

Supplemental material:

Long-term trends in urban NO₂ concentrations and associated pediatric asthma incidence: estimates from global datasets

Susan C. Anenberg, Arash Mohegh, Daniel L. Goldberg, Gaige H. Kerr, Michael Brauer, Katrin Burkart, Perry Hystad, Andrew Larkin, Sarah Wozniak, Lok Lamsal

Methods for adjusting surface NO₂ concentrations in rural areas

For grid cells >5km away from roadways and in rural areas, we developed new NO₂ concentration estimates using NO₂ column observations from the OMI satellite instrument with some adjustments to fill spatial and temporal gaps in the OMI satellite record, and to estimate 24-hour averages from the early afternoon OMI overpass time (Figure S1). We use an OMI NO₂ version 3, level 4 surface concentration dataset (0.1° x 0.1° resolution) for 2011, which followed methods described by Lamsal et al.¹ and was obtained from the NASA Goddard Space Flight Center (GSFC). The newer version 4 OMI retrieval uses enhanced surface reflectivities in the calculation of the tropospheric column amounts, but surface concentrations prepared by NASA GSFC are not currently available from the version 4 product. Due to the lack of satellite dataset coverage over snow/ice covered areas, some gridcells (mostly in higher latitudes) have no OMI observations in some months. We used the MERRA-2 reanalysis product (0.625° x 0.5° resolution) to generate a correction factor to ensure availability of NO₂ concentrations in all locations and months, as follows:

$$\text{Equation S1: } \textit{Correction factor\#1} = \frac{\textit{MERRA2}_{\textit{annual average}}}{\textit{MERRA2}_{\textit{average for same months that OMI level 4 is available}}}$$

We also applied a second correction factor to convert surface NO₂ concentrations from the early afternoon OMI overpass time (13:00 local time) to 24-hour averages. Following Anenberg et al.², we used NO₂ surface concentrations from the GMI-Replay chemical transport model^{3,4} (2° x 2.5° resolution) simulations to generate these correction factors, as follows:

$$\text{Equation S2: } \textit{Correction factor\#2} = \frac{\textit{GMI}_{\textit{24 hour average}}}{\textit{GMI}_{\textit{13:00}}}$$

The NO₂ surface concentration estimates used for gridcells >5km away from roads and in rural areas were then generated using the following formula:

$$\text{Equation S3: } \textit{Adjusted rural concentrations} = \textit{OMI level 4} \times \textit{Correction factor\#1} \times \textit{Correction factor \#2}$$

For rural gridcells within 5km of major roadways, we linearly scaled between Larkin et al.⁵ values and the new adjusted rural concentrations in the span of the 5 km distance. The result of these steps is a 1km x 1km annual average surface NO₂ concentration dataset for 2011 that uses Larkin et al.⁵ values in gridcells that are categorized as urban or over roads, and a new concentration dataset derived from OMI satellite observations in rural areas (Figure S2).

Methods for scaling NO₂ concentrations from 2011 to 1990-2019

The GBD requires NO₂ concentrations for each year included in the comparative risk assessment, from 1990-2019. We therefore scaled the new 2011 surface NO₂ concentration dataset to each year in this time period, in five-year

increments from 1990-2005, and annually from 2010-2019. For the years 2005-2019, we scaled surface NO₂ concentrations from 2011 to each year using 3-year rolling averages of annual average NO₂ columns from the OMI version 4.0 level 2 product (13 km x 25 km resolution at nadir; Figure S3) at the gridcell level. We use NO₂ columns because surface concentrations derived from the version 4 OMI retrieval are not yet available. We oversampled the column NO₂ dataset to 0.1° x 0.1° resolution and re-gridded to 0.0083° x 0.0083° (approximately 1km x 1km). The 3-year rolling averages remove noise from the satellite data. For 2005 and 2019, we did not have data to create 3-year rolling averages, so we used that year's NO₂ columns directly. The years 1990, 1995, and 2000 predated the OMI observational record. We therefore used NO₂ concentrations from the MERRA-2 reanalysis product to scale 2011 NO₂ concentrations to those years.⁶ To remove model noise, we created the MERRA-2 scaling factors across broad world regions (Figure S5), as opposed to applying scaling factors on a gridcell by gridcell basis as we did for the OMI scaling.

The final result used for estimating the global burden of disease from NO₂ is a global, 0.0083° x 0.0083° (approximately 1km x 1km) resolution dataset of annual average surface NO₂ concentrations from 1990-2019 (Figure 1).

Evaluation of NO₂ concentration dataset

The Larkin et al.⁵ NO₂ concentration dataset was evaluated extensively in that work and agreed well with ground observations in urban areas. Here we add two limited new analyses to evaluate the changes we made in rural areas and the scaling other years (focusing specifically on the latest year, 2019).

We evaluated the rural NO₂ concentration estimates using the European Monitoring and Evaluation Program (EMEP) ground monitoring dataset, which has a large number of stations in rural areas (Table S1 and Figure S6). Other ground monitoring datasets (e.g. from EPA) may have rural sites, but we found that most were located directly downwind from urban areas. For example, the average surface annual mean NO₂ concentration in rural areas in 2011 from the EPA network is 4.3 ppb, likely too high to represent true background concentrations. We aggregated the available monitoring stations for the year 2011 to calculate annual averages and used a set of criteria to filter for stations that mostly closely represent background concentrations: 1) Stations with >300 days of data (the threshold was selected based on the distribution in days available for stations); 2) Stations that are at least 500m away from roads; 3) Stations that are not in urban and suburban areas. After applying these criteria, 67 stations across Europe remained. The evaluation is performed based on the aggregated annual average surface NO₂ concentrations for each monitor, and the value of the gridcell corresponding to that monitor for both original exposure dataset and final product.

The evaluation results show that the newly developed NO₂ surface concentrations outperformed the Larkin et al.⁵ concentrations in rural areas, based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias (MB), and correlation with ground observations (Table S1). The slope of the best fitted line is improved from 1.41 to 1.10, and the mean ratio of estimated to observed concentrations is improved from 1.81 to 1.32 (Table S1 and Figure S6). The RMSE is reduced from 3.37 ppb to 2.26 ppb, and MAE is improved from 2.74 ppb to 1.72 ppb, and the MB is reduced from 2.40 ppb to 1.02 ppb. The correlation between the estimated surface concentrations and ground measurements is improved from Pearson correlation coefficient (R) of 0.51 in the original product to 0.58.

In addition to the comparison of our dataset with previously published NO₂ concentrations during their overlapping time period, we further test the fidelity of our dataset for a more recent year (Figure S7). We obtain annual average observations for 2019 from three different networks: the National Air Pollution Surveillance (NAPS) program in Canada⁷, the Air Quality System (AQS) in the United States⁸, and the European Environment Agency (EEA) in Europe.⁹ These networks provide data from 4,348 individual monitors (181 NAPS, 466 AQS, and 3,701 EEA), and we compare each monitor's NO₂ concentration to the concentration in the gridcell co-located with each monitor for

2019. All monitor data for 2019 have passed several quality control tests and quality assurance assessments by the entities that disseminate these data.

The mean bias (normalized mean bias) was -2.27 ppb (-20.40%) in Europe, 1.69 ppb in the U.S. (20.79%), and 3.34 ppb (49.56%) in Canada across the three networks in 2019 (Figure S7b,e,h). In urban areas specifically, the mean bias was -2.87 (-22.51%) in Europe, 1.26 ppb (12.21%) in the U.S., and 3.93 ppb (50.14%) in Canada for an average of 0.77 ppb across these three networks (Figure S7c,f,i). The mean bias at rural sites averaged over the three networks is 1.56 ppb, similar in magnitude to the rural bias reported for 2010-2012 (compare Figure S7c,f,i with Table S1).

