Supplemental material:

Long-term trends in urban NO_2 concentrations and associated pediatric asthma cases: estimates from global datasets

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

Methods for generating surface NO_2 concentrations

We estimated surface annual average NO_2 concentrations at 1km x 1km resolution in five-year increments from 1990-2010 and annually from 2010-2019. To begin, we used a previously published NO_2 concentration dataset for the average of 2010-2012, which used land use regression modeling with inputs from road networks and other land use variables, as well as satellite NO_2 column observations from SCIAMACHY and GOME-2 (Larkin et al. 2017, Geddes et al. 2016). We aggregated NO_2 concentrations from this dataset from its native 100m x 100m resolution globally to 1km x 1km, which is still a high enough resolution to avoid substantially underestimating NO_2-attributable asthma impacts (Mohegh et al. 2020). Due to the lack of ground measurements in rural areas, the Larkin et al. (2017) NO_2 dataset is fine-tuned towards the urban areas and overestimates NO_2 concentrations in rural areas, likely due to a high sensitivity to the normalized difference vegetation index (NDVI).

Given the good performance in urban areas, we applied the Larkin et al. (2017) NO_2 concentrations in all 1km x 1km grid cells globally that are categorized as "urban" according to the Global Human Settlement Model grid (Pesaresi et al. 2019), as well as those grid cells that include major roadways (Larkin et al. 2017). For grid cells >5km away from roadways and in rural areas, we developed new NO_2 concentration estimates using NO_2 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. We use an OMI NO_2 version 3, level 4 surface concentration dataset (0.1 x 0.1 degree resolution) for 2011, which followed methods described by Lamsal et al. (2008) 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 (Figure S1). We used the MERRA 2 reanalysis product (0.625 x 0.5 degree resolution) to generate a correction factor to ensure availability of NO_2 concentrations in all locations and months, as follows:

$$Correction\ factor\#1 = \frac{MERRA2_{annual\ averge}}{MERRA2_{average\ for\ same\ months\ that\ OMI\ level\ 4\ is\ available}}$$

We also applied a second correction factor to convert surface NO_2 concentrations from the early afternoon OMI overpass time (13:00 local time) to 24-hour averages. Following Anenberg et al. (2018), we used NO_2 surface concentrations from the GMI-Replay chemical transport model (2 x 2.5 degree resolution) simulations to generate these correction factors, as follows:

$$Correction\ factor\#2 = \frac{GMI_{24\ hour\ averge}}{GMI_{13:00}}$$

The NO_2 surface concentration estimates used for gridcells >5km away from roads and in rural areas were then generated using the following formula:

$$Adjusted\ rural\ concentrations = OMI\ level\ 4\ \times Correction\ factor\#1 \times Correction\ factor\ \#2$$

For rural gridcells within 5km of major roadways, we linearly scaled between Larkin et al. (2017) 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 NO_2 concentration dataset for 2011 that uses Larkin et al. (2017) 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).

The GBD requires NO_2 concentrations for each year included in the comparative risk assessment, from 1990-2019. We therefore scaled the new 2011 surface NO_2 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 NO_2 concentrations from 2011 to each year using 3-year rolling averages of annual average NO_2 columns from the OMI version 4.0 level 2 product (13 km x 25 km resolution at nadir). We use NO_2 columns because surface concentrations derived from the version 4 OMI retrieval are not yet available. We oversampled the column NO_2 dataset to 0.1 x 0.1 degree resolution and re-gridded to 0.0083 x 0.0083 degrees (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 NO_2 columns directly. The years 1990, 1995, and 2000 predated the OMI observational record. We therefore use NO_2 concentrations from the MERRA2 reanalysis product to scale 2011 NO_2 concentrations to those years (Gelaro et al. 2017). To remove model noise, we created the scaling factors across broad world regions, as opposed to applying scaling factors on a gridcell by gridcell basis.

The final result used for estimating the global burden of disease from NO_2 is a global, 0.0083 x 0.0083 degree (approximately 1km x 1km) resolution dataset of annual average surface NO_2 concentrations from 1990-2019 (Figure S3).

