

<https://doi.org/10.1038/s41612-024-00683-5>

# Integrated satellite observations unravel the relationship between urbanization and anthropogenic non-methane volatile organic compound emissions globally

Check for updates

Dongchuan Pu<sup>1,2</sup>, Lei Zhu<sup>2,3,4</sup> , Huizhong Shen<sup>2,3,4</sup>, Isabelle De Smedt<sup>5</sup>, Jianhui Ye<sup>2,3,4</sup>, Juan Li<sup>1,2</sup>, Lei Shu<sup>2,6</sup>, Dakang Wang<sup>2,7</sup>, Xicheng Li<sup>2</sup>, Xiaoxing Zuo<sup>2</sup>, Xin Yang<sup>2,3,4</sup> & Tzung-May Fu<sup>2,3,4</sup>

As urban areas expand globally, human activities are leading to a sustained increase in non-methane volatile organic compound (NMVOC) emissions, escalating both environmental and health-related concerns. Given their diverse origins, estimating anthropogenic NMVOC emissions levels from global urban areas remains challenging. Here, we integrate TROPospheric Monitoring Instrument (TROPOMI) formaldehyde (HCHO) column data, Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime light (NTL) radiance data, and the Emission Database for Global Atmospheric Research (EDGAR) to develop a method for estimating global anthropogenic NMVOC emissions. Furthermore, we construct a linear model to analyze the relationship between urbanization and anthropogenic NMVOC emissions. Our research reveals that meticulously filtered TROPOMI HCHO columns have a high Pearson correlation coefficient ( $r = 0.91$ ) with anthropogenic NMVOC emissions, indicating its reliability as an indicator reflecting the levels of anthropogenic NMVOC emissions. We establish linear models at various scales, including global, continental, and national, linking HCHO columns (as indicators of anthropogenic NMVOC emissions) and NTL radiance (as an indicator of urbanization). The global-scale linear model exhibits an  $r$  of 0.81, with a *slope* of  $0.42 \times 10^{15}$  molec.  $\text{cm}^{-2}$  nanoWatts $^{-1}$   $\text{cm}^2$  sr and an *intercept* of  $9.26 \times 10^{15}$  molec.  $\text{cm}^{-2}$ . This linear model reflects a positive correlation between urbanization and anthropogenic NMVOC emissions, also serving as a tool for estimating the levels of anthropogenic NMVOC emissions in urban areas. This study offers valuable insights for real-time monitoring of extensive anthropogenic NMVOC emissions.

Among the key issues exacerbated by urbanization is air pollution<sup>1,2</sup>, such as the emission of large quantities of non-methane volatile organic compounds (NMVOCs) from human activities<sup>3,4</sup>. From 1970 to 2012, global anthropogenic NMVOC emissions increased from 119 Tg to 169 Tg<sup>5</sup>. NMVOCs are crucial in atmospheric chemistry, acting as precursors to ozone<sup>6</sup> and secondary organic aerosols<sup>7</sup>. Additionally, they pose health risks, including respiratory ailments and elevated cancer

risks<sup>8</sup>. The emitting activities of anthropogenic NMVOC emissions range from vehicular traffic, industrial processes, and biomass burning to smaller scales yet ubiquitously present actions such as cooking and the use of volatile chemical products<sup>9,10</sup>. The complexity of these sources makes anthropogenic NMVOC emissions especially hard to monitor, regulate, and control. Estimating the anthropogenic NMVOC emissions in urban areas globally still poses a challenge.

<sup>1</sup>School of Environment, Harbin Institute of Technology, Harbin, China. <sup>2</sup>School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>3</sup>Guangdong Provincial Observation and Research Station for Coastal Atmosphere and Climate of the Greater Bay Area, Shenzhen, Guangdong, China. <sup>4</sup>Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks, School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>5</sup>Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3, 1180 Uccle, Belgium. <sup>6</sup>School of Geographical Sciences, Fujian Normal University, Fuzhou, Fujian, China. <sup>7</sup>School of Geography and Remote Sensing, Guangzhou University, Guangzhou, Guangdong, China. e-mail: [zhul3@sustech.edu.cn](mailto:zhul3@sustech.edu.cn)

Recent studies underscore the significant potential of satellites in offering an expansive geographical and temporal perspective for NMVOC emissions monitoring<sup>3,11–14</sup>. Formaldehyde (HCHO), identified as an intermediary compound in the atmospheric oxidation of various NMVOCs<sup>15</sup>, positions satellite HCHO columns as effective indicators for terrestrial NMVOC emissions<sup>16,17</sup>. Satellites, such as the Global Ozone Monitoring Experiment<sup>18</sup> (GOME), the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography<sup>19</sup> (SCIAMACHY), the Ozone Monitoring Instrument<sup>20</sup> (OMI), and the TROPospheric Monitoring Instrument<sup>21</sup> (TROPOMI), play pivotal roles in retrieving HCHO columns. TROPOMI, a satellite payload of the European Space Agency's (ESA) S-5P mission, delivers daily global information on trace gases and aerosol concentrations at a resolution of 5–7 km crucial for air quality, climate forcing, and global environmental changes. Specifically, TROPOMI delivers high signal-to-noise ratio data on HCHO columns with superior spatial resolution<sup>22,23</sup>, which is pivotal for more precise NMVOC emissions monitoring. Integrating Visible Infrared Imaging Radiometer Suite<sup>24</sup> (VIIRS) nighttime light (NTL) radiance data<sup>25</sup>, which are indicative of urbanization, enhances our comprehension of the nexus between urbanization and anthropogenic NMVOC emissions.

Urban areas stand as focal points driving environmental changes from local to global scales<sup>26–28</sup>. The interplay between urbanization and ecological degradation, such as atmospheric pollution, has emerged as a critical topic in urban sustainable development<sup>29</sup>. Different air pollutants and their precursors exhibit varied responses to urbanization<sup>30–32</sup>. For instance, NO<sub>2</sub> emissions displays an inverted U-shaped pattern in relation to urbanization, as represented by per capita income<sup>33</sup>, while PM<sub>2.5</sub> shows a linear trend with respect to time<sup>34</sup>. Globally, the relationship between anthropogenic NMVOC emissions and urbanization remains ambiguous<sup>35</sup>, due largely to uncertainties in estimating the anthropogenic NMVOC emissions in urban areas<sup>36–38</sup>. While some studies have expanded our grasp of biogenic NMVOC emissions through satellite HCHO columns<sup>15,39,40</sup>, current scholarly work has yet to adequately meld global HCHO columns with anthropogenic NMVOC emissions in urban areas. Global NMVOC emissions predominantly originate from plants<sup>16,41–43</sup>, making the precise identification of anthropogenic sources from satellites a challenge.

