × 0.1°."

<u>4. Reviewer #1 comment:</u> L152: please provide a little bit detail about the calculation of 8-fold cross-validation.

Authors' response: Thanks for your comment. We have added the description and calculations of 8-fold in Text S5 in the SI (Page 6).

"*Cross-validation is a resampling process used to evaluate machine/deep learning models on a limited sample of data,*[19, 20] *which has also been widely used in the verification of machine learning models for air pollution prediction.* [21, 22] *The core idea of cross-validation lies in dividing the dataset into several groups and averaging the results of multiple evaluations, thus thoroughly evaluating the proposed model and fully implementing all data for validation.*[17]

The procedure of 8-fold cross-validation is as follows.

1. *Full set of the input data is divided into 8 parts.*
2. *In the training process, each time, one part of the input data is withheld, and the remaining 7 parts are input to the CNN framework to train the model.*
3. *The correlation coefficient ($CORR_i$) and RMSEi are calculated by validation with the withheld one-part data.*
4. *Steps 2 and 3 are performed 8 times, and then achieve 8 times of $RMSE_i$ and $CORR_i$.*
5. *The final RMSE and CORR are calculated by averaging the 8 different $RMSE_i$ and $CORR_i$.*"

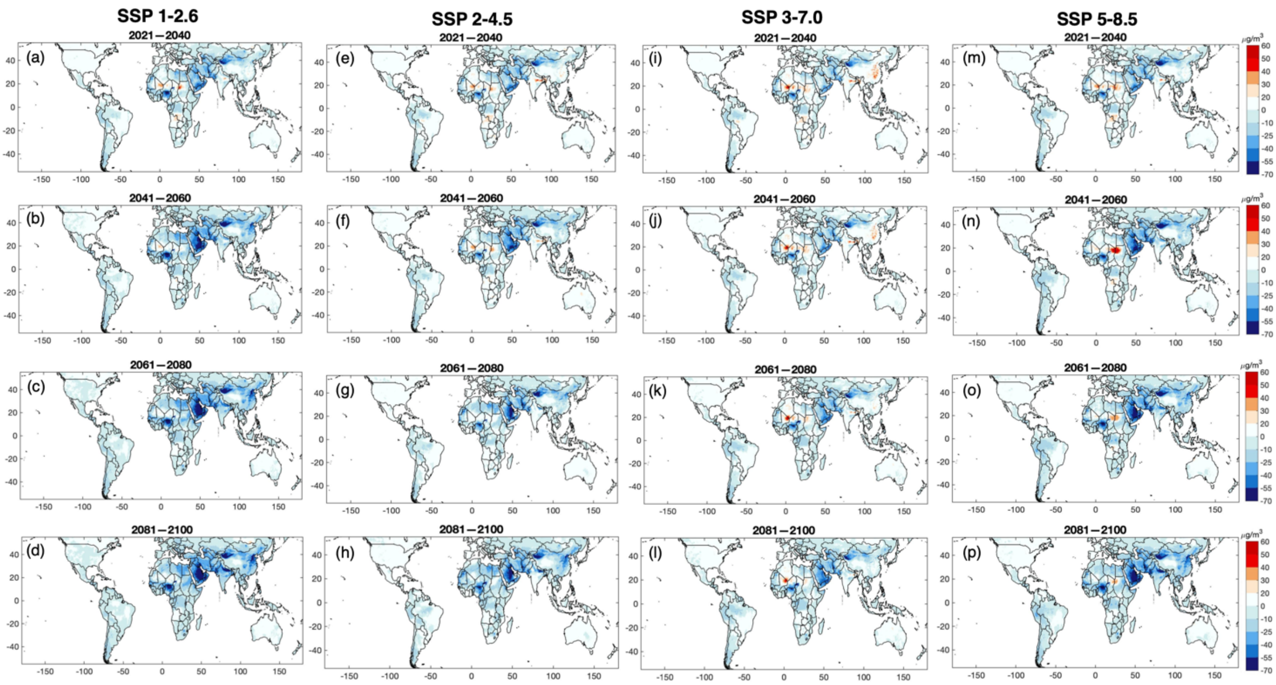
5. Reviewer #1 comment: L176: Figure2: What are the time period and scope of the validation presented in the figure?

Authors' response: Thanks for your comment. We have included the time period and scope in the caption of Figure 2.

"Figure 2. 8-fold cross-validation of the global $PM_{2.5}$ concentrations predicted by the U-Net CNN model during 1998-2014. The color represents the sample density."

6. Reviewer #1 comment: L195: Figure 3: Each panel subplot includes a letter label. It is necessary to include them in the figure caption.

Authors' response: Thank you for the suggestion. We have included a letter label for each panel subplot in this figure caption.



"Figure 3. Spatial distribution of changes in projected global $PM_{2.5}$ concentrations relative to the baseline period (2010–2019) under different climate change scenarios. Panels (a)-(d) represent the changes in $PM_{2.5}$ concentration for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) under SSP1-2.6 scenarios compared to the baseline condition. Panels (e)-(h) represent that of the same period but under SSP2-4.5. Panels (i)-(l) represent that of the same period but under SSP3-7.0. Panels (m)-(p) represent that of the same period but under SSP5-8.5."

Reference

1. Feng, L.; Smith, S. J.; Braun, C.; Crippa, M.; Gidden, M. J.; Hoesly, R.; Klimont, Z.; Van Marle, M.; Van Den Berg, M.; Van Der Werf, G. R., The generation of gridded emissions data for CMIP6. *Geoscientific Model Development* **2020,** *13*, (2), 461-482.
2. Arguez, A.; Vose, R. S., The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society* **2011,** *92*, (6), 699-704.
3. Arguez, A.; Durre, I.; Applequist, S.; Vose, R. S.; Squires, M. F.; Yin, X.; Heim, R. R.; Owen, T. W., NOAA's 1981–2010 US climate normals: an overview. *Bulletin of the American Meteorological Society* **2012,** *93*, (11), 1687-1697.
4. Navarro-Racines, C.; Tarapues, J.; Thornton, P.; Jarvis, A.; Ramirez-Villegas, J., High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific data* **2020,** *7*, (1), 1-14.
5. Ramanswamy, V.; Shine, K.; Leovy, C.; Wang, W.-C.; Rodhe, H.; Wuebbles, D. J.; Ding, M.; Lelieveld, J.; Edmonds, J. A.; Mccormick, M. P., Radiative forcing of climate. *NASA, Washington, Scientific Assessment of Ozone Depletion: 1991* **1991**.
6. Colette, A.; Bessagnet, B.; Vautard, R.; Szopa, S.; Rao, S.; Schucht, S.; Klimont, Z.; Menut, L.; Clain, G.; Meleux, F., European atmosphere in 2050, a regional air quality and climate perspective under CMIP5 scenarios. *Atmospheric Chemistry and Physics* **2013,** *13*, (15), 7451-7471.
7. Van Vuuren, D. P.; Stehfest, E.; Gernaat, D. E.; Doelman, J. C.; Van den Berg, M.; Harmsen, M.; de Boer, H. S.; Bouwman, L. F.; Daioglou, V.; Edelenbosch, O. Y., Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. *Global Environmental Change* **2017,** *42*, 237-250.
8. Kriegler, E.; Bauer, N.; Popp, A.; Humpenöder, F.; Leimbach, M.; Strefler, J.; Baumstark, L.; Bodirsky, B. L.; Hilaire, J.; Klein, D., Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century. *Global environmental change* **2017,** *42*, 297-315.
9. Van Vuuren, D. P.; Stehfest, E.; den Elzen, M. G.; Kram, T.; van Vliet, J.; Deetman, S.; Isaac, M.; Klein Goldewijk, K.; Hof, A.; Mendoza Beltran, A., RCP2. 6: exploring the possibility to keep global mean temperature increase below 2 C. *Climatic change* **2011,** *109*, (1), 95-116.
10. Riahi, K.; Rao, S.; Krey, V.; Cho, C.; Chirkov, V.; Fischer, G.; Kindermann, G.; Nakicenovic, N.; Rafaj, P., RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic change* **2011,** *109*, (1), 33-57.
11. Thomson, A. M.; Calvin, K. V.; Smith, S. J.; Kyle, G. P.; Volke, A.; Patel, P.; Delgado-Arias, S.; Bond-Lamberty, B.; Wise, M. A.; Clarke, L. E., RCP4. 5: a pathway for stabilization of radiative forcing by 2100. *Climatic change* **2011,** *109*, (1), 77-94.
12. Fujimori, S.; Hasegawa, T.; Masui, T.; Takahashi, K.; Herran, D. S.; Dai, H.; Hijioka, Y.; Kainuma, M., SSP3: AIM implementation of shared socioeconomic pathways. *Global Environmental Change* **2017,** *42*, 268-283.
13. Cha, Y. J.; Choi, W.; Büyüköztürk, O., Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering* **2017,** *32*, (5), 361-378.
14. Yamashita, R.; Nishio, M.; Do, R. K. G.; Togashi, K., Convolutional neural networks: an overview and application in radiology. *Insights into imaging* **2018,** *9*, (4), 611-629.
15. Chae, S.; Shin, J.; Kwon, S.; Lee, S.; Kang, S.; Lee, D., PM10 and PM2. 5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports* **2021,** *11*, (1), 1-9.
16. Huang, C.-J.; Kuo, P.-H., A deep CNN-LSTM model for particulate matter (PM2. 5) forecasting in smart cities. *Sensors* **2018,** *18*, (7), 2220.
17. Vanwinckelen, G.; Blockeel, H. In *On estimating model accuracy with repeated cross-validation*, BeneLearn 2012: Proceedings of the 21st Belgian-Dutch conference on machine learning, 2012; 2012; pp 39-44.
18. Pahikkala, T.; Boberg, J.; Salakoski, T. In *Fast n-fold cross-validation for regularized least-squares*, Proceedings of the ninth Scandinavian conference on artificial intelligence (SCAI 2006), 2006; Espoo Otamedia: 2006; p 90.

19. Allen, D. M., The relationship between variable selection and data agumentation and a method for prediction. *technometrics* **1974,** *16*, (1), 125-127.
20. Stone, M., Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)* **1974,** *36*, (2), 111-133.
21. Lu, X.; Yuan, D.; Chen, Y.; Fung, J. C.; Li, W.; Lau, A. K., Estimations of Long-Term nss-SO42–and NO3–Wet Depositions over East Asia by Use of Ensemble Machine-Learning Method. *Environmental Science & Technology* **2020,** *54*, (18), 11118-11126.
22. Chen, Y.; Yuan, D.; Chen, W.; Hu, M.; Fung, J. C.; Sun, H.; Lu, X., Estimation and variation analysis of secondary inorganic aerosols across the Greater Bay Area in 2005 and 2015. *Chemosphere* **2022,** *292*, 133393.
23. Jerrett, M.; Burnett, R. T.; Pope III, C. A.; Ito, K.; Thurston, G.; Krewski, D.; Shi, Y.; Calle, E.; Thun, M., Long-term ozone exposure and mortality. *New England Journal of Medicine* **2009,** *360*, (11), 1085-1095.
24. Janssen, N.; Hoek, G.; Harssema, H.; Brunekreef, B., Childhood exposure to PM10: relation between personal, classroom, and outdoor concentrations. *Occupational and environmental medicine* **1997,** *54*, (12), 888-894.
25. Lin, C.; Li, Y.; Lau, A. K.; Deng, X.; Tim, K.; Fung, J. C.; Li, C.; Li, Z.; Lu, X.; Zhang, X., Estimation of long-term population exposure to PM2. 5 for dense urban areas using 1-km MODIS data. *Remote sensing of environment* **2016,** *179*, 13-22.
26. Jenkins, P., Personal exposure to airborne particles and metals: results from the Particle TEAM study in Riverside, California. *Journal of Exposure Anahsis and Environmental Epidemiolog* **1996,** *6*, (1), 57.
27. Tie, X.; Zhang, Q.; He, H.; Cao, J.; Han, S.; Gao, Y.; Li, X.; Jia, X. C., A budget analysis of the formation of haze in Beijing. *Atmospheric Environment* **2015,** *100*, 25-36.
28. Du, C.; Liu, S.; Yu, X.; Li, X.; Chen, C.; Peng, Y.; Dong, Y.; Dong, Z.; Wang, F., Urban boundary layer height characteristics and relationship with particulate matter mass concentrations in Xi'an, central China. *Aerosol and Air Quality Research* **2013,** *13*, (5), 1598-1607.
29. Zheng, C.; Zhao, C.; Zhu, Y.; Wang, Y.; Shi, X.; Wu, X.; Chen, T.; Wu, F.; Qiu, Y., Analysis of influential factors for the relationship between PM 2.5 and AOD in Beijing. *Atmospheric Chemistry and Physics* **2017,** *17*, (21), 13473-13489.
30. Shi, L.; Liu, P.; Zanobetti, A.; Schwartz, J., Climate penalty: climate-driven increases in ozone and PM2. 5 levels and mortality. *Environmental Epidemiology* **2019,** *3*, 365.

Reviewer: 2

Comments:

The authors have developed a machine learning-based approach to predicting monthly average $PM_{2.5}$ concentrations at high spatial resolution (0.1 degree), trained on MERRA2 reanalysis fields, historical emissions from CMIP6, and $PM_{2.5}$ data based on satellite observations and GEOS-Chem simulations. This trained model is then applied to estimate monthly average $PM_{2.5}$ concentrations from 28 CMIP6 global climate models for each month from 2015 to 2100, for each of four SSPs. The resulting $PM_{2.5}$ concentrations are then used with future projections of population distributions to estimate $PM_{2.5}$-associated mortality. The subject is certainly of interest for ES&T. However, I cannot recommend this manuscript for publication in its current form.

Authors' response: We are grateful for the constructive comments. Substantial revisions have been made for the manuscript based on your comments. The U-Net model has been described and elaborated in more detail in the revised manuscript, accompanied by the replotting of the U-Net architecture with additional detailed information on model structure and parameters. More discussions of the model results, such as generalisation and standard error, have been added. In addition, the implications of the study have been further elaborated in the revised manuscript, with a focus on the economic impact of future premature mortality associated with $PM_{2.5}$. The point-by-point response to your questions are listed as follows. The page numbers and line numbers refer to the revised manuscript file with tracked changes.

1. Reviewer #2 comment: First, the most interesting and novel aspect of this analysis, and where the authors clearly have expertise, is the U-Net machine learning model. However, the description of this model is inadequate for me to follow, and I suspect inadequate for the majority of readers of this journal. The authors do provide many references, but in my opinion more details in this paper are warranted, at least in the Supplement. Figure 1 raises more questions for me than it answers, and the supplementary text S1 does not have much depth. Are w and s equal to 3600 and 1800 (representing 0.1 degrees longitude and latitude?). What value is used for the shuffle factor s? What is ADMM, and how does the learning rate enter into the equation S2? Did the authors use their own custom code, or

did they use existing machine learning packages? What is the reason for subtracting 2.4 ug/m^3 from the $PM_{2.5}$ concentration (Text S3)? Why 64 channels? The SI includes this unhelpful fragment: "For more details about these parameters" without referring the reader to where those details may be obtained.

Authors' response: Thank you very much for the comments.

In this work, we aimed to predict $PM_{2.5}$ concentration based on meteorological and emission data, as shown in Eq. (S1):

$$\boldsymbol{\mathcal{X}} \xrightarrow{f:U-Net} \boldsymbol{\mathcal{Y}} \qquad \text{Eq. (S1)}$$

where $\boldsymbol{\mathcal{X}} = (\boldsymbol{X_1}, \dots, \boldsymbol{X_T})$ represents the meteorological and pollutant emissions data over T months, $\boldsymbol{\mathcal{Y}} = (\boldsymbol{Y_1}, \dots \boldsymbol{Y_T})$ represents surface $PM_{2.5}$ concentrations in the corresponding T months, and f is the well-trained U-net CNN model.

The U-net CNN model is used to learn the relationship between meteorological variables, emissions ($\boldsymbol{\mathcal{X}}$) and $PM_{2.5}$ concentration ($\boldsymbol{\mathcal{Y}}$). CNN model we used is an image-based multilayer feedforward neural network. In this work, the meteorological and emission fields ($\boldsymbol{\mathcal{X}}$) and $PM_{2.5}$ concentration fields ($\boldsymbol{\mathcal{Y}}$) were treated as 2-D images. The geographic range of the meteorological fields, emissions, and global $PM_{2.5}$ concentrations in this study is 180°W to 180°E, 69.75°N to 54.75°S, with a resolution of 0.1° × 0.1°. Therefore, the size of 2-D images is equal to 3600×1246.

The architecture of U-net CNN consists of a contracting path to capture the context in images and a symmetric expansive path to expand the image size.[1] In the contracting path, convolutional layers and max-pooling layers are included. The convolutional layer, which can extract data features, is the key to CNN. The convolutional layer relies on convolutional nuclei to extract the features from the input (meteorological and emissions data) and the target ($PM_{2.5}$ concentration), and to obtain a complete feature map of the input and target data by sliding the convolutional kernel.[2] The max-pooling layer is mainly used for feature dimensionality reduction and also for improving the quality of the extracted information.[3] In the expansive path, up-sampling layers, convolutional layers, copy connection, and a final output layer are included. The up-sampling layer is designed to expand the image size so that the expanded features can be concatenated with feature maps from the corresponding max-pooling layer. The convolutional layers in the expansive path are used to combine the information from the contracting path (including high-level features extracted from the original image) and the information from the copy connections (including detailed features copied from the contracting path). After the above-mentioned operations, the U-net CNN model can mine the deep features implied between $PM_{2.5}$ concentration and meteorological and emissions data.

To clarify the architecture of the U-net model and explain the parameters in the supplementary information (SI), we replotted the model architecture and added more detailed descriptions of the U-net model in Text S2 (SI Page 2–3).

"In this study, the meteorological, emission fields, and $PM_{2.5}$ concentration fields are treated as 2-D images. The detailed architecture of our proposed U-Net for $PM_{2.5}$ projection is shown in Figure 1. There are mainly three parts in U-Net, i.e., contracting path, bottleneck, and expansive path. In the contracting path, the multichannel input images first went through two contraction blocks. Each contraction block, aimed to capture and aggregate the spatial features, is comprised of two convolution layers (represented as yellow cuboid) and one max-pooling layer (represented as orange cuboid). All the convolution layers share the same kernel size of 3×3 and utilize Rectified Linear Unit (ReLU) as the activation function. Max-pooling layers are adopted for adjusting the size of images in order to obtain better bottleneck information. After each block, the image size will be halved by using the max pooling layer with kernel size of 2×2, but the number of channels will be doubled. In the bottleneck part, combined with two convolution layers, the most crucial spatial information was automatically extracted and refined with the smallest intermediate images, which is crucial to ensure the accuracy of the final prediction. In the expansive path, each block consists of one up-sampling layers (represented as black cuboid) and three 3×3 convolution layers, and one copy connection layer (represented as blue cuboid). The up-sampling layer (with kernel size 2×2) is designed to expand the image size so that the expanded features could be concatenated with feature maps from the corresponding contraction layer. The copy connection is to ensure that the detailed features learned by contraction can be directly used for reconstruction. The arrows with different colors in Figure 1 indicate the information flow in different parts of the model. At the end of the expansion part, a final layer with kernel size 1×1 helps make the final prediction of the estimated $PM_{2.5}$ concentration."

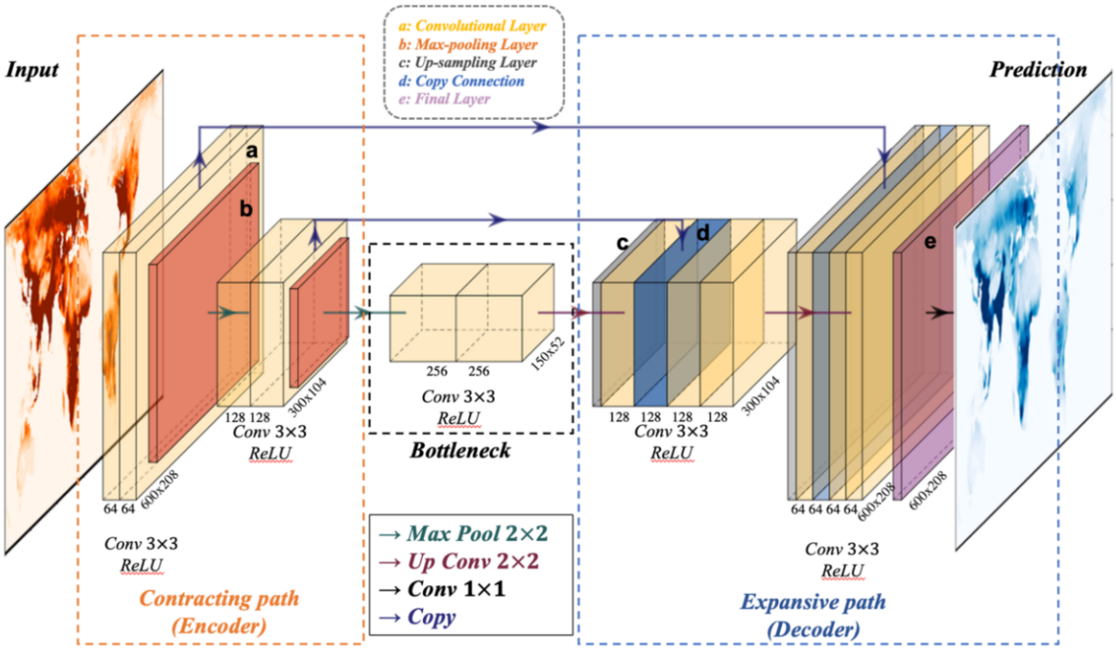


Figure 1. Architecture of the U-Net model

Are w and s equal to 3600 and 1800 (representing 0.1 degrees longitude and latitude?). What value is used for the shuffle factor s?

Authors' response: In this work, the meteorological and emission fields and $PM_{2.5}$ concentration fields are treated as 2-D images. Since the initial data of $PM_{2.5}$ surface concentrations are not available in the Arctic and Antarctic Circle, the geographic range of global $PM_{2.5}$ concentrations predicted in this study is 180°W to 180°E, 69.75°N to 54.75°S, with a resolution of 0.1° × 0.1°. Therefore, w (width) and h (hight) are equal to 3600 and 1246, respectively.

Directly inputting data with a 0.1° × 0.1° grid resolution would lead to a large computation load for deep learning. However, reducing the input size would cause inevitable information loss for the deep learning model, which would degrade the prediction performance. To overcome this barrier, the inverse pixel shuffle strategy was used to rearrange images to reduce the computational load without losing image information.

The steps of inverse pixel shuffle strategy are as follows. First, the original high-resolution images were separated into s^2 parts. Secondly, the inverse pixel shuffle rearranges the raw input with size $w \times h$ to form s^2 sample images with a size of $\frac{w}{s} \times \frac{h}{s}$. We set the s (shuffle factor) to 6, which means that the original input (3600 * 1246) is divided into 6^2 samples with the size of 600 × 208 by inverse pixel shuffle.

We have revised the descriptions in Text S2 (SI Page 4).

"*The inverse pixel shuffle strategy was used to reduce the computational load without information loss, as shown in Figure S1. In this way, the computational load was reduced, and the training data could be fully processed. For the inverse pixel shuffle operation with shuffle factor* s, *the raw input with size* $w \times h$ *was separated and reorganized into* s^2 *sample images with a size of* $\frac{w}{s} \times \frac{h}{s}$, *where* w *and* h *are equal to 3600 and 1246 (representing longitude 180°W to 180°E, latitude 69.75°N to 54.75°S with a resolution of 0.1° × 0.1°), and* s *set to 6.*"

What is ADMM, and how does the learning rate enter into the equation S2?

Authors' response: Thanks for your comments. We have clarified the ADMM and learning rate in Text S2 (SI Page 3–4).

"Alternating Direction Method of Multipliers (ADMM) algorithm is a widely used optimization method for the constrained problems in machine learning technique. The principle and detailed information of ADMM can be found in Boyd, et al. (2011) [4] From the perspective of solving constrained problems, this method is generated mainly to compensate for the disadvantages of quadratic penalties. In some problems, approximating constrained problems with quadratic penalties near the optimal point requires the coefficients of the penalty terms to converge to infinity, which will make the Heisenberg matrix very large, so the approximate objective function is very unstable. To solve this limitation, a linear approximation is introduced, in which the coefficients of the linear terms continuously approach the optimal solution (pairwise ascent), so that the solution can be obtained with the required accuracy even if the coefficients of the quadratic penalty terms are small. Moreover, from the perspective of solving distributed problems, ADMM decomposes the large global problem into several smaller and easier-to-solve local subproblems through the Decomposition-Coordination process. Boyd, et al. (2011) [4] have proved its applicability to large-scale distributed optimization

problems. Since ADMM is a mature and popular general framework for constraint optimization, we used the ADMM as the optimizer in this study.

The learning rate (0.001 in this study) is a tunable parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum loss function (Eq. S2). The learning rate is usually determined from the gradient of the loss function."

$$\min_{f} \sum_{t=1}^{T} \left(\boldsymbol{Y_t} - f(\boldsymbol{X_t}) \right)^2 \qquad \text{Eq. (S2)}$$

Did the authors use their own custom code, or did they use existing machine learning packages?

Authors' response: The U-net CNN model was built by using our own custom code in this study.

What is the reason for subtracting 2.4 ug/m^3 from the $PM_{2.5}$ concentration (Text S3)?

Authors' response: According to the GEMM model[5], 2.4 $\mu g/m^3$ represents the counterfactual $PM_{2.5}$ concentration. If the $PM_{2.5}$ concentration is lower than 2.4 $\mu g/m^3$, no adverse health impact is assumed. In other word, the health impact of $PM_{2.5}$ are only estimated above the counterfactual concentration, therefore the $PM_{2.5}$ concentration should be subtracted by 2.4 $\mu g/m^3$.

To explain the reason for subtracting 2.4 $\mu g/m^3$, we added more information in Text S6 (SI Page 7).

"*where z = max (0, $PM_{2.5}$ − 2.4 μg/m^3), 2.4 $\mu g/m^3$ represents the counterfactual $PM_{2.5}$ concentration. No adverse health impact is assumed when the $PM_{2.5}$ concentration is lower than this counterfactual concentration."*

Why 64 channels?

Authors' response: 64 channel is a popular model setting and it is based on the trial and error from previous studies.[6] In machine/deep learning research, it is customary to use the multiple of 2 as the channel number.

The SI includes this unhelpful fragment: "For more details about these parameters" without referring the reader to where those details may be obtained.

Authors' response: We are sorry for the typo and we have revised the sentence.