The high mean bias evident in the NAPS and AQS datasets (Figure S7b-c, e-f) could, in part, reflect known issues with NO₂ monitors, which have been reported to overestimate NO₂ concentrations by up to ~50% due to interference from reactive nitrogen compounds, especially at locations distant from NO_x sources.¹⁰ Additionally, this high bias could also stem from monitors sited near traffic or other sources of NO_x emissions that may not be resolved in our ~1 km² dataset.

The paucity of monitors throughout large swaths of Canada and the United States (Figure S7a, d) and throughout the rest of the world inhibits a more in-depth performance assessment of our dataset and highlights the urgent need for more strategic and equitable monitoring of ambient air pollution (e.g. ¹¹).

Methods for decomposing parameter contributions to NO₂-attributable asthma trends

We calculate the contribution of each parameter used in health impact assessment (population, baseline asthma rates, and concentrations) using four sets of simulations:

- Control scenario, where we calculated the asthma cases for each year.
- Three “parameter rollback” simulations in which we revert one of the parameters (population, baseline asthma rates, or concentrations) to the base year 2000.

By comparing each of the three parameter rollback scenarios to the control scenario, we calculate the contribution of each parameter to the change in asthma cases between 2000 and all other years. We use the following set of equations to calculate the contribution of each parameter.

We use Equation S4 to calculate pediatric asthma incidence attributable to NO₂ for the control scenario. This equation is the same as Equation 1 in the main text, but we denote the parameters differently here to make it easier to compare with the control scenario equations.

$$\text{Equation S4: } \textit{Control}_t(x_t, y_t, z_t) = x_t \times y_t \times z_t$$

Where $\textit{Control}_t$ is the NO₂-attributable pediatric asthma incidence for year t , x_t is the baseline pediatric asthma rate for year t , y_t is the pediatric population for year t , and z_t is the fraction of pediatric asthma incidence that is attributable to NO₂ for year t .

We then calculate NO₂-attributable pediatric asthma incidence for each simulation, replacing one parameter with its value in the year 2000 while holding the other two parameters at the same value used in the control scenario (Equations S5-S7).

$$\text{Equation S5: } \textit{Simulation}_{x,t}(x_0, y_t, z_t) = x_0 \times y_t \times z_t$$

Equation S6: $Simulation_{y,t}(x_t, y_0, z_t) = x_t \times y_0 \times z_t$

Equation S7: $Simulation_{z,t}(x_t, y_t, z_0) = x_t \times y_t \times z_0$

Where $Simulation_{i,t}$ is the estimated NO₂-attributable pediatric asthma incidence for year t, where we have reverted one parameter back to the base year of 2000.

We then calculate the ratio of estimated NO₂-attributable pediatric asthma incidence in the control scenario versus in each of the parameter rollback scenarios, as shown in Equation S8.

Equation S8: $Ratio_{i,t} = \frac{Control_t}{Simulation_{i,t}}$

Since NO₂-attributable pediatric asthma incidence is calculated by multiplying three parameters, we assume that the ratio of NO₂-attributable asthma incidence between year t and base year 2000 would be equivalent to the multiplication of the three rollback scenario ratios calculated in Equations S5-S7 (Equation S9). In this step we assume that aggregating the three parameter rollbacks separately is equivalent to reverting all of them together.

Equation S9: $\frac{Asthma_t}{Asthma_0} \approx Ratio_{x,t} \times Ratio_{y,t} \times Ratio_{z,t}$

To calculate the contribution of each parameter individually, we need to transform the parameter ratios so that they add up to 1 when summed. We therefore calculate a logarithm in the base of the left side of Equation S9 ($Asthma_t/Asthma_0$); since the logarithm of every number in its own base equals 1, this step makes the left side equal to 1 (Equation S10).

Equation S10: $1 = \log(Ratio_{x,t}) + \log(Ratio_{y,t}) + \log(Ratio_{z,t})$

Finally, we multiply each of the three log-transformed parameter rollback ratios by the total percentage change in NO₂-attributable asthma incidence between years 2000 and t to calculate the percent contribution of each parameter to that total change (Equation S11).

Equation S11: $Contribution_{i,t} = \frac{Asthma_t}{Asthma_0} \times \log(Ratio_{i,t})$

Using this methodology, we calculated percent contributions for each of the three health impact function parameters (concentration, population, asthma rates) that add up to the total percentage changes between the two years, while remaining loyal to the multiplicative nature of the original health impact assessment function.

Supplemental Tables and Figures

Table S1. Statistical parameters for NO₂ concentrations from the Larkin et al.⁵ dataset and our new concentration estimates for rural areas compared with EMEP rural observations. Values reported here for Larkin et al.⁵ differ from those reported in their paper because here we are only evaluating predicted concentrations in rural areas at the EMEP monitor locations.

	Root Mean Square Error (RMSE) (ppb)	Mean Absolute Error (MAE) (ppb)	Mean Bias (MB) (ppb)	Pearson coefficient (R)	Mean ratio: Estimate/obs	Slope of best fitted line
New product	2.26	1.72	1.02	0.58	1.32	1.10
Larkin et al. ⁵	3.37	2.74	2.40	0.51	1.81	1.41

Table S2. Count of urban clusters in each GBD region/super region. The total does not match the total urban clusters in analysis (13,189) since some urban clusters are located at the border between two regions.

Super region name	Region name	Count
Central Europe, Eastern Europe, and Central Asia		628
	Central Asia	148
	Central Europe	161
	Eastern Europe	319
High-income		1280
	Australasia	35
	High-income Asia Pacific	148
	High-income North America	389
	Southern Latin America	115
	Western Europe	593
Latin America and Caribbean		968
	Andean Latin America	93
	Caribbean	75
	Central Latin America	438
	Tropical Latin America	362
North Africa and Middle East		1231
	North Africa and Middle East	1231
South Asia		3899
	South Asia	3899
Southeast Asia, East Asia, and Oceania		2904
	East Asia	1955
	Oceania	49
	Southeast Asia	900
Sub-Saharan Africa		2313
	Central Sub-Saharan Africa	250
	Eastern Sub-Saharan Africa	1024
	Southern Sub-Saharan Africa	124
	Western Sub-Saharan Africa	915
Grand Total		13223

Table S3. Population-weighted NO₂ concentrations (ppb), NO₂-attributable pediatric asthma incidence (95% uncertainty interval in parentheses), NO₂-attributable pediatric asthma incidence rate (per 100,000), and NO₂ attributable fraction (%) for each super-region in 2000, 2005, 2010, and 2019.