Here, we utilize data from TROPOMI HCHO columns, VIIRS NTL radiance, the Emission Database for Global Atmospheric Research (EDGAR)<sup>44</sup>, and the Global Earth Observing System Chemistry (GEOS-Chem)<sup>45</sup> to monitor the global levels of anthropogenic NMVOC emissions and the impact of urbanization on these emissions. We develop a satellite-based technical approach to estimate the levels of anthropogenic NMVOC emissions; furthermore, we establish linear models relating anthropogenic NMVOC emissions to urbanization from a national to a global scale and conduct sensitivity experiments. Our aim is to provide references for global air quality assessments and promote sustainable development of the environment.

## Results and discussion

### Global HCHO columns, HCHO emission rates, and NTL radiance

Figure 1 shows the global HCHO columns from TROPOMI, HCHO emission rates based on the EDGAR inventory, and NTL radiance from VIIRS. Our research primarily focuses on urban areas located in the northern hemisphere. We use TROPOMI HCHO column data from May to October in both 2019 and 2021 to ensure data quality, as satellite HCHO columns exhibit weaker signals during the in other months (winter) in the northern hemisphere<sup>46</sup> (Supplementary Figure 1). The 2020 outbreak of coronavirus disease (COVID-19) significantly impacted the global economy, leading to marked variations in the HCHO columns<sup>11,47</sup>. As a result, we exclude the 2020 data. TROPOMI Level 2 column pixels are filtered based on cloud fraction (< 30%), solar zenith angle (< 60°), and quality assurance values (> 0.5). We then resample all eligible level-2 pixels onto a 0.05° × 0.05° (~5 × 5 km<sup>2</sup>) grid for the periods mentioned, built on our previous oversampling approach<sup>36</sup>. The TROPOMI HCHO columns (Fig. 1a) provide a top-down proxy of anthropogenic NMVOC emissions.

Using the EDGAR inventory, we estimated the HCHO emission rates (Fig. 1b), both primary and secondary, based on five highly reactive volatile organic compounds (VOCs) with atmospheric lifetimes of less than 3 hours (Supplementary Table 1), employing the methodology proposed by Palmer et al.<sup>17,42</sup>. When not constrained by real-time constraints, the HCHO emission rates estimated from the EDGAR inventory reflect the more tangible characteristics of anthropogenic NMVOC emissions from a bottom-up perspective.

Here, we employ VIIRS NTL radiance data as an indicator of urbanization levels (Fig. 1c). NTL radiance data are collected through remote sensing and capture the amount of artificial lighting during nighttime. One common application of nightlight data is to explore urbanization, for instance, by mapping urban areas<sup>48–51</sup> and estimating economic parameters<sup>52,53</sup>. We synthesize the 2019 VIIRS NTL monthly radiation data based on the Google Earth Engine<sup>54,55</sup> platform with a previously used median composite method<sup>55,56</sup> that can remove some interference from wildfires and ice.

Globally, evident relationship exists among HCHO columns, HCHO emission rates, and NTL radiance. Developed regions and major urban centers typically exhibit elevated nighttime radiance, reflecting heightened human activities and urban dynamism, often corresponding to increased NMVOC anthropogenic emissions. However, this relationship varies across geographic regions. In Asia, large regions such as eastern China and northern India, where nighttime radiance has significantly increased, also demonstrate high HCHO columns and HCHO emission rates, underscoring their rapid urbanization. In contrast, Africa and South America primarily feature natural sources of NMVOC emissions<sup>57</sup>, making anthropogenic emissions challenging to discern based on HCHO columns.

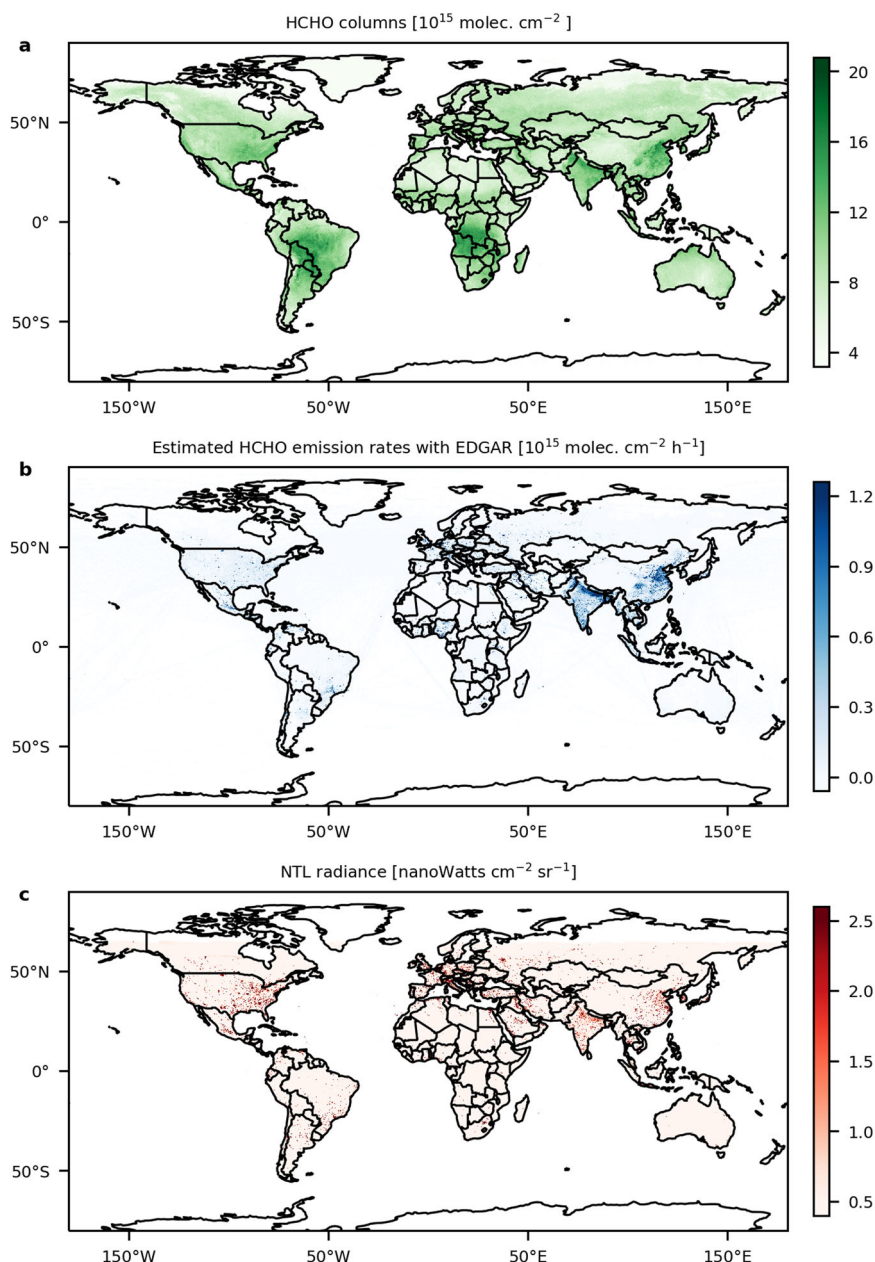
Furthermore, Supplementary Figure 2 shows the variations at the city scale in HCHO columns, HCHO emission rates, and NTL radiance for Paris (Europe), New York (North America), and Tokyo (Asia). The selected cities are developed metropolises with the highest per capita GDP<sup>58</sup> in their respective continents. The normalized latitudinally averaged results (Supplementary Figure 2j–l) present a comprehensive comparison of these three cities. The HCHO columns, HCHO emission rates, and NTL radiance in Paris show a relatively consistent pattern of change (Supplementary Figure 2j), whereas New York and Tokyo exhibit lower consistency than Paris. Globally, using HCHO columns as indicators of anthropogenic NMVOC emissions presents certain challenges, even in developed urban areas, due to the need to consider the effects of factors such as NO<sub>x</sub> and biological NMVOC emissions.