"*For more details about these parameters, please refer to Burnett et al. (2018)* [7]"

2. Reviewer #2 comment: Second, more discussion is warranted as to the capabilities of the model. The average results shown (Fig S2) are impressive in terms of how well they match the satellite data, but what is the standard error of the monthly estimates? How does the variability seen in the training set compare to the future conditions, with changes in emissions and meteorology? Do any of the CMIP models include atmospheric chemistry components, which could be used as an independent check on the U-Net model results? What accounts for the anomalous increases in $PM_{2.5}$ in isolated regions of central Africa (Fig. 3)? Are wildfire emissions included in the historical CMIP emissions data? They almost certainly are not included in the future SSP emissions data. The same is true for windblown dust. What are the implications of these limitations for the analysis?

Authors' response: We very much appreciate the helpful comments and have added more discussions on the capabilities of the model.

The average results shown (Fig S2) are impressive in terms of how well they match the satellite data, but what is the standard error of the monthly estimates?

Authors' response: Thanks for your question, if we understand correctly, the reviewer means the standard error of performance for monthly estimates. We have added the Root Mean Squared Error (RMSE), which is related to the standard error of the mean, in Table 1. Table 1 represents the 8-fold cross-validation for the monthly mean, where the standard errors of the relevant metric are included. Values in the first row indicate the means of performance metrics (NMB, NME, MB, MAGE, RMSE, and R) on the monthly estimates, and the values in the second row indicate the standard error of statistics (NMB, NME, MB, MAGE, RMSE, and R) of the monthly estimates. The standard error of performance metrics was used to evaluate the variability of performance metrics on monthly estimates.

Table 1. 8-fold cross-validation of U-Net CNN model performance

	NMB*	NME*	MB* ($\mu g/m^3$)	MAGE* ($\mu g/m^3$)	RMSE* ($\mu g/m^3$)	R
Average	−0.01	0.22	−0.05	1.36	4.02	0.987
Standard error	0.01	0.03	0.08	0.11	0.39	0.010

***NMB:** normalized mean bias; **NME:** normalized mean error; **MB:** mean bias; **MAGE:** mean absolute gross error; **RMSE:** Root Mean Squared Error

How does the variability seen in the training set compare to the future conditions, with changes in emissions and meteorology?

Authors' response: To elucidate the changes in the inputs, we plotted the data distribution of the training dataset and inputs under the four SSPs with the data from 2071-2100 as an example. We have described the dataset distributions in Text S1 (SI Page 2).

Text S1. Frequency distribution of current and future data

Figure S1 and Figure S2 show the frequency distributions of current and future meteorological and emission data. For the meteorological inputs in Figure S1, overall, the distributions for current and future scenarios are similar. For example, the PBLH within the ranges of 350m to 1200m appears in high frequency under current and future scenarios. The temperature and specific humidity under current and future scenarios are skewed to the right and contain similar distribution patterns.

For emissions inputs, since the difference between the maximum (up to 8×10^6 g km^{-2} yr^{-1}) and minimum (0 g km^{-2} yr^{-1}) values is very large, in order to be able to compactly display numerical data with such a wide range of values, we used logarithmic scale here to present the distribution. In Figure S2, the change in OC emission is the most noticeable and the peaks in future scenarios tend to shift towards higher values. However, overall, few significant changes were found in the shape of the distribution (peaks, symmetry, skewness, uniformity) for either meteorology or emission input. Therefore, although the values between current periods and future scenarios differ, the distribution variations are subtle and the U-Net deep learning model applied in this work has the capability to digest the data of future scenarios.

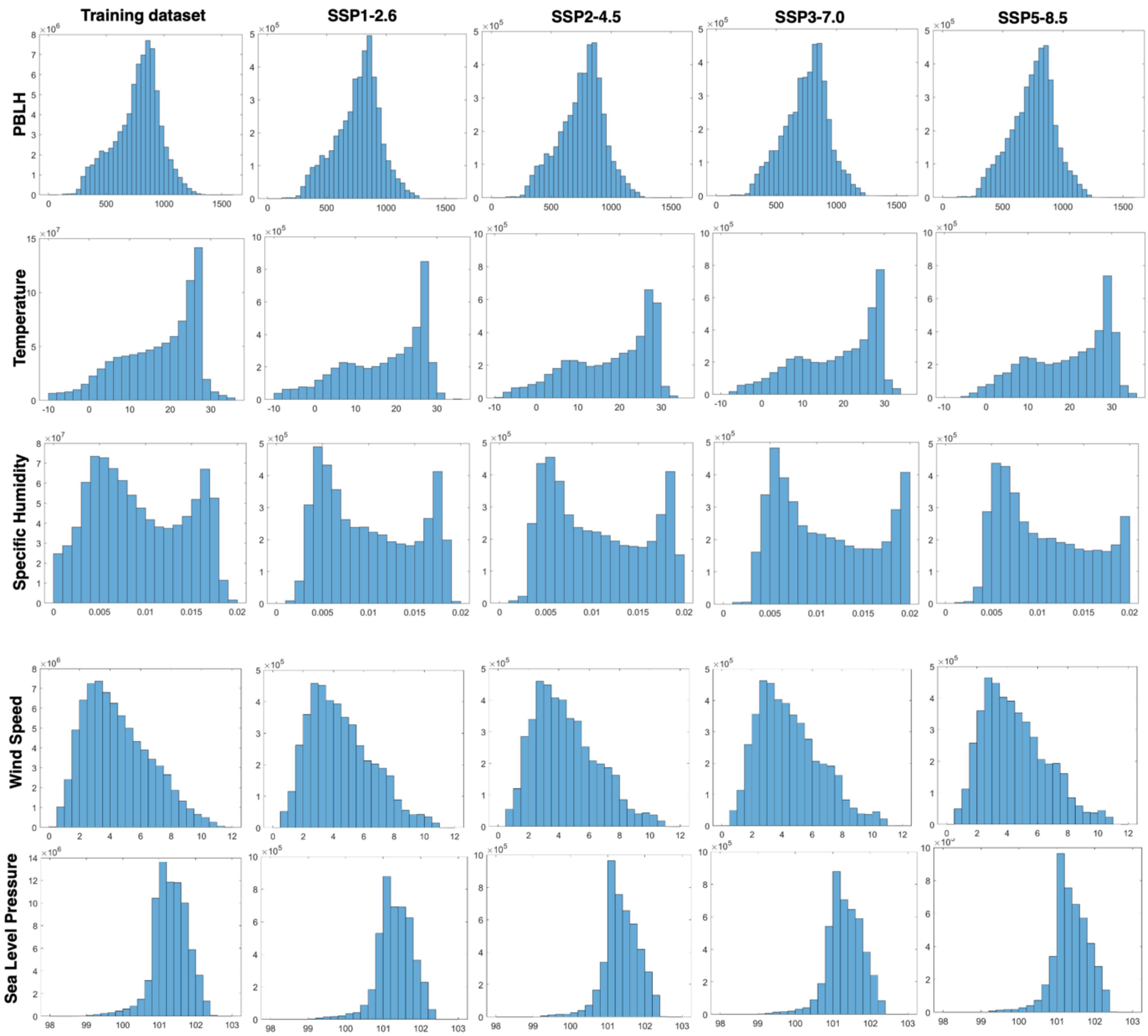


Figure S1. Frequency distribution of meteorological inputs for the four SSPs in the 2080s (2071-2100) compared with the training dataset (1998-2014). The x-axis corresponds to the value of meteorological inputs, and the y-axis represents the frequency counts falling in each interval. Units for the x-axis of PBLH, temperature, specific humidity, wind speed, and sea level pressure are m, °C, kg/kg, m/s, and kPa, respectively.

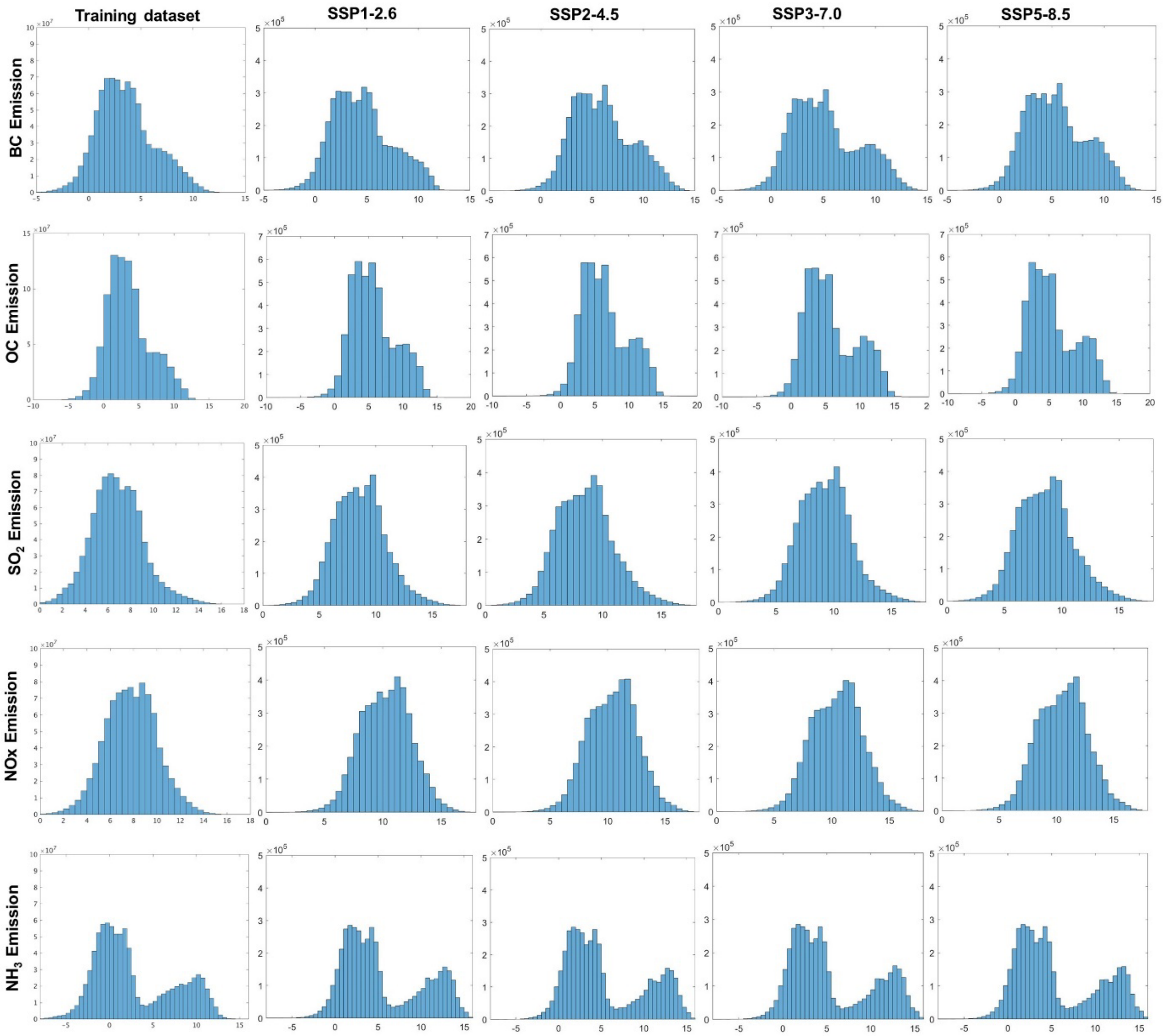


Figure S2. Frequency distribution of emission (g km^{-2} yr^{-1} in log scale) for the four SSPs in the 2080s (2071-2100) compared with the training dataset (1998-2014). The x-axis corresponds to the value of emission inputs in the log scale, and the y-axis represents the frequency counts falling in each interval. Units for the x-axis are g km^{-2} yr^{-1} in log scale.

Do any of the CMIP models include atmospheric chemistry components, which could be used as an independent check on the U-Net model results?

Authors' response: Thanks for your comments. Some CMIP6 models do include atmospheric chemical components of $PM_{2.5}$ (e.g., organic aerosol and sulfate), however, the dataset from some of models in CMIP6 platform do not contain $PM_{2.5}$. For those models that contain the $PM_{2.5}$ data, the calculations of $PM_{2.5}$ are inconsistent due to the different treatments of aerosols and their components. For example, only Goddard Institute for Space Studies ModelE2.1 (GISS-E2-1-H) and Geophysical Fluid Dynamics Laboratory's Earth System Model Version 4 (GFDL-ESM4) have provided nitrate mass mixing ratios. Therefore, the CMIP6 dataset does not contain the complete set of aerosol component data, and the aerosol data from this dataset cannot be used as an independent check for the U-Net model results.

In some of previous studies,[8, 9] future $PM_{2.5}$ concentration projection was generally based on a simple equation. Surface $PM_{2.5}$ was estimated as the sum of the mass of individual species of black carbon (BC), OA, sulphate (SO_4), sea salt (SS) and dust (DU). All BC, OA and SO_4 aerosol masses were assumed to be presented in the fine-size fraction (< 2.5 μm), while a factor of 0.25 for SS and 0.1 for DU was used to calculate the approximate contribution of these components to the fine particulate matter (Eq. A1).

$$PM_{2.5} = BC + OA + SO_4 + NH_4 + (0.25\times SS) + (0.1\times dust) \qquad \text{Eq. (A1)}$$

A comprehensive study has systematically compared simulated $PM_{2.5}$ concentrations based on Eq. A1 with observed surface $PM_{2.5}$ concentrations (obtained from ground-based observations and reanalysis products).[8] Compared to ground-based observations from the Global Aerosol Synthesis and Science Project database, the $PM_{2.5}$ concentration estimated from Eq. A1 by inputting the component data from CMIP6 dataset was underestimated by up to 10 $\mu g\ m^{-3}$. [10, 11]

Therefore, the $PM_{2.5}$ concentrations estimated by this equation will differ numerically from the actual concentrations and from our U-Net results, so it is not reliable to use the output from CMIP6 for the direct verification.

What accounts for the anomalous increases in $PM_{2.5}$ in isolated regions of central Africa (Fig. 3)?

Authors' response: Thank you for your comment. The increases in $PM_{2.5}$ in central Africa compared with the baseline values are mainly caused by the emission changes in these regions. To illustrate this issue more clearly, we have provided spatial emission maps for Africa. The anomalous increases in central Africa occurred in the SSP3-7.0 and SSP5-8.5 scenarios, so we compare emissions of SSP3-7.0 and SSP5-8.5 scenarios with those of the 1990s. As shown in Figure 3, $PM_{2.5}$ concentrations continued to exceed baseline values in central Africa (e.g., The Republic of Niger, The Republic of Mali, The Republic of Chad, and The Republic of Sudan) under SSP3-7.0 scenario until the end of the 21st century. This area is quite consistent with the emission-increasing region, as shown in Figures S5 and S6, which means that the emission change should be the major driver for the $PM_{2.5}$ concentration increase in Central Africa.

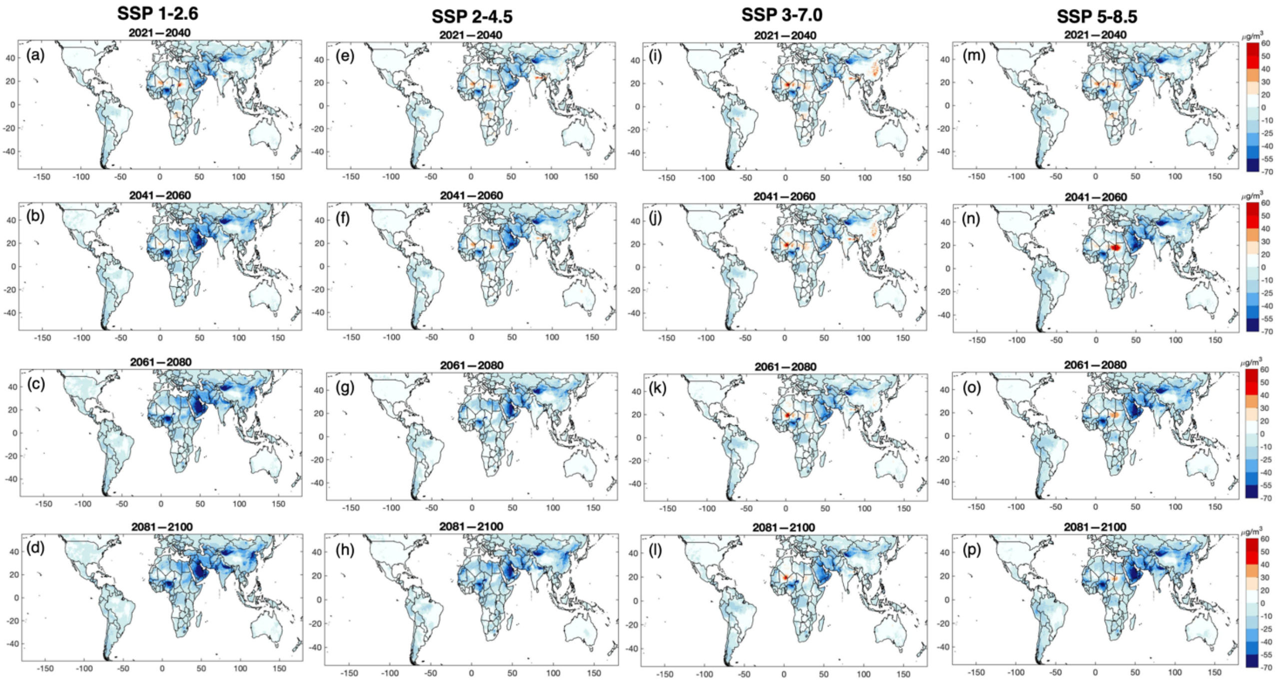


Figure 3. Spatial distribution of changes in projected global $PM_{2.5}$ concentrations relative to the baseline period (2010–2019) under different climate change scenarios. Panels (a)-(d) represent the changes in $PM_{2.5}$ concentration for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) under SSP1-2.6 scenarios compared to the baseline condition. Panels (e)-(h) represent that of the same period but under SSP2-4.5. Panels (i)-(l) represent that of the same period but under SSP3-7.0. Panels (m)-(p) represent that of the same period but under SSP5-8.5.

We have added Figure S5 and Figure S6 into the supplementary material and provided an explanation in the revised manuscript (Lines 295–298):

"$PM_{2.5}$ concentrations continued to exceed baseline values in central Africa under SSP3-7.0 and SSP5-8.5, which is quite consistent with the emission-increasing trends in the region as shown in

Figure S22 and S23, implying that the emission change should be the major driver for the $PM_{2.5}$ concentration increase in Central Africa."

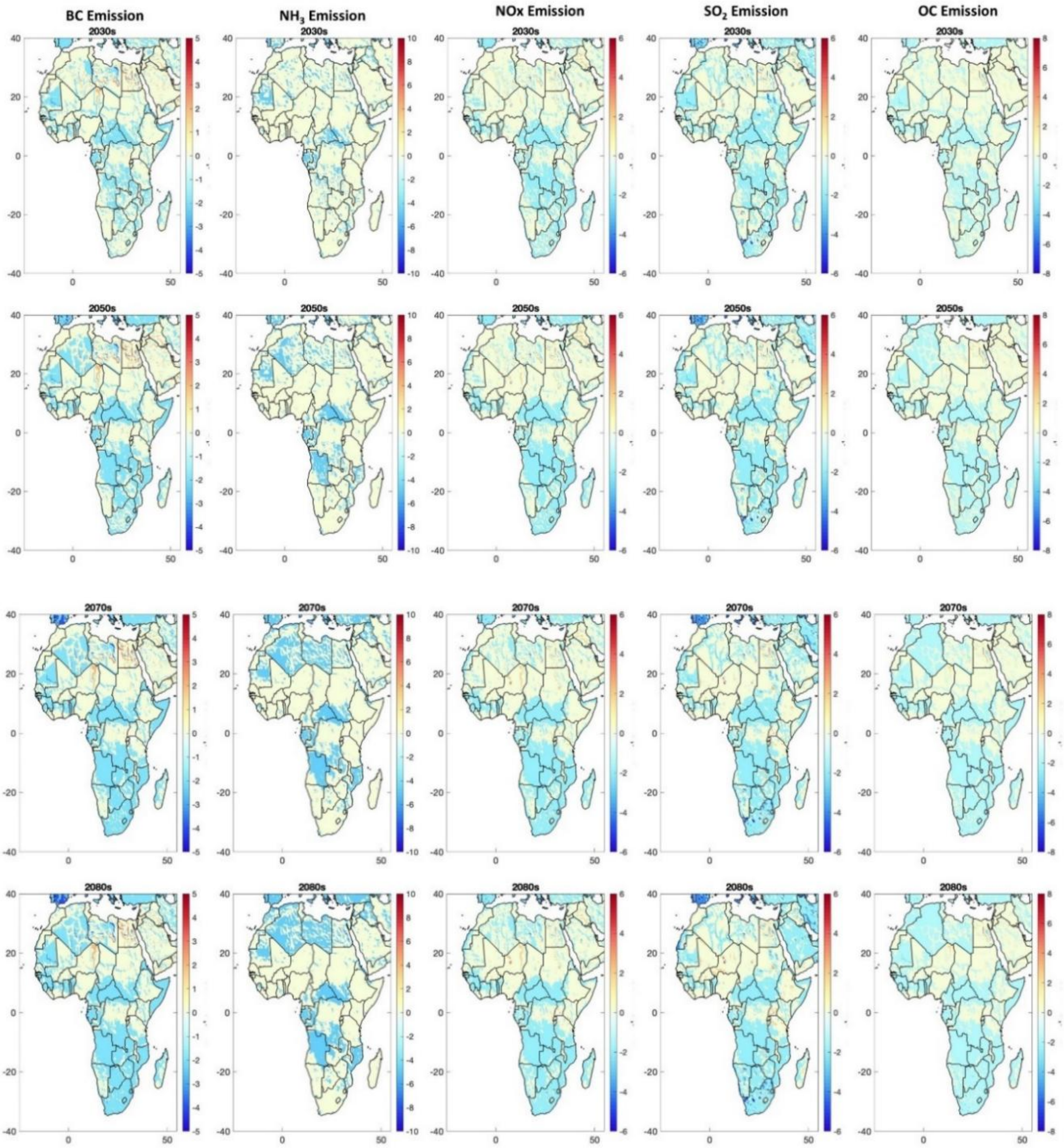


Figure S22. Changes in BC, NH_3, NOx, SO_2, and OC emissions (g km^{-2} yr^{-1} in log scale) under SSP3-7.0 scenario for 2030s (average of 2021-2050), 2050s (average of 2041-2070), 2070s (average of 2061-2090), and 2080s (average of 2071-2100) compared to that in 1990s (1981-2010).

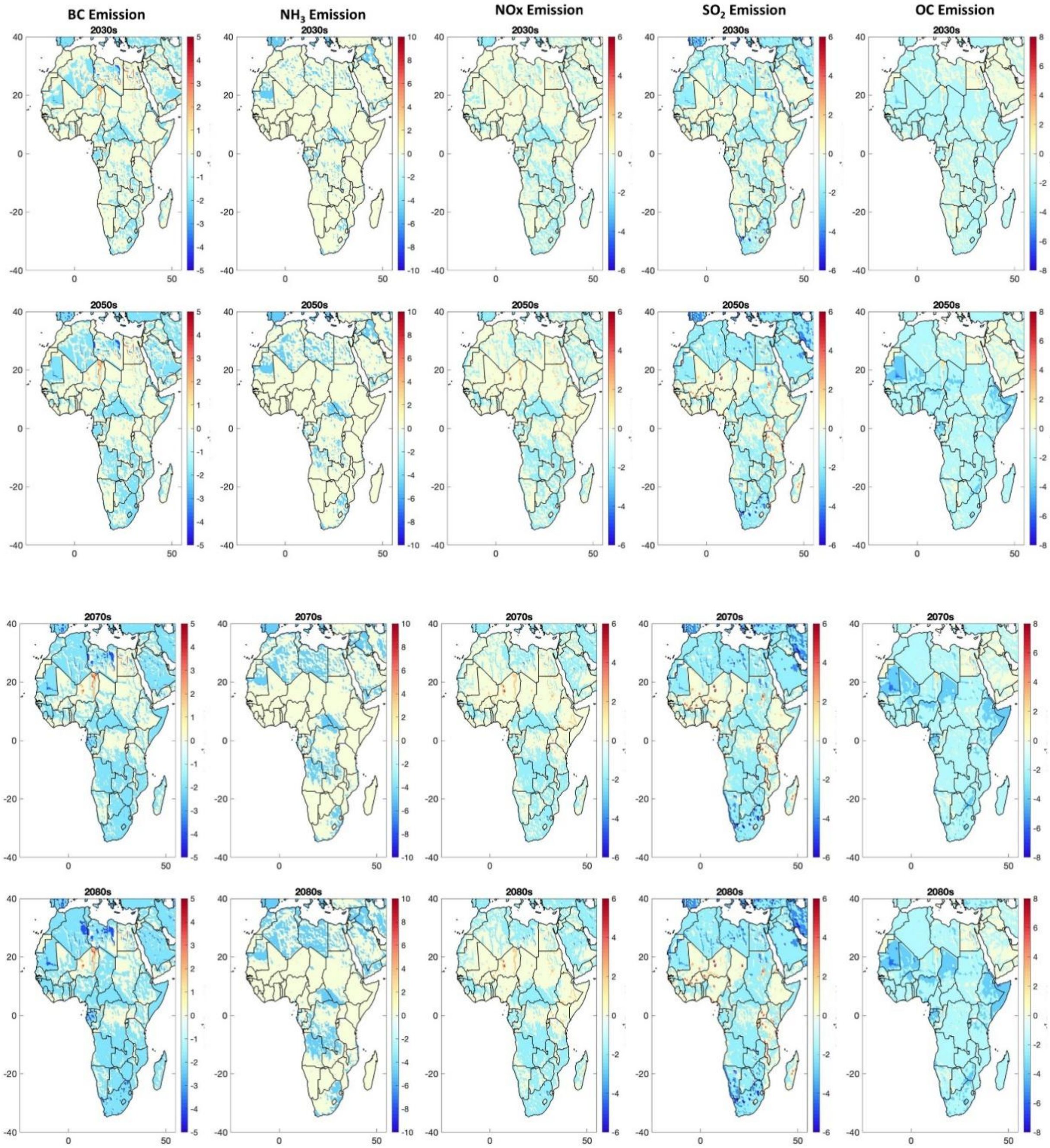


Figure S23. Changes in BC, NH_3, NOx, SO_2, and OC emissions (g km^{-2} yr^{-1} in log scale) under SSP5-8.5 scenario for 2030s (average of 2021-2050), 2050s (average of 2041-2070), 2070s (average of 2061-2090), and 2080s (average of 2071-2100) compared to that in 1990s (1981-2010).