Super region	Year	Pop-wt NO ₂ concentration (ppb)	NO ₂ -attributable pediatric asthma incidence	NO ₂ -attributable asthma rate (per 100,000)	NO ₂ attributable fraction
Central Europe, Eastern Europe, Central Asia	2000	15.1	65,000 (33,400 – 107,500)	202	23
	2010	13.6	44,300 (26,400 – 85,300)	172	20
	2019	12.7	54,900 (26,400 – 87,000)	188	20
High-income	2000	17.3	464,800 (264,700 – 725,800)	428	29
	2010	15.1	433,300 (255,500 – 655,000)	404	24
	2019	11.1	340,900 (191,400 – 523,200)	310	17
Latin America and Caribbean	2000	12.8	256,100 (147,300 – 397,000)	332	19
	2010	12.4	236,900 (138,000 – 380,300)	302	19
	2019	10.6	233,800 (125,300 – 375,800)	281	16
North Africa and Middle East	2000	12.1	111,800 (63,300 – 176,300)	180	19
	2010	13.2	120,100 (65,000 – 189,200)	189	20
	2019	12.8	157,600 (67,000 – 202,200)	203	20
South Asia	2000	8.6	50,000 (28,600 – 78,100)	33	13
	2010	9.9	106,500 (55,900 – 161,900)	65	15
	2019	10.1	90,400 (40,700 – 130,700)	50	16
Southeast Asia, East Asia, and Oceania	2000	11.1	211,300 (113,900 – 343,100)	109	16
	2010	14	226,600 (129,500 – 377,600)	128	18
	2019	10.6	240,900 (122,100 – 378,000)	129	15
Sub-Saharan Africa	2000	6.4	49,100 (27,400 – 78,700)	89	9
	2010	6.9	65,700 (26,000 – 74,700)	84	9
	2019	7.1	102,900 (30,200 – 90,900)	97	10

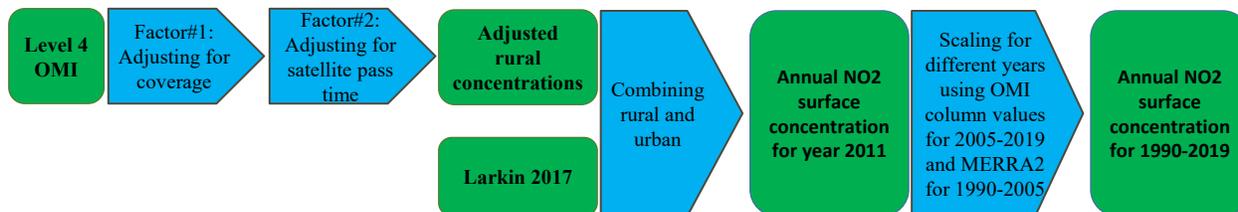


Figure S1. Schematic of datasets used and the process of combining them. Blue arrows represent applied processes.

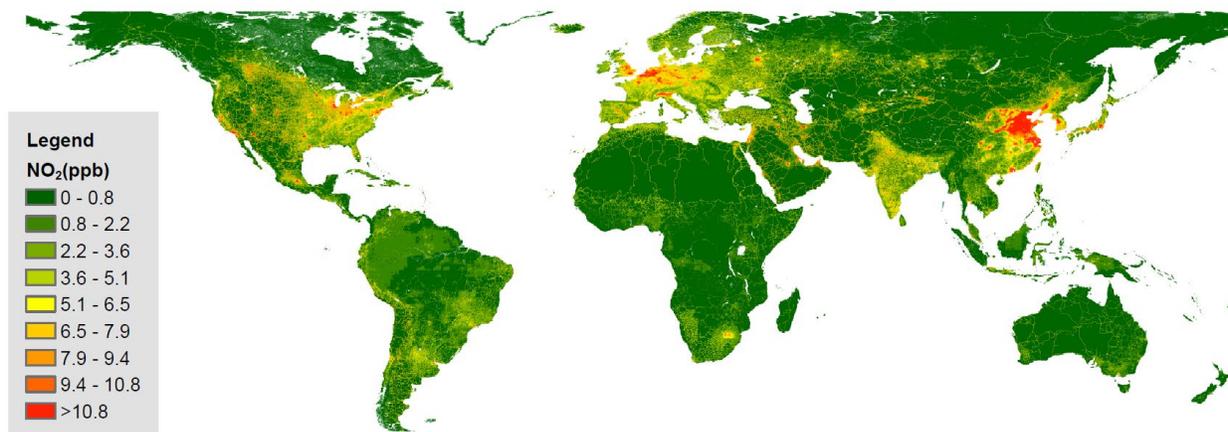


Figure S2. Annual average surface NO₂ concentration estimates for 2011 at ~1km x 1km resolution globally from this work, using a combination of Larkin et al.⁵ land use regression estimates, OMI satellite observations, and chemical transport modeling.

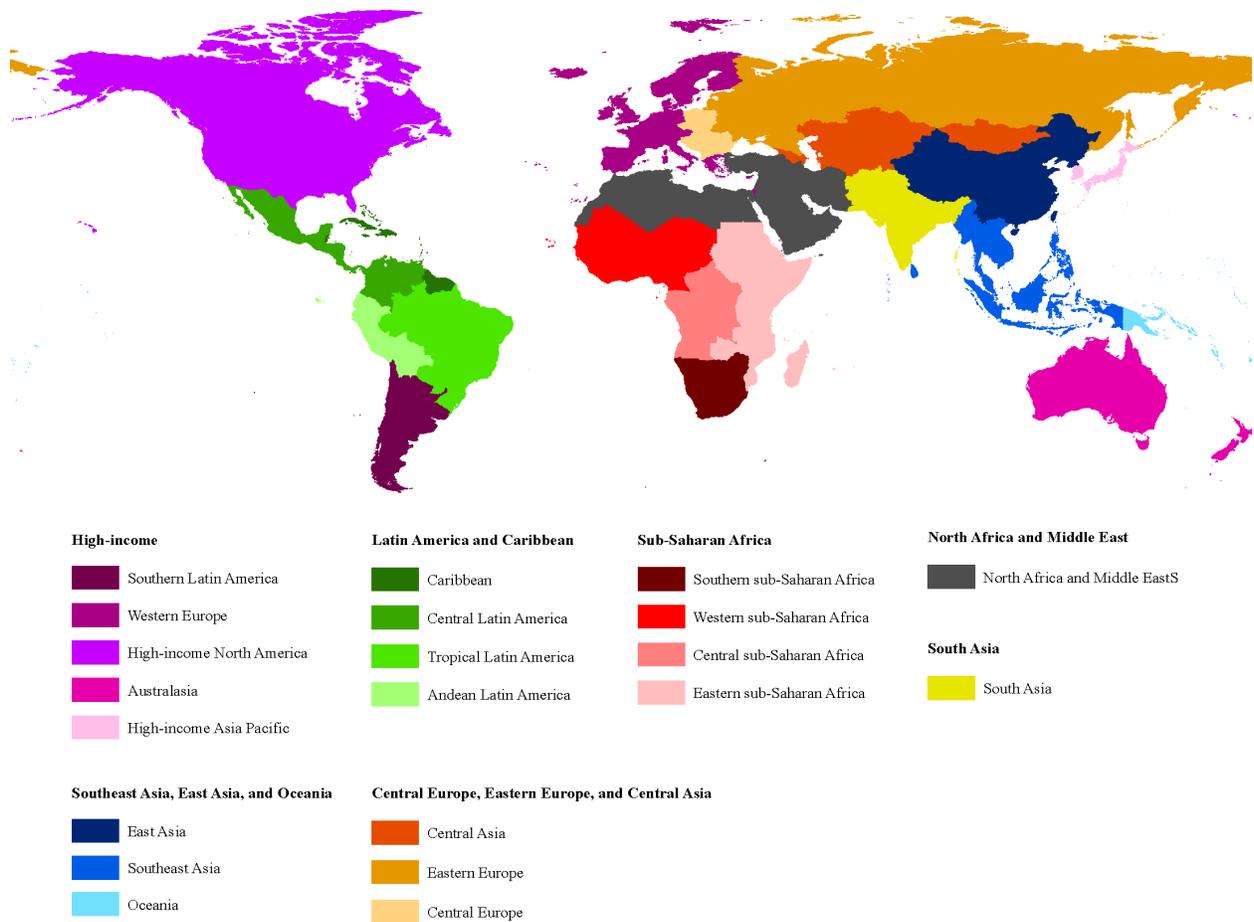


Figure S4. Countries and territories included in each region and super region, using regional definitions from the GBD 2019 Study.