Evaluation of NO_2 concentration dataset

The Larkin et al. (2017) NO_2 concentration dataset was evaluated extensively in that work and agreed well with ground observations in urban areas. Since we used the Larkin et al. (2017) concentrations for urban areas, we focus here on evaluating the Larkin et al. (2017) and the newly developed NO_2

concentrations for rural areas for the year 2011. We evaluated the rural NO_2 concentration estimates using the European Monitoring and Evaluation Program (EMEP) ground monitoring dataset, since it has a large number of stations in rural areas. 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 (Figure S4). The evaluation is performed based on the aggregated annual average surface NO_2 concentrations for each monitor, and the value of the grid-cell corresponding to that monitor for both original exposure dataset and final product.

The evaluation results (Figure S5) show that the newly developed NO_2 surface concentrations outperformed the Larkin et al. (2017) concentrations in rural areas, based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and correlation with ground observations. The slope of the best fitted line is improved from 1.41 to 1.10, and the mean ratio is improved from 1.81 to 1.32 (Table S1). The RMSE is reduced from 3.37 ppb to 2.26 ppb, and MAE is improved from 2.74 ppb to 1.72 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. Figure S5 shows the improvement for each station in a scatter plot.

Methods for decomposing parameter contributions to NO_2-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 NO_2 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.

Eq S4: $$Control_t(x_t, y_t, z_t) = x_t \times y_t \times z_t$$

Where $Control_t$ is the NO_2-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 NO_2 for year *t*.

We then calculate NO_2-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).

Eq S5: $$Simulation_{x,t}(x_0, y_t, z_t) = x_0 \times y_t \times z_t$$

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

Eq 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 NO_2-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 NO_2-attributable pediatric asthma incidence in the control scenario versus in each of the parameter rollback scenarios, as shown in Equation S8.

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

Since NO_2-attributable pediatric asthma incidence is calculated by multiplying three parameters, we assume that the ratio of NO_2-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.

Eq 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).

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

Finally, we multiply each of the three log-transformed parameter rollback ratios by the total percentage change in NO_2-attributable asthma incidence between years 2000 and *t* to calculate the percent contribution of each parameter to that total change (Equation S11).

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

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 multiplication nature of the original health impact assessment function.

Supplemental Tables and Figures

Table S1. Statistical parameters for NO_2 concentrations from the Larkin et al. (2017) dataset and our new concentration estimates.

	Root Mean Square Error (RMSE) (ppb)	Mean Absolute Error (MAE) (ppb)	Pearson coefficient (R)	Mean ratio: Estimate/obs	Slope of best fitted line
New product	2.26	1.72	0.58	1.32	1.10
Larkin et al. (2017)	3.37	2.74	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**