Through spatial sampling and filtering, we have identified urban sites globally dominated by anthropogenic NMVOC emissions. Using NTL radiance with intervals of 0.1 log (nanoWatts cm<sup>-2</sup> sr<sup>-1</sup>), we perform a linear correlation analysis. Supplementary Figure 3a shows the relationship between HCHO columns and HCHO emission rates for the selected months (May to October). It is observed that higher levels of urbanization correlate with higher anthropogenic NMVOC emissions, as evidenced by the concurrent increase in HCHO columns and HCHO emission rates. Moreover, HCHO columns and HCHO emission rates exhibit a high positive correlation, with a Pearson correlation coefficient (*r*) of 0.91. Supplementary Figure 3b presents the relationship between HCHO columns and HCHO emission rates for other months, showing a lower *r*, which indicates a weaker capability to indicate anthropogenic NMVOC emissions.

### Global anthropogenic NMVOC emissions and urbanization

Figure 2 shows the linear models between TROPOMI HCHO columns (indicating anthropogenic NMVOC emissions) and VIIRS NTL radiance (indicating urbanization) for the entire globe and four continents: Asia, Africa, Europe, and North America. These representations are derived from linear models applied to selected urban sites. Considering the high skewness in NTL radiance (*skewness* = 3.31), the linear models are conducted in logarithmic space. VIIRS NTL radiance values are categorized into bins with a width of 0.1 log (nanoWatts cm<sup>-2</sup> sr<sup>-1</sup>). Mean values for VIIRS NTL radiance and TROPOMI HCHO columns in each bin are calculated and

**Fig. 1 | Global TROPOMI HCHO columns, estimated HCHO emission rates with EDGAR, and VIIRS NTL radiance.** **a** Shows the TROPOMI HCHO columns for May through October 2019 and 2021, resampled to a grid resolution of  $0.05^\circ \times 0.05^\circ$  ( $\sim 5 \times 5 \text{ km}^2$ ) grid resolution. Panel **(b)** shows HCHO emission rates in 2012 at a resolution of  $0.1^\circ \times 0.1^\circ$  ( $\sim 10 \times 10 \text{ km}^2$ ) estimated from anthropogenic NMVOC emissions of five species (ethene, propene, isoprene, monoterpenes, and primary formaldehyde) from EDGAR using the methodology proposed by Palmer et al.<sup>42</sup>. Panel **(c)** shows VIIRS annual NTL radiance in 2019 at a resolution of  $0.05^\circ \times 0.05^\circ$  ( $\sim 5 \times 5 \text{ km}^2$ ), computed based on the median synthesis method provided by the Google Earth Engine.