Are wildfire emissions included in the historical CMIP emissions data? They almost certainly are not included in the future SSP emissions data. The same is true for windblown dust. What are the implications of these limitations for the analysis?

Authors' response: The wildfire emissions, defined as the emissions from forest, grassland, and peatland fires, along with agricultural waste burning (AWB) on fields, have been included in the historical CMIP6 gridded emissions dataset. The development of the historical open burning emissions in CMIP6 can be found in Marle et al. (2017).[12] CMIP6 also contains the future open burning emissions which are derived from integrated assessment models (IAMs). However, unlike historical open burning emissions, the spatial distribution of open burning emissions of a given category (e.g., forest burning) within each country does not vary in future scenarios.[13]

The wind-blown dust emission has not been included in this study due to the relatively coarse resolution (100km) and large uncertainties of dust emission in the CMIP6 dataset.[14] The deficiency of dust emissions may have an influence on our results, especially for the Sahara and Middle East, which are close to large sources of dust emissions. However, wind-blown dust has a limited impact on our health burden assessment due to the fact that these areas only contain small number of the population. In addition, a study conducted in Spokane found that particulate matter, composed mostly of dust, generally had limited association with premature mortalities.[15]

We have added more descriptions of CMIP6 emissions in the manuscript from Line 101 to Line 104.

"*Covering the period of 1750–2100 (historical dataset: 1750–2014, future emissions dataset: 2015–2100), CMIP6 gridded emissions dataset includes aviation emissions, all other anthropogenic emissions sectors, and total open burning emissions. This gridded dataset has previously been used for global model simulation and for emission scenarios comparisons.*[48, 49]"

For the implications of these limitations, we have discussed more in the revised manuscript (Lines 488–494).

"Finally, the biases in emissions data (e.g., bias in future wildfires and missing windblown dust), can be directly propagated to the air pollution concentration estimation. Thus, $PM_{2.5}$ projections in this work contain unavoidable uncertainty. The spatial pattern of windblown dust was not included in this study, which may have an influence on our results, especially for the Sahara and Middle East. Given

the proximity of these regions to large sources of dust emissions, there is a possibility that an underestimation of $PM_{2.5}$ concentrations would occur in these regions. However, the impact on the mortality estimations is limited since these regions are more sparsely populated."

3. Reviewer #2 comment: Finally, there is not much discussion as to the meaning or utility of these results. Ultimately, the authors conclude that SSP1-2.6 produces the lowest $PM_{2.5}$ concentrations and would lead to a significantly reduced mortality burden. This seems rather weak: policymakers and the scientific community do not need this paper to reach that conclusion. One possibility would be to extend the mortality burdens to consider costs and compare this to estimates of costs for controls needed to reach different SSP scenarios.

Authors' response: Thank you for your comments. In our revised manuscript, we have highlighted the meaning of our work. We have calculated the economic impact of $PM_{2.5}$-related mortality as suggested by the reviewer in Text S9. (SI Page 8–10)

Lines 442–453:

"*From the methodology and dataset perspectives, this work provides a new set of global-scale future $PM_{2.5}$ dataset in 10km spatial resolution. This dataset can be used by others for air quality-related studies at the national and even regional scales. The dataset can be downloaded from the link listed in Text S11. The method developed in this work can also be implemented for other air pollution-related research. Researchers can further develop other more advanced deep learning frameworks for relevant studies based on the design of the method proposed in this work.*

From the results perspective, this work has quantified how the future $PM_{2.5}$ and its associated adverse health impacts will change based on different SSP scenarios. Based on our results, governments and relevant stakeholders from different countries can generally understand to what extent can $PM_{2.5}$ influence their specific local health burdens. This can provide useful scientific references for future air pollution control policy design. In addition, when other studies come out in the future, the results from this work can also be used for the comparison. For example, compared to the adverse effects caused by other pollutants, such as O_3, which pollutant should the government put onto the priority position under different SSP scenarios."

We agree with the reviewer that it is also important to present the results from the economic loss perspective. We have calculated the economic burdens caused by future $PM_{2.5}$ pollution, as shown in Text S9. The health-economic impact has been discussed from Line 454 to Line 461 in the manuscript. "*From the health-economic impact perspective, the results of economic burdens shed light on the relationship between mortality cost that is associated with $PM_{2.5}$ pollution and economic development in various countries under different future scenario. The economic burdens related to future $PM_{2.5}$ pollution are discussed in Text S9. Figure S18 and Figure S19 show the economic loss that is associated with $PM_{2.5}$-related health burdens. SSP3-7.0 and SSP5-8.5 scenarios result in the lightest and heaviest economic loss, respectively. As shown in Figure S19, for most OECD countries, China, and Central Asia, air pollution mitigation and economic development can have a beneficial synergistic effect. The ratio of economic loss associated with $PM_{2.5}$ pollution to the total GDP (PPP based) is minimal in the sustainable development scenario (SSP1-2.6). For Central Africa and South America, from $PM_{2.5}$ associated economic loss perspective, these countries may consider choosing SSP2-4.5 pathway as their development modes.*"

Text S9. Health-economic impact

The cost of premature deaths is estimated from the value of statistical life (VSL). VSL is a measure of how much that individuals are willing to pay for a reduction in the risk or likelihood of premature death. [52] *This methodology has been applied by the World Bank and IHME (2016)*[53] *to estimate the economic loss that is associated with air pollution. The proposed benefit-transfer function is shown below:*

$$VSL_{c,n} = VSL_{OECD} \times \left(\frac{Y_{c,n}}{Y_{OECD}}\right)^{\in} \qquad \text{Eq. (S9)}$$

where $VSL_{c,n}$ *is the estimated VSL for country c in year n,* VSL_{OECD} *is the average base VSL in Organisation for Economic Co-operation and Development (OECD) countries ($3.83 million),* $Y_{c,n}$ *is Gross Domestic Product (GDP) per capita based on Purchasing Power Parity (PPP) in country c in year n,* Y_{OECD} *is the average GDP per capita based on PPP for the OECD countries ($37,000), and* $\in$ *is the income elasticity (1.2 for low- and middle-income countries and 0.8 for high-income countries).* $\in$ *is assumed as constant over the future scenarios in line with other literatures.*[16, 17]

The economic loss that is associated with the $PM_{2.5}$ pollution for country c in year n can be calculated as follow:

$$Cost_{c,n} = M_{c,n} \times VSL_{c,n} \qquad \text{Eq. (S10)}$$

The economic burden of premature death (Eco_{burden}) on the overall economy of the country can be calculated as follows:

$$Eco_{burden} = \frac{Cost_{c,n}}{GDP_{c,n}} \times 100\% \qquad \text{Eq. (S11)}$$

where $M_{c,n}$ is the $PM_{2.5}$-associated premature mortality, $GDP_{c,n}$ is the country's total GDP (PPP based). GDP (PPP based) for different future scenarios at national level can be obtained from https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=citation, which is produced by the OECD Env-Growth model.

As shown in Figure S31, for almost all of the representative regions, the SSP3-7.0 and SSP5-8.5 scenarios result in the lightest and heaviest economic costs, respectively. The global cost of $PM_{2.5}$-related mortality burdens will reach 8.9×10^7 (95% CI: $6.3\text{-}11.2\times10^7$) million USD by 2100 for the SSP5-8.5 scenario, but only 2.2×10^7 (95% CI: $1.5\text{-}2.7\times10^7$) million USD for SSP3-7.0. Even though SSP3-7.0 is the scenario with the highest premature mortalities, it has the lowest economic losses given that its income level is much lower than the other scenarios.

Figure S32 represents the ratio of economic loss caused by $PM_{2.5}$-related mortality to the total GDP (PPP based) for each country, indicating the extent to which can the $PM_{2.5}$-associated health burden affect the national economy. For most OECD countries, China, and Central Asia, air pollution mitigation and economic development can have a beneficial synergistic effect. In the sustainable development scenario (SSP1-2.6), the ratio of economic loss that is associated with $PM_{2.5}$ pollution to the total GDP (PPP based) is minimal. By 2100, the ratios of the economic burdens that are associated with $PM_{2.5}$ pollution to total GDP (PPP based) are lower than 6%, 12% and 8% for OECD countries, China, and Central Asia, respectively. However, as shown in Figure S32, the economic burdens for Central Africa and South America under SSP1-2.6 scenario are quite excessive when compared with the overall GDP (PPP based). By incorporating the factors of economic growth and economic loss associated with $PM_{2.5}$ pollution, these countries may consider choosing SSP2-4.5 pathway as their development mode.

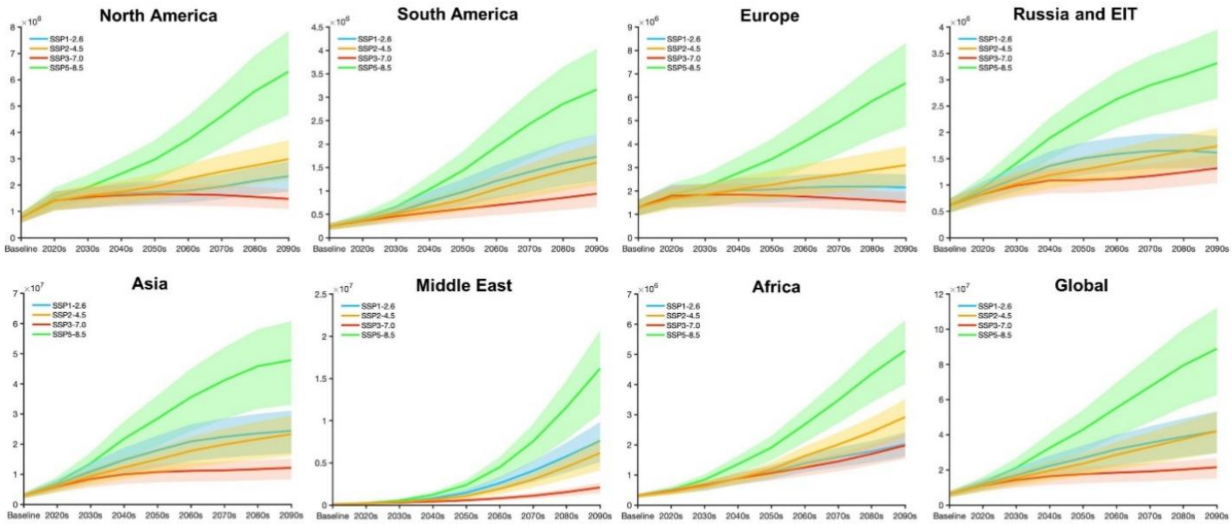


Figure S31. Economic loss caused by $PM_{2.5}$-related mortality over different regions around the world for SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios (Units: million US$).

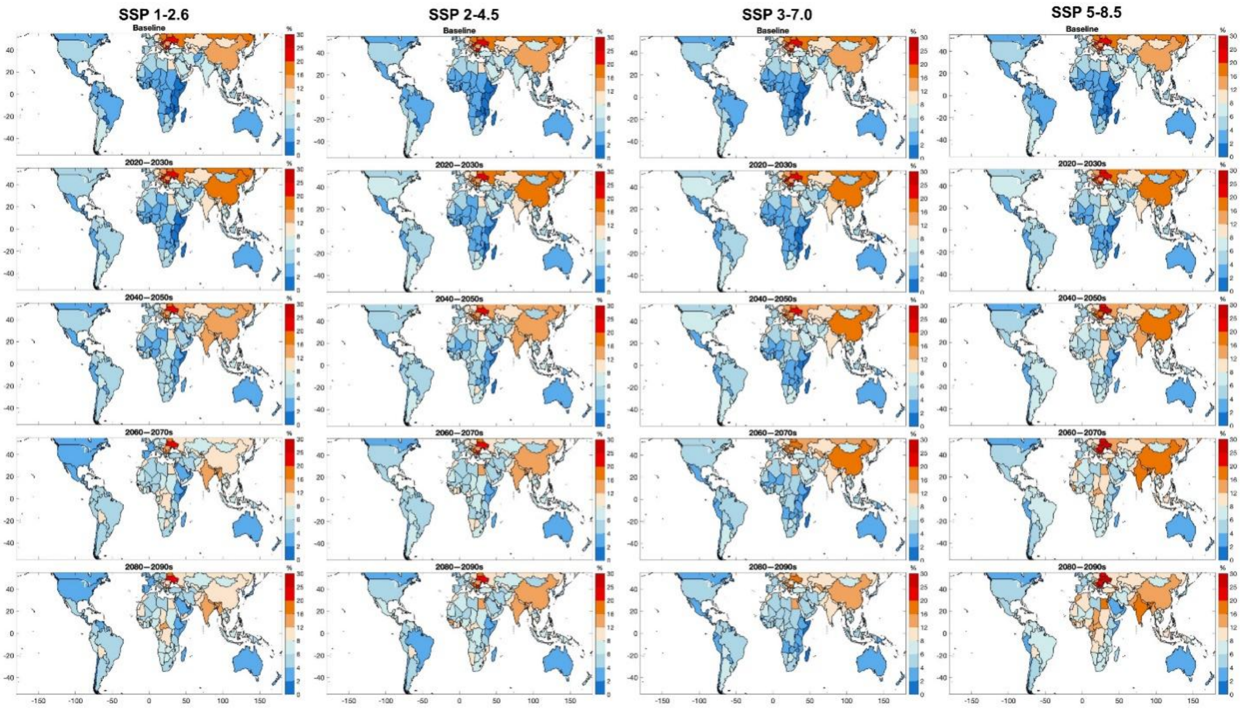


Figure S32. The ratio of economic loss caused by $PM_{2.5}$-associated premature mortality to the total GDP (PPP based) for each country. The first row represents the ratios in the 2010s. (Units: %).

Minor comments by line number

1. Reviewer #2 comment: Line 68 conscientious is not the right word here. Perhaps the authors mean "comprehensive".

Authors' response: Thank you very much. We have revised the word "conscientious" to "comprehensive". (Line 69)

2. Reviewer #2 comment: Line 116, 147, 154: "well-trained" should be just be "trained".

Authors' response: Thank you. We have corrected "well-trained" to "trained".

3. Reviewer #2 comment: 158: "the model gave a well-fitted in the areas": this phrase is missing a noun ("estimate" or "result"). Alternatively, "The results show that the model produced good fits in areas with both ...". Reading this a few times, I believe it is the case that this section refers to how well the model reproduces the training data. Are the values shown in Figure S2 averages across all months for all 17 years?

Authors' response: Thank you for your comments. We are sorry for the typo, and we have corrected the sentence in the manuscript. (Line 198–199)

"*The results show that the model produced good fits in the areas with both low ($\leq 35\ \mu g/m^3$) and high ($> 35\ \mu g/m^3$) $PM_{2.5}$ concentration.*"

Yes, Figure S2 presents the monthly comparison between satellite-retrieved $PM_{2.5}$ data and the average 8-fold cross-validation data predicted using the U-net convolutional neural network for the years of 1998-2014. We have revised the caption of Figure S2 (Figure S5 in the revised SI).

"Figure S5. Monthly spatial comparison between satellite-retrieved $PM_{2.5}$ data and the average 8-fold cross-validation data predicted using the U-Net convolutional neural network for the years of 1998-2014."

4. Reviewer #2 comment: 160-161: are these stated results ("all grid cells within +/- 12 ug/m3") true for all months, or just for the long-term averages?

Authors' response: Thank you for your comment. This sentence means the monthly average during the period of 1998 to 2014. We have revised the sentences in Line 199–201 and Figure S6–S7.

"*As demonstrated in Figures S6 and S7, the errors between the simulated and target monthly average $PM_{2.5}$ concentrations for all grid cells were within* $\pm 12\ \mu g/m^3$ *for 1998-2014. The monthly average relative errors specific to each country were within* $\pm$ *10%.*"

"***Figure S6. Absolute error between monthly satellite retrieved $PM_{2.5}$ and the data predicted by using U-net convolutional neural networks for 1998-2014***

Figure S7. Country-specific relative error between annual satellite retrieved $PM_{2.5}$ and the data predicted by using U-net convolutional neural networks"

5. Reviewer #2 comment: 165-167 There is no need to put these statistics in both the text and in Table 1. Four significant digits seems too many.

Authors' response: Thank you for the comment. We have shortened the significant digits in Table 1 (as shown below) and delated the statistics in the text.

We have revised the sentence to:

"*The statistical evaluation metrics (A1–A6 in the supplemental material) shown in Table 1 were further used to verify the model performance.*"

Table 1. 8-fold cross-validation of U-Net CNN model performance

	NMB*	**NME***	**MB*** **($\mu g/m^3$)**	**MAGE*** **($\mu g/m^3$)**	**RMSE*** **($\mu g/m^3$)**	**R**
Average	−0.01	0.22	−0.05	1.36	4.02	0.987
Standard error	0.01	0.03	0.08	0.11	0.39	0.010

***NMB:** normalized mean bias; **NME:** normalized mean error; **MB:** mean bias; **MAGE:** mean absolute gross error; **RMSE:** Root Mean Squared Error

6. Reviewer #2 comment: 167-168: how do you determine whether the model is or is not overfitted?

Authors' response: The model we proposed in this study can overcome the overfitting issue. First, from the model design perspective, we utilized the data augmentation and dropout layer in each block of the U-Net CNN structure. Data Augmentation is a technique to extend the training data by using flipping, rotating, scaling, etc., so that the generalization ability can be improved by increasing the diversity of learning samples.[18-20] In our work, the specific operations of data augmentation were to randomly rotate 2D-array images with a rotation angle ranging from -10 degrees to 10 degrees, and to flip 2D-array images with 50% probability randomly.

The dropout layer is able to reduce overfitting in the deep neural network model. Dropout strategy solves the overfitting problem from two main aspects.[21] One is to drop units from neural networks and average them randomly. The dropout regularization is to randomly discard units (together with their connections) from the neural network during training in each iteration. When different sets of units are dropped, it is equivalent to training a different neural network. Thus, the dropout procedure is like averaging the effects of a large number of different networks. This strategy is often effective in preventing overfitting problems. [22, 23] The second is to reduce the complex co-adaptation relationship between different units. The dropout regularization allows hidden units to be randomly omitted from the network, so that a hidden unit cannot depend on the existence of other hidden units. In this way, the update of weights no longer relies on the co-adaptation relationship between different units, which forces the network to learn more robust features. If our neural network targets to make predictions, it should not be too sensitive to some specific cue fragments, and it should be able to learn some common patterns from many other cues (robustness).[24, 25] Srivastava et al.(2014)[21] compared the performance of dropout with other techniques and found that the dropout strategy could reduce the generalization error effectively, a measure of how accurately an algorithm is able to predict results for data not previously involved in training.

In addition, as shown in Figure S4, the monotonic decrease in training and validation loss shows that our model is not overfitted.

The overfitting issue and generalization ability of our model have been discussed in the manuscript from Line 135 to Line 137.

"The data augmentation and dropout regularization have been applied to improve the model generalization ability, as discussed in Text S7. And the monotonic decreases in training and validation loss (Figure S4) have proved that no overfitting was detected."

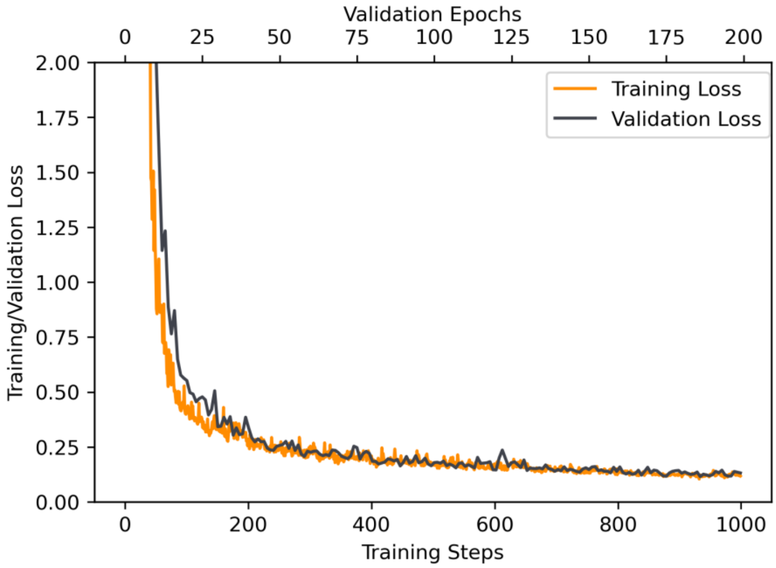


Figure S4. Illustration of training loss and validation loss.

Text S3. Generalization ability of deep learning model

In this work, data augmentation and regularization technique were used to improve the model generalization. Data Augmentation is a technique to extend the training data by using flipping, rotating, scaling, etc., so that the generalization ability can be improved by increasing the diversity of learning samples and increasing the difficulty of learning samples. [18-20] *Taylor and Nitschke (2018)*[26] *conducted a comparative study on the effectiveness of geometric and photometric (color space) transformations, showing that flipping, rotating, and color jittering can improve the accuracy of the results by 2.86%, 4.91%, and 2.68%, respectively. In our work, the specific operations of data augmentation were to randomly rotate 2D-array images with a rotation angle ranging from -10 degrees to 10 degrees, and to flip 2D-array images with 50% probability randomly.*

Compared with data augmentation, the dropout strategy is a type of regularization technique. The dropout layer within the model is able to reduce the overfitting and thus improve the generalization ability in the deep neural network model. Srivastava et al.(2014)[21] *compared the performance of the dropout method with other techniques and found that the dropout strategy could reduce the* *generalization error effectively, a measure of how accurately an algorithm is able to predict results for data not previously involved in training. Thus, dropout regularization has proven to be successful in reducing overfitting.*

Equipped with data augmentation and regularization, our network can predict the $PM_{2.5}$ concentration in the future scenario better, even if the data in the future scenario are slightly different from the training input data.

7. Reviewer #2 comment: 188-193: Is there enough variability in the test dataset (2015-2019) to represent changes to 2100? Alternatively, are the changes in meteorological and emissions inputs under the SSPs outside the range of the training data?

Authors' response: Thanks for your comment. This is a good point and requires a lengthy explanation as described below. On the one hand, our results are based on the underlying assumption that the relationship between $PM_{2.5}$ concentrations and other independent variables (e.g., meteorological factors and emission) in the historical period is also suitable for future scenarios. In other words, this study assumes that the current atmospheric physical and chemical mechanisms also hold for future scenarios.

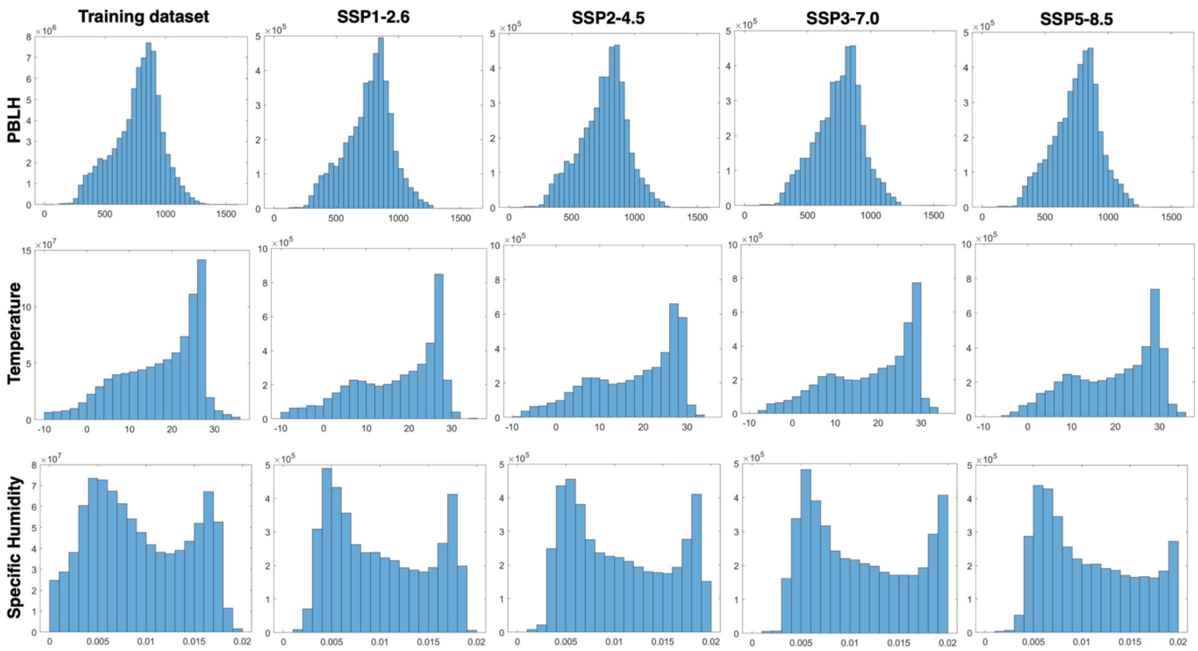
On the other hand, deep learning models have strong generalization ability, which refers to the ability to adapt to new data with similar distributions as the ones used to create the models. Dinal et al. (2017)[27] and Kawaguchi et al. (2017)[28] provide theoretical insights into why and how deep learning can generalize well on previously unseen data. To improve the model generalization, we applied data augmentation and dropout regularization in this work, as discussed in Text S3. Equipped with data augmentation and regularization, our network can predict the $PM_{2.5}$ concentration in the future scenario better, even if the data in the future scenario are slightly different from the training input data.

We plotted the data distribution of the training dataset and inputs for the period of 2071-2100 under the four SSPs as an example. We have described the dataset distributions in Text S1.

Text S1. Frequency distribution of current and future data

Figure S1 and Figure S2 show the frequency distributions of current and future meteorological and emission data. For the meteorological inputs in Figure S1, overall, the distributions for current and *future scenarios are similar. For example, the PBLH within the ranges of 350m to 1200m appears in high frequency under current and future scenarios. The temperature and specific humidity under current and future scenarios are skewed to the right and contain similar distribution patterns.*

For emissions inputs, since the difference between the maximum (up to 8×10^6 g km^{-2} yr^{-1}) and minimum (0 g km^{-2} yr^{-1}) values is very large, in order to be able to compactly display numerical data with such a wide range of values, we used logarithmic scale here to present the distribution. In Figure S2, the change in OC emission is the most noticeable and the peaks in future scenarios tend to shift towards higher values. However, overall, few significant changes were found in the shape of the distribution (peaks, symmetry, skewness, uniformity) for either meteorology or emission input. Therefore, although the values between current periods and future scenarios differ, the distribution variations are subtle, and the U-Net deep learning model applied in this work has the capability to digest the data of future scenarios.