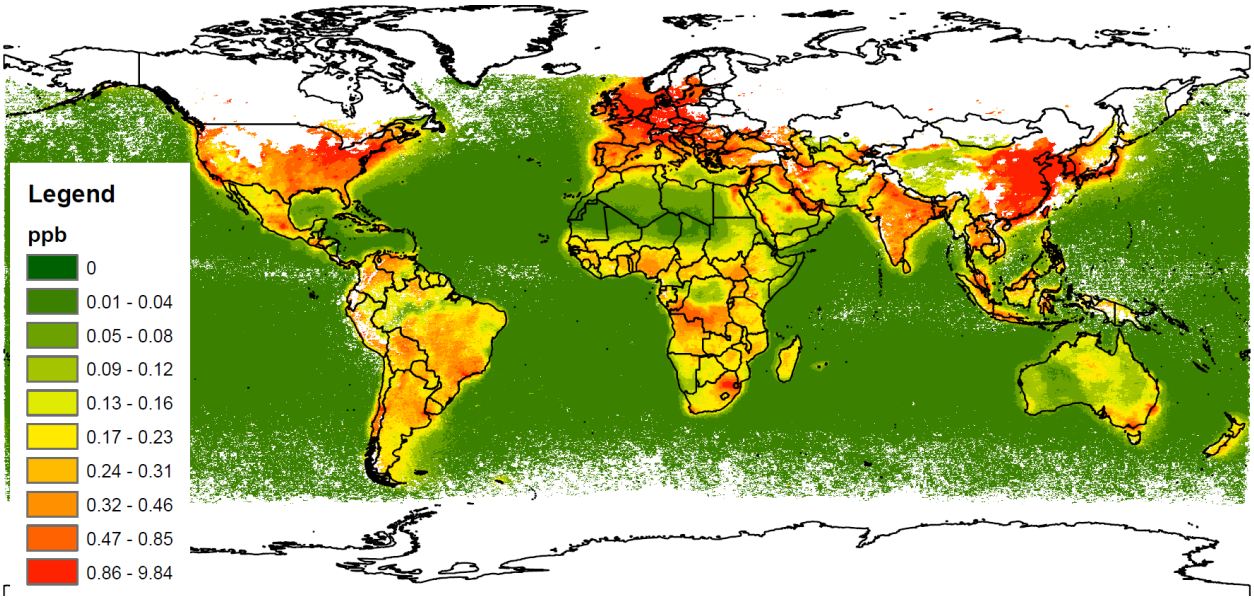

Figure S1. Spatial coverage of the OMI level 4 NO_2 product (aggregated for all months of 2011), before adjustments. White spaces indicate areas with missing data for some months.

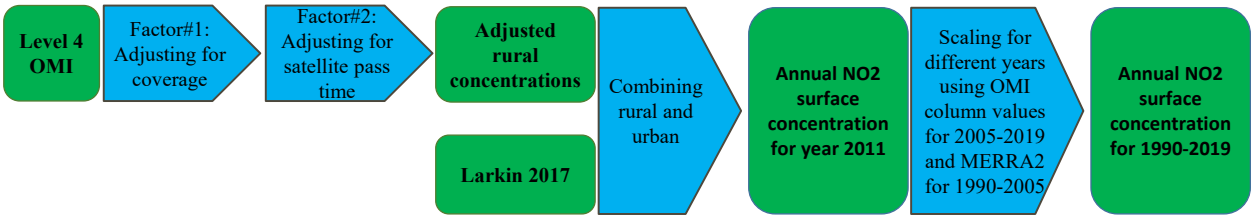

Figure S2. Schematic of datasets used and the process of combining them. Blue arrows represent applied processes.

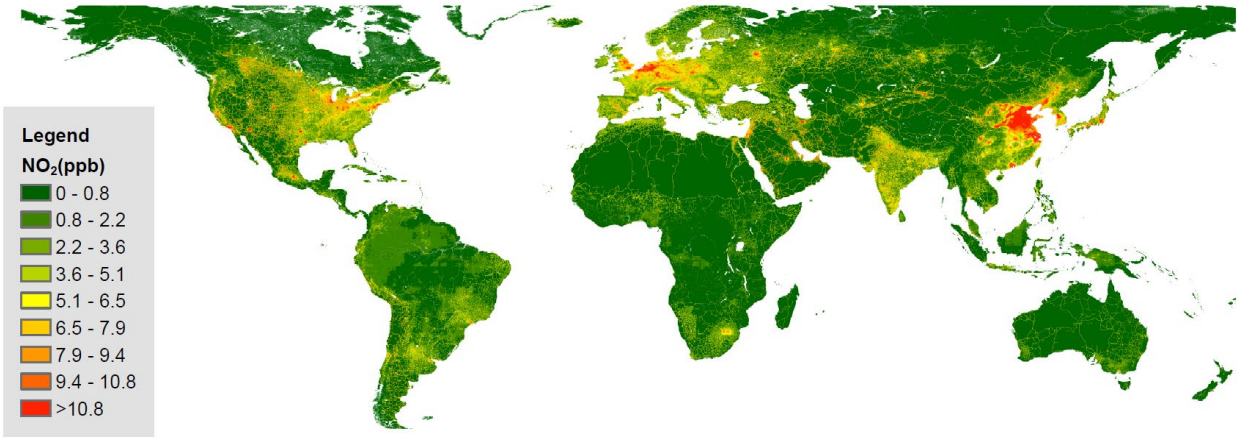

Figure S3. Annual average surface NO_2 concentration estimates for 2011 at ~1km x 1km resolution globally from this work, using a combination of Larkin et al. (2017) land use regression estimates, OMI satellite observations, and chemical transport modeling.