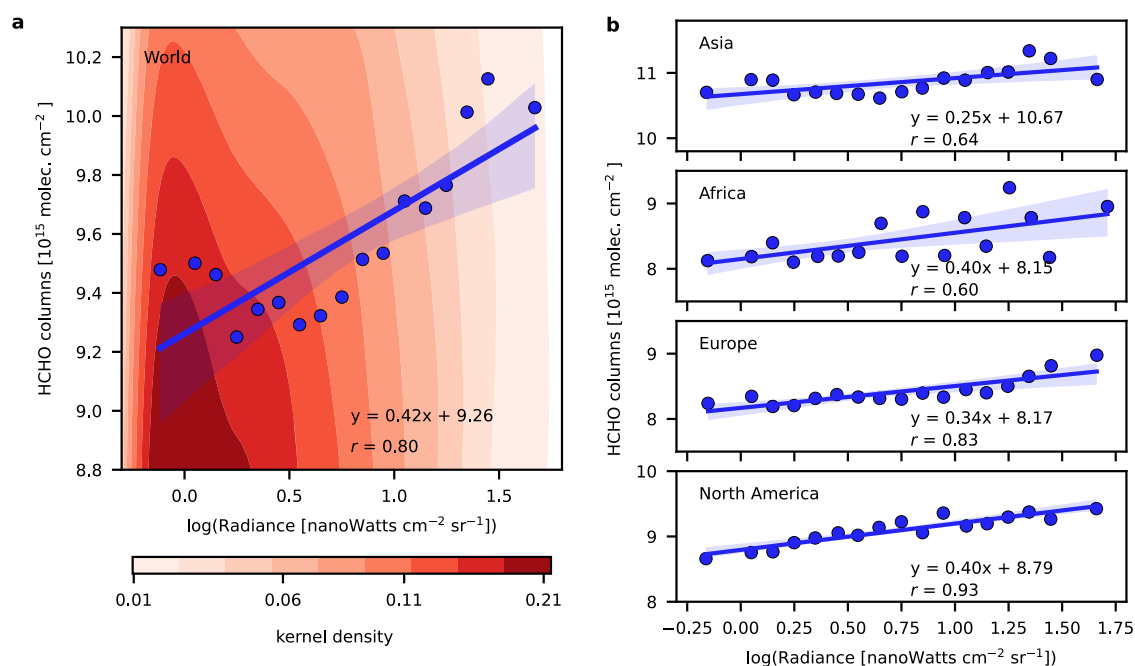
then subjected to linear fitting. The global-scale linear model (Fig. 2a) exhibits an  $r$  of 0.81, with a *slope* of  $0.42 \times 10^{15} \text{ molec. cm}^{-2} \text{ nanoWatts}^{-1} \text{ cm}^2 \text{ sr}$  and an *intercept* of  $9.26 \times 10^{15} \text{ molec. cm}^{-2}$ . The results reveal a significant positive correlation ( $p$  value ranging from 0 to 0.03), with  $r$  ranging from 0.60 to 0.93, across the four continents. We obtained the uncertainty of the models based on the 1000 Monte Carlo analysis (Supplementary Fig. 4). The results show the distribution of slopes and intercepts, which implies the stability of the method used for the linear model.

A particularly insightful metric is the *intercept*, which reflects the background HCHO columns in urban areas, primarily derived from the oxidation of methane and long-lived VOCs. Through spatial sampling and filtering, NMVOC emissions at urban sites globally are mainly from anthropogenic sources, although there is also an influence from biological emissions. Urban sites in Asia record the highest *intercept*, possibly indicating high emissions of methane and long-lived VOCs.

The high  $r$  and *slope* suggest that urbanization more strongly drives anthropogenic NMVOC emissions, possibly due to regional differences in industrial activities and energy use patterns. Figure 3a illustrates the proportion of sector emission contributions for urban sites on the four

continents, as estimated from the EDGAR inventory. Across the four continents, industrial and transportation contributions to anthropogenic NMVOC emissions exceed 75%, with other sectors contributing a smaller proportion of anthropogenic NMVOC emissions. Supplementary Table 2 shows the proportion of species emission contributions for urban sites on the four continents, as estimated from the EDGAR inventory. For Asia, Europe, and North America, ethene and propene dominate urban anthropogenic NMVOC emissions, accounting for more than 60%. In contrast, primary formaldehyde is the predominant component in Africa, constituting 63% of anthropogenic NMVOC emissions. Generally, the structure of emission proportions across continents appears similar, suggesting that industrial and transportation NMVOC emissions likely drive the linear model.

On the one hand, as the VIIRS NTL radiance values for urban sites increase, the five anthropogenic NMVOC emissions also rise (Supplementary Figure 5). On the other hand, the structure (or proportionate contributions) of anthropogenic NMVOC emissions across various sectors changes (Fig. 3b). The VIIRS NTL radiance values are still categorized into 0.1 log ( $\text{nanoWatts cm}^{-2} \text{ sr}^{-1}$ ) bins, with the average percentage



**Fig. 2 | Comparison of TROPOMI HCHO column with VIIRS NTL radiance grouped by global regions.** In panel (a), a point represents the mean HCHO columns at a specific NTL radiance bin for all urban grid cells. The blue line shows the simple linear regression line, with a blue area enveloping the 95% confidence

interval of the mean response. Pearson correlation coefficients ( $r$ ) and the linear model equation are also inserted. Kernel density estimates for all global urban sites are plotted in the background. Panel (b) shows the results of the linear models for the four continents, including Asia, Africa, Europe, and North America.

contribution of each sector’s anthropogenic NMVOC emissions calculated for each bin. The standard deviation of the percentages for each bin is then computed to represent structural changes. From the smallest to the largest bin, the standard deviations for the four continents vary by 10.8%, 3.3%, 14.2%, and 4.2%, respectively.

We also analyze the relationship between anthropogenic NMVOC emissions (indicated by TROPOMI HCHO columns) and urbanization (indicated by VIIRS NTL radiance) in the 12 major countries (defined as those with more than 500 urban sites), as shown in Supplementary Table 4. Overall, we find significant linear relationships in 8 of the 12 major countries, with  $r$  ranging from 0.59 to 0.95. We examine the possible driving factors for the differences in response relationships between major countries by the contributions of various sectors and species (Supplementary Table 5). For example, the correlation coefficient between anthropogenic NMVOC emissions and urbanization in Brazil ( $r = 0.95$ ) is much higher than that in India ( $r = 0.59$ ) (Supplementary Table 4). We hypothesize that this difference is caused by the contribution of emission sectors (72% vs. 56%), including industry and ground transport. The response of NMVOC to urbanization is more likely related to industry and ground transport emissions rather than residential and agricultural sources. In short, this is consistent with the comparison of some local cities (Supplementary Fig. 2) and the global comparison result (Fig. 2 and Supplementary Table 3), indicating that there are similar positive patterns in the responses of anthropogenic NMVOC emissions to urbanization at the scales of cities, countries, continents, and even globally.