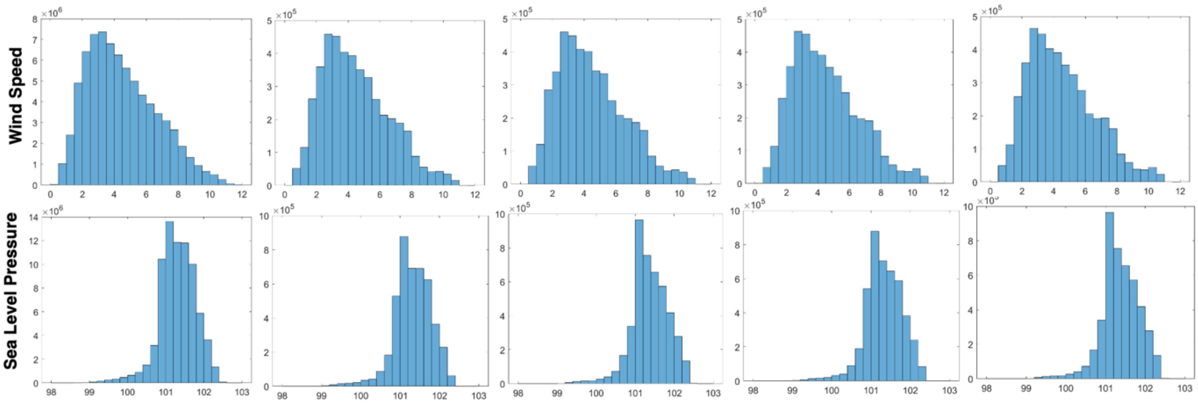


Figure S3. Frequency distribution of meteorological inputs for the four SSPs in the 2080s (2071-2100) compared with the training dataset (1998-2014). The x-axis corresponds to the value of meteorological inputs, and the y-axis represents the frequency counts falling in each interval. Units for the x-axis of PBLH, temperature, specific humidity, wind speed, and sea level pressure are m, °C, kg/kg, m/s, and kPa, respectively.

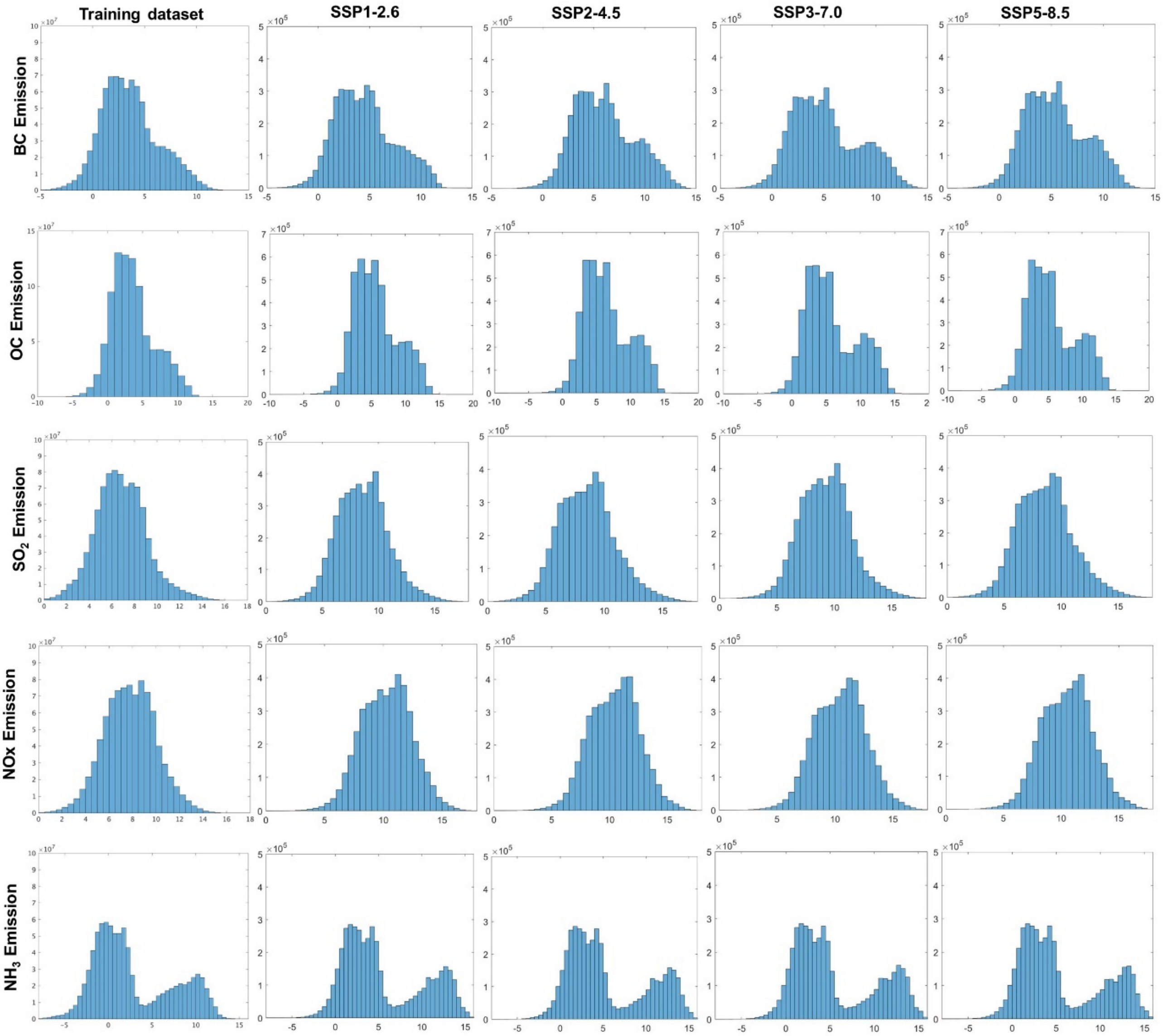


Figure S4. Frequency distribution of emission (g km^{-2} yr^{-1} in log scale) for the four SSPs in the 2080s (2071-2100) compared with the training dataset (1998-2014). The x-axis corresponds to the value of emission inputs in the log scale, and the y-axis represents the frequency counts falling in each interval. Units for the x-axis are g km^{-2} yr^{-1} in log scale.

8. Reviewer #2 comment: 208-209 This study is using future climate and emissions projections to estimate future air quality. Both emissions and climate are changing. In order for the authors to attribute the increases in $PM_{2.5}$ at mid-century under SSP5-8.5 specifically to "climate change", they should show that emissions changes are not affecting the concentration changes in central Africa.

Authors' response: We agreed with the reviewer that both future climate and emissions will change under the SSP5-8.5 scenario, leading to the changes in $PM_{2.5}$. We have revised the description in the manuscript. (Line 292–293)

"*Thus, in the middle of the 21st century, climate and emission change would considerably increase $PM_{2.5}$ concentrations and cause considerable damage to human health in central Africa.*"

<u>9. Reviewer #2 comment:</u> 213-216 This sentence does not make sense. What asymmetry is being referred to, and how is it disproportionate? "Intimidation" is clearly not the word the authors intend, but I am unable to tell what is intended.

Authors' response: We apologize for causing the misunderstanding.

The meaning we intended to convey is as follows. Due to the urbanization process, residents usually gather in the city where $PM_{2.5}$ concentration is relatively high.[29, 30] Since a large percentage of people are living in urban areas, using the average $PM_{2.5}$ concentration of rural and urban areas cannot reveal the exact population exposure level.

We have revised the sentences in the manuscript from Line 300 to Line 304 and added the population density (persons/km^2) in the SI (Page 31).

"*SSPs narratives gave rise to spatial and temporal differences in the demographic projections. The projected population density (persons/km^2) and the corresponding variations (compared to the situation in the 2010s) in four SSP scenarios are shown in Figure S26. Combined with demographic projections, the exposure concentration can be estimated and used to assess the $PM_{2.5}$ exposure associated health impacts.*[31, 32]"

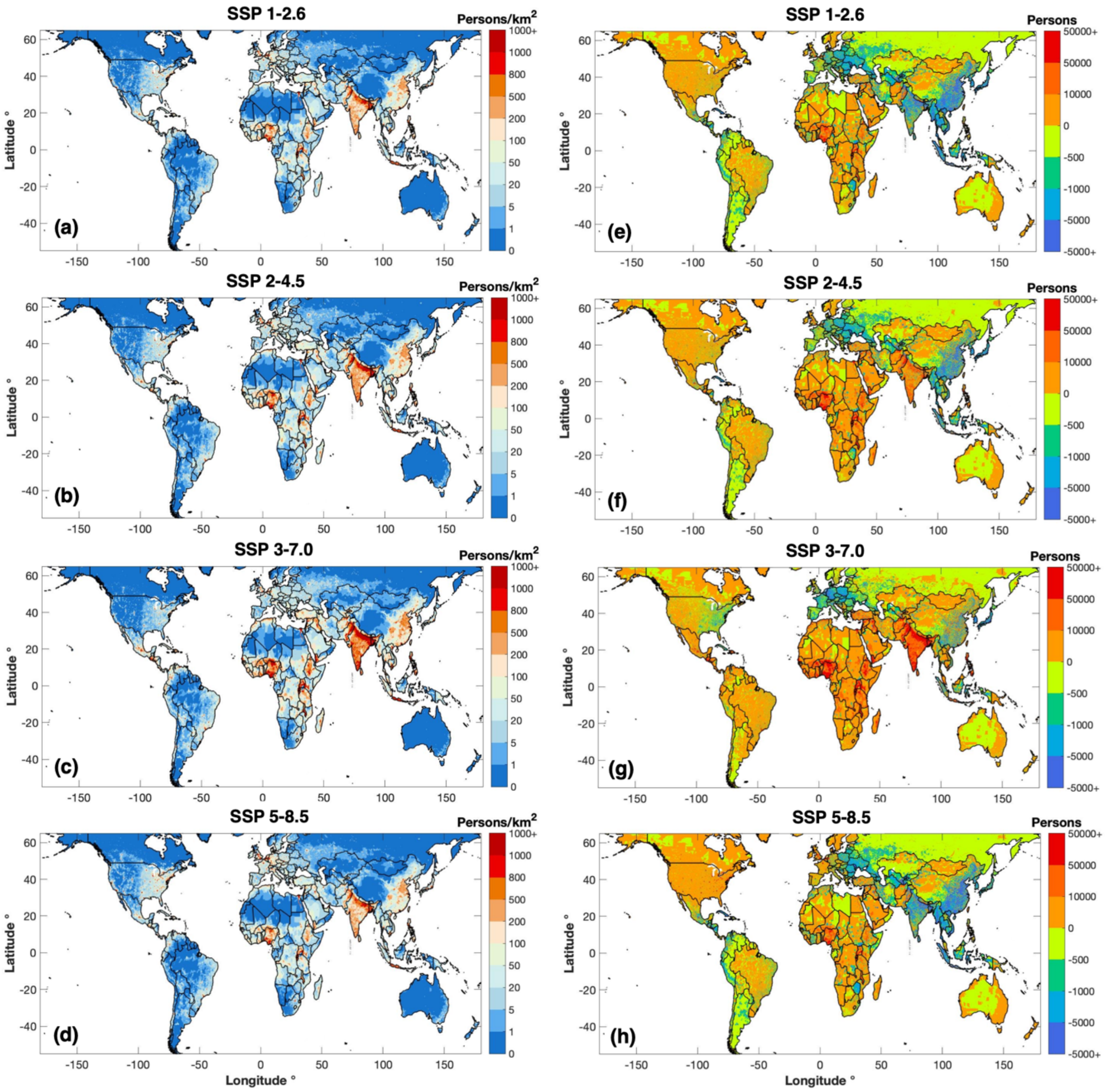


Figure S26. (a)–(d) Projected population density (Persons for 25+ years/km^2) for the four SSPs in the 2091-2100, and (e)–(g) corresponding projected population change compared with that in 2010s (2010-2019 average).

<u>10. Reviewer #2 comment:</u> 217-230 and Fig 4: I believe the authors are using the word "exposure" to mean population-weighted $PM_{2.5}$ concentrations. If so, please state this is the case. The authors use the term "population-weighted" on line 225 but it is not clear whether this is intended to be synonymous with the earlier definition of "exposure".

Authors' response: Thanks for your comment. Yes, the "population-weighted" is used to represent "exposure" in this study. To clarify, we have added the description and calculation results of "exposure" in Text S5. (SI Page 6)

"*Population-weighted concentration can better reflect the impact of $PM_{2.5}$ pollution on the exposed population, and thus build a more reliable theoretical basis for public health assessment. Eq. (S5) and Eq. (S6) were used to calculate the exposure concentration ($\boldsymbol{C_{PM_{2.5}}^{i-pop}}$) and spatially averaged ambient levels ($\boldsymbol{C_{PM_{2.5}}^{i-space}}$) of $PM_{2.5}$, respectively.*

$$\boldsymbol{C_{PM_{2.5}}^{i-pop} = \left(\sum_{k=1}^{N} C_k^i * P_k^i\right) / \sum_{k=1}^{N} P_k^i} \qquad \textit{Eq. (S5)}$$

$$\boldsymbol{C_{PM_{2.5}}^{i-space} = \frac{1}{N}\left(\sum_{k=1}^{N} C_k^i\right)} \qquad \textit{Eq. (S6)}$$

Here, different countries and the total grid number of the country are represented by i and N, respectively. C_k^i and P_k^i are the $PM_{2.5}$ concentration and corresponding population of grid k in country i, respectively."

<u>11. Reviewer #2 comment:</u> 223-224 "Space-weighted" should be "Spatially-averaged". More importantly, the claims that $PM_{2.5}$ concentrations are lower under SSP5-8.5 than under SSP2-4.5, and that this is due to tighter pollution controls in the former, are counterintuitive. My understanding is that air pollutants are often co-emitted with CO_2 during combustion, so that scenarios in which CO_2 emissions are reduced tend to have lower air pollutant emissions also. Please expand on this point, or provide a reference, or possibly additional figures in the SI showing changes in emissions of the species considered in your model.

Authors' response: Thanks for your comments. We have corrected the word "space-weighted" to "spatially-averaged" in the revised manuscript.

It should be noted that unlike the general greenhouse gases (GHG), such as CH_4 and CO_2, aerosols and their precursors (BC, OC, NH_3, NOx, SO_2) are short-lived climate forcers (SLCFs) pollutants, with lifetimes ranging from minutes to weeks. Therefore, the impacts of SLCFs on radiative forcing are relatively limited. In contrast, these SLCF pollutants are often tied to regional air quality and can be effectively controlled. Decisive and rapid action to address SLCF pollutants will have an immediate impact on improving air quality. As described in the IPCC AR6 report, in the SSP scenarios, SLCF emissions trajectories are steered by different levels of air pollution controls, independently from climate change mitigation (i.e., GHG mitigation).[33] Thus, the emission trajectories of SLCF in the SSPs scenario are independent of CO_2 emissions. In the SSP5-8.5 scenario, extremely stringent pollution control policies will be implemented for SLCFs control.[34]

SSP5-8.5 shows a transition to less polluting fuels and technologies, leading to a rapid and sustained reduction in emission intensities of air pollutants. SSP2-4.5, on the other hand, emphasizes large-scale electrification and modest technology advances. As illustrated by Rao et al.(2017), SSP2 shows a continued decline in all pollutants throughout the century, while SSP5 exhibits a faster decline due to more effective pollution controls resulting in the lowest emissions of air pollutants in the second half of the century. [34]

We have further added a summarized table of key differences among the four SSP scenarios relevant to air pollution in Table S5 in the supplemental material. We have also provided below figures (Figure S13 – Figure S17) to reveal the emission changes under different scenarios.

Table S5. Descriptions of the critical elements for SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 relevant to air pollution

Scenario	Description	Land-use Change Regulation	Environmental Development	Pollution Control Policy	Energy Tech Change	Population Growth
SSP1-2.6	Sustainability	Strong regulation to avoid environmental tradeoffs	High environmental awareness; more inclusive development that respects perceived environmental boundaries	Strong	Directed away from fossil fuels, toward efficiency and renewables	Relatively Low
SSP2-4.5	Middle-of-the-road	Medium regulation; slow decline in the rate of deforestation	Medium environmental awareness; work toward but make slow progress in achieving sustainable development goals	Medium	Some investment in renewables but continued reliance on fossil fuels	Medium
SSP3-7.0	Regional rivalry	Limited regulation; continued deforestation	Low environmental awareness; strong environmental degradation in some regions	Weak	Slow tech change, directed toward domestic energy sources	High in developing countries; low in industrialized countries
SSP5-8.5	Fossil-fueled development	Medium regulation; slow decline in the rate of deforestation	High environmental awareness; local environmental problems are successfully managed	Strong	Directed toward fossil fuels; alternative sources not actively pursued	Relatively Low

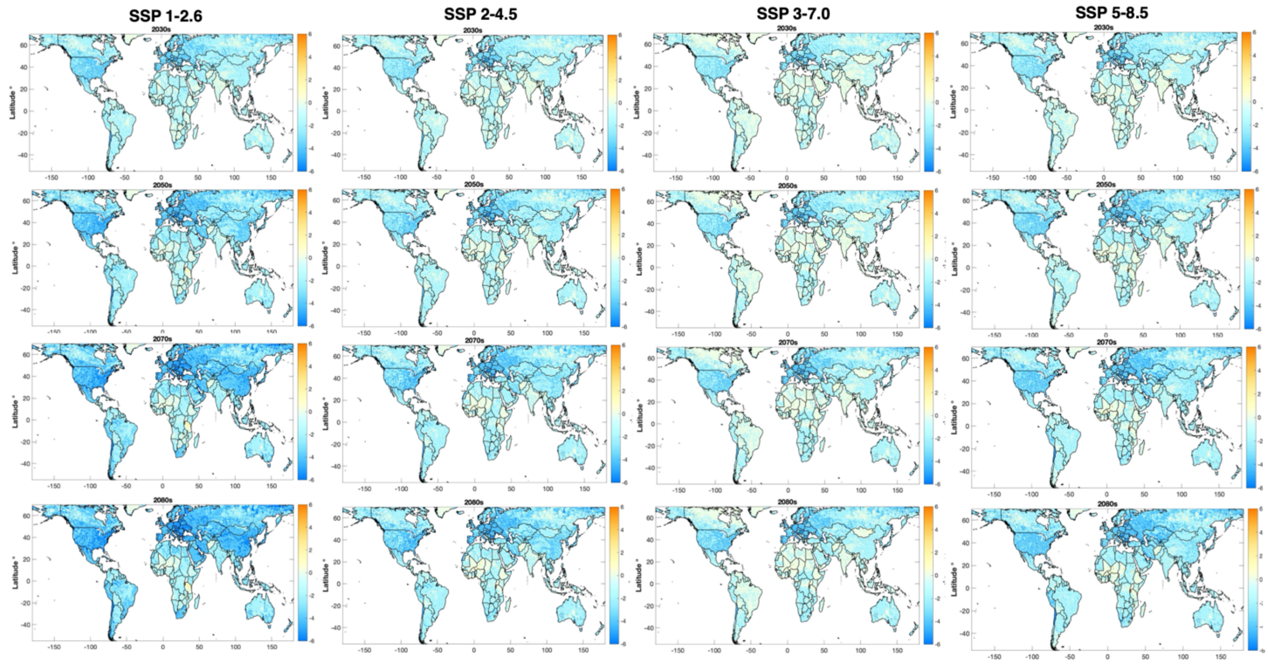


Figure S13. Changes in SO_2 emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

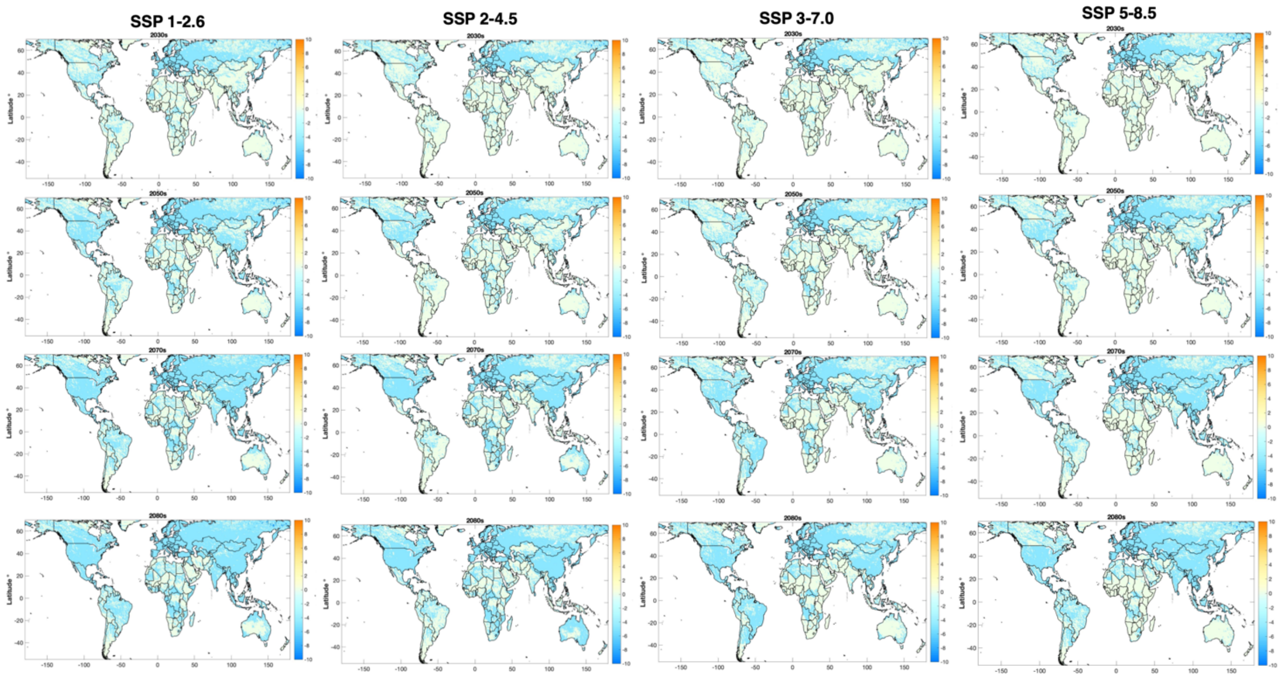


Figure S14. Changes in NH_3 emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

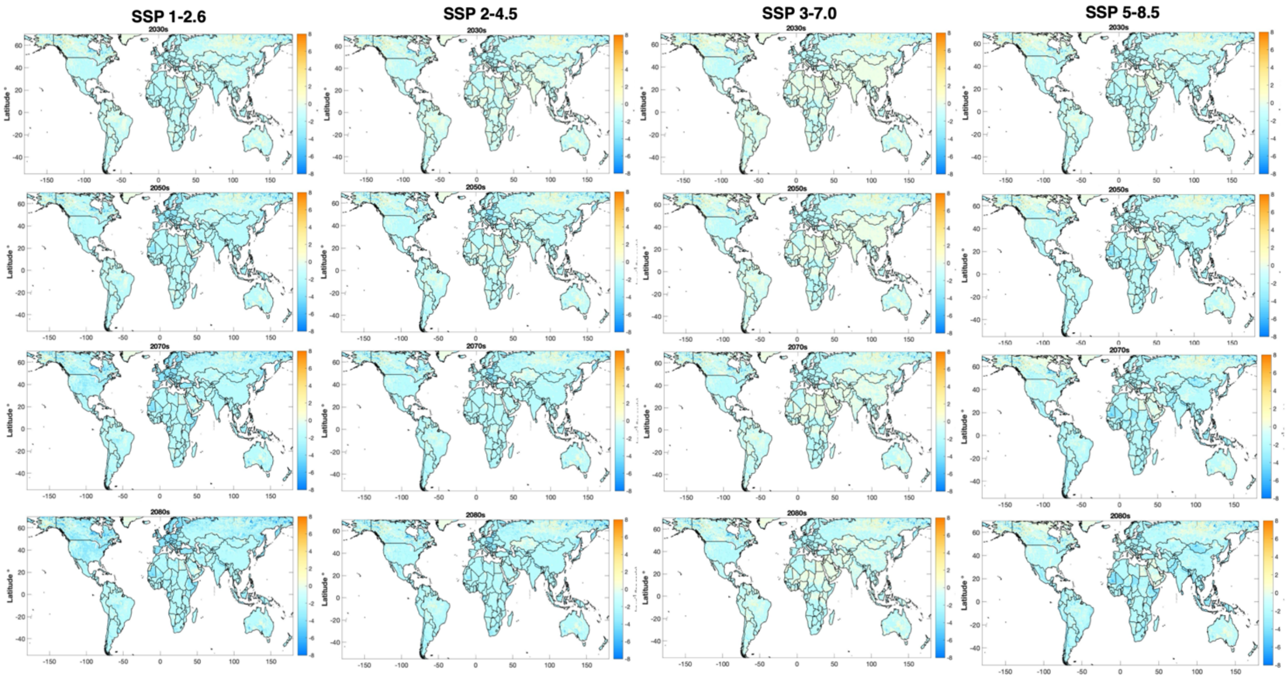


Figure S15. Changes in OC emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

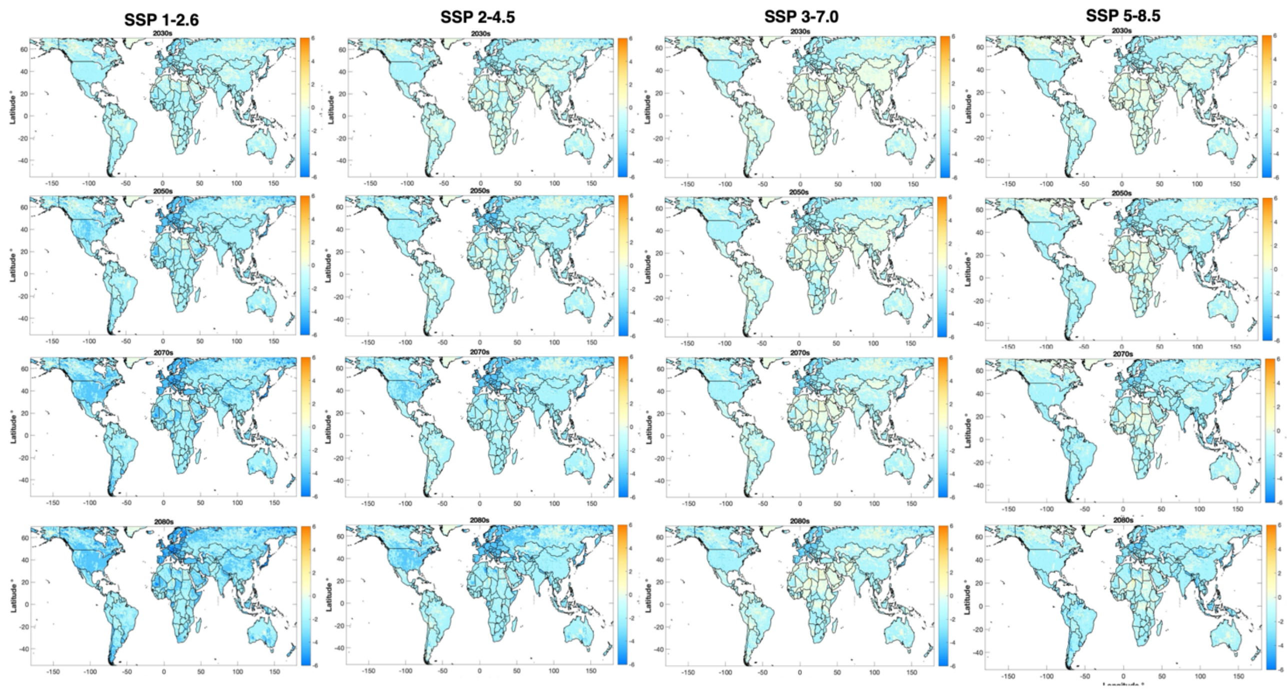


Figure S16. Changes in BC emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