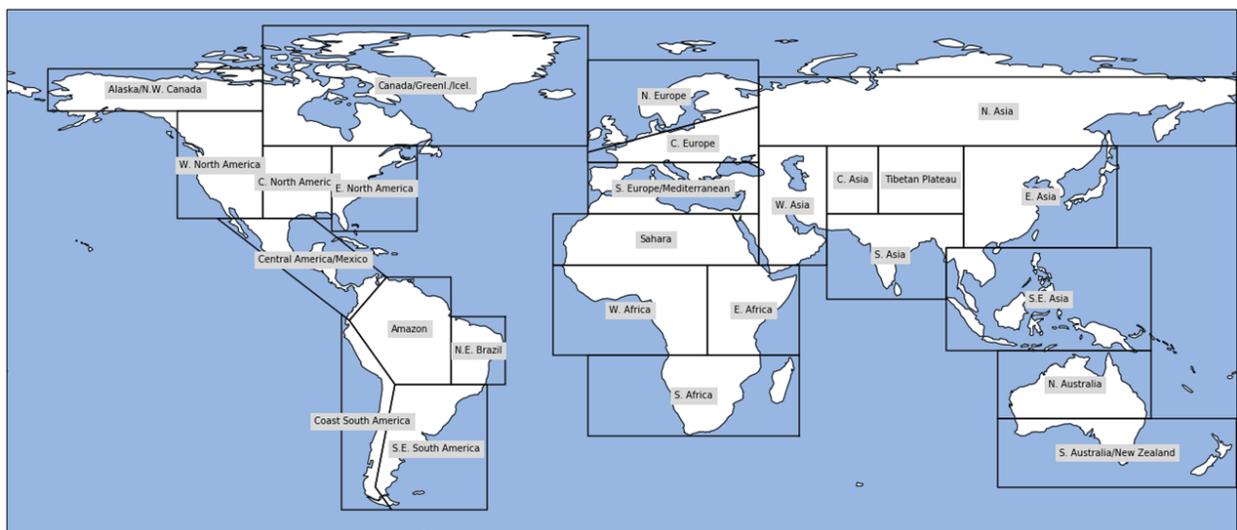


Figure S5. World regions used to generate the MERRA-2 scaling factors for NO₂ in 1990, 1995, and 2000.

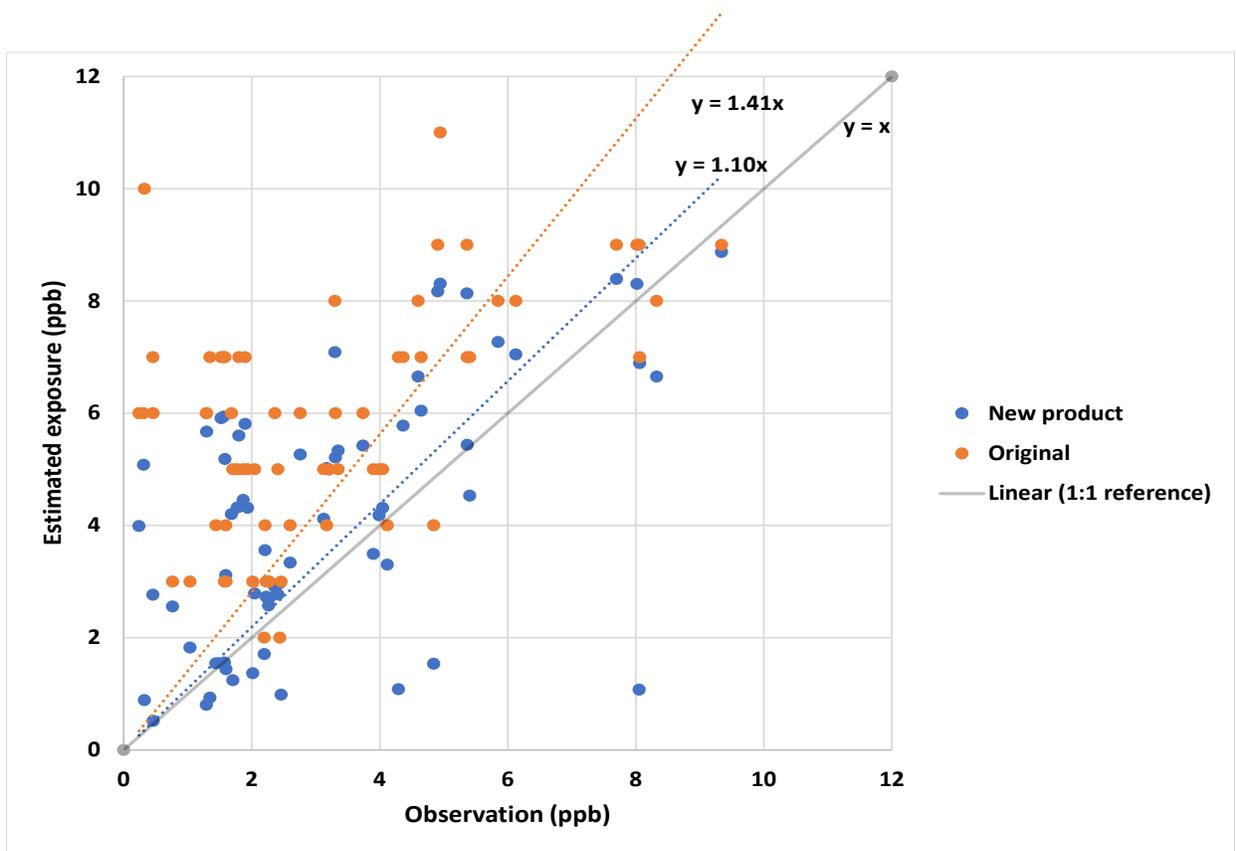


Figure S6. Comparison between annual average NO₂ concentrations from the original Larkin et al.⁵ product (orange) and our new NO₂ concentration product (blue), versus concentrations from ground measurements for 2011 in rural areas. A 1:1 reference line is added for comparison. Each point represents a monitor. Monitor data source: European Monitoring and Evaluation Programme (EMEP).