### Sensitivity simulations for the linear models

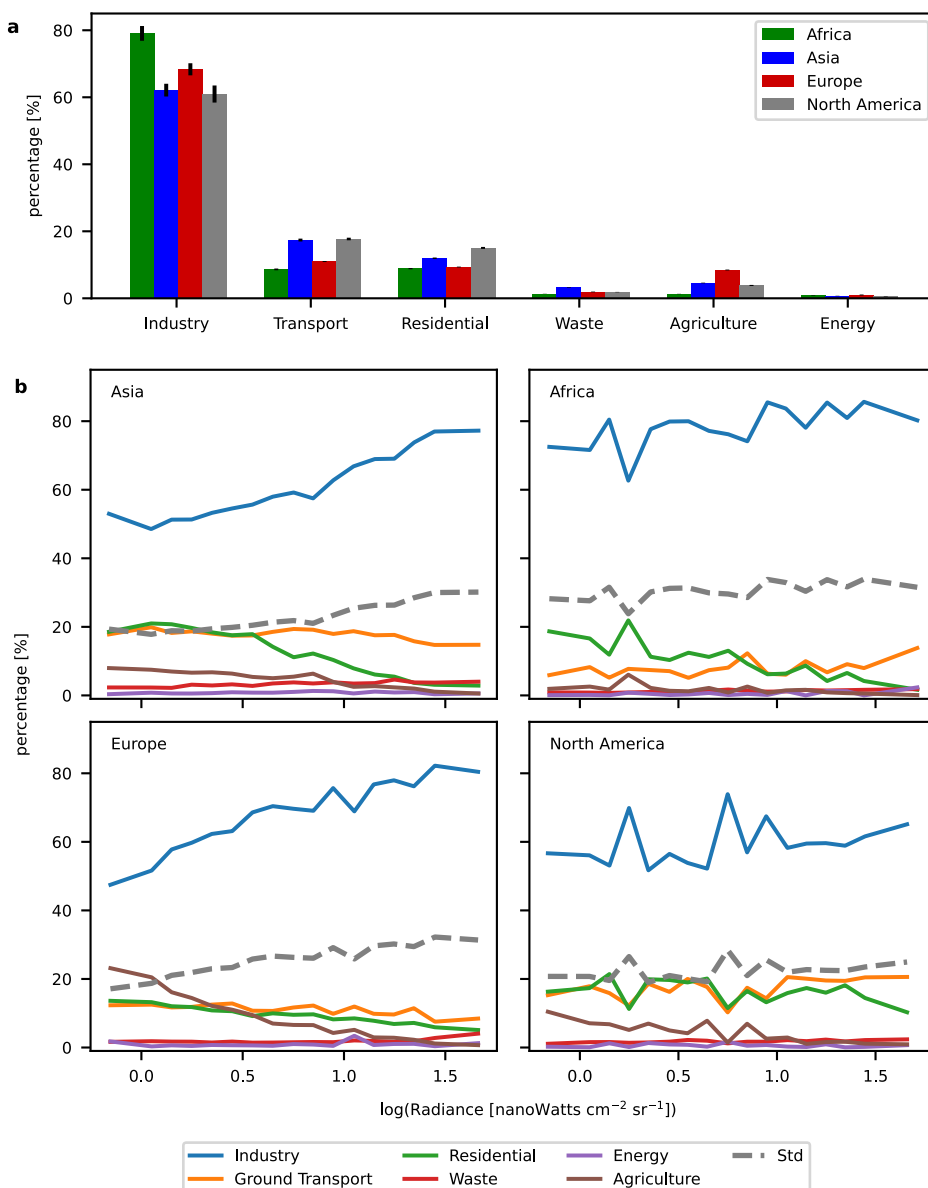
Factors such as  $\text{NO}_x$  and biogenic NMVOC emissions can affect HCHO columns indicative of anthropogenic NMVOC emissions, which can further interfere with the reliability of the linear models. One of the interference factors in the linear models is the biogenic NMVOC emissions. Subsequently, using the Olson terrestrial ecoregions<sup>59</sup>, we assess the influence of biological NMVOC emissions on those linear models. Each urban site is marked according to its ecoregion, revealing that the five most prevalent

biomes are tropical and subtropical moist broadleaf forests, temperate grasslands, savannas and shrublands, deserts and xeric shrublands, temperate broadleaf and mixed forests, and boreal forests or taigas.

By excluding urban sites from each of the five ecoregions and re-establishing the global-scale linear model, we compare the subsequent changes in the  $r$ , *slope*, and *intercept*, as shown in Fig. 4a. The exclusion of sites from the temperate broadleaf and mixed forest biomes led to a small increase in the  $r$  of the linear model, indicating that these sites are influenced to a lesser degree by biological NMVOC emissions. When excluding sites from the other four ecoregions, the  $r$  decreases slightly, indicating that there are no strong biological NMVOC emissions. Following the exclusion of sites from the temperate broadleaf and mixed forests biome for each continent, linear models are established and are presented in Supplementary Table 3. Notably, the standard deviations (Fig. 3b) exhibit a Spearman rank correlation of 0.9 with linear models’  $r$  and 0.8 with the *slope*. The variation in the structure of sectoral anthropogenic NMVOC emissions reflects the differences in the various linear models (both the *slope* and  $r$ ).

Subsequently, we establish relationships between EDGAR HCHO emission rates and NTL radiance for four continents (Supplementary Table 3). In Europe, the linear relationship between HCHO columns, HCHO emission rates, and NTL radiation displays a high degree of consistency. The model’s  $r$  (0.91 vs. 0.90) and *slope* (0.58 vs. 0.59) are nearly identical. There is also a noticeable consistency in Asia and North America. However, a significant divergence is observed in Africa. We hypothesize that this disparity stems from more potent biological NMVOC emissions and varied emission species contributions within the continent (Supplementary Table 2). We also establish relationships between HCHO emission rates and NTL radiance for the 12 major countries (Supplementary Table 4). The linear model between HCHO columns, HCHO emission rates, and NTL radiation displays some degree of consistency. The model’s  $r$  in Indonesia (0.74 vs. 0.71) and Russia (0.68 vs. 0.59) are similar. The model’s *slope*, in Indonesia (0.80 vs. 1.96) and Russia (0.32 vs. 0.62) have similar patterns. This suggests that our top-down approach for monitoring anthropogenic NMVOC emissions remains broadly reliable.

**Fig. 3 | Evaluating the structure of anthropogenic NMVOC emissions from various sectors.** Panel (a) shows the share of emission contributions by sector for urban sites on four continents based on EDGAR inventory statistics. The error bar indicates the standard deviation of 1000 simulations by the Montecarlo method. Panel (b) depicts the variations (standard deviation) in the structure (contribution percentage) of anthropogenic NMVOC emissions from different sectors with increasing VIIRS NTL radiance at urban sites.