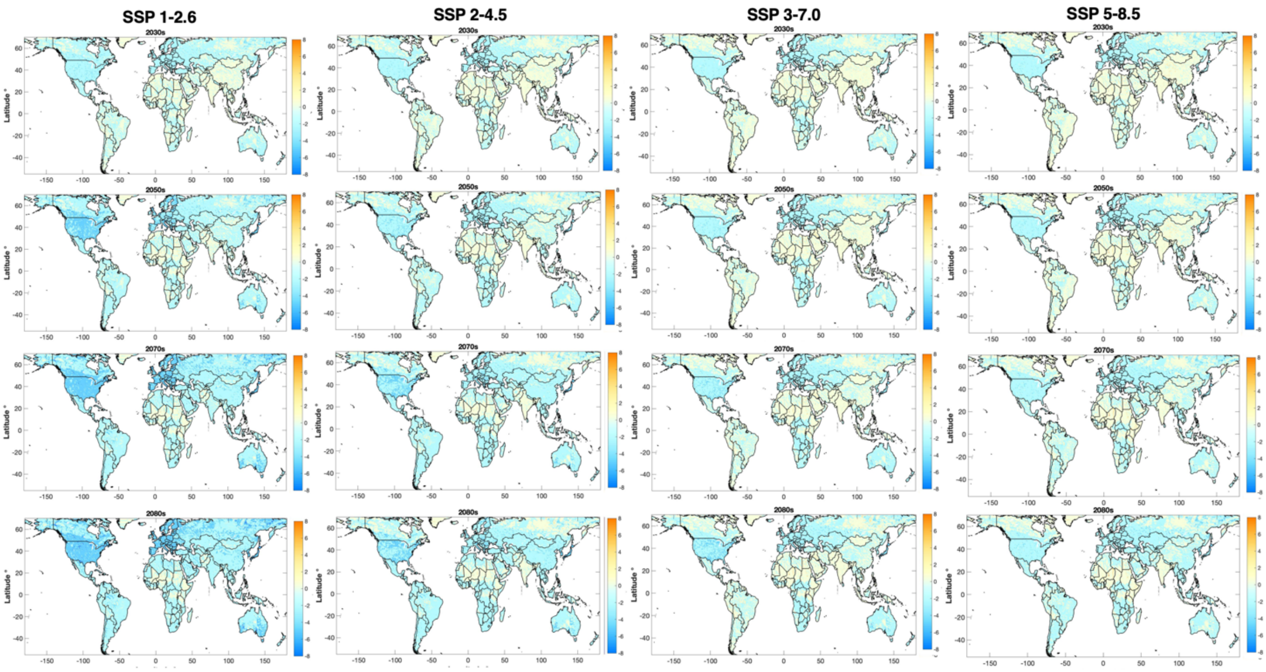


Figure S17. Changes in NOx emission (g km^{-2} yr^{-1} in log scale) for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010) under the four scenarios.

12. Reviewer #2 comment: 239-247 The tenor of this paragraph is strange, especially "emerge victorious". The differences between scenarios are significant, even if they are not very different in terms of the fraction of the population meeting the WHO's stringent 5 ug/m3 AQ guideline.

Authors' response: We agree with the reviewer, and we have reworded the paragraph and emphasized the difference between scenarios. (Lines 339–346)

"The proportion of the population that would be exposed to the $PM_{2.5}$ concentration below previous and current Air Quality Guideline (AQG) values is also estimated under future climate change scenarios. As shown in Figure S29, the differences between SSP1-2.6 scenario and the other three scenarios are considerable. Compared with the other three scenarios, SSP1-2.6 would result in the largest fraction of the population exposed to the $PM_{2.5}$ level that is lower than 5 μg/m³. In the SSP1-2.6 scenario, 3.5% of the world's population will live in areas that have $PM_{2.5}$ concentrations lower than 5 μg/m³ by 2100, which is well above the baseline (2.0%). The other scenarios are comparable

in terms of the proportion of the population exposed to the $PM_{2.5}$ concentration that is below the current AQG values."

<u>13. Reviewer #2 comment: </u>254-255 The text and the caption to Fig S19 indicates that the first row shows baseline PM2.5 mortality burden for 2010s, but the titles of the plots suggests otherwise. The caption to Fig 5 says "premature mortality rate"; a mortality rate usually means "mortalities per 100k" or similar. If instead the authors intend "rate" to mean "per year", then please state that.

Authors' response: We apologize for the mistake and Figure S19 has been revised (Figure S30 in the revised SI). We also corrected the caption for Fig. 5.

"Figure 5. $PM_{2.5}$-associated premature deaths (> 25 years old) in different regions. The red bars represent premature deaths, and the vertical black lines indicate the 95% empirical confidence intervals."

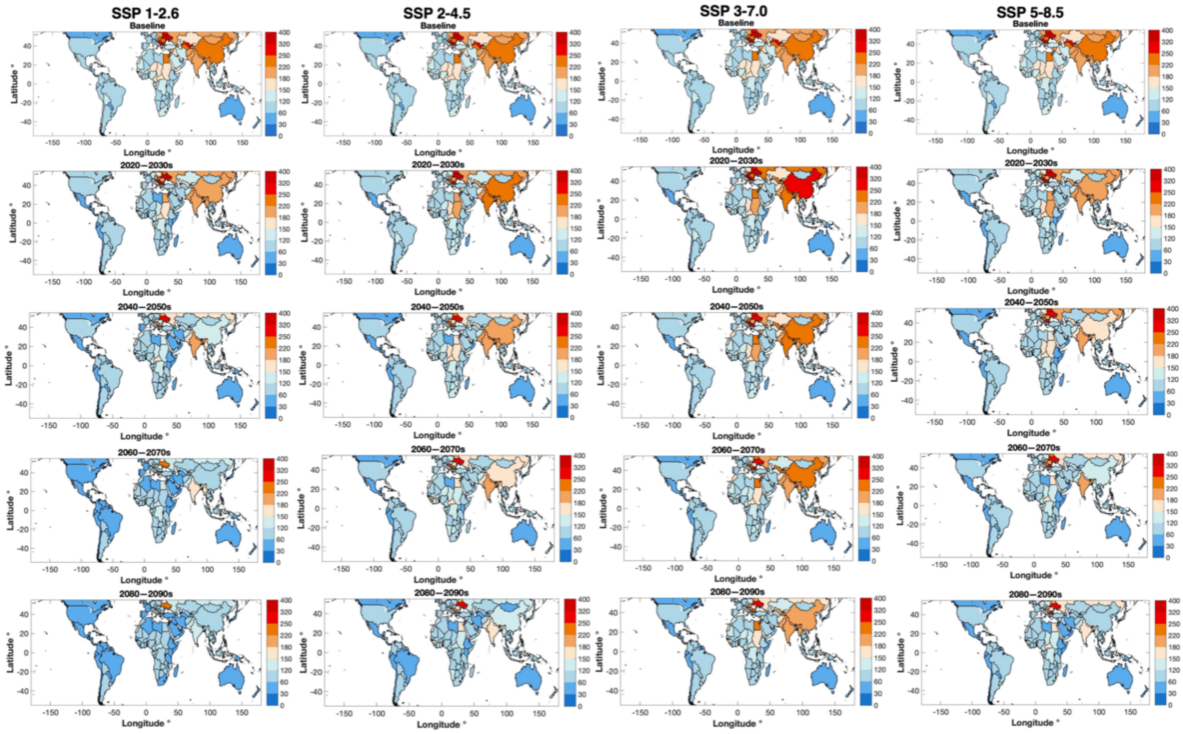


Figure S30. Projected premature mortality rate (per 100,000 population) in 184 countries under four SSP scenarios. The first row represents the baseline mortality rate in the 2010s.

14. Reviewer #2 comment: 259-269 The number of significant digits given for PM2.5-associated mortalities is unscientifically precise.

Authors' response: We have revised the significant digits in Lines 377–389.

"Given the middle-road development pattern of SSP2-4.5, premature deaths in this scenario would peak at 9,024,000 (95% Confidence Interval (CI): 6,352,000–11,236,000) in the 2060s and then steadily decline to 7,394,000 (95% CI: 5,202,000–9,291,000) in the final decade of the century, which is a less rapid decline than in the SSP1-2.6 scenario. SSP3-7.0 assumes weak pollution control in which the implementation of pollution mitigation measures is delayed and less ambitious in the long term. In this scenario, premature deaths would spike dramatically in all regions except North America, Europe, and Russia and would not decrease until the end of the century. The global number of $PM_{2.5}$-associated premature deaths would reach 11,149,000 (95% CI: 7,877,000–13,800,000) in 2091–2100, an increase of 63% from the baseline period. In the SSP5-8.5 scenario, which emphasizes technological progress and rapid economic growth through human capital development, environmental issues become a priority health concern, and ambitious air quality goals result in pollutant levels well below current levels in the medium to long term.[16,72] Therefore, in SSP5-8.5, global premature deaths would peak at 8,509,000 (95% CI: 5,981,000–10,617,000) in the 2040s and then decline to 6,258,000 (95% CI: 4,410,000–7,887,000) in the second half of the 21st century as high-performance pollution control technologies are developed. This decrease would result in a smaller premature death burden than in the baseline period."

15. Reviewer #2 comment: 285-287 This sentence appears to be referring to the numbers in the first row of Table S4, which are all positive. These are for SA1, which is described as a "constant population" scenario, but the text indicates it is considering only population changes. This is contradictory. If population changes are exacerbating the burden of premature deaths, then the values from the table (holding population constant) should be negative.

Authors' response: We apologize for the typo, and we have revised the statement in Lines 409–411.

"*In the first sensitivity experiment (SA1), the only contributor to the difference in the estimated premature deaths from the baseline period is the demographic transition, while the contributor to the difference in the second sensitivity experiment (SA2) is the $PM_{2.5}$ variation.*"

16. Reviewer #2 comment: 293-294 This sentence says SA2 is assuming constant population distribution, the opposite from what is written in 282-284.

Authors' response: We apologize for this mistake, and the statements in Lines 409–411 have been corrected.

"*In the first sensitivity experiment (SA1), the only contributor to the difference in the estimated premature deaths from the baseline period is the demographic transition, while the contributor to the difference in the second sensitivity experiment (SA2) is the $PM_{2.5}$ variation.*"

Technical Corrections

Several references are being generated incorrectly by your reference software.

The authors of ref 2 are given as "Collaborators, G.R.F."

Ref. 33 appears to have an additional reference embedded within it.

Refs 50-56 appear to be using just the first initials for the journal titles.

Authors' response: We are sorry for the typo, and it has been corrected.

"2. Forouzanfar, M. H.; Afshin, A.; Alexander, L. T.; Anderson, H. R.; Bhutta, Z. A.; Biryukov, S.; Brauer, M.; Burnett, R.; Cercy, K.; Charlson, F. J., Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet* **2016**, 388, (10053), 1659-1724.

33. Wen, A.; Wu, T.; Wu, X.; Zhu, X.; Li, R.; Ni, J.; Hu, G.; Qiao, Y.; Zou, D.; Chen, J., Evaluation of MERRA-2 land surface temperature dataset and its application in permafrost mapping over China. *Atmospheric Research* **2022,** *279*, 106373.

50. Eyring, V.; Bony, S.; Meehl, G. A.; Senior, C. A.; Stevens, B.; Stouffer, R. J.; Taylor, K. E., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **2016,** *9*, (5), 1937-1958.

51. Gidden, M. J.; Riahi, K.; Smith, S. J.; Fujimori, S.; Luderer, G.; Kriegler, E.; Van Vuuren, D. P.; Van Den Berg, M.; Feng, L.; Klein, D., Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development* **2019,** *12*, (4), 1443-1475.

52. Ramírez Villegas, J.; Jarvis, A., Downscaling global circulation model outputs: the delta method decision and policy analysis Working Paper No. 1. *International Center for Tropical Agriculture (CIAT). Cali. CO.* **2010**, 18.

53. Hay, L. E.; Wilby, R. L.; Leavesley, G. H., A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States 1. *JAWRA Journal of the American Water Resources Association* **2000,** *36*, (2), 387-397.

54. Navarro-Racines, C.; Tarapues, J.; Thornton, P.; Jarvis, A.; Ramirez-Villegas, J., High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific data* **2020,** *7*, (1), 1-14.

55. Hawkins, E.; Osborne, T. M.; Ho, C. K.; Challinor, A. J., Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agricultural and forest meteorology* **2013,** *170*, 19-31.

56. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope Iii, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S., Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* **2018,** *115*, (38), 9592-9597.”

Comments on Supplement

1. Reviewer #2 comment: Table S4: Suggest adding "(constant population)" and "(constant air pollution)", or similar, under the names SA1 and SA2 (though these designations appear to be reversed). There are far too many significant digits included in this table. Clarify whether "baseline" here refers to the 2010-2020 period or whether it refers to the "base" (non-sensitivity analysis) simulation.

Authors' response: Thanks for your comment. The "baseline" refers to the 2010-2019 period. We have revised Table S4 to adjust the significant digits and make it clear.

Table S8. Changes in projected premature deaths (> 25 years old) for two sensitivity studies (SA1 and SA2) relative to the premature deaths in the baseline period (2010-2019). Values are in 1000s.

	Time frame	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
SA1* (constant air pollution)	2021–2040	1594.31	1458.18	1340.62	1581.58
	2041–2060	3016.17	2916.32	2829.60	3000.95
	2061–2080	3214.96	3550.96	4024.10	3228.01
	2081–2100	2398.21	3505.78	5151.04	2483.66
SA2* (constant population)	2021–2040	−521.75	−50.25	403.24	−282.53
	2041–2060	−1819.82	−688.96	205.40	−1089.66
	2061–2080	−2583.48	−1129.58	−225.21	−1642.07
	2081–2100	−3083.89	−1800.52	−607.03	−2236.05

*SA1 focused on the impact of future demographic changes on the number of premature deaths.

*SA2 targets quantifying the future changes in the premature mortality burden due to climate change and emissions.

2. Reviewer #2 comment: Figure S9: There appears to be almost the same pattern in each panel. Recommend checking closely for a scripting error. Also, this strikes me as a very large decrease in average wind speeds, much larger than I have seen in my own climate modeling.

Authors' response: Thanks for your comment. We apologize for the script mistake, and Figure S9 of the wind speed comparison has been replotted in the SI (Page 22).

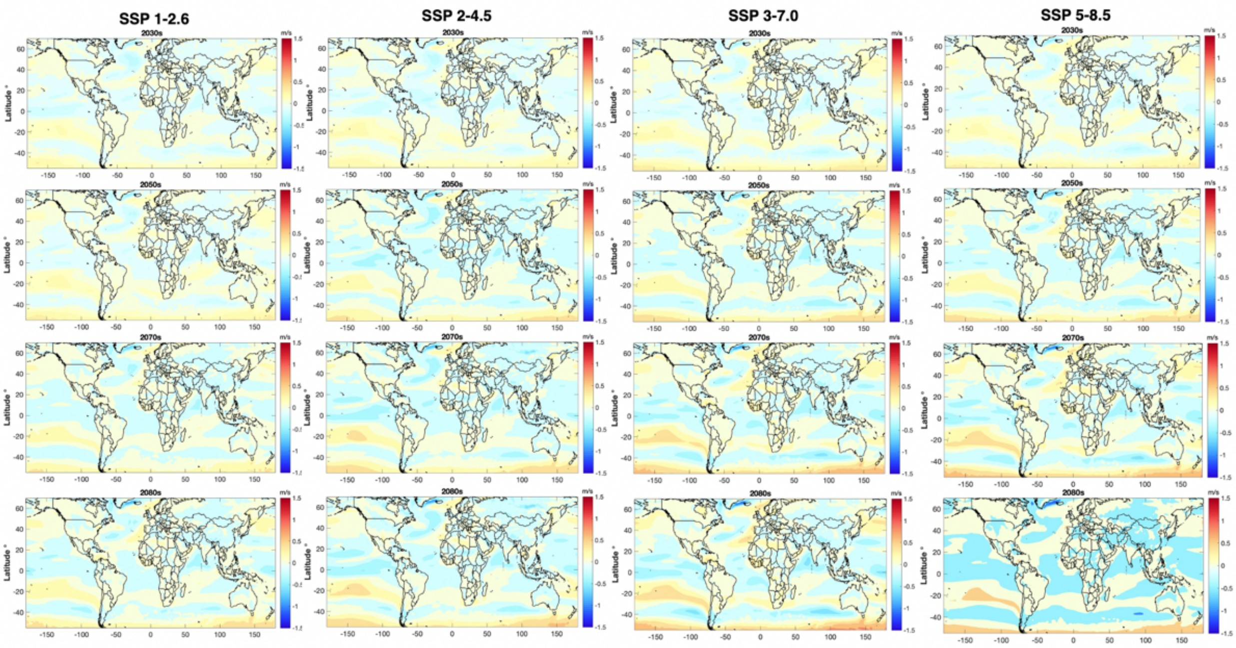


Figure S12. Changes in multi-model ensembles of surface wind speed using delta change downscaling for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) compared to that in 1990s (1981-2010).

3. Reviewer #2 comment: Fig S14: Change the y-axis on so that it is units of billions, so that you can show 5.0, 5.5, etc., rather than repeating "5E+09" twice.

Authors' response: Thanks for your suggestions. We have changed the y-axis in Figure S24.

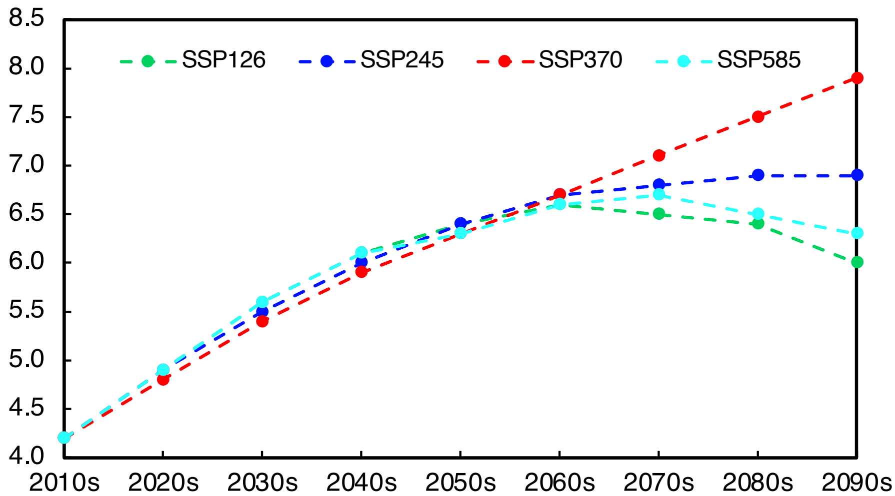


Figure S24. Total population size (values in billion) aged 25 years and above for major world regions under SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 until the end of the 21st century.

4. Reviewer #2 comment: Fig S17: What are the units of population growth rate? Percent per year? It's striking that on this scale North America is the same for all scenarios and times despite significant variation shown in Fig S15. Perhaps instead of showing the growth rate, the authors should show the population density at selected time points for each SSP.

Authors' response: Thanks for your comments.

The population growth rates for the decadal mean population compared to that in the 2010s (2010-2019 average) are calculated by using the following formula. The units are persons/persons.

$$\boldsymbol{Population\ growth\ rate} = \frac{\boldsymbol{Future\ decadal\ country\_level\ pop. - 2010s\ pop.}}{\boldsymbol{2010s\ pop.}}$$

In fact, the population growth pattern in North America and three countries in North America (Canada, U.S.A., and Mexico) is different in the four future SSP scenarios, as shown in Table A1-A4.

Table A1. Population aged 25 years and above (values in million) in North America (the sum of Canada, Mexico, and U.S.A.) under the four SSP scenarios.

	Baseline	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
SSP1-2.6	312.0	353.3	392.8	425.8	453.3	476.9	495.2	503.8	500.4
SSP2-4.5	312.0	349.6	384.6	413.9	439.8	462.9	481.9	495.4	503.4
SSP3-7.0	312.0	341.9	367.3	383.7	393.4	399.6	401.8	399.2	393.9
SSP5-8.5	312.0	357.4	401.7	444.5	488.2	532.1	575.5	613.7	641.8

Table A2. Population aged 25 years and above (values in million) in Canada under the four SSP scenarios.

	Baseline	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
SSP1-2.6	25.9	29.6	32.8	36.1	39.1	41.6	43.5	44.3	43.7
SSP2-4.5	25.9	29.3	32.1	34.8	37.3	39.4	41.2	42.4	42.7
SSP3-7.0	25.9	27.9	29.2	29.8	29.5	28.7	27.6	25.9	24.0
SSP5-8.5	25.9	30.6	35.0	40.0	45.5	50.9	56.1	60.4	63.1

Table A3. Population aged 25 years and above (values in million) in U.S.A under the four SSP scenarios.

	Baseline	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
SSP1-2.6	217.5	244.6	269.5	291.5	312.1	332.3	350.1	361.5	364.9
SSP2-4.5	217.5	241.9	263.2	281.3	298.6	315.6	330.0	340.6	347.7
SSP3-7.0	217.5	234.3	246.5	251.3	250.7	247.8	241.0	229.8	216.1
SSP5-8.5	217.5	249.5	279.9	311.9	348.3	387.6	428.2	466.0	497.3

Table A4. Population aged 25 years and above (values in million) in Mexico under the four SSP scenarios.

	Baseline	2020s	2030s	2040s	2050s	2060s	2070s	2080s	2090s
SSP1-2.6	68.7	79.1	90.5	98.2	102.1	103.0	101.6	98.1	91.8
SSP2-4.5	68.7	78.5	89.2	97.8	103.9	107.9	110.7	112.4	113.0
SSP3-7.0	68.7	79.7	91.6	102.6	113.3	123.1	133.2	143.4	153.8
SSP5-8.5	68.7	77.3	86.8	92.5	94.3	93.5	91.2	87.3	81.5

We agree with the reviewers that changes in population density need to be presented, as shown in Figure S26.

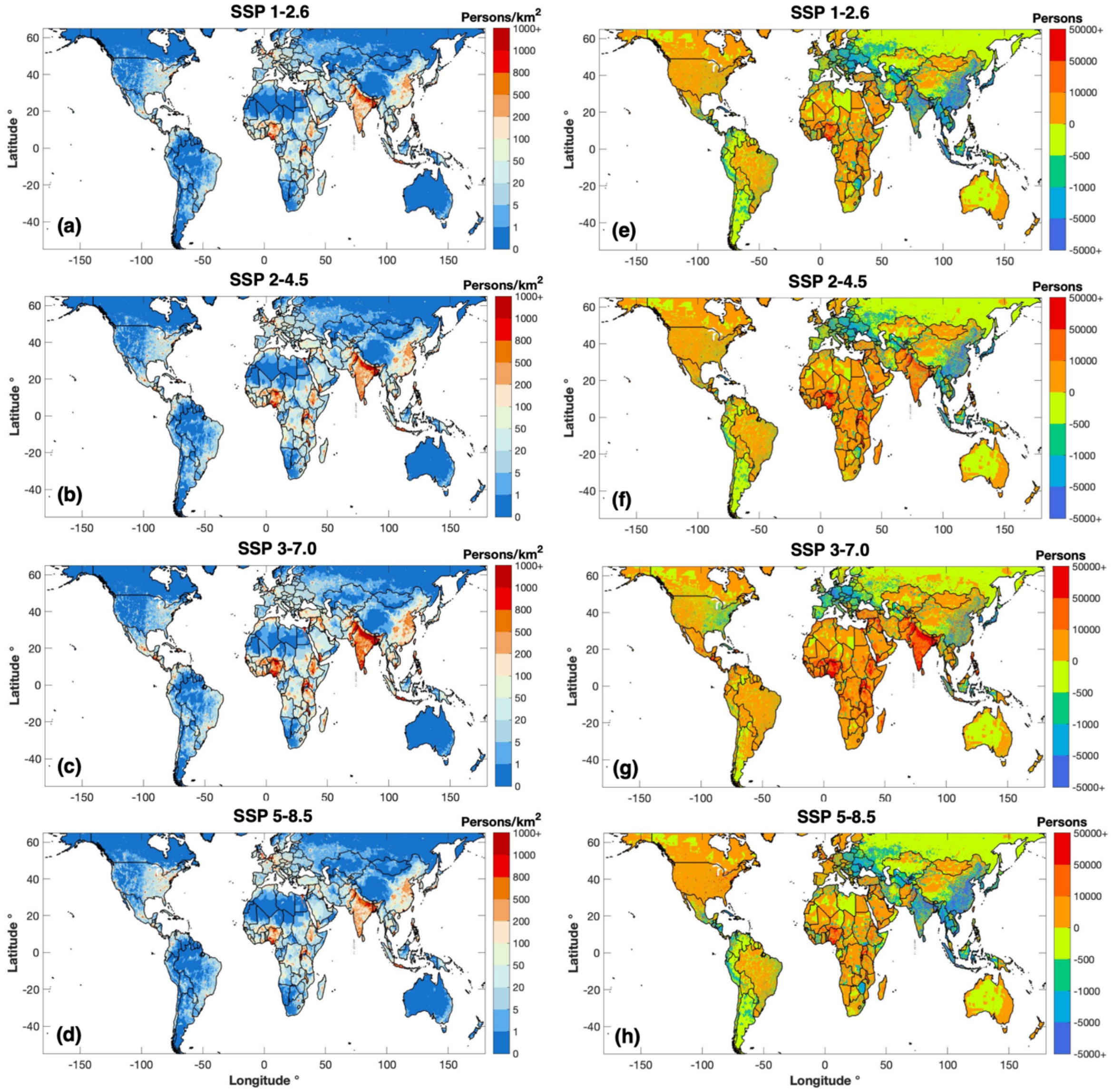


Figure S26. (a)–(d) Projected population density (persons for 25+ years old/km^2) for the four SSPs in 2091-2100, and (e)–(g) corresponding projected population change compared with that in 2010s (2010-2019 average).