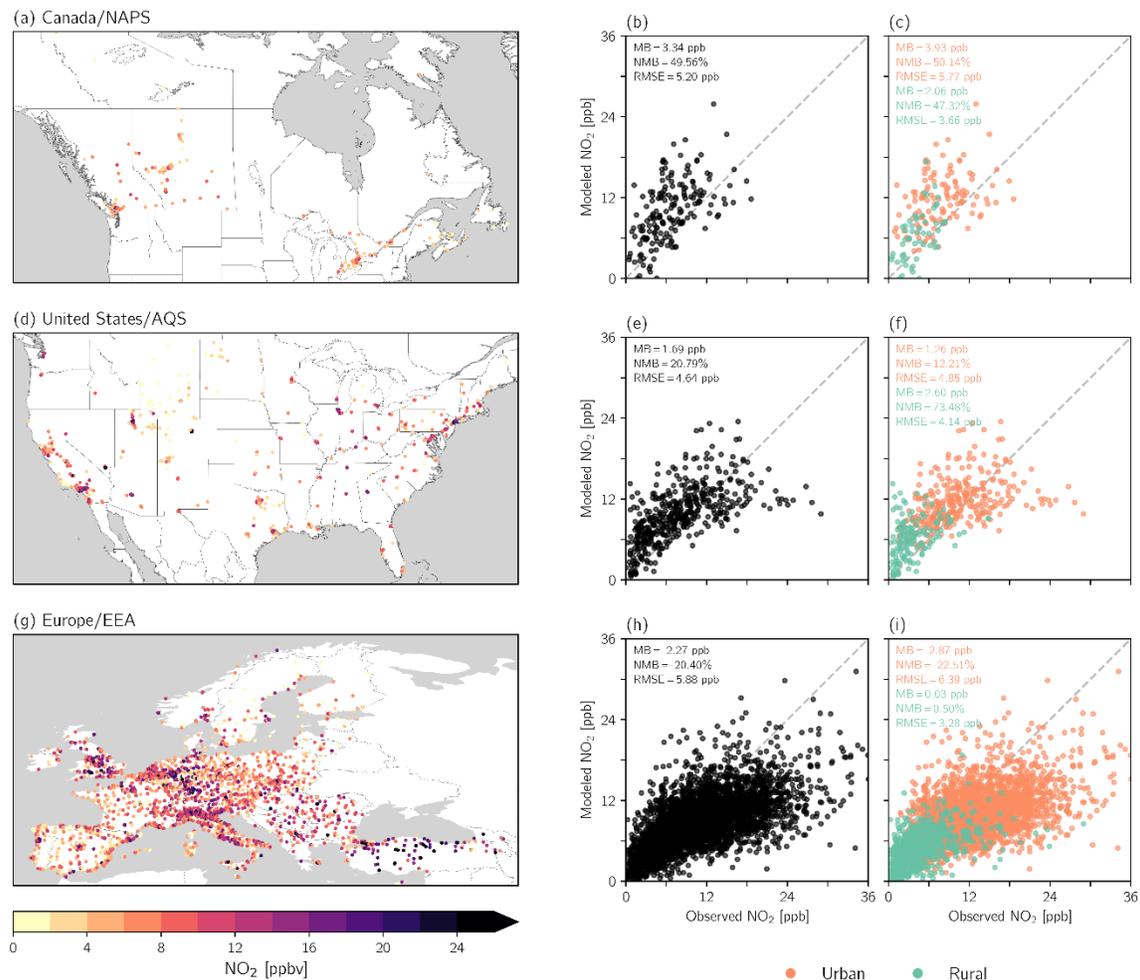


Figure S7. (a,d,g) The location of NO₂ monitors from the NAPS, AQS, and EEA networks and their annual average 2019 concentrations. (b,e,h) Annual average 2019 NO₂ concentrations from the gridcells co-located with each monitor versus the monitor concentrations. (c,f,i) are the same as (b,e,h) but for urban versus rural monitors. Rurality in (c,f,i) is determined with the GHS-SMOD dataset. The mean bias (MB; $= \bar{M} - \bar{O}$), normalized mean bias (NMB; $= (\frac{\bar{M}}{\bar{O}} - 1) \times 100\%$), and RMSE ($= \left[\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2 \right]^{\frac{1}{2}}$) are indicated in each scatterplot. Here, M corresponds to the new dataset and O corresponds to observed concentrations. A 1:1 reference line is included in scatterplots for comparison.

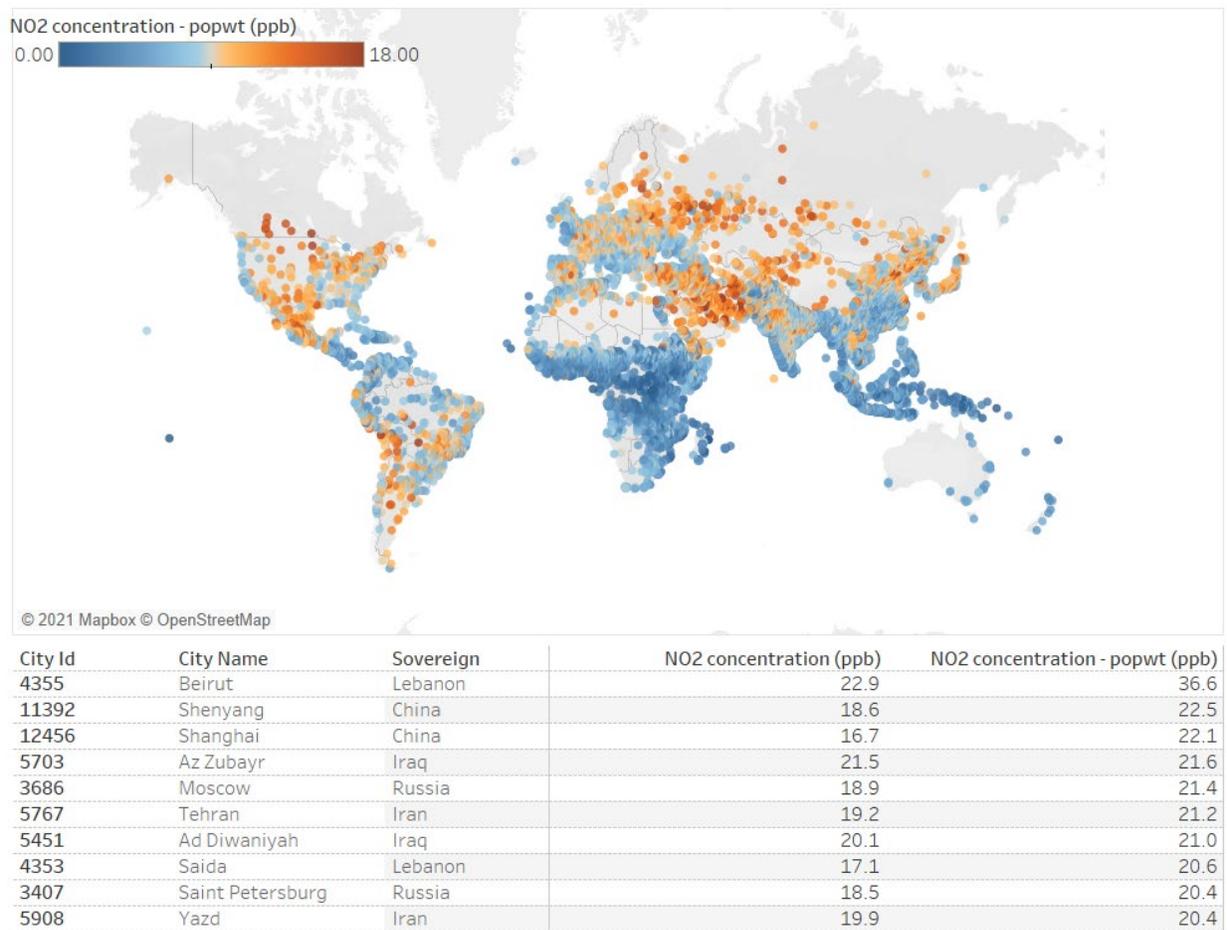


Figure S8. Population-weighted annual average NO₂ concentrations (ppb) for 13,189 urban areas (top) and the cities with the top 10 concentrations (bottom) in 2019. Color bar saturates at 18 ppb for greater contrast.