Another interference factor in the linear model is  $\text{NO}_x$  emissions. As  $\text{NO}_x$  an important participant in the photochemical reaction that oxidizes NMVOCs to HCHO<sup>60,61</sup>, we base five sensitivity simulations with GEOS-Chem (nested version,  $0.5^\circ \times 0.625^\circ$ ) and TROPOMI  $\text{NO}_2$  column data to quantify this primary interference. Figure 4b shows the differences in the linear models established using the aforementioned methodology between GEOS-Chem simulated HCHO columns and VIIRS NTL radiances under varying  $\text{NO}_x$  levels. It is found that when  $\text{NO}_x$  levels are not zero, the influence of  $\text{NO}_x$  on the linear models (*r*, *slope*, and *intercept*) is generally less than 20%, indicating the reliability of the linear model. We also use the previously proposed method<sup>35</sup> to quantify this dependency among 4 major countries from four continents by TROPOMI  $\text{NO}_2$  columns. For 4 countries with a significant positive correlation (*p*-value = 0.01) between anthropogenic NMVOC emissions and urbanization, we found that this effect exists in most countries but does not alter this positive linear relationship. For example, after considering the spatial variation in  $\text{NO}_x$  emissions, the correlation coefficient of Brazil decreases by 8%, and the *slope* decreases by 19% (Supplementary Figure 6). Therefore, we believe that the linear pattern between HCHO columns and VIIRS NTL radiance is mainly driven by the NMVOC emission gradient within each country, but the impact of  $\text{NO}_x$  emissions still needs to be noted.

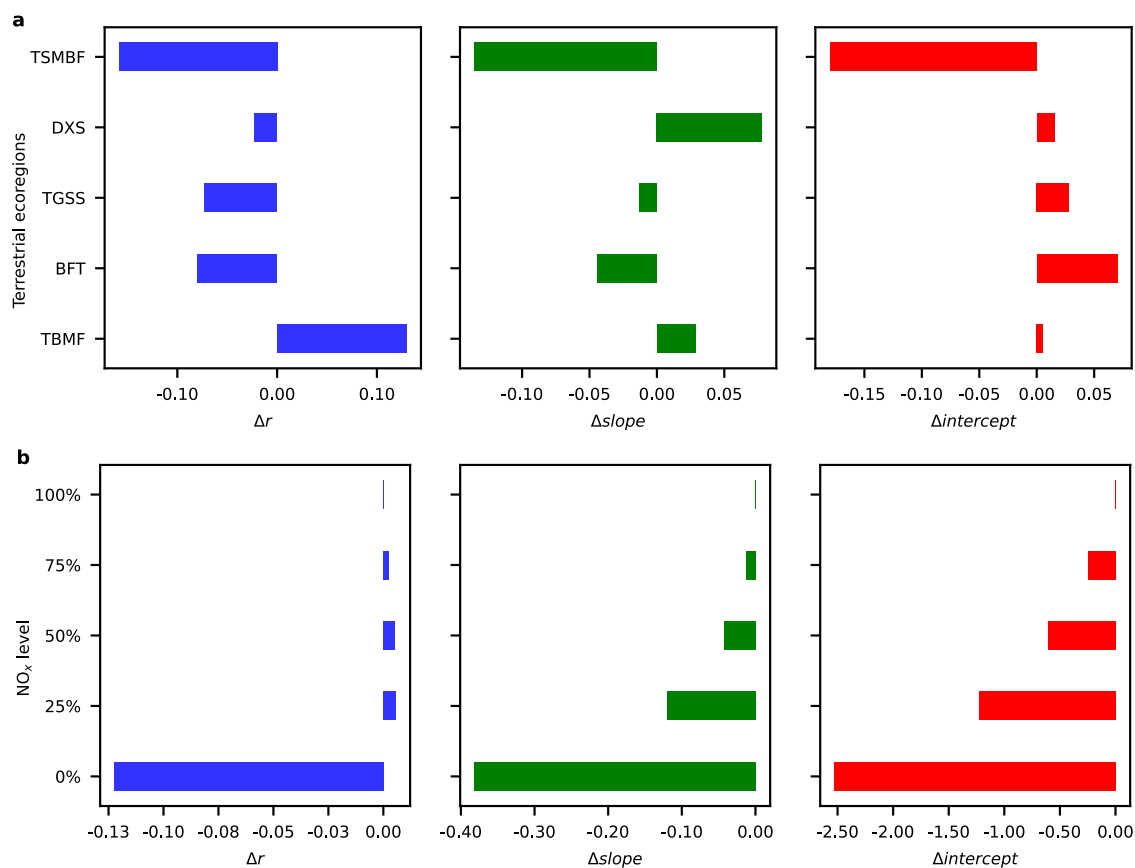
## Methods

### TROPOMI HCHO and $\text{NO}_2$ columns

TROPOMI is a space-based observational hyperspectral imager that measures atmospheric properties and composition<sup>21</sup>, developed jointly by the Netherlands and European Space Agency (ESA). The instrument uses passive remote sensing technology to retrieve atmospheric trace gas columns by measuring solar radiation reflected and radiated at the top of the atmosphere<sup>22,62</sup>. TROPOMI provides daily products of HCHO and  $\text{NO}_2$  columns with global coverage and a high spatial resolution<sup>22,63</sup> of up to  $5.0 \times 3.5 \text{ km}^2$ . Vigouroux et al.<sup>23</sup> based on Fourier transform infrared (FTIR) observations, validated the accuracy of the TROPOMI HCHO column product and demonstrated its reliability. The TROPOMI HCHO<sup>64</sup> and  $\text{NO}_2$ <sup>63,65</sup> column products have passed an intercomparison check with the OMI product.

### Quantitative analysis and spatial sampling

To quantify the response patterns of anthropogenic NMVOC emissions to urbanization, we eliminate the influences of wildfires and biogenic sources on NMVOCs. Air temperature is the main driving factor for the seasonal variations in NMVOCs from biogenic sources<sup>40,66,67</sup>. To mitigate the impact of biogenic sources and wildfires, we select grid cells based on two criteria.



**Fig. 4 | Sensitivity simulations for the linear model.** Panel (a) illustrates the comparison of the linear model ( $r$ ,  $slope$ , and  $intercept$ ) before and after excluding urban site data from five terrestrial ecoregions. Panel (b) shows the differences

between the linear models established from GEOS-Chem simulated HCHO columns and VIIRS NTL radiance at different  $NO_x$  levels, relative to when the  $NO_x$  level is at 100%.