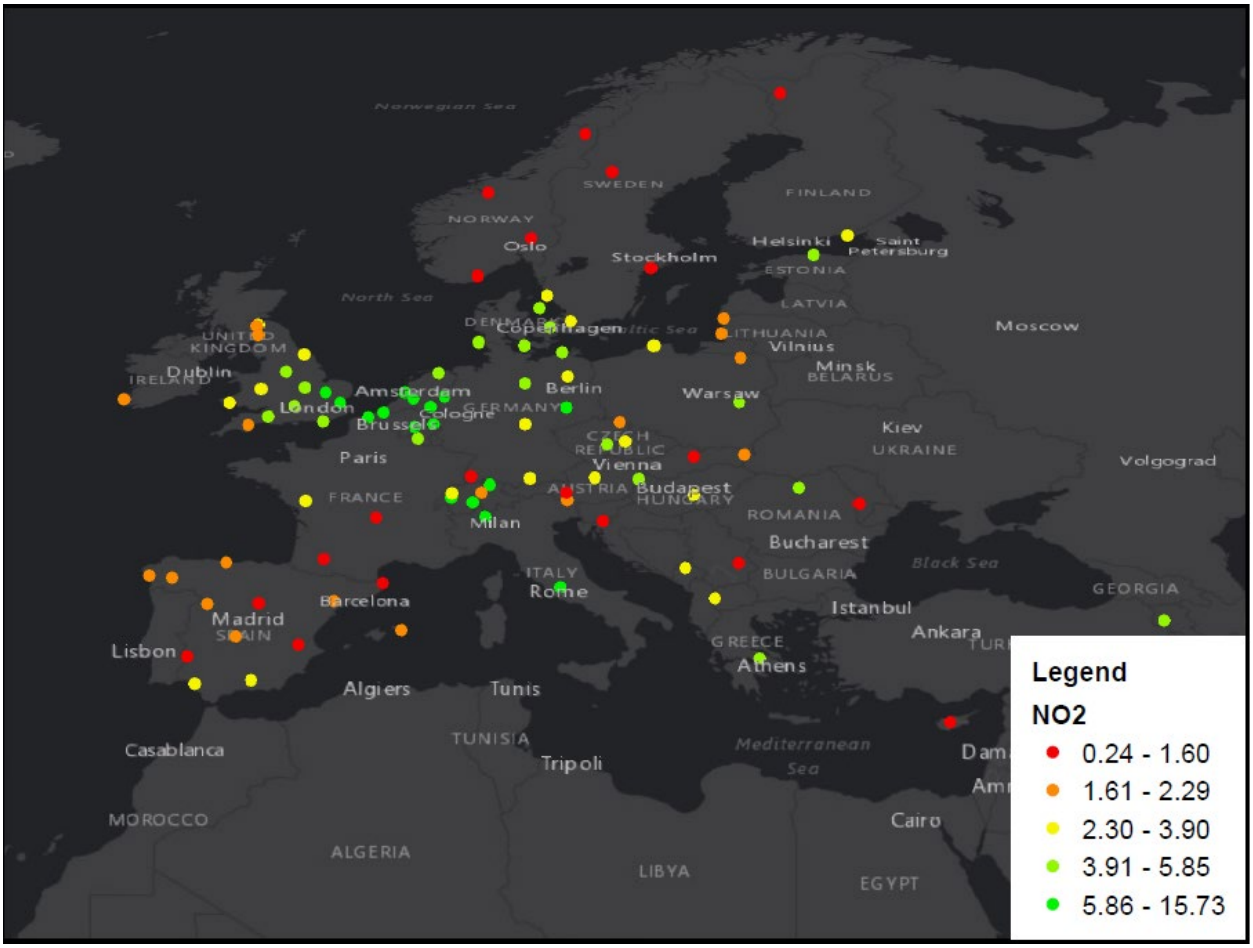

Figure S4. Location of all EMEP monitoring stations across Europe, color coded based on annual average NO_2 (ppb) concentrations for 2011.