Reference

1. Ronneberger, O.; Fischer, P.; Brox, T. In *U-net: Convolutional networks for biomedical image segmentation*, International Conference on Medical image computing and computer-assisted intervention, 2015; Springer: 2015; pp 234-241.
2. Yamashita, R.; Nishio, M.; Do, R. K. G.; Togashi, K., Convolutional neural networks: an overview and application in radiology. *Insights into imaging* **2018,** *9*, (4), 611-629.
3. Nagi, J.; Ducatelle, F.; Di Caro, G. A.; Cireşan, D.; Meier, U.; Giusti, A.; Nagi, F.; Schmidhuber, J.; Gambardella, L. M. In *Max-pooling convolutional neural networks for vision-based hand gesture recognition*, 2011 IEEE international conference on signal and image processing applications (ICSIPA), 2011; IEEE: 2011; pp 342-347.
4. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J., Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* **2011,** *3*, (1), 1-122.
5. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope Iii, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S., Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* **2018,** *115*, (38), 9592-9597.
6. Hou, R.; Chen, C.; Shah, M. In *Tube convolutional neural network (T-CNN) for action detection in videos*, Proceedings of the IEEE international conference on computer vision, 2017; 2017; pp 5822-5831.
7. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S., Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* **2018,** *115*, (38), 9592-9597.
8. Turnock, S. T.; Allen, R. J.; Andrews, M.; Bauer, S. E.; Deushi, M.; Emmons, L.; Good, P.; Horowitz, L.; John, J. G.; Michou, M., Historical and future changes in air pollutants from CMIP6 models. *Atmospheric Chemistry and Physics* **2020,** *20*, (23), 14547-14579.
9. Chowdhury, S.; Dey, S.; Smith, K. R., Ambient PM2. 5 exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nature communications* **2018,** *9*, (1), 1-10.
10. Glotfelty, T.; He, J.; Zhang, Y., Impact of future climate policy scenarios on air quality and aerosol-cloud interactions using an advanced version of CESM/CAM5: Part I. model evaluation for the current decadal simulations. *Atmospheric environment* **2017,** *152*, 222-239.
11. Solazzo, E.; Bianconi, R.; Hogrefe, C.; Curci, G.; Tuccella, P.; Alyuz, U.; Balzarini, A.; Baró, R.; Bellasio, R.; Bieser, J., Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown. *Atmospheric chemistry and physics* **2017,** *17*, (4), 3001-3054.
12. Van Marle, M. J.; Kloster, S.; Magi, B. I.; Marlon, J. R.; Daniau, A.-L.; Field, R. D.; Arneth, A.; Forrest, M.; Hantson, S.; Kehrwald, N. M., Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development* **2017,** *10*, (9), 3329-3357.

13. Feng, L.; Smith, S. J.; Braun, C.; Crippa, M.; Gidden, M. J.; Hoesly, R.; Klimont, Z.; Van Marle, M.; Van Den Berg, M.; Van Der Werf, G. R., The generation of gridded emissions data for CMIP6. *Geoscientific Model Development* **2020,** *13*, (2), 461-482.
14. Zhao, A.; Ryder, C. L.; Wilcox, L. J., How well do the CMIP6 models simulate dust aerosols? *Atmospheric Chemistry and Physics* **2022,** *22*, (3), 2095-2119.
15. Schwartz, J.; Norris, G.; Larson, T.; Sheppard, L.; Claiborne, C.; Koenig, J., Episodes of high coarse particle concentrations are not associated with increased mortality. *Environmental Health Perspectives* **1999,** *107*, (5), 339.
16. Reis, L. A.; Drouet, L.; Tavoni, M., Internalising health-economic impacts of air pollution into climate policy: a global modelling study. *The Lancet Planetary Health* **2022,** *6*, (1), e40-e48.
17. Markandya, A.; Sampedro, J.; Smith, S. J.; Van Dingenen, R.; Pizarro-Irizar, C.; Arto, I.; González-Eguino, M., Health co-benefits from air pollution and mitigation costs of the Paris Agreement: a modelling study. *The Lancet Planetary Health* **2018,** *2*, (3), e126-e133.
18. Miao, Y.; Luo, W. In *Improve Generalization Ability of CNN by Data Augmentation and SE Block in Landmark Classification*, 2022 14th International Conference on Computer Research and Development (ICCRD), 2022; IEEE: 2022; pp 250-255.
19. Volpi, R.; Namkoong, H.; Sener, O.; Duchi, J. C.; Murino, V.; Savarese, S., Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems* **2018,** *31*.
20. Shorten, C.; Khoshgoftaar, T. M., A survey on image data augmentation for deep learning. *Journal of big data* **2019,** *6*, (1), 1-48.
21. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R., Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **2014,** *15*, (1), 1929-1958.
22. Srivastava, N., Improving neural networks with dropout. *University of Toronto* **2013,** *182*, (566), 7.
23. Dahl, G. E.; Sainath, T. N.; Hinton, G. E. In *Improving deep neural networks for LVCSR using rectified linear units and dropout*, 2013 IEEE international conference on acoustics, speech and signal processing, 2013; IEEE: 2013; pp 8609-8613.
24. Jain, P.; Kulkarni, V.; Thakurta, A.; Williams, O., To drop or not to drop: Robustness, consistency and differential privacy properties of dropout. *arXiv preprint arXiv:1503.02031* **2015**.
25. Park, S.; Kwak, N. In *Analysis on the dropout effect in convolutional neural networks*, Asian conference on computer vision, 2017; Springer: 2017; pp 189-204.
26. Taylor, L.; Nitschke, G. In *Improving deep learning with generic data augmentation*, 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018; IEEE: 2018; pp 1542-1547.
27. Dinh, L.; Pascanu, R.; Bengio, S.; Bengio, Y. In *Sharp minima can generalize for deep nets*, International Conference on Machine Learning, 2017; PMLR: 2017; pp 1019-1028.
28. Kawaguchi, K.; Kaelbling, L. P.; Bengio, Y., Generalization in deep learning. *arXiv preprint arXiv:1710.05468* **2017**.
29. Wang, S.; Zhao, Y.; Chen, G.; Wang, F.; Aunan, K.; Hao, J., Assessment of population exposure to particulate matter pollution in Chongqing, China. *Environmental Pollution* **2008,** *153*, (1), 247-256.
30. Shan, Z.; Li, H.; Pan, H.; Yuan, M.; Xu, S., Spatial Equity of PM2. 5 Pollution Exposures in High-Density Metropolitan Areas Based on Remote Sensing, LBS and GIS Data: A Case Study in Wuhan, China. *International Journal of Environmental Research and Public Health* **2022,** *19*, (19), 12671.

31. Wang, Y.; Hu, J.; Huang, L.; Li, T.; Yue, X.; Xie, X.; Liao, H.; Chen, K.; Wang, M., Projecting future health burden associated with exposure to ambient PM2. 5 and ozone in China under different climate scenarios. *Environment International* **2022,** *169*, 107542.
32. Chowdhury, S.; Dey, S.; Smith, K. R., Ambient PM2. 5 exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nature communications* **2018,** *9*, (1), 318.
33. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S. L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M., Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* **2021,** *2*.
34. Rao, S.; Klimont, Z.; Smith, S. J.; Van Dingenen, R.; Dentener, F.; Bouwman, L.; Riahi, K.; Amann, M.; Bodirsky, B. L.; van Vuuren, D. P., Future air pollution in the Shared Socio-economic Pathways. *Global Environmental Change* **2017,** *42*, 346-358.

Global $PM_{2.5}$ prediction and associated mortality to 2100 under different climate change scenarios

Wanying Chen [1,2], Xingcheng Lu [3*], Dehao Yuan [4], Yiang Chen [1,2], Zhenning Li [1], Yeqi Huang [1], Haochen Sun [5,6], Jimmy C.H. Fung [1,2,6*]

[1] Division of Environment and Sustainability, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

[2] Atmospheric Research Center, Guangzhou HKUST Fok Ying Tung Research Institute, Guangzhou, China

[3] Department of Geography and Resource Management, Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

[4] Department of Computer Science, University of Maryland, College Park, Maryland, USA

[5] Department of Mathematics, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

[6] Department of Computer Science and Engineering, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

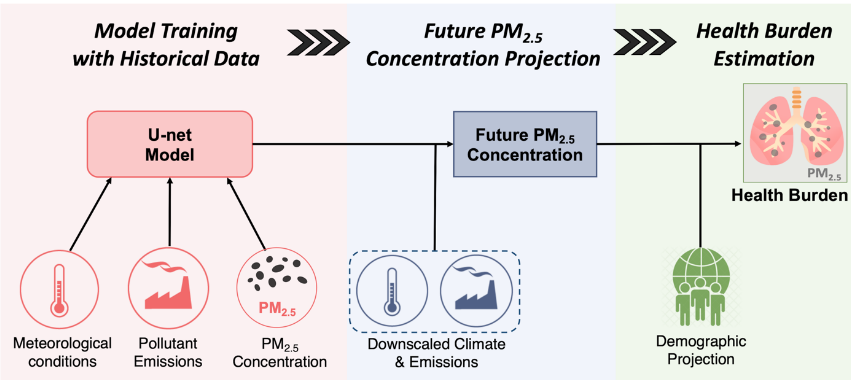
Correspondence to: Xingcheng Lu (xingchenglu2011@gmail.com), J.C.H. Fung (majfung@ust.hk).

Abstract

Ambient fine particulate matter ($PM_{2.5}$) can cause severe adverse health impacts in humans. Thus, reducing $PM_{2.5}$ exposure is an important public health focus. Meteorological and emissions factors, which considerably affect the $PM_{2.5}$ concentrations in air, vary significantly under different climate change scenarios. However, $PM_{2.5}$ concentrations and their associated disease burden under future climate scenarios are not well clarified. In this work, the global $PM_{2.5}$ concentrations from 2021 to 2100 were estimated by combining the U-Net convolutional neural network deep learning technique, reanalysis data, emissions data, and bias-corrected Coupled Model Intercomparison Project Phase 6 future climate scenario data. Based on the estimated $PM_{2.5}$ concentrations, the future premature mortality burden associated with $PM_{2.5}$ exposure was assessed using the Global Exposure Mortality Model. Ambient $PM_{2.5}$ exposure is expected to be highest in the SSP3-7.0 scenario and lowest in the SSP1-2.6 scenario in the major representative regions of the world. The global mortality rate (per 100,000 exposed population) associated with $PM_{2.5}$ under the four different scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, ranging from 84.6 (95% Confidence Interval (CI): 59.6–107.0) to 150.0 (95% CI: 106.2–185.0)) at the end of this century is expected to be lower than the baseline (the 2010s, 161.1 (95% CI: 113.3–199.9)). Among all four scenarios, the sustainable development scenario (SSP1-2.6) results in the lowest $PM_{2.5}$ concentrations and the lowest premature mortality burden, which indicates that this is the pathway that countries should strive for. Our work helps to advance the scientific understanding of the global geo-climatic system and provides suggestions for scientists and policymakers.

Keywords: Climate change; Global; $PM_{2.5}$; Mortality; Deep learning

Synopsis: Future $PM_{2.5}$ pollution and its associated health burden have not been well clarified. In this study, a new set of global-scale, spatially explicit $PM_{2.5}$ concentration from 2021 to 2100 with a spatial resolution of 0.1°×0.1° was estimated, and associated $PM_{2.5}$ exposure and premature mortality burden was calculated.



Graphic for Table of Contents (TOC)

1. Introduction

Ambient particulate matter ($PM_{2.5}$) poses a considerable global threat to human health. Exposure to outdoor $PM_{2.5}$ caused 4.14 million deaths in 2019, accounting for 62% of all global deaths attributable to air pollution estimated by the Global Burden of Disease Project.[1-4] Unmitigated climate change is projected to exacerbate inevitable challenges and threats to global air quality and increase its attributable adverse health impacts.[5-7] Therefore, it is necessary to understand how future climate change scenarios will influence surface $PM_{2.5}$ concentrations and propose appropriate climate mitigation measures.

Most studies[7, 8] on $PM_{2.5}$ concentration estimation under different climate scenarios have been based on the Coupled Model Intercomparison Project 5 (CMIP5) Representative Concentration Pathways scenarios. However, with the release of the CMIP6 simulation results, the Scenario Model Intercomparison Project provides new alternative scenarios that are intimately connected with societal concerns regarding climate change mitigation, adaptation, and impacts.[9, 10] Some studies have estimated future air quality based on CMIP6 climate projections;[11, 12] however, these studies either investigated the $PM_{2.5}$ exposure in only one country or region,[11-13] or the predicted periods were shorter than 50 years.[14, 15] Although future global-scale $PM_{2.5}$ simulations are available,[12, 16] the low model spatial resolution (e.g., 1.875° × 1.25°) prevents a clear understanding of how this pollutant will evolve over the next several decades and hampers reliable estimations of how this pollutant will influence human health in the future. As yet, no comprehensive study has estimated the global mortality burden associated with ambient $PM_{2.5}$ based on high-resolution (e.g., 0.1° × 0.1°) and bias-corrected future climate projections that incorporate demographic and emissions data. Such a study is urgently needed to understand how the $PM_{2.5}$ concentration and the associated health burden in each country will vary under different climate scenarios.

In this study, we estimated $PM_{2.5}$ exposure and its associated mortality burden over the 2021–2100 period under the SSP1-2.6[17], SSP2-4.5[18], SSP3-7.0,[19] and SSP5-8.5[20] scenarios (SSP: Shared Socioeconomic Pathway). The relationships between critical meteorological variables and $PM_{2.5}$ concentrations were constructed using a U-Net convolutional neural network[21] based on Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2)[22], CMIP6 global emissions data,[23] and satellite-retrieved $PM_{2.5}$ data.[24] $PM_{2.5}$ exposure and the associated premature mortality over the 2021–2100 period were estimated based on the constructed relationships between the $PM_{2.5}$ concentrations, meteorological variables, emissions, the high-resolution and bias-corrected CMIP6 future climate SSP scenario data (adjusted using the delta downscaling method), and future SSP demographic

Deleted: and

projections. Our work endeavored to elucidate how and through what pathways $PM_{2.5}$ exposure would influence the premature mortality burden in 184 countries and regions worldwide over the forthcoming 80 years, spanning the space of challenges to mitigation and adaptation to climate change, which can exhibit a more expansive and comprehensive blueprint to air quality projection.

Deleted: conscientious

2. Methods

2.1. Data acquisition

2.1.1. Surface $PM_{2.5}$ data for training

High-resolution and highly accurate global surface $PM_{2.5}$ data are required to examine the relationships between $PM_{2.5}$ concentrations and meteorological and emissions conditions. Therefore, global surface $PM_{2.5}$ data at 0.1° × 0.1° combining AOD retrievals from the NASA MODIS, MISR, and SeaWIFS instrument, GEOS-Chem chemical transport model, and ground-based observations calibrated by geographically weighted regression were selected for the study.[24] Compared with previous global surface $PM_{2.5}$ concentration datasets,[25-27] this set of $PM_{2.5}$ values contained finer resolution data and compensated for missing or limited monthly measurements. This $PM_{2.5}$ dataset was highly consistent with collocated ground-based observations from monitoring networks $PM_{2.5}$ (R^2 = 0.84), with a root mean square error (RMSE) of 8.4 μg m^{-3}, and thus can accurately represent the surface $PM_{2.5}$ concentrations.

2.1.2. Meteorological and emissions data for model input

To train the deep learning model, the following monthly average meteorological data were taken from the MERRA-2 dataset:[28] surface temperature, wind speed, specific humidity, planetary boundary layer height, and sea level pressure; these parameters can strongly influence the $PM_{2.5}$ concentration.[29] Several studies have contrasted the MERRA-2 dataset with ground-based observations and other reanalysis datasets and have shown that the MERRA-2 data better represent the surface meteorological conditions.[22, 30-33] For example, when compared with the ground observation data from China, the RMSE, MB (mean bias), and R value for temperature were 3.62 K, −2.14 K, and 0.95, respectively.[33] These three statistical metrics for humidity were 5%, 0.63%, and 0.89.[34]

Since the deficiency in the emissions of primary $PM_{2.5}$ components (except organic carbon (OC) and black carbon (BC)) in the CMIP6 datasets, future $PM_{2.5}$ concentrations are driven by changes to precursor emissions (ammonia (NH_3), nitrogen oxides (NOx), and sulfur dioxide (SO_2)), BC, OC and climate in this study. Global emission amounts and percentages of the five species are presented in Table S1 and S2. Based on existing global emission inventory,

Deleted: Primary $PM_{2.5}$ emissions data are not available in the CMIP6 dataset, we used the emissions of five pollutants (ammonia, nitrogen oxides, organic carbon, black carbon, and sulfur dioxide) as the emissions input for the deep learning model because these pollutants can have a marked influence on surface $PM_{2.5}$ concentrations.[35, 36]

such as PKU-FUEL, primary $PM_{2.5}$ emission has high correlation with the emissions of these five pollutants.[35, 36] Covering the period of 1750–2100 (historical dataset: 1750–2014, future emissions dataset: 2015–2100), CMIP6 gridded emissions dataset includes aviation emissions, all other anthropogenic emissions sectors, and total open burning emissions. This gridded dataset has previously been used for global model simulation and for emission scenario comparisons.[13, 37, 38] CMIP6 dataset contains historical emissions (1750 to 2014) and future emission data for SSP scenarios (2015-2100). CMIP6 emissions data were utilized in both the training and prediction processes. CMIP6 historical emissions data (1998–2014, the historical emissions data are available till 2014) were used to build the deep learning model, and the emission data of SSP scenarios for 2015 to 2019 were used in the deep learning model verification process. Future emission data for 2021–2100 were input into the trained deep learning model for prediction. The detailed narratives of emission inventory used in this study are summarized in Table S3.

Deleted: Covering the 1750–2100 period (historical dataset: 1750–2014, future emissions dataset: 2015–2100), the CMIP6 gridded emissions dataset has previously been used for assessing air control policies and for comparing divergent emissions scenarios.

The monthly MERRA-2 meteorological data and CMIP6 emissions data from 1998 to 2019 were input into the deep learning model for training and validation. The frequency distribution of meteorological and emission data is presented in Figure S1–S2 and discussed in Text S1. Before training the deep learning model, all meteorological and emissions data were re-interpolated from their original spatial resolutions (meteorological data with 0.625° × 0.500° and emission data with 0.5° × 0.5°) into the same grid as the surface $PM_{2.5}$ data with a resolution of 0.1° × 0.1°. The bilinear interpolation technique was applied in this work, which has been widely used to interpolate climate data into different resolutions in previous studies.[39,40]

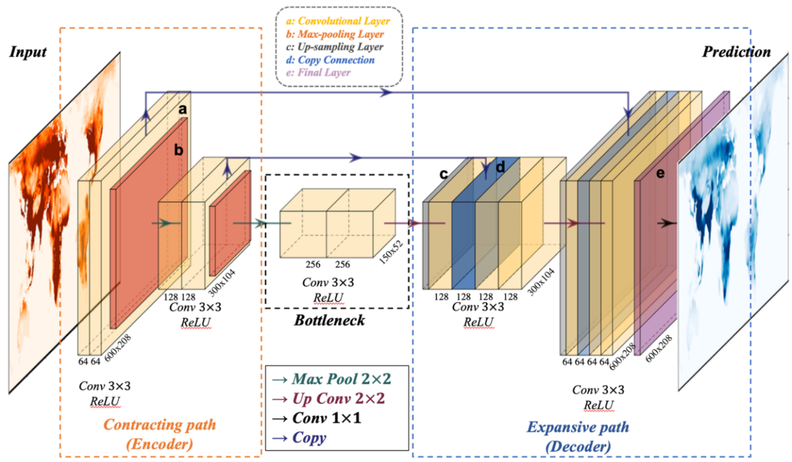
Deleted: primary

2.2. U-Net convolutional neural networks

Tremendous advances in computer vision have led to convolutional neural networks (CNNs) being widely used for 2D data analysis.[41] We built a CNN-based U-Net framework to construct relationships between $PM_{2.5}$ concentrations and predictor variables.[21] First proposed for medical segmentation,[21] U-Net assumes that local information and global information are both essential, which is also apposite for $PM_{2.5}$ prediction. Equipped with flexible global aggregation blocks, U-Net can sufficiently consider non-local influences from other grid cells to local $PM_{2.5}$ concentration. In addition, multiple layers of U-Net CNNs make it possible to elucidate nonlinear relationships among critical meteorological variables, ambient pollutant emissions, and surface $PM_{2.5}$ concentrations; these relationships can be too complex to be delineated through traditional regression methods.[42-44]

All of the predictor variables (meteorological and emission data) and the $PM_{2.5}$ concentrations were treated as 2D images. The detailed architecture of our U-Net model, including the number of channels for each convolution layer, the size of the convolution kernel, the activation function of the convolution layer, and the image size are provided in Figure 1. The description of the U-Net model can be found in Text S2. The data augmentation and dropout regularization have been applied to improve the model generalization ability, as discussed in Text S3. And the monotonic decreases in training and validation loss (Figure S4) have proved that no overfitting was detected.

Deleted: 1



Deleted:

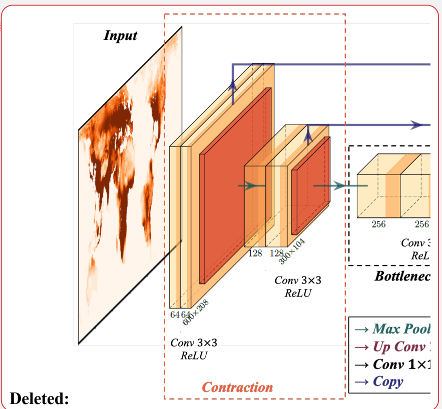


Figure 1. Architecture of the U-Net model

2.3. Future climate data under different Shared Socioeconomic Pathway (SSP) scenarios

The trained model was used to predict the 2021–2100 $PM_{2.5}$ concentrations using the meteorological variables from the CMIP6 future climate scenarios dataset. As shown in Table S4, historical simulations (1981–2010) and future projections (2021–2100) of global climate multiple-model ensemble results from 28 global climate models (GCMs) and four SSPs (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) were utilized. The four SSPs are classified by socioeconomic, land use, and environmental development assumptions and represent conceivable future scenarios that capture distinctive climate mitigation and adaptation challenges. SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 represent low, medium, medium to high, and high radiative forcing by the end of the century, respectively.[45-47] Changes in greenhouse gas concentrations in the atmosphere affect radiative forcing; thus, 'radiative forcing' mentioned in this work corresponds to different greenhouse gas (GHG) emission scenarios and the resulting climate

Deleted: well-

Deleted: 1

change. Both SSP1-2.6 and SSP5-8.5 represent strong climate mitigation scenarios,[17, 48] with the distinction that the anthropogenic radiative forcings by 2100 are 2.6 Watt/m^2 and 8.5 Watt/m^2, respectively.[20, 49] SSP2-4.5 represents a moderate mitigation scenario and the radiative forcing is stabilized at 4.5 Watt/m^2 until 2100 by implementing moderately restrictive emission reduction measures and strategies.[18] SSP3-7.0 is the weakest climate mitigation scenario with an anthropogenic radiative forcing of 7.0 Watt/m^2 by 2100.[19] The critical elements relevant to air pollution among four SSP scenarios are summarized in Table S5. Further information and the assumptions used in the future scenarios are provided in Eyring et al. (2016)[50] and Gidden et al. (2019).[51] The SSPs explored in this study cover a wide range of plausible socioeconomic trends for this century.

2.4. Bias correction and downscaling

Before being fed into the trained U-Net model, the meteorological variables from CMIP6 were corrected and downscaled to achieve reliable climate change impact metrics. To produce high-resolution and bias-corrected future climate information, we used the delta change (DC) method, which applies a change factor (i.e., delta) derived from GCMs to historical observations.[52, 53] Studies have found the DC method to be robust for downscaling climate data.[54, 55] Our implementation of the DC method was intended to correct the simulated climate data while providing results at high spatial resolution. The details of the DC method are described in the Text S4.

Deleted: 2

2.5. Mortality calculation

The Global Exposure Mortality Model (GEMM) proposed by Burnett et al. (2018)[56] was used as a hazard ratio model to estimate the premature mortality burden associated with $PM_{2.5}$ exposure (i.e., population-weighted $PM_{2.5}$ concentration, Text S5). GEMM has relieved some of the contentious assumptions that are stipulated by other disease-specific hazard ratio models, such as the Integrated Exposure Response Model.[56] The detailed of the GEMM model are provided in the Text S6.

Formatted: Subscript

Deleted: 3

The baseline mortality rates for different countries in 2015 obtained from the Global Health Data Exchange data catalog were used for estimating premature mortality. The gridded population projections for all SSPs during 2021–2100 at a resolution of 1 km × 1 km are available from the Spatial Population Scenario database. This demographic projection dataset has been previously verified[57] and has been used to project heat-related excess mortality[58, 59] and to model future patterns of urbanization.[60] In this work, we calculated the $PM_{2.5}$-associated premature mortality in accordance with the projected population, but the baseline mortality rate was assumed to be that of 2015 owing to a lack of credible alternatives. Constant baseline mortality has been applied in other works that have projected the future environmental burdens of disease[61, 62].

3. Results

3.1. Performance evaluation

To verify that the trained U-Net model could generate accurate $PM_{2.5}$ concentration predictions, we validated the model performance from the spatial, scatter point, and statistical matrix perspectives. CMIP6 historical emissions data are available through 2014, while the data from 2015 to 2019 were from the CMIP6 future scenario emissions dataset. In the CMIP6 future scenario emissions dataset, the different scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) have their own emissions datasets, but the differences are very limited throughout the 2015–2019 period. Because of this, we separated the verification by (1) implementing 8-fold cross-validation to verify the performance for the estimations from 1998 to 2014 and (2) inputting the future emissions datasets (2015–2019) of the four scenarios together with other independent variables into the trained model to output the $PM_{2.5}$ estimation for the comparison. For the 8-fold cross-validation, 15 years of data were used for training and 2 years of data were used for comparison in each fold. The calculation of 8-fold cross-validation is further described in Text S7.