One criterion is that HCHO columns weakly depend on air temperature<sup>36</sup>. This is done by using the exponential dependence (Supplementary Fig. 7) of biogenic HCHO columns on air temperature from NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)<sup>55</sup>. We then exclude grid cells with strong biogenic influence ( $R^2 > 0.5$  and  $slope > 0.05$ )<sup>35</sup>. The second criterion is to eliminate grid cells (carbon monoxide emission flux  $> 1 \times 10^{-6} \text{ kg m}^{-2} \text{ year}^{-1}$ ) affected by wildfires based on the Global Fire Emissions Database version 4 (GFED 4) in 2019<sup>68</sup>.

We restrict our analysis to urban sites. The main data issue with VIIRS NTL radiance is background noise, such as the high reflection of ice and snow<sup>69</sup>. To reduce its impact, we survey 119,670 global urban sites (NTL radiance  $> 0.5 \text{ nanoWatts cm}^{-2} \text{ sr}^{-1}$ ) by using information about impervious surfaces, population density, and industrial and commercial centers<sup>56</sup>. The accuracy (95%) of the defined urban sites is verified by visually inspecting 1000 randomly selected sites against corresponding high-resolution remote sensing images in Google Earth. Some impacts of biogenic sources and wildfires have been further eliminated to limit the selection of urban sites.

### Sensitivity tests with the GEOS-Chem model

GEOS-Chem (<http://www.geos-chem.org>), a global three-dimensional (3-D) model of tropospheric chemistry, has been previously used to simulate HCHO under various  $NO_x$  conditions<sup>42,67</sup>. To explore the impact of  $NO_x$  emissions on HCHO columns in major countries, we perform five simulations using the nested version ( $0.5^\circ \times 0.625^\circ$ ) of the GEOS-Chem model (version 12.9.3) globally, driven by the MERRA-2<sup>70</sup> meteorological fields. In the control run, we use anthropogenic emissions from the EDGAR inventory, biogenic VOC emissions from the MEGAN v2.1<sup>71</sup>, and open fire emissions from the fourth-generation global fire emissions database

(GFED4). In the sensitivity runs, we reduce anthropogenic  $NO_x$  emissions in EDGAR by 25%, 50%, 75%, and 100%, and keep other emissions the same as in the control run. All runs are for July 2019, with a spinning-up time of 1 month.

We use TROPOMI tropospheric  $NO_2$  data<sup>65</sup> along with GEOS-Chem results to quantify the impact of  $NO_x$  emissions on the spatial distribution of HCHO columns within a certain country. We filter, regrid, and sample  $NO_2$  TROPOMI pixels following the same methods as HCHO. Second, we compute the relative change in the tropospheric  $NO_2$  columns at grid cell  $i$ ,  $\Delta NO_2(i)$ , from the background, defined as the minimum  $NO_2$  columns in that country. As tropospheric  $NO_2$  columns and  $NO_x$  emissions are linearly related, such a relative change in tropospheric  $NO_2$  columns,  $\Delta NO_2(i)$ , can be roughly regarded as the relative change in  $NO_x$  emissions,  $\Delta NO_x(i)$ , from a baseline emission of that country. Third, we build a 4-segmented function at grid cell  $i$ ,  $\delta(\Delta NO_x(i), i)$ , to describe the relative change in the HCHO columns,  $\Delta HCHO(i)$ , in response to the reduction in  $NO_x$  emissions,  $\Delta NO_x(i)$ , based on localized results from GEOS-Chem sensitivity simulations, which are run with 25%, 50%, 75%, and 100% reductions in  $NO_x$  emissions (Supplementary Fig. 8). Finally, we substitute TROPOMI-based  $\Delta NO_2(i)$  to quantify the impact of spatial variations of  $NO_x$  emissions on HCHO columns at each grid cell  $i$  by interpolating  $\delta(\Delta NO_x(i), i)$  between the two nearest reductions.

### Data availability

The TROPOMI HCHO,  $NO_2$ , and MERRA-2 products used in this study were downloaded from the NASA Goddard Earth Sciences Data and Information Services Center ([https://disc.gsfc.nasa.gov/datasets/S5P\\_L2\\_HCHO\\_\\_\\_1/summary?keywords=TROPOMI%20HCHO](https://disc.gsfc.nasa.gov/datasets/S5P_L2_HCHO___1/summary?keywords=TROPOMI%20HCHO), [https://disc.gsfc.nasa.gov/datasets/S5P\\_L2\\_NO2\\_\\_\\_HiR\\_1/summary?keywords=](https://disc.gsfc.nasa.gov/datasets/S5P_L2_NO2___HiR_1/summary?keywords=)

NO<sub>2</sub>, and [https://disc.gsfc.nasa.gov/datasets/M2T1NXSLV\\_5.12.4/summary?keywords=MERRA-2](https://disc.gsfc.nasa.gov/datasets/M2T1NXSLV_5.12.4/summary?keywords=MERRA-2)). The VIIRS night light data are from NOAA National Centers for Environmental Information (NCEI) ([https://ngdc.noaa.gov/eog/viirs/download\\_ut\\_mos.html](https://ngdc.noaa.gov/eog/viirs/download_ut_mos.html)). The EDGAR data are from the European Union Joint Research Centre ([https://edgar.jrc.ec.europa.eu/dataset\\_htap\\_v3](https://edgar.jrc.ec.europa.eu/dataset_htap_v3)).

### Code availability

The code used to oversample satellite data on: <https://zenodo.org/record/6843869#.YtjMLXZByUk>.