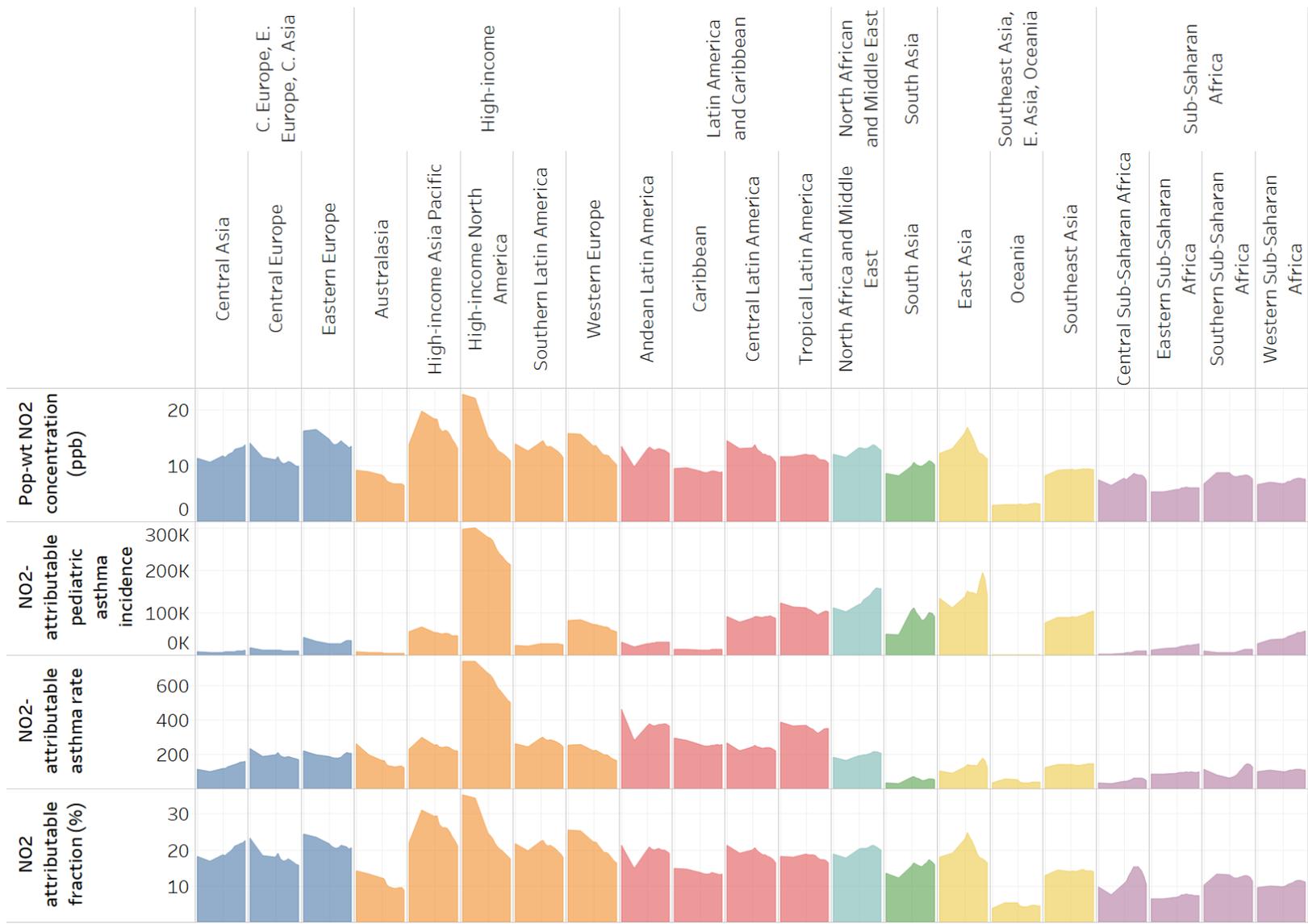


Figure S9. As for Figure 3, but for each subregion within each super-region.

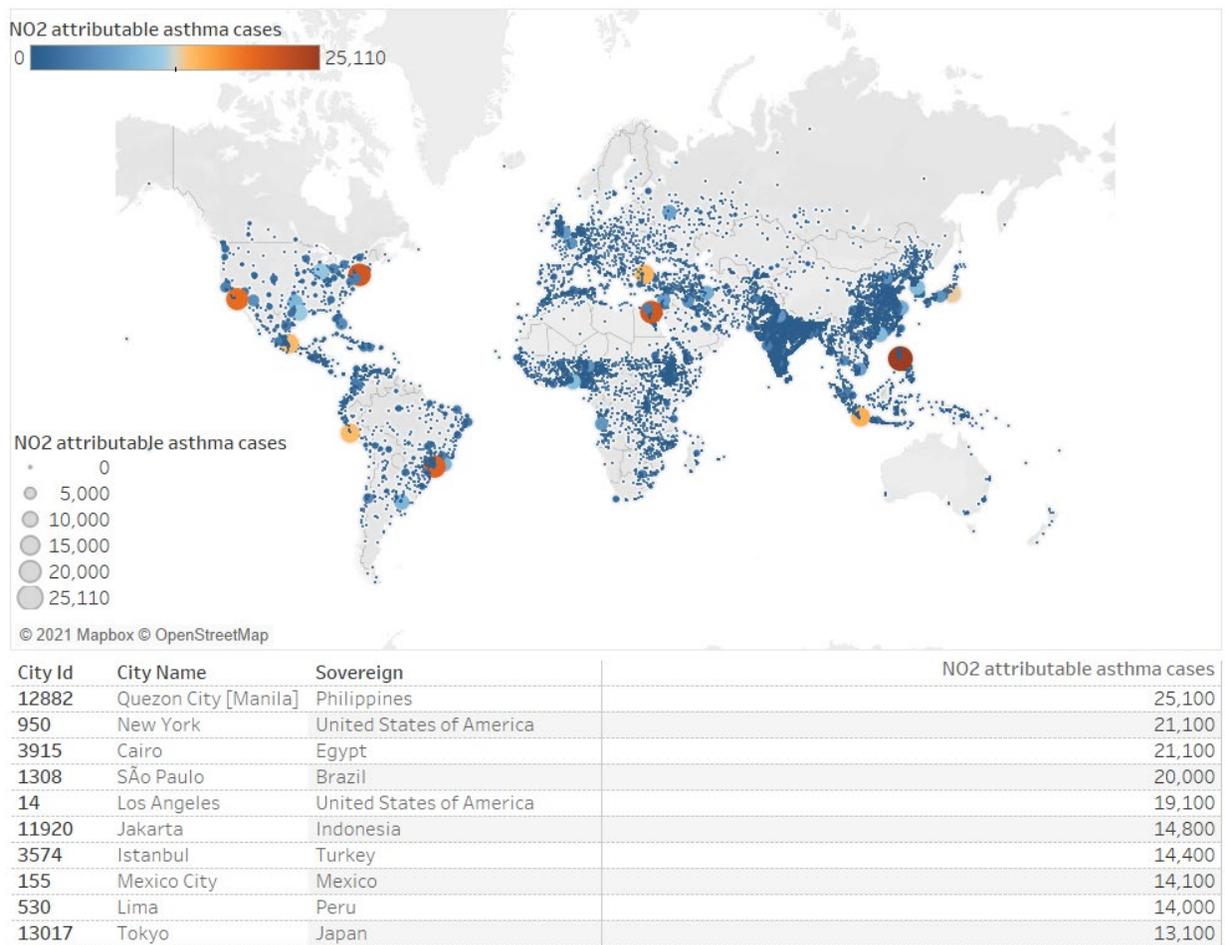


Figure S10. As for Figure S8, but for NO₂-attributable pediatric asthma incidence.