Figure S5 shows a spatial comparison between the satellite-retrieved $PM_{2.5}$ data and the values predicted by the U-Net CNN using 8-fold validation. The results show that the model produced good fits in the areas with both low ($\leq$ 35 μg/m^3) and high ($>$ 35 μg/m^3) $PM_{2.5}$ concentration. As demonstrated in Figures S6 and S7, the errors between the simulated and target monthly average $PM_{2.5}$ concentrations for all grid cells were within $\pm$ 12 μg/m^3 for 1998-2014. The monthly average relative errors specific to each country were within $\pm$ 10%.

Figure 2 shows the scatter plots of the satellite-retrieved $PM_{2.5}$ concentrations and the 8-fold average predicted concentrations. The strong correlation coefficient (R, 0.987) was better than that of previous studies[63-65] and indicates that the model could accurately predict all of the 8-fold cross-validation data. The statistical evaluation metrics (A1–A6 in the supplemental material) shown in Table 1 were further used to verify the model performance. The relatively small standard deviation of error indicates that our trained model has considerable stability. From these statistical matrix perspectives, the $PM_{2.5}$ concentrations estimated by our proposed deep learning model were also better than those of previous studies.[29, 66, 67] In addition to the comparison with the satellite-retrieved $PM_{2.5}$ data, we compared the annual model-predicted $PM_{2.5}$ concentrations with the monitor-based observations in China, the United States,

Deleted: well-

Deleted: well-

Deleted: 2

Deleted: The results show that the model gave a well-fitted in the areas with both low ($\leq$ 35 μg/m^3) and high ($>$ 35 μg/m^3) $PM_{2.5}$ concentration. As demonstrated in Figures S3 and S4, the error between the simulated and target $PM_{2.5}$ concentrations for all grid cells was within $\pm$ 12 μg/m^3. The annual relative errors specific to each country were within $\pm$ 10%.

Deleted: A5

Deleted: The NMB, NME, MB, and MAGE of the average 8-fold cross-validation were −0.0073 $\pm$ 0.0138, 0.2211 $\pm$ 0.0274, −0.0469 $\pm$ 0.0848 μg/m^3, and 1.3622 $\pm$ 0.1059 μg/m^3, respectively.

and Europe because these regions have well-established ground-based observation networks (Table S6). The R values for China, the United States, and Europe were 0.91, 0.80, and 0.81, respectively. These results show that the $PM_{2.5}$ estimates from our method were also in general agreement with the ground-based observations in these regions.

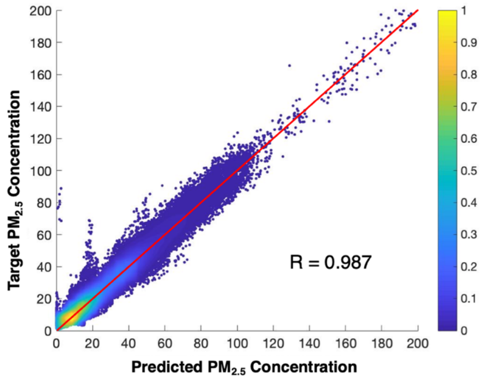


Figure 2. 8-fold cross-validation of the global $PM_{2.5}$ concentrations predicted by the U-Net CNN model during 1998-2014. The color represents the sample density.

Table 1. 8-fold cross-validation of U-Net CNN model performance

	NMB*	NME*	MB* ($\mu g/m^3$)	MAGE* ($\mu g/m^3$)	RMSE* ($\mu g/m^3$)	R
Average	−0.01	0.22	−0.05	1.36	4.02	0.987
Standard error	0.01	0.03	0.08	0.11	0.39	0.010

***NMB:** normalized mean bias; **NME:** normalized mean error; **MB:** mean bias; **MAGE:** mean absolute gross error; **RMSE:** Root Mean Squared Error

As mentioned above, the CMIP6 emissions from 2015–2019 under the four SSP scenarios together with other input data were fed into the trained deep learning model to estimate the $PM_{2.5}$ concentrations for these 5 years, as shown in Table S7. These metrics indicate the good feasibility and generalizability of our model in predicting the $PM_{2.5}$ concentrations. In summary, the satisfactory performance indicated that the trained U-Net model was able to identify the relationships between $PM_{2.5}$ and the influencing factors, which demonstrates that this model could be used for future $PM_{2.5}$ pollution estimation in the 2021–2100 period under different climate scenarios.

Deleted: 2

Deleted: 8-fold cross-validation of the $PM_{2.5}$ concentrations predicted by the U-Net convolutional neural network model. The color represents the sample density.

Deleted: Table 1. 8-fold cross-validation of U-Net neural network model performance¶

... [1]

Deleted: 3

Deleted: The NMB ranged from 0.146 to 0.157 with an average of 0.148, the NME ranged from 0.338 to 0.341 with an average of 0.339, the MB ranged from 0.824 to 0.828 $\mu g/m^3$, and the MAGE ranged from 1.911 to 1.957 $\mu g/m^3$.

3.2. Projection of future ambient $PM_{2.5}$ concentrations

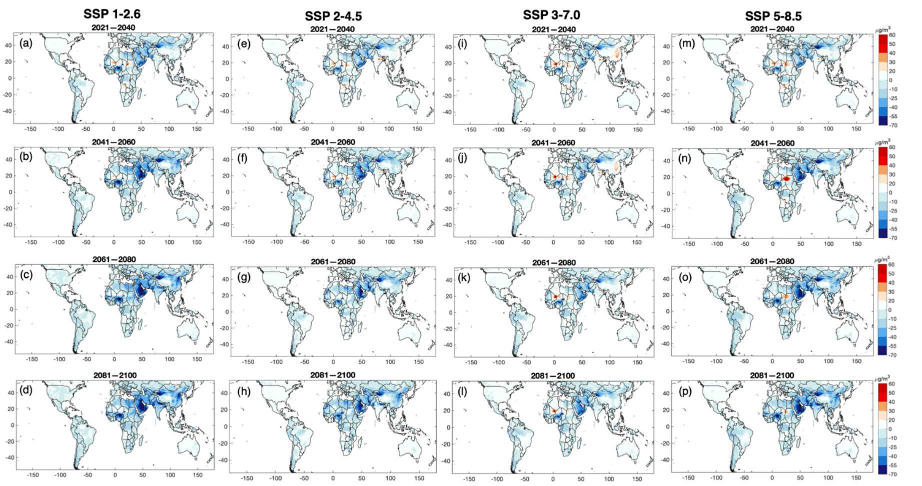
The built U-Net deep learning model was used to project future $PM_{2.5}$ concentrations under the SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios. Changes in the downscaled multi-model ensembles of critical meteorological variables are shown in Figures S8–S12. Emissions (SO_2, NH_3, OC, BC, NOx) for the four SSP scenarios up to 2100 are shown in Figures S13-S17. The projected $PM_{2.5}$ concentrations were compared with the baseline concentration (the average $PM_{2.5}$ concentration from 2010 to 2019), as shown in Figure 3. The $PM_{2.5}$ decadal average concentrations for the different SSP scenarios are shown in Figures S18–S21.

Deleted: 5

Deleted: 9

Deleted: 0

Deleted: 13



Deleted:

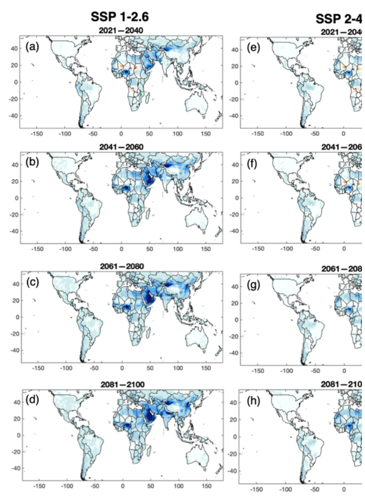


Figure 3. Spatial distribution of changes in projected global $PM_{2.5}$ concentrations relative to the baseline period (2010–2019) under different climate change scenarios. Panels (a)-(d) represent the changes in $PM_{2.5}$ concentration for 2030s (2021-2050 average), 2050s (2041-2070 average), 2070s (2061-2090 average), and 2080s (2071-2100 average) under SSP1-2.6 scenarios compared to the baseline condition. Panels (e)-(h) represent that of the same period but under SSP2-4.5. Panels (i)-(l) represent that of the same period but under SSP3-7.0. Panels (m)-(p) represent that of the same period but under SSP5-8.5.

Deleted: **Figure 3. Spatial distribution of changes in projected global $PM_{2.5}$ concentrations relative to the baseline period (2010–2019) under different climate change scenarios.**

Formatted: Space After: 0.5 line

Based on the deep learning model estimations, the $PM_{2.5}$ concentrations are projected to decrease in almost all regions in all scenarios; however, there are some notable differences among the projections. In SSP1-2.6, the projected $PM_{2.5}$ concentration will decrease consistently from 2030 to 2100. Among the investigated regions, the Middle East, Eastern

China, and India will undergo the most significant decline in $PM_{2.5}$ concentrations under this scenario. SSP2-4.5 represents the middle range of plausible future pathways. In this scenario, although the furthest projection into the 2090s showed a decline compared with the baseline level (that of the 2010s), this reduction was much smaller than the corresponding changes under SSP1-2.6. The projections are different for SSP3-7.0, which assumes more pessimistic development strategies, such as less investment in the environment and health care and a fast-growing population.[17, 19, 68] This would lead to an apparent increase in $PM_{2.5}$ concentrations in Asia and Africa before the 2050s. After meeting economic development needs and implementing environmental control measures, the $PM_{2.5}$ concentrations would decrease to a level similar to the baseline period. In SSP5-8.5, fossil fuels are heavily relied on to achieve rapid economic growth. Thus, in the middle of the 21st century, climate and emission change would considerably increase $PM_{2.5}$ concentrations and cause considerable damage to human health in central Africa. Nevertheless, with the rapid development of society and pollution mitigation policies, the overall $PM_{2.5}$ concentrations will undergo a sharper reduction after the 2050s. $PM_{2.5}$ concentrations continued to exceed baseline values in central Africa under SSP3-7.0 and SSP5-8.5, which is quite consistent with the emission-increasing trends in the region as shown in Figure S22 and S23, implying that the emission change should be the major driver for the $PM_{2.5}$ concentration increase in Central Africa.

3.3. Projection of future ambient $PM_{2.5}$ exposure

SSPs narratives gave rise to spatial and temporal differences in the demographic projections. Figures S24 and S25 show the demographic projections for the four SSPs scenarios for the world and for different regions, respectively. The projected population density (persons/km^2) and the corresponding variations (compared to the situation in the 2010s) in four SSP scenarios are shown in Figure S26. Combined with demographic projections, the exposure concentration can be estimated and used to assess the $PM_{2.5}$ exposure associated health impacts.[69, 70]

Figure 4 shows the projected $PM_{2.5}$ exposure concentrations in several representative regions (North America, South America, Europe, Africa, the Middle East, Russia and Economies in Transition [EIT], Asia, and the rest of the world) under the various SSP scenarios. The region boundaries are shown in Figure S27. Overall, $PM_{2.5}$ exposure is highest in the SSP3-7.0 scenario and lowest in the SSP1-2.6 scenario for the major representative regions of the world, although the main drivers for the projected outcomes differ.

Deleted: Thus, in the middle of the 21st century, climate change would considerably increase the $PM_{2.5}$ concentrations and cause considerable damage to human health in central Africa.

Deleted: Derived from the disproportionate spatial and temporal asymmetry under four SSP scenarios, the $PM_{2.5}$ exposure concentrations that coalesced with the future geographically demographic information can reveal the health intimidation to people from the future. Figures S14 and S15 show the demographic projections for the four SSPs scenarios for the world and for different regions, respectively.

Deleted: 16

In Europe and North America, where $PM_{2.5}$ concentrations will be relatively low, the population distribution is the main determinant of $PM_{2.5}$ exposure. Spatially-averaged $PM_{2.5}$ concentrations will be lower in the SSP5-8.5 scenario owing to the stronger pollution control measures than in the "middle of the road" SSP2-4.5 scenario, but the population-weighted $PM_{2.5}$ concentrations in SSP5-8.5 will slightly exceed those of SSP2-4.5 and even surpass those of SSP3-7.0 after the 2060s. These trends will be caused by the higher birthrate in Europe and North America in SSP5-8.5 driven by economic optimism and international migration, leading to accelerated population growth in these two regions (Figures S25 and S28).[71] This implies that a greater share of the population will be concentrated in areas with higher levels of social development and education. Therefore, compared with SSP2-4.5, the SSP5-8.5 scenario will result in a higher population-weighted $PM_{2.5}$ exposure in North America and Europe after the 2060s.

Deleted: Space-weighted

Deleted: 1

Deleted: 17

In both Asia and Africa, $PM_{2.5}$ exposure will decline steadily over time, reaching −58.2% (−47.3%) and −52.5% (−32.0%) for Asia (Africa) by the end of the century under the SSP1-2.6 and SSP5-8.5 scenarios, respectively, compared with the baseline period. However, there will be no significant decline under the SSP3-7.0 scenario, and before the 2060s, the exposure levels will be even higher than in the baseline period. Two explanations can be offered for the persistently high exposure concentrations in Asia and Africa under the SSP3-7.0 scenario. The emissions and unfavorable meteorological factors will lead to increased $PM_{2.5}$ pollution under this scenario before the 2030s. Meanwhile, population increase due to high fertility accompanied by slow urbanization in these regions will intensify the density of urban and rural settlement patterns, thereby increasing $PM_{2.5}$ exposure.[71]

The proportion of the population that would be exposed to the $PM_{2.5}$ concentration below previous and current Air Quality Guideline (AQG) values is also estimated under future climate change scenarios. As shown in Figure S29, the differences between SSP1-2.6 scenario and the other three scenarios are considerable. Compared with the other three scenarios, SSP1-2.6 would result in the largest fraction of the population exposed to the $PM_{2.5}$ level that is lower than 5 μg/m³. In the SSP1-2.6 scenario, 3.5% of the world's population will live in areas that have $PM_{2.5}$ concentrations lower than 5 μg/m³ by 2100, which is well above the baseline (2.0%). The other scenarios are comparable in terms of the proportion of the population exposed to the $PM_{2.5}$ concentration that is below the current AQG values.

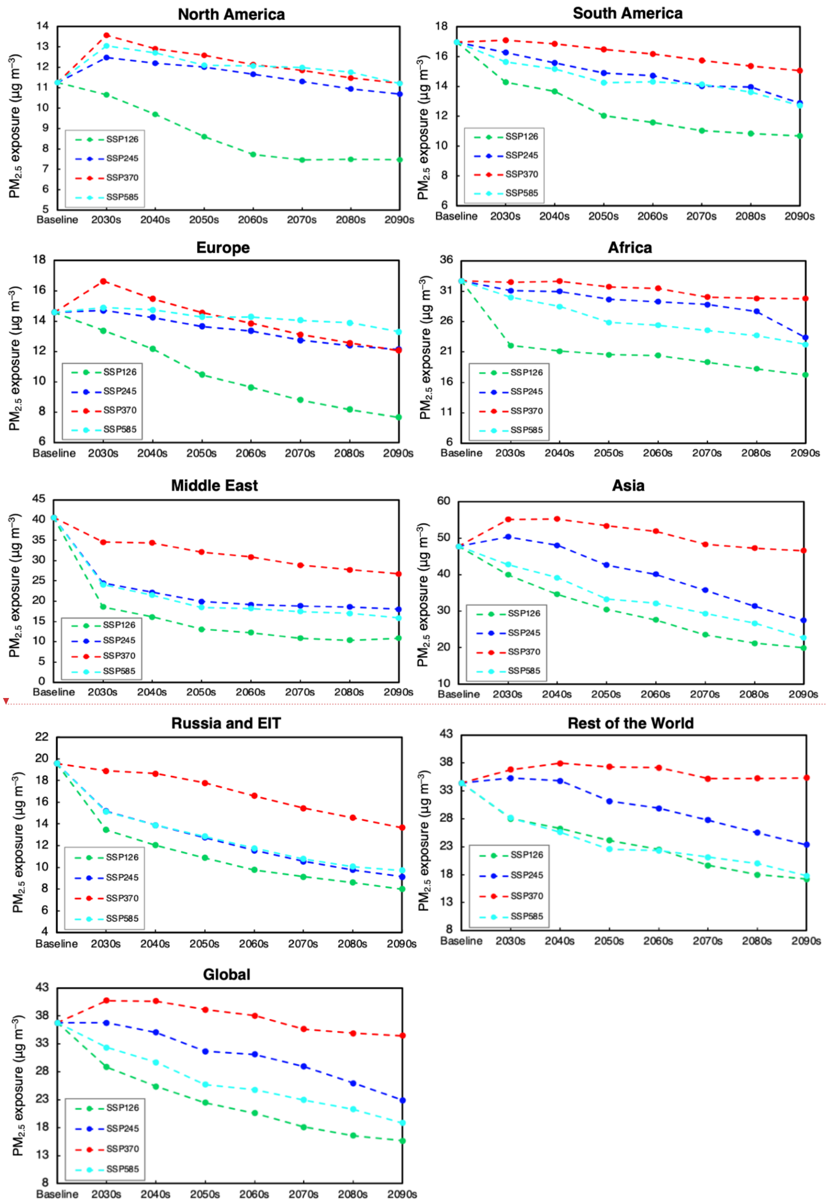


Figure 4. Projected ambient $PM_{2.5}$ exposure concentrations for 2030–2100 under different climate change scenarios.

Deleted: We also estimated the proportion of the population that would be exposed to the previous and current Air Quality Guideline (AQG) values under future climate change scenarios. As shown in Figure S18, the trends in the population fraction exposed to the AQG values of 10 μg/m³ and 5 μg/m³ are similar for the four climate change scenarios, although there are considerable differences in the magnitude of the population fraction that would be exposed. By 2100, in the SSP1-2.6 scenario, 3.5% of the world's population will live in areas that have $PM_{2.5}$ concentrations lower than 5 μg/m³, which is well above the baseline population fraction of 2.0%. Compared with the other three scenarios, SSP1-2.6 would emerge victorious with tremendous benefits to global public health. Once SSP1-2.6 is not approachable, the other scenarios are comparable in terms of the proportion of the population exposed to the two AQG values.

3.4. Projection of premature mortality burden

The global premature mortality burden associated with future $PM_{2.5}$ concentrations under the different SSP scenarios was also analyzed. Figure 5 shows the $PM_{2.5}$-associated premature deaths for the baseline (2010–2019) and future (2030–2100) periods in several representative regions. The green growth and sustainable development assumptions in the SSP1-2.6 scenario would lead to a rapid reduction in air pollution emissions globally. Therefore, the number of $PM_{2.5}$-associated premature deaths worldwide would start to decline in the near future (2031–2040) before the population growth turning point (2071–2080). Given the middle-road development pattern of SSP2-4.5, premature deaths in this scenario would peak at 9,024,000 (95% Confidence Interval (CI): 6,352,000–11,236,000) in the 2060s and then steadily decline to 7,394,000 (95% CI: 5,202,000–9,291,000) in the final decade of the century, which is a less rapid decline than in the SSP1-2.6 scenario. SSP3-7.0 assumes weak pollution control in which the implementation of pollution mitigation measures is delayed and less ambitious in the long term. In this scenario, premature deaths would spike dramatically in all regions except North America, Europe, and Russia and would not decrease until the end of the century. The global number of $PM_{2.5}$-associated premature deaths would reach 11,149,000 (95% CI: 7,877,000–13,800,000) in 2091–2100, an increase of 63% from the baseline period. In the SSP5-8.5 scenario, which emphasizes technological progress and rapid economic growth through human capital development, environmental issues become a priority health concern, and ambitious air quality goals result in pollutant levels well below current levels in the medium to long term.[16, 72] Therefore, in SSP5-8.5, global premature deaths would peak at 8,509,000 (95% CI: 5,981,000–10,617,000) in the 2040s and then decline to 6,258,000 (95% CI: 4,410,000–7,887,000) in the second half of the 21st century as high-performance pollution control technologies are developed. This decrease would result in a smaller premature death burden than in the baseline period.

Deleted: 9,023,922 (95% CI: 6,352,113–11,236,028)

Deleted: 7,393,925 (95% CI: 5,202,070–9,290,539)

Deleted: 11,148,502 (95% CI: 7,876,580–13,800,471)

Deleted: 8,508,685 (95% CI: 5,980,955–10,617,435)

Deleted: 6,257,869 (95% CI: 4,410,176–7,886,851)

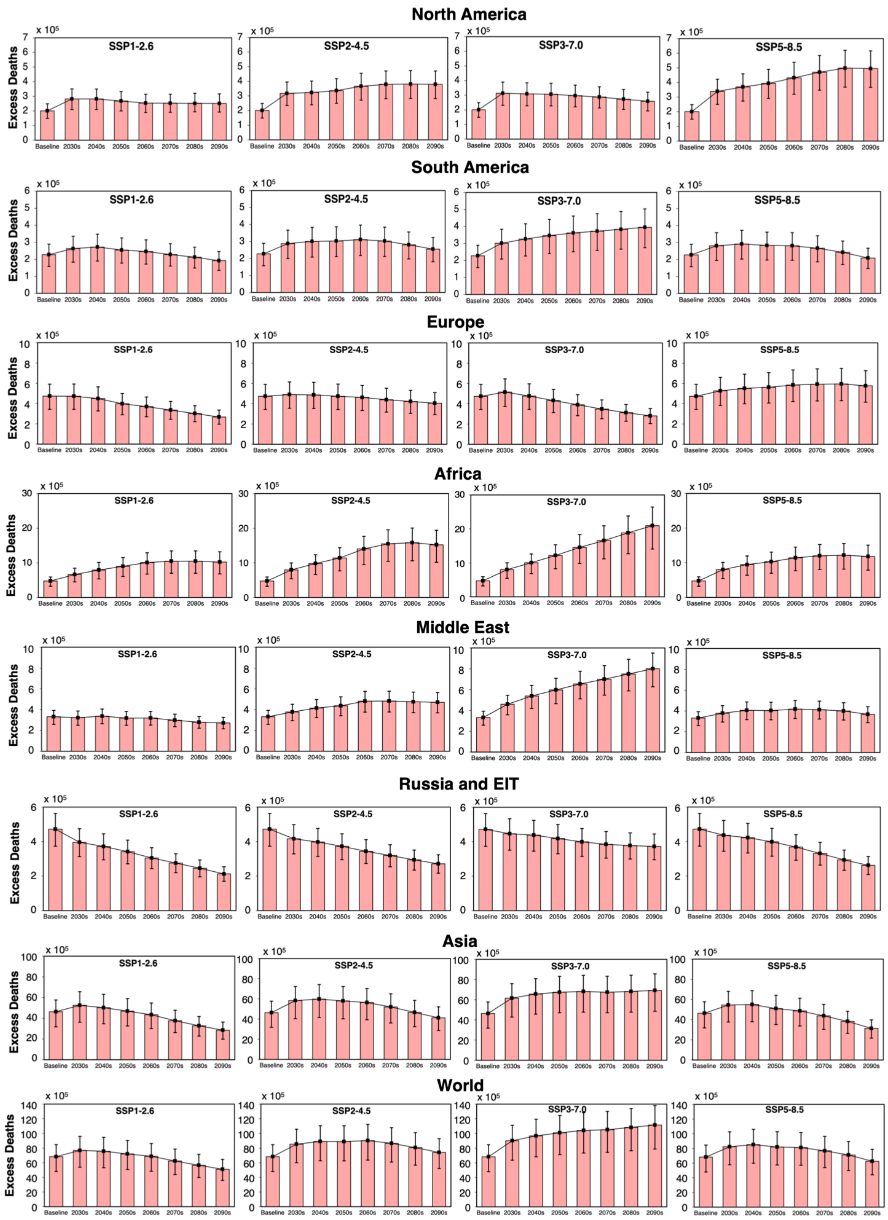


Figure 5. $PM_{2.5}$-associated premature deaths (> 25 years old) in different regions. The red bars represent premature deaths, and the vertical black lines indicate the 95% empirical confidence intervals.

Deleted: Figure 5. $PM_{2.5}$-associated premature deaths (> 25 years old) in different regions. The red bars represent the premature mortality rate, and the vertical black lines indicate the 95% empirical confidence intervals.

Because the size of the population will determine the absolute number of premature deaths, country-specific mortality rates per 100,000 people were used to describe the $PM_{2.5}$-associated mortality burden. Figure S30 shows the mortality rate per 100,000 people for 184 countries or districts. Overall, countries in North America, Western Europe, and Oceania will have the lowest mortality rates. The mortality rates in Eastern European countries (e.g., Ukraine and Serbia) will be the highest, followed by some countries in Asia, such as China and India.

Deleted: 19

3.5. Key factors that influence the premature mortality burden

Two sensitivity studies were conducted to explore how future population distributions and $PM_{2.5}$ concentrations would affect the burden of $PM_{2.5}$-associated premature mortality (Table S8). In the first sensitivity experiment (SA1), the only contributor to the difference in the estimated premature deaths from the baseline period is the demographic transition, while the contributor to the difference in the second sensitivity experiment (SA2) is the $PM_{2.5}$ variation.

Deleted: 4

Deleted: In the first (SA1) and second (SA2) sensitivity experiment, the population size and the $PM_{2.5}$ concentration was the same as that in 2010–2019, respectively.

When considering only the future demographic projections (i.e., demographic changes and changes in total population by age), the changes in the population distribution over the coming decades (2021–2040) will exacerbate the burden of premature deaths in all four scenarios, but the magnitude of the effect differs among the scenarios. These differences are reflected in the demographic assumptions about the birthrate, mortality, and migration.[73] SSP1-2.6 and SSP5-8.5 both envision a development path of increased investment in education and health, thereby accelerating the demographic transition.[71] Therefore, in these two scenarios, the demographic turning point in population decline will be reached earlier, in the medium term (2050s) (Figure S25), after which the impact of demographics on the burden of premature mortality will gradually decrease.

Deleted: .

Deleted: 13

In the second sensitivity experiment, we explore the effect of the $PM_{2.5}$ concentration on premature mortality assuming a constant future population distribution and size. Disease burden alleviation resulting from implementing air pollution control measures will become apparent in the near future under the SSP1-2.6 and SSP5-8.5 scenarios. SSP1-2.6 is the only scenario in which the effect of the $PM_{2.5}$ concentration will be greater than the effect of population size by the end of the century. Rapidly declining emissions would successfully offset the burden of premature mortality resulting from population growth by the end of the century. Under the SSP3-7.0 scenario, the planetary boundary layer height (PBLH) exerts strong influence on $PM_{2.5}$ dispersion, and thus its decreases in East Asia, South Asia, and eastern Africa (Figure S10) will increase the $PM_{2.5}$ concentrations. Besides PBLH, other meteorological conditions, such as higher temperature,[74] are also favorable for $PM_{2.5}$ accumulation in these regions and therefore exacerbate the $PM_{2.5}$-associated mortality burden until the 2050s.