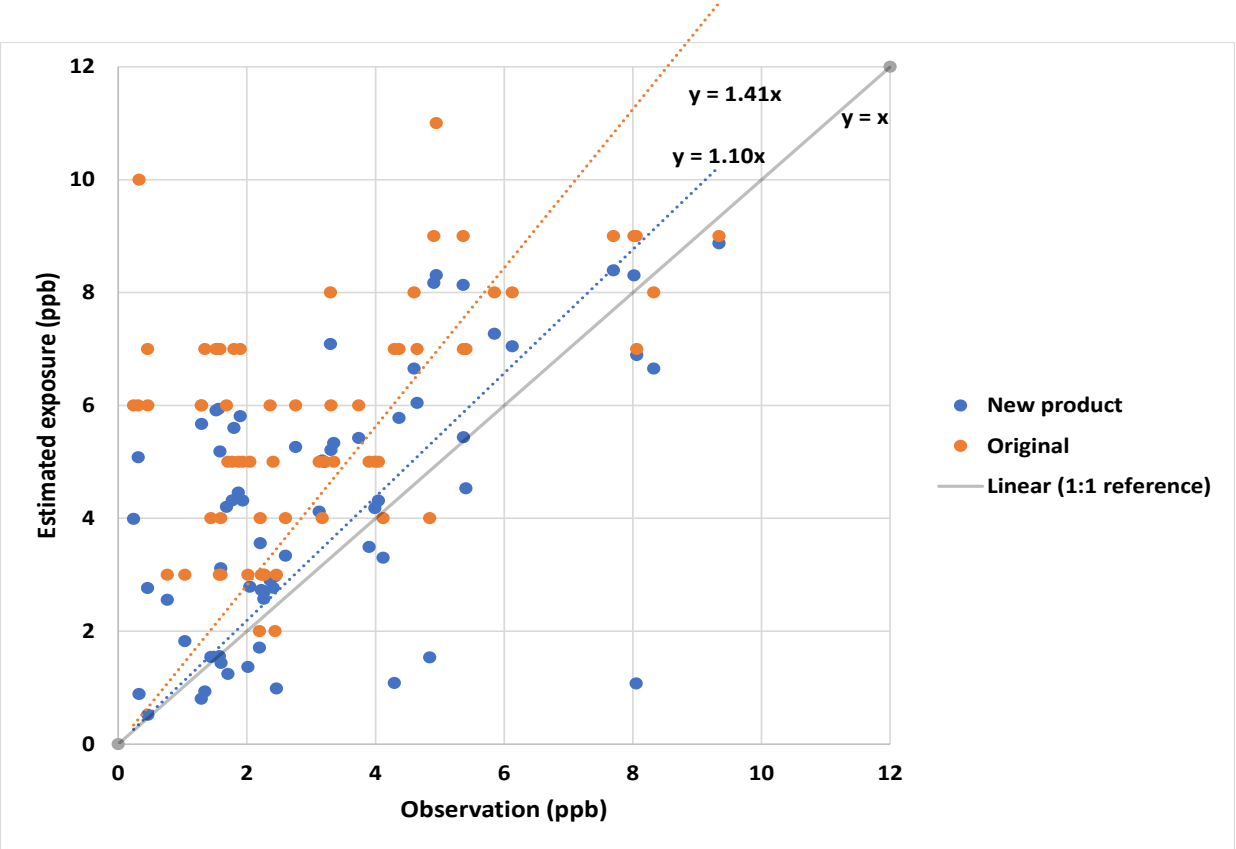

Figure S5. Comparison between annual average NO_2 concentrations from the "original" Larkin et al. (2017) product (orange) and our new NO_2 concentration product (blue), versus concentrations from ground measurements for 2011. A 1:1 reference line is added for comparison. Each point represents a monitor. Monitor data source: European Monitoring and Evaluation Programme (EMEP).