Received: 21 November 2023; Accepted: 5 June 2024;

Published online: 11 June 2024

### References

- Lelieveld, J., Evans, J., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**, 367–371 (2015).
- Collaborators, GBDRF. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1923–1994 (2018).
- Cao, H. S. et al. Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal. *Atmos. Chem. Phys.* **18**, 15017–15046 (2018).
- Curci, G., Palmer, P., Kurosu, T., Chance, K. & Visconti, G. Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument. *Atmos. Chem. Phys.* **10**, 11501–11517 (2010).
- Huang, G. L. et al. Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970–2012. *Atmos. Chem. Phys.* **17**, 7683–7701 (2017).
- Wei, W. et al. A WRF-Chem model study of the impact of VOCs emission of a huge petrochemical industrial zone on the summertime ozone in Beijing, China. *Atmos. Environ.* **175**, 44–53 (2018).
- McDuffie, E. et al. A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS). *Earth Syst. Sci. Data* **12**, 3413–3442 (2020).
- Strum, M. & Scheffe, R. National review of ambient air toxics observations. *J. Air Waste Manag. Assoc.* **66**, 120–133 (2016).
- Streets, D. G. et al. Emissions estimation from satellite retrievals: A review of current capability. *Atmos. Environ.* **77**, 1011–1042 (2013).
- Martin, R. V. Satellite remote sensing of surface air quality. *Atmos. Environ.* **42**, 7823–7843 (2008).
- Pakkattil, A., Muhsin, M. & Varma, M. K. R. COVID-19 lockdown: Effects on selected volatile organic compound (VOC) emissions over the major Indian metro cities. *Urban Clim.* **37**, 135–149 (2021).
- Souri, A. H. et al. An inversion of NO<sub>x</sub> and non-methane volatile organic compound (NMVOC) emissions using satellite observations during the KORUS-AQ campaign and implications for surface ozone over East Asia. *Atmos. Chem. Phys.* **20**, 9837–9854 (2020).
- Wells, K. C. et al. Satellite isoprene retrievals constrain emissions and atmospheric oxidation. *Nature* **585**, 225–233 (2020).
- Fu, D. et al. Direct retrieval of isoprene from satellite-based infrared measurements. *Nat. Commun.* **10**, 3811 (2019).
- Abbot, D. S. et al. Seasonal and interannual variability of North American isoprene emissions as determined by formaldehyde column measurements from space. *Geophys. Res. Lett.* **30**, n/a–n/a (2003).
- Millet, D. B. et al. Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission. *J. Geophys. Res. Atmos.* **111**, 1129–1138 (2006).
- Palmer, P. et al. Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column. *J. Geophys. Res. Atmos.* **111**, n/a–n/a (2006).
- Chance, K. et al. Satellite observations of formaldehyde over North America from GOME. *Geophys. Res. Lett.* **27**, 3461–3464 (2000).
- De Smedt, I. et al. Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors. *Atmos. Chem. Phys.* **8**, 4947–4963 (2008).
- Abad, G. et al. Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval. *Atmos. Meas. Tech.* **8**, 19–32 (2015).
- Veeffkind, J. et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **120**, 70–83 (2012).
- De Smedt, I. et al. Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. *Atmos. Meas. Tech.* **11**, 2395–2426 (2018).
- Vigouroux, C. et al. TROPOMI-Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations. *Atmos. Meas. Tech.* **13**, 3751–3767 (2020).
- Cao, C., De Luccia, F., Xiong, X., Wolfe, R. & Weng, F. Early On-Orbit Performance of the Visible Infrared Imaging Radiometer Suite Onboard the Suomi National Polar-Orbiting Partnership (S-NPP) Satellite. *IEEE Trans. Geosci. Remote Sens.* **52**, 1142–1156 (2014).
- Elvidge, C., Baugh, K., Zhizhin, M., Hsu, F. & Ghosh, T. VIIRS nighttime lights. *Int. J. Remote Sens.* **38**, 5860–5879 (2017).
- Li, G. et al. Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nat. Commun.* **13**, 1628 (2022).
- Li, Y., Schubert, S., Kropp, J. & Rybski, D. On the influence of density and morphology on the Urban Heat Island intensity. *Nat. Commun.* **11**, 2647 (2020).
- Yang, C. et al. Human expansion into Asian highlands in the 21st Century and its effects. *Nat. Commun.* **13**, 4955 (2022).
- Ding, L., Zhao, W., Huang, Y., Cheng, S. & Liu, C. Research on the Coupling Coordination Relationship between Urbanization and the Air Environment: A Case Study of the Area of Wuhan. *Atmos.* **6**, 1539–1558 (2015).
- Sinha, A. & Bhattacharya, J. Environmental Kuznets curve estimation for NO<sub>2</sub> emission: A case of Indian cities. *Ecol. Indic.* **67**, 1–11 (2016).
- Li, G., Fang, C., Wang, S. & Sun, S. The Effect of Economic Growth, Urbanization, and Industrialization on Fine Particulate Matter (PM<sub>2.5</sub>) Concentrations in China. *Environ. Sci. Technol.* **50**, 11452–11459 (2016).
- Stern, D. The rise and fall of the environmental Kuznets curve. *World Dev.* **32**, 1419–1439 (2004).
- Huang, T. et al. Spatial and Temporal Trends in Global Emissions of Nitrogen Oxides from 1960 to 2014. *Environ. Sci. Technol.* **51**, 7992–8000 (2017).
- van Donkelaar, A., Martin, R., Brauer, M. & Boys, B. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* **123**, 135–143 (2015).
- Pu, D. C. et al. Response of Anthropogenic Volatile Organic Compound Emissions to Urbanization in Asia Probed With TROPOMI and VIIRS Satellite Observations. *Geophys. Res. Lett.* **49**, n/a–n/a (2022).
- Zhu, L. et al. Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns. *Environ. Res. Lett.* **9**, 114004 (2014).
- Wei, W., Wang, S., Hao, J. & Cheng, S. Projection of anthropogenic volatile organic compounds (VOCs) emissions in China for the period 2010–2020. *Atmos. Environ.* **45**, 6863–6871 (2011).

38. Fu, T. M. et al. Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone. *J. Geophys. Res. Atmos.* **112**, n/a–n/a (2007).
39. Barkley, M. et al. Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns. *J. Geophys. Res. Atmos.* **118**, 6849–6868 (2013).
40. Kaiser, J. et al. High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US. *Atmos. Chem. Phys.* **18**, 5483–5497 (2018).
41. Xing, C. et al. Identifying the wintertime sources of volatile organic compounds (VOCs) from MAX-DOAS measured formaldehyde and glyoxal in Chongqing, southwest China. *Sci. Total Environ.* **715**, 136258 (2020).
42. Palmer, P. et al. Mapping isoprene emissions over North America using formaldehyde column observations from space. *J. Geophys. Res. Atmos.* **108**, n/a–n/a (2003).
43. Hewitt, C. et al. Ground-level ozone influenced by circadian control of isoprene emissions. *Nat. Geosci.* **4**, 671–