Deleted: Under the SSP3-7.0 scenario, the planetary boundary layer height critically influences $PM_{2.5}$ dispersion, and it decreases in East Asia, South Asia, and eastern Africa (Figure S7). The decrease in the planetary boundary layer height will increase the $PM_{2.5}$ concentrations and therefore exacerbate the $PM_{2.5}$-associated mortality burden until the 2050s in these regions.

3.6. Implications and limitations

Global climate change is a significant challenge for society, and its impact on future air pollution is a critical perspective that requires quantitative assessment. Herein, a global $PM_{2.5}$ concentration dataset with a spatial resolution of 0.1° × 0.1° was estimated based on SSP scenarios.

Deleted: The results showed that the global $PM_{2.5}$ concentrations and the associated $PM_{2.5}$ exposure and premature mortality burden vary considerably under the four SSP scenarios. Among the scenarios, SSP1-2.6 would have the earliest inflection point for $PM_{2.5}$-associated premature deaths and the lowest mortality burden. This scenario presents an ideal target pathway that governments should strive to achieve.

From the methodology and dataset perspectives, this work provides a new set of global-scale future $PM_{2.5}$ dataset in 10km spatial resolution. This dataset can be used by others for air quality-related studies at the national and even regional scales. The dataset can be downloaded from the link listed in Text S8. The method developed in this work can also be implemented for other air pollution-related research. Researchers can further develop other more advanced deep learning frameworks for relevant studies based on the design of the method proposed in this work.

From the results perspective, this work has quantified how the future $PM_{2.5}$ and its associated adverse health impacts will change based on different SSP scenarios. Based on our results, governments and relevant stakeholders from different countries can generally understand to what extent can $PM_{2.5}$ influence their specific local health burdens. This can provide useful scientific references for future air pollution control policy design. In addition, when other studies come out in the future, the results from this work can also be used for the comparison. For example, compared to the adverse effects caused by other pollutants, such as O_3, which pollutant should the government put onto the priority position under different SSP scenarios.

From the health-economic impact perspective, the results of economic burdens shed light on the relationship between mortality cost that is associated with $PM_{2.5}$ pollution and economic development in various countries under different future scenarios. The economic burdens related to future $PM_{2.5}$ pollution are discussed in Text S9. Figure S31 and Figure S32 show the economic loss that is associated with $PM_{2.5}$-related health burdens. For most OECD countries, China, and Central Asia, air pollution mitigation and economic development can have a beneficial synergistic effect. The ratio of economic loss associated with $PM_{2.5}$ pollution over the total GDP (PPP based) is minimal in the sustainable development scenario (SSP1-2.6). For Central Africa and South America, from $PM_{2.5}$ associated economic loss perspective, these countries may consider choosing SSP2-4.5 pathway as their development modes.

This work has some limitations. First, satellite-retrieved $PM_{2.5}$ datasets were used as training targets, but according to the results in Table S6, there were discrepancies with the observational data obtained from ground measurements. Second, future climate, emission, and population projections harbor relatively large uncertainty, even if they have been calibrated against observed patterns of changes using historical data.[23, 71] Third, there are no generalizable and accurate findings that indicate how baseline mortality rates will change in the future. Therefore, in accordance with previous studies[75, 76] in the projection literature, we assumed that the nonlinear relationship between $PM_{2.5}$ concentrations and the baseline mortality rate would also be consistent. Fourth, the $PM_{2.5}$ projections derived in this study were based on several underlying assumptions. Primarily, in line with previous works,[75, 76] the relationships between $PM_{2.5}$ concentrations and meteorological conditions and precursor emissions explored in this study were assumed to be true for future climate and emissions scenarios. Fifth, our predictions were based on the premise that the world is steadily developing, and our method cannot predict the effects of uncontrollable factors (such as war and strong earthquakes) on $PM_{2.5}$ and population distributions. Finally, the biases in emissions data (e.g., bias in future wildfires and missing windblown dust), can be directly propagated to the air pollution concentration estimation. Thus, $PM_{2.5}$ projections in this work contain unavoidable uncertainty. The spatial pattern of windblown dust was not included in this study, which may have an influence on our results, especially for the Sahara and Middle East. Given the proximity of these regions to large sources of dust emissions, there is a possibility that an underestimation of $PM_{2.5}$ concentrations would occur in these regions. However, the impact on the mortality estimations is limited since these regions are more sparsely populated. Despite these limitations, this work helps quantify the extent to which climate change will influence $PM_{2.5}$ concentrations worldwide. The results can contribute to the ongoing assessment of $PM_{2.5}$-associated exposure and vulnerability under different climate change scenarios, and governments can use this information to design useful strategies to reduce pollution.

Deleted: 2

Deleted: Relative uncertainty generally increases over time because detailed spatial and temporal information is unavailable.

Deleted: Finally

Deleted: the

Acknowledgment

The work described in this paper was supported by National Natural Science Foundation of China (Project No. 42007203), Guangzhou Scientific and Technological Planning Project (Project No. 202102021297), the Research Grants Council of Hong Kong Government (Project No.16305921), the Environment and Conservation Fund (ECF 2020123), and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. AoE/E-603/18 and T31-603/21-N).

Reference

1. Cohen, A. J.; Brauer, M.; Burnett, R.; Anderson, H. R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R., 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. *The Lancet* **2017,** *389*, (10082), 1907-1918.
2. Forouzanfar, M. H.; Afshin, A.; Alexander, L. T.; Anderson, H. R.; Bhutta, Z. A.; Biryukov, S.; Brauer, M.; Burnett, R.; Cercy, K.; Charlson, F. J., Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* **2016,** *388*, (10053), 1659-1724.
3. Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; AlMazroa, M. A.; Amann, M.; Anderson, H. R.; Andrews, K. G., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **2012,** *380*, (9859), 2224-2260.
4. Murray, C. J.; Aravkin, A. Y.; Zheng, P.; Abbafati, C.; Abbas, K. M.; Abbasi-Kangevari, M.; Abd-Allah, F.; Abdelalim, A.; Abdollahi, M.; Abdollahpour, I., Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **2020,** *396*, (10258), 1223-1249.
5. Costello, A.; Abbas, M.; Allen, A.; Ball, S.; Bell, S.; Bellamy, R.; Friel, S.; Groce, N.; Johnson, A.; Kett, M., Managing the health effects of climate change: lancet and University College London Institute for Global Health Commission. *The Lancet* **2009,** *373*, (9676), 1693-1733.
6. Wang, H.-J.; Chen, H.-P., Understanding the recent trend of haze pollution in eastern China: roles of climate change. *Atmospheric Chemistry and Physics* **2016,** *16*, (6), 4205-4211.
7. Hong, C.; Zhang, Q.; Zhang, Y.; Davis, S. J.; Tong, D.; Zheng, Y.; Liu, Z.; Guan, D.; He, K.; Schellnhuber, H. J., Impacts of climate change on future air quality and human health in China. *Proceedings of the National Academy of Sciences* **2019,** *116*, (35), 17193-17200.
8. Chowdhury, S.; Dey, S.; Smith, K. R., Ambient PM2. 5 exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nature communications* **2018,** *9*, (1), 1-10.
9. O'Neill, B. C.; Tebaldi, C.; Vuuren, D. P. v.; Eyring, V.; Friedlingstein, P.; Hurtt, G.; Knutti, R.; Kriegler, E.; Lamarque, J.-F.; Lowe, J., The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development* **2016,** *9*, (9), 3461-3482.
10. Gidden, M. J.; Riahi, K.; Smith, S. J.; Fujimori, S.; Luderer, G.; Kriegler, E.; Vuuren, D. P. v.; Berg, M. v. d.; Feng, L.; Klein, D., Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development* **2019,** *12*, (4), 1443-1475.
11. Shim, S.; Sung, H.; Kwon, S.; Kim, J.; Lee, J.; Sun, M.; Song, J.; Ha, J.; Byun, Y.; Kim, Y., Regional Features of Long-Term Exposure to PM2. 5 Air Quality over Asia under SSP Scenarios Based on CMIP6 Models. *International journal of environmental research and public health* **2021,** *18*, (13), 6817.
12. Turnock, S. T.; Allen, R. J.; Andrews, M.; Bauer, S. E.; Deushi, M.; Emmons, L.; Good, P.; Horowitz, L.; John, J. G.; Michou, M., Historical and future changes in air pollutants from CMIP6 models. *Atmospheric Chemistry and Physics* **2020,** *20*, (23), 14547-14579.
13. Cheng, J.; Tong, D.; Liu, Y.; Yu, S.; Yan, L.; Zheng, B.; Geng, G.; He, K.; Zhang, Q., Comparison of current and future PM2. 5 air quality in China under CMIP6 and DPEC emission scenarios. *Geophysical Research Letters* **2021,** *48*, (11), e2021GL093197.

Deleted: 2. Collaborators, G. R. F., Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet (London, England) 2016, 388, (10053), 1659.

14. Cheng, J.; Tong, D.; Zhang, Q.; Liu, Y.; Lei, Y.; Yan, G.; Yan, L.; Yu, S.; Cui, R. Y.; Clarke, L., Pathways of China's PM2. 5 air quality 2015–2060 in the context of carbon neutrality. *National science review* **2021,** *8*, (12), nwab078.
15. Liu, S.; Xing, J.; Westervelt, D. M.; Liu, S.; Ding, D.; Fiore, A. M.; Kinney, P. L.; Zhang, Y.; He, M. Z.; Zhang, H., Role of emission controls in reducing the 2050 climate change penalty for PM2. 5 in China. *Science of the Total Environment* **2021,** *765*, 144338.
16. Rao, S.; Klimont, Z.; Smith, S. J.; Van Dingenen, R.; Dentener, F.; Bouwman, L.; Riahi, K.; Amann, M.; Bodirsky, B. L.; van Vuuren, D. P., Future air pollution in the Shared Socio-economic Pathways. *Global Environmental Change* **2017,** *42*, 346-358.
17. Van Vuuren, D. P.; Stehfest, E.; Gernaat, D. E.; Doelman, J. C.; Van den Berg, M.; Harmsen, M.; de Boer, H. S.; Bouwman, L. F.; Daioglou, V.; Edelenbosch, O. Y., Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. *Global Environmental Change* **2017,** *42*, 237-250.
18. Thomson, A. M.; Calvin, K. V.; Smith, S. J.; Kyle, G. P.; Volke, A.; Patel, P.; Delgado-Arias, S.; Bond-Lamberty, B.; Wise, M. A.; Clarke, L. E., RCP4. 5: a pathway for stabilization of radiative forcing by 2100. *Climatic change* **2011,** *109*, (1), 77-94.
19. Fujimori, S.; Hasegawa, T.; Masui, T.; Takahashi, K.; Herran, D. S.; Dai, H.; Hijioka, Y.; Kainuma, M., SSP3: AIM implementation of shared socioeconomic pathways. *Global Environmental Change* **2017,** *42*, 268-283.
20. Riahi, K.; Rao, S.; Krey, V.; Cho, C.; Chirkov, V.; Fischer, G.; Kindermann, G.; Nakicenovic, N.; Rafaj, P., RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic change* **2011,** *109*, (1), 33-57.
21. Ronneberger, O.; Fischer, P.; Brox, T., U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, Springer: 2015; pp 234-241.
22. Bosilovich, M. G., *MERRA-2: Initial evaluation of the climate*. National Aeronautics and Space Administration, Goddard Space Flight Center: 2015.
23. Feng, L.; Smith, S. J.; Braun, C.; Crippa, M.; Gidden, M. J.; Hoesly, R.; Klimont, Z.; Van Marle, M.; Van Den Berg, M.; Van Der Werf, G. R., The generation of gridded emissions data for CMIP6. *Geoscientific Model Development* **2020,** *13*, (2), 461-482.
24. van Donkelaar, A.; Hammer, M. S.; Bindle, L.; Brauer, M.; Brook, J. R.; Garay, M. J.; Hsu, N. C.; Kalashnikova, O. V.; Kahn, R. A.; Lee, C., Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology* **2021,** *55*, (22), 15287-15300.
25. Hammer, M. S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A. M.; Hsu, N. C.; Levy, R. C.; Garay, M. J.; Kalashnikova, O. V.; Kahn, R. A., Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environmental Science & Technology* **2020,** *54*, (13), 7879-7890.
26. Van Donkelaar, A.; Martin, R. V.; Brauer, M.; Hsu, N. C.; Kahn, R. A.; Levy, R. C.; Lyapustin, A.; Sayer, A. M.; Winker, D. M., Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology* **2016,** *50*, (7), 3762-3772.
27. Van Donkelaar, A.; Martin, R. V.; Brauer, M.; Boys, B. L., Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environmental health perspectives* **2015,** *123*, (2), 135-143.
28. Gelaro, R.; McCarty, W.; Suárez, M. J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C. A.; Darmenov, A.; Bosilovich, M. G.; Reichle, R., The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate* **2017,** *30*, (14), 5419-5454.
29. Lu, X.; Yuan, D.; Chen, Y.; Fung, J. C., Impacts of urbanization and long-term meteorological variations on global PM2. 5 and its associated health burden. *Environmental Pollution* **2021,** *270*, 116003.
30. Theresa, N. M.; Sempewo, J. I.; Tumutungire, M. D.; Byakatonda, J., Performance evaluation of CFSR, MERRA-2

and TRMM3B42 data sets in simulating river discharge of data-scarce tropical catchments: a case study of Manafwa, Uganda. *Journal of Water and Climate Change* **2021**.

31. Huang, L.; Guo, L.; Liu, L.; Chen, H.; Chen, J.; Xie, S., Evaluation of the ZWD/ZTD values derived from MERRA-2 global reanalysis products using GNSS observations and radiosonde data. *Sensors* **2020,** *20*, (22), 6440.

32. Carvalho, D., An assessment of NASA's GMAO MERRA-2 reanalysis surface winds. *Journal of Climate* **2019,** *32*, (23), 8261-8281.

33. Wen, A.; Wu, T.; Wu, X.; Zhu, X.; Li, R.; Ni, J.; Hu, G.; Qiao, Y.; Zou, D.; Chen, J., Evaluation of MERRA-2 land surface temperature dataset and its application in permafrost mapping over China. *Atmospheric Research* **2022,** *279*, 106373.

34. Zhang, J.; Zhao, T.; Li, Z.; Li, C.; Li, Z.; Ying, K.; Shi, C.; Jiang, L.; Zhang, W., Evaluation of Surface Relative Humidity in China from the CRA-40 and Current Reanalyses. *Advances in Atmospheric Sciences* **2021,** *38*, (11), 1958-1976.

35. Meng, J.; Liu, J.; Xu, Y.; Guan, D.; Liu, Z.; Huang, Y.; Tao, S., Globalization and pollution: tele-connecting local primary PM2. 5 emissions to global consumption. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016,** *472*, (2195), 20160380.

36. Huang, Y.; Shen, H.; Chen, H.; Wang, R.; Zhang, Y.; Su, S.; Chen, Y.; Lin, N.; Zhuo, S.; Zhong, Q., Quantification of global primary emissions of PM2. 5, PM10, and TSP from combustion and industrial process sources. *Environmental science & technology* **2014,** *48*, (23), 13834-13843.

37. Cheng, J.; Tong, D.; Liu, Y.; Bo, Y.; Zheng, B.; Geng, G.; He, K.; Zhang, Q., Air quality and health benefits of China's current and upcoming clean air policies. *Faraday Discussions* **2021,** *226*, 584-606.

38. Hsieh, I.-Y. L.; Chossière, G. P.; Gençer, E.; Chen, H.; Barrett, S.; Green, W. H., An Integrated Assessment of Emissions, Air Quality, and Public Health Impacts of China's Transition to Electric Vehicles. *Environmental Science & Technology* **2022**.

39. Almazroui, M.; Saeed, S.; Saeed, F.; Islam, M. N.; Ismail, M., Projections of precipitation and temperature over the South Asian countries in CMIP6. *Earth Systems and Environment* **2020,** *4*, (2), 297-320.

40. Almazroui, M.; Saeed, S.; Islam, M. N.; Khalid, M. S.; Alkhalaf, A. K.; Dambul, R., Assessment of uncertainties in projected temperature and precipitation over the Arabian Peninsula: a comparison between different categories of CMIP3 models. *Earth Systems and Environment* **2017,** *1*, (2), 1-21.

41. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P., Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998,** *86*, (11), 2278-2324.

42. Rybarczyk, Y.; Zalakeviciute, R., Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences* **2018,** *8*, (12), 2570.

43. Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A., Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment* **2012,** *60*, 632-655.

44. Catalano, M.; Galatioto, F.; Bell, M.; Namdeo, A.; Bergantino, A. S., Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environmental science & policy* **2016,** *60*, 69-83.

45. Meehl, G. A.; Moss, R.; Taylor, K. E.; Eyring, V.; Stouffer, R. J.; Bony, S.; Stevens, B., Climate model intercomparisons: Preparing for the next phase. *Eos, Transactions American Geophysical Union* **2014,** *95*, (9), 77-78.

46. O'Neill, B. C.; Kriegler, E.; Riahi, K.; Ebi, K. L.; Hallegatte, S.; Carter, T. R.; Mathur, R.; van Vuuren, D. P., A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic change* **2014,** *122*, (3), 387-400.

47. Van Vuuren, D. P.; Kriegler, E.; O'Neill, B. C.; Ebi, K. L.; Riahi, K.; Carter, T. R.; Edmonds, J.; Hallegatte, S.; Kram,

Deleted: 33. Wen, A.; Wu, X.; Zhu, X.; Hu, G.; Qiao, Y.; Wang, D.; Lou, P., Evaluation of Merra-2 Land Surface Temperature Dataset and its Application in Permafrost Mapping Over China. *Xiadong and Zhu, Xiaofan and li, ren and ni, jie and Hu, Guojie and Qiao, Yongping and zou, defu and chen, Jie and Wang, Dong and Lou, Peiqin, Evaluation of Merra-2 Land Surface Temperature Dataset and its Application in Permafrost Mapping Over China.*

T.; Mathur, R., A new scenario framework for climate change research: scenario matrix architecture. *Climatic Change* **2014,** *122*, (3), 373-386.

48. Kriegler, E.; Bauer, N.; Popp, A.; Humpenöder, F.; Leimbach, M.; Strefler, J.; Baumstark, L.; Bodirsky, B. L.; Hilaire, J.; Klein, D., Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century. *Global environmental change* **2017,** *42*, 297-315.

49. Van Vuuren, D. P.; Stehfest, E.; den Elzen, M. G.; Kram, T.; van Vliet, J.; Deetman, S.; Isaac, M.; Klein Goldewijk, K.; Hof, A.; Mendoza Beltran, A., RCP2. 6: exploring the possibility to keep global mean temperature increase below 2 C. *Climatic change* **2011,** *109*, (1), 95-116.

50. Eyring, V.; Bony, S.; Meehl, G. A.; Senior, C. A.; Stevens, B.; Stouffer, R. J.; Taylor, K. E., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **2016,** *9*, (5), 1937-1958.

51. Gidden, M. J.; Riahi, K.; Smith, S. J.; Fujimori, S.; Luderer, G.; Kriegler, E.; Van Vuuren, D. P.; Van Den Berg, M.; Feng, L.; Klein, D., Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development* **2019,** *12*, (4), 1443-1475.

52. Ramírez Villegas, J.; Jarvis, A., Downscaling global circulation model outputs: the delta method decision and policy analysis Working Paper No. 1. *International Center for Tropical Agriculture (CIAT). Cali. CO.* **2010,** 18.

53. Hay, L. E.; Wilby, R. L.; Leavesley, G. H., A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States 1. *JAWRA Journal of the American Water Resources Association* **2000,** *36*, (2), 387-397.

54. Navarro-Racines, C.; Tarapues, J.; Thornton, P.; Jarvis, A.; Ramirez-Villegas, J., High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. *Scientific data* **2020,** *7*, (1), 1-14.

55. Hawkins, E.; Osborne, T. M.; Ho, C. K.; Challinor, A. J., Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agricultural and forest meteorology* **2013,** *170*, 19-31.

56. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope Iii, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S., Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* **2018,** *115*, (38), 9592-9597.

57. Gao, J., Downscaling global spatial population projections from 1/8-degree to 1-km grid cells. *National Center for Atmospheric Research, Boulder, CO, USA* **2017,** *1105*.

58. Yang, J.; Zhou, M.; Ren, Z.; Li, M.; Wang, B.; Liu, D. L.; Ou, C.-Q.; Yin, P.; Sun, J.; Tong, S., Projecting heat-related excess mortality under climate change scenarios in China. *Nature communications* **2021,** *12*, (1), 1-11.

59. Rohat, G.; Wilhelmi, O.; Flacke, J.; Monaghan, A.; Gao, J.; Dao, H.; van Maarseveen, M., Characterizing the role of socioeconomic pathways in shaping future urban heat-related challenges. *Science of the total environment* **2019,** *695*, 133941.

60. Luo, C.; Posen, I. D.; Hoornweg, D.; MacLean, H. L., Modelling future patterns of urbanization, residential energy use and greenhouse gas emissions in Dar es Salaam with the Shared Socio-Economic Pathways. *Journal of Cleaner Production* **2020,** *254*, 119998.

61. Wang, Q.; Wang, J.; Zhou, J.; Ban, J.; Li, T., Estimation of PM2· 5-associated disease burden in China in 2020 and 2030 using population and air quality scenarios: a modelling study. *The Lancet Planetary Health* **2019,** *3*, (2), e71-e80.

62. Ingole, V.; Dimitrova, A.; Sampedro, J.; Sacoor, C.; Acacio, S.; Juvekar, S.; Roy, S.; Moraga, P.; Basagaña, X.; Ballester, J., Local mortality impacts due to future air pollution under climate change scenarios. *Science of the Total Environment* **2022,** *823*, 153832.

Deleted: 52. Ramírez Villegas, J.; Jarvis, A., Downscaling global circulation model outputs: the delta method decision and policy analysis Working Paper No. 1. **2010.**
53. Hay, L. E.; Wilby, R. L.; Leavesley, G. H. J. J. o. t. A. W. R. A., A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States 1. **2000,** *36*, (2), 387-397.
54. Navarro-Racines, C.; Tarapues, J.; Thornton, P.; Jarvis, A.; Ramirez-Villegas, J. J. S. d., High-resolution and bias-corrected CMIP5 projections for climate change impact assessments. **2020,** *7*, (1), 1-14.
55. Hawkins, E.; Osborne, T. M.; Ho, C. K.; Challinor, A. J. J. A.; meteorology, f., Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. **2013,** *170*, 19-31.
56. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S. J. P. o. t. N. A. o. S., Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. **2018,** *115*, (38), 9592-9597.51. Gidden, M. J.; Riahi, K.; Smith, S. J.; Fujimori, S.; Luderer, G.; Kriegler, E.; Vuuren, D. P. v.; Berg, M. v. d.; Feng, L.; Klein, D. J. G. m. d., Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. **2019,** *12*, (4), 1443-1475.50. Eyring, V.; Bony, S.; Meehl, G. A.; Senior, C. A.; Stevens, B.; Stouffer, R. J.; Taylor, K. E. J. G. M. D., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. 2016, 9, (5), 1937-1958.

63. Park, Y.; Kwon, B.; Heo, J.; Hu, X.; Liu, Y.; Moon, T., Estimating PM2. 5 concentration of the conterminous United States via interpretable convolutional neural networks. *Environmental Pollution* **2020,** *256*, 113395.
64. Zheng, T.; Bergin, M. H.; Hu, S.; Miller, J.; Carlson, D. E., Estimating ground-level PM2. 5 using micro-satellite images by a convolutional neural network and random forest approach. *Atmospheric Environment* **2020,** *230*, 117451.
65. Li, S.; Xie, G.; Ren, J.; Guo, L.; Yang, Y.; Xu, X., Urban pm2. 5 concentration prediction via attention-based cnn–lstm. *Applied Sciences* **2020,** *10*, (6), 1953.
66. Xiao, F.; Yang, M.; Fan, H.; Fan, G.; Al-Qaness, M. A., An improved deep learning model for predicting daily PM2. 5 concentration. *Scientific reports* **2020,** *10*, (1), 1-11.
67. Yan, X.; Zang, Z.; Luo, N.; Jiang, Y.; Li, Z., New interpretable deep learning model to monitor real-time PM2. 5 concentrations from satellite data. *Environment International* **2020,** *144*, 106060.
68. Riahi, K.; Van Vuuren, D. P.; Kriegler, E.; Edmonds, J.; O'neill, B. C.; Fujimori, S.; Bauer, N.; Calvin, K.; Dellink, R.; Fricko, O., The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global environmental change* **2017,** *42*, 153-168.
69. Wang, Y.; Hu, J.; Huang, L.; Li, T.; Yue, X.; Xie, X.; Liao, H.; Chen, K.; Wang, M., Projecting future health burden associated with exposure to ambient PM2. 5 and ozone in China under different climate scenarios. *Environment International* **2022,** *169*, 107542.
70. Chowdhury, S.; Dey, S.; Smith, K. R., Ambient PM2. 5 exposure and expected premature mortality to 2100 in India under climate change scenarios. *Nature communications* **2018,** *9*, (1), 318.
71. Jones, B.; O'Neill, B. C., Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environmental Research Letters* **2016,** *11*, (8), 084003.
72. Kriegler, E.; Bauer, N.; Popp, A.; Humpenöder, F.; Leimbach, M.; JS, B. L.; Bodirsky, B.; Hilaire, J.; Klein, D.; Mouratiadou, I., Fossil-fueled development (SSP5): an emission, energy and resource intensive reference scenario for the 21 st century. *Global Environ. Change* **2016**.
73. Samir, K.; Lutz, W., The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change* **2017,** *42*, 181-192.
74. Shi, L.; Liu, P.; Zanobetti, A.; Schwartz, J., Climate penalty: climate-driven increases in ozone and PM2. 5 levels and mortality. *Environmental Epidemiology* **2019,** *3*, 365.
75. Liu, S.; Xing, J.; Wang, S.; Ding, D.; Cui, Y.; Hao, J., Health benefits of emission reduction under 1.5 C pathways far outweigh climate-related variations in China. *Environmental Science & Technology* **2021,** *55*, (16), 10957-10966.
76. Aleluia Reis, L.; Drouet, L.; Van Dingenen, R.; Emmerling, J., Future global air quality indices under different socioeconomic and climate assumptions. *Sustainability* **2018,** *10*, (10), 3645.

Page 7: [1] Deleted Wanying CHEN 02/02/2023 21:00:00