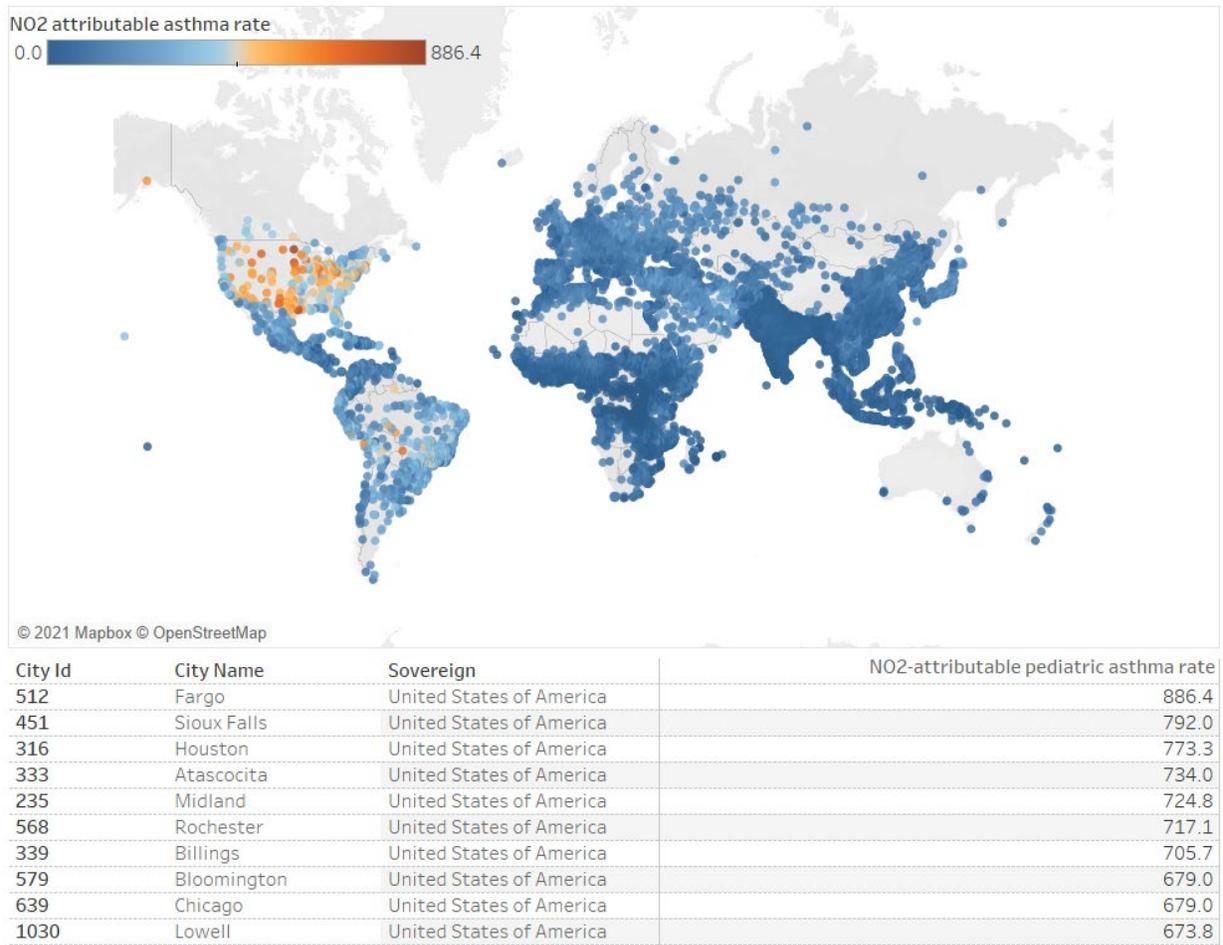


Figure S11. As for Figure S8, but for NO₂-attributable pediatric asthma incidence rate per 100,000 children.

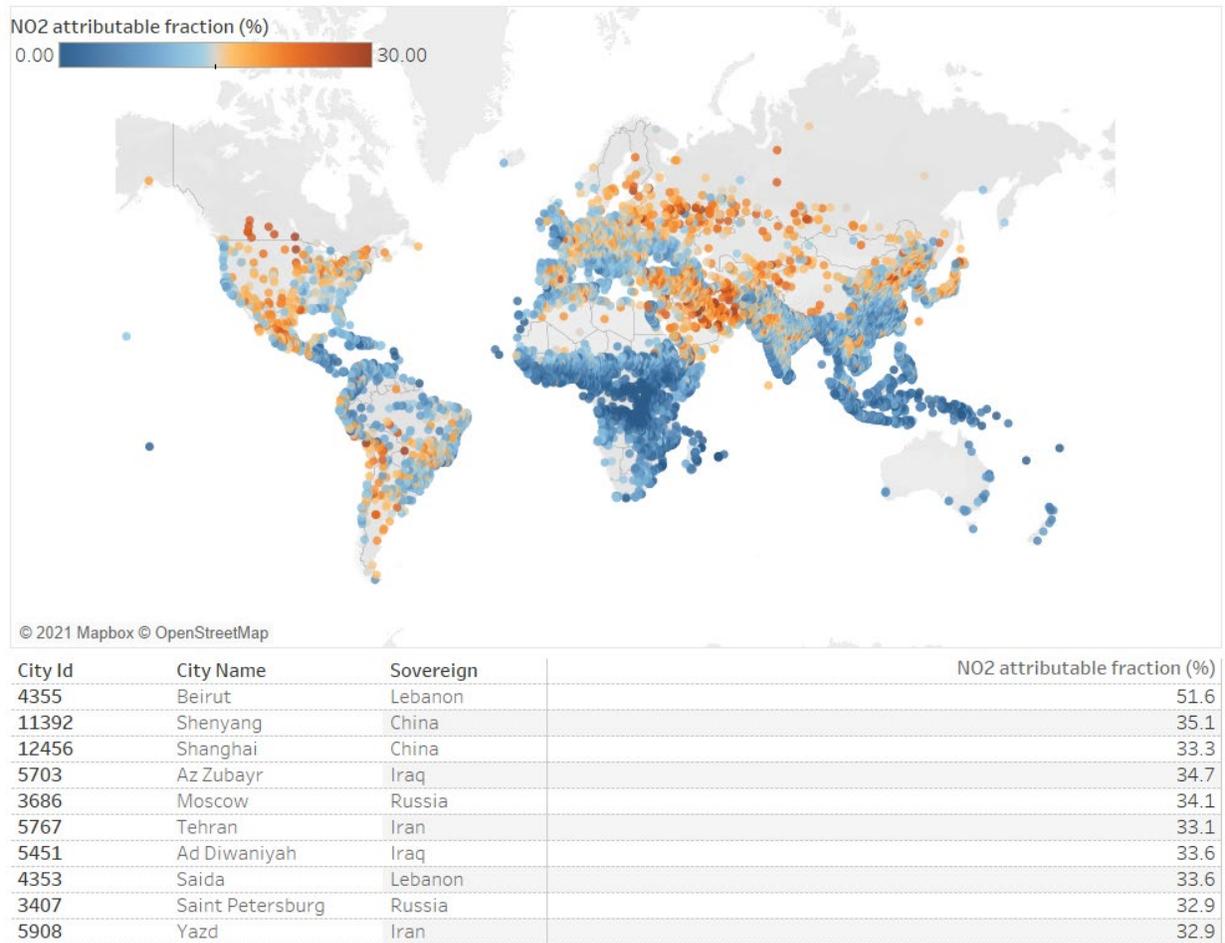


Figure S12. As for Figure S8, but for the fraction of pediatric asthma incidence attributable to NO₂ (%). Color bar saturates at 30% for greater contrast.