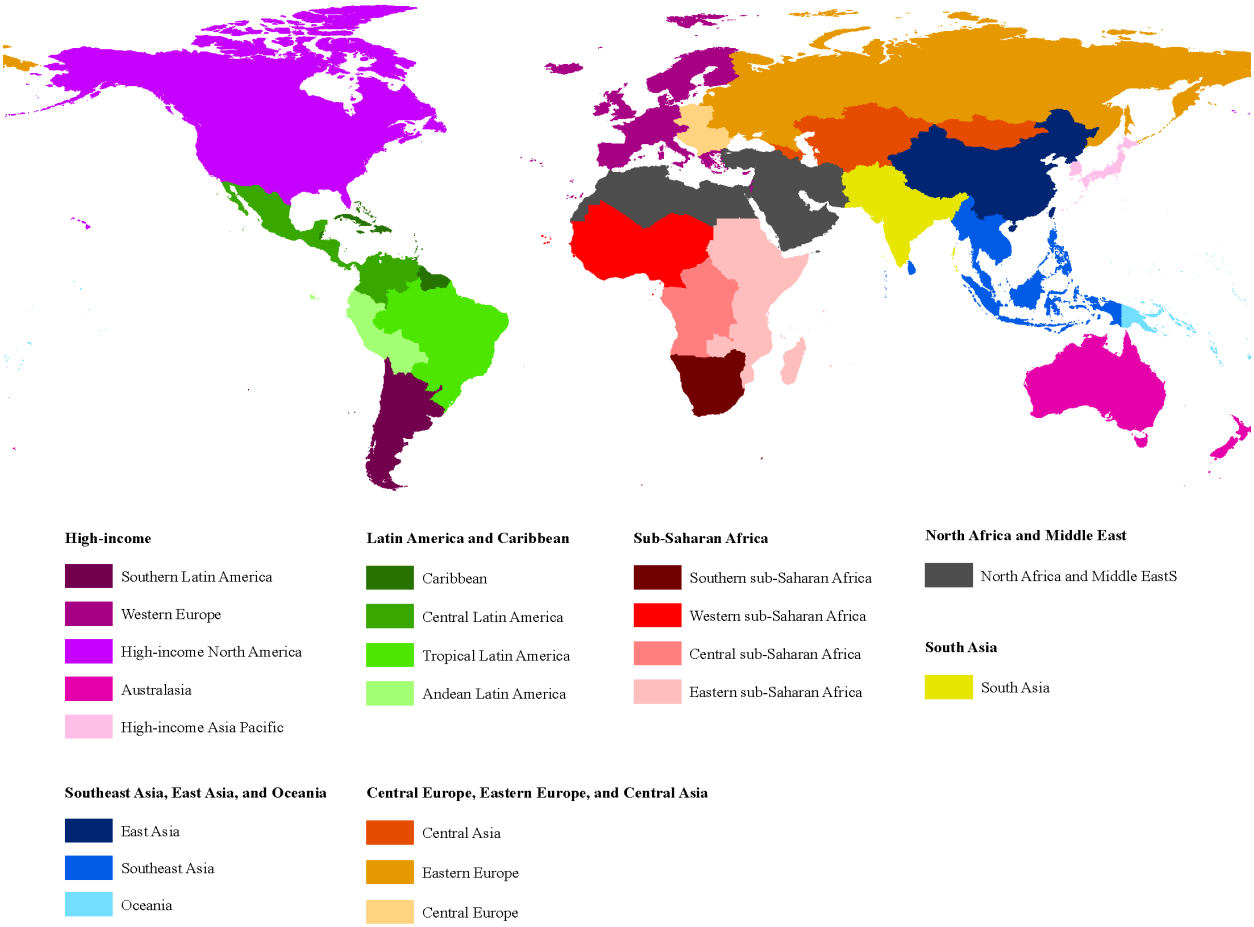

Figure S6. Countries and territories included in each region and super region, using regional definitions from the GBD 2019 Study.

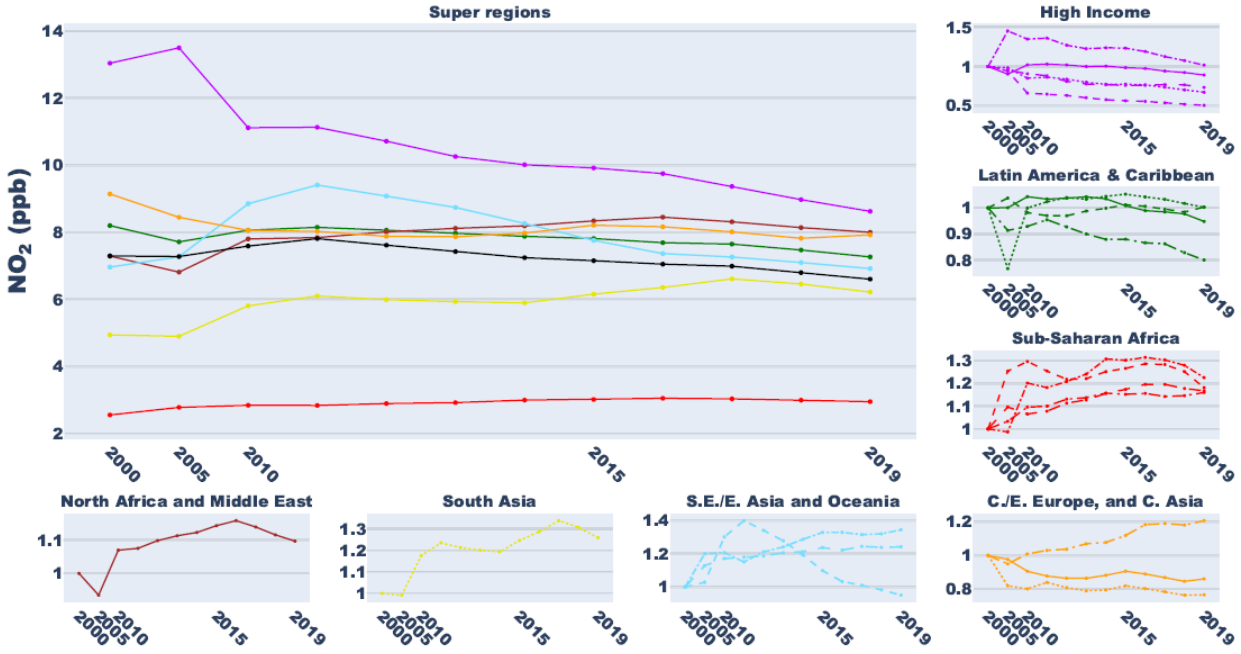

Figure S7. As for Figure 1, but including both urban and rural areas, whereas Figure 1 is only presenting the results for urban areas. Legend is the same as Figure 1.

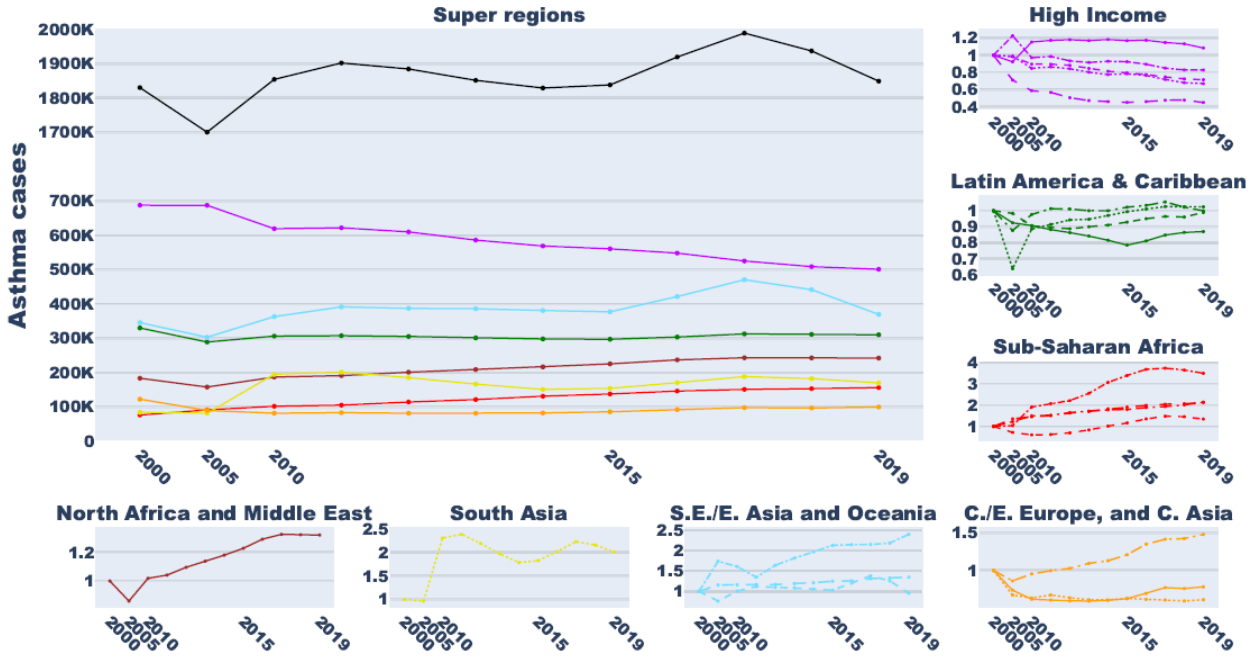

Figure S8. As for Figure 2, but including both urban and rural areas, whereas Figure 2 is only presenting the results for urban areas. Legend is the same as Figure 1.

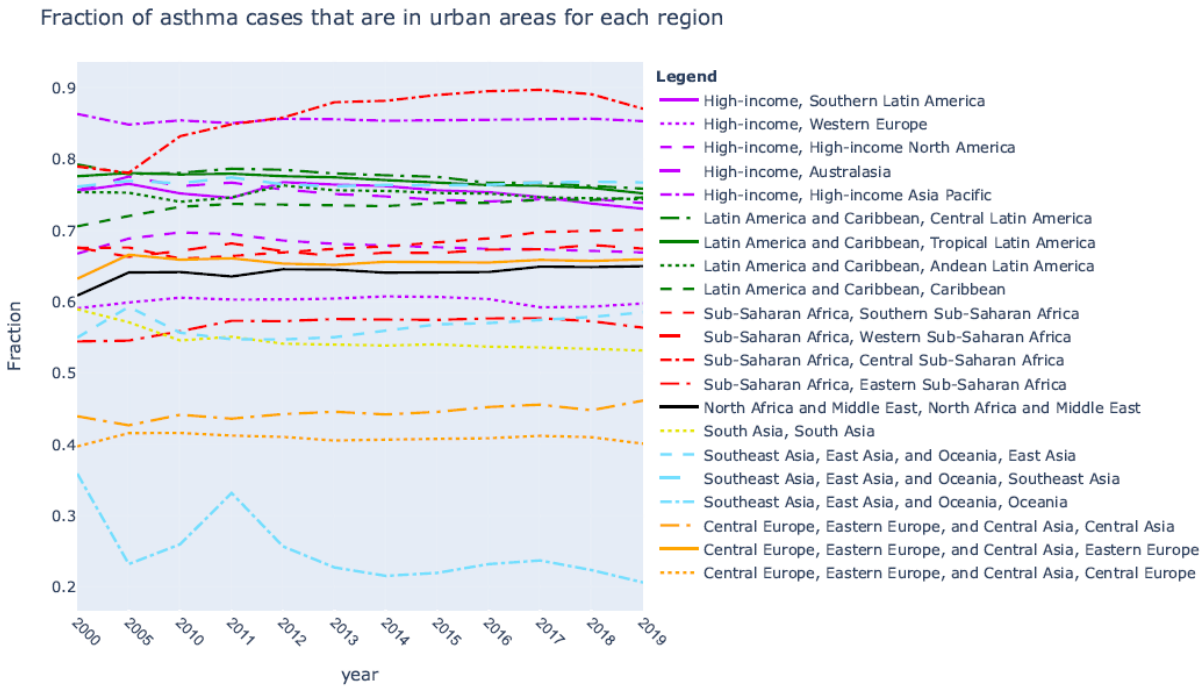

Figure S9. Trend of the fraction of regional estimated NO_2-attributable pediatric asthma incidence that occurs in urban areas.

Figure S10. As for Figure 4, but showing the full time trend for selected sub-regions.

References

Geddes, J. A., Martin, R. V., Boys, B.L., & van Donkelaar, A. (2016). Long-term trends worldwide in ambient NO_2 concentrations inferred from satellite observations. Environmental Health Perspectives, 124, 281–289. doi:10.1289/ehp.1409567

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., … Zhao, B. (2017). The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). Journal of Climate, 30, 5419–5454. doi:10.1175/JCLI-D-16-0758.1

Lamsal, L. N., Martin, R.V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, J. & Pinto, P. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. Journal of Geophysical Research Atmospheres, 113, 1–15. doi:10.1029/2007JD009235

Larkin, A., Geddes, J. A., Martin, R. V, Xiao, Q., Liu, Y., Marshall, J. D., … Hystad, P. (2017). Global land use regression model for nitrogen dioxide air pollution. Environmental Science & Technology, 51(12), 6957–6964. doi:10.1021/acs.est.7b01148

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., & Maffenini, L. (2019). GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A. European Commission, Joint Research Centre (JRC). doi:10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218