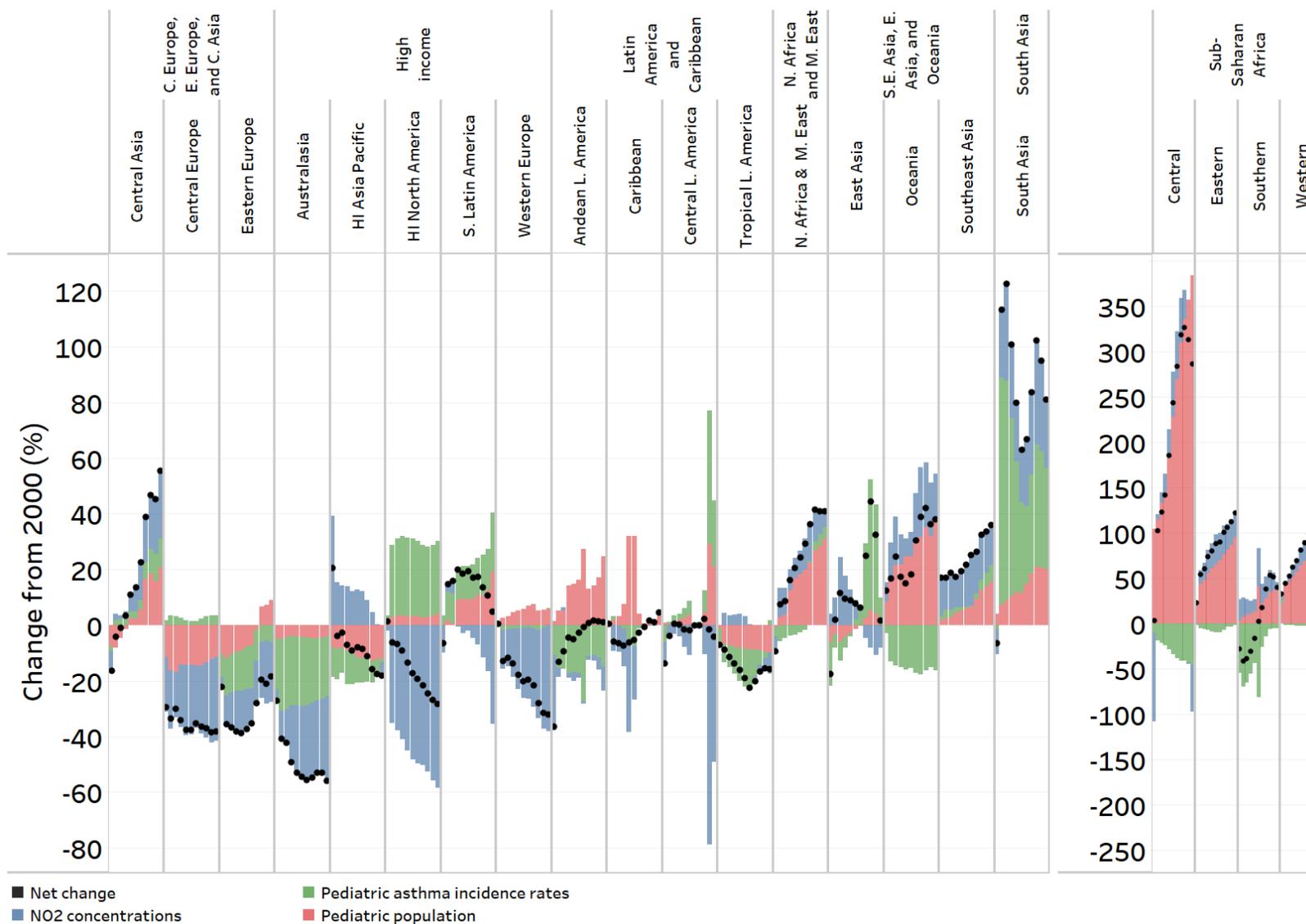


Figure S13. As for Figure 5, but for the change from 2000 to 2005 and 2010-2019 annually (represented left to right by the bars in each panel).

Supplemental References

- (1) Lamsal, L. N.; Martin, R. V.; van Donkelaar, A.; Steinbacher, M.; Celarier, E. A.; Bucsela, E.; Dunlea, E. J.; Pinto, J. P. Ground-Level Nitrogen Dioxide Concentrations Inferred from the Satellite-Borne Ozone Monitoring Instrument. *J. Geophys. Res.* **2008**, *113* (D16). <https://doi.org/10.1029/2007JD009235>.
- (2) Anenberg, S. C.; Henze, D. K.; Tinney, V.; Kinney, P. L.; Raich, W.; Fann, N.; Malley, C. S.; Roman, H.; Lamsal, L.; Duncan, B.; Martin, R. V.; van Donkelaar, A.; Brauer, M.; Doherty, R.; Jonson, J. E.; Davila, Y.; Sudo, K.; Kuynlenstierna, J. C. I. Estimates of the Global Burden of Ambient PM_{2.5}, Ozone, and NO₂ on Asthma Incidence and Emergency Room Visits. *Environ. Health Perspect.* **2018**, *126* (10), 107004. <https://doi.org/10.1289/EHP3766>.
- (3) Strode, S. A.; Ziemke, J. R.; Oman, L. D.; Lamsal, L. N.; Olsen, M. A.; Liu, J. Global Changes in the Diurnal Cycle of Surface Ozone. *Atmos. Environ.* **2019**, *199*, 323–333. <https://doi.org/10.1016/j.atmosenv.2018.11.028>.
- (4) Orbe, C.; Oman, L. D.; Strahan, S. E.; Waugh, D. W.; Pawson, S.; Takacs, L. L.; Molod, A. M. Large-Scale Atmospheric Transport in GEOS Replay Simulations: TRANSPORT IN GEOS REPLAY SIMULATIONS. *J. Adv. Model. Earth Syst.* **2017**, *9* (7), 2545–2560. <https://doi.org/10.1002/2017MS001053>.
- (5) Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P. Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environ. Sci. Technol.* **2017**, *51* (12), 6957–6964. <https://doi.org/10.1021/acs.est.7b01148>.
- (6) Gelaro, R.; McCarty, W.; Suárez, M. J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C. A.; Darmenov, A.; Bosilovich, M. G.; Reichle, R.; Wargan, K.; Coy, L.; Cullather, R.; Draper, C.; Akella, S.; Buchard, V.; Conaty, A.; da Silva, A. M.; Gu, W.; Kim, G.-K.; Koster, R.; Lucchesi, R.; Merkova, D.; Nielsen, J. E.; Partyka, G.; Pawson, S.; Putman, W.; Rienecker, M.; Schubert, S. D.; Sienkiewicz, M.; Zhao, B. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30* (14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- (7) Environment and Climate Change Canada/Environnement et Changement climatique Canada. National Air Pollution Surveillance Program <https://open.canada.ca> (accessed May 11, 2021).
- (8) U.S. Environmental Protection Agency. Air Quality System Data Mart <https://www.epa.gov/airdata> (accessed Nov 24, 2020).
- (9) European Environment Agency. Air Quality e-Reporting Air quality annual statistics <https://www.eea.europa.eu/> (accessed May 11, 2021).
- (10) Lamsal, L. N.; Duncan, B. N.; Yoshida, Y.; Krotkov, N. A.; Pickering, K. E.; Streets, D. G.; Lu, Z. U.S. NO₂ Trends (2005–2013): EPA Air Quality System (AQS) Data versus Improved Observations from the Ozone Monitoring Instrument (OMI). *Atmos. Environ.* **2015**, *110*, 130–143. <https://doi.org/10.1016/j.atmosenv.2015.03.055>.
- (11) Martin, R. V.; Brauer, M.; van Donkelaar, A.; Shaddick, G.; Narain, U.; Dey, S. No One Knows Which City Has the Highest Concentration of Fine Particulate Matter. *Atmospheric Environ. X* **2019**, *3*, 100040. <https://doi.org/10.1016/j.aeaoa.2019.100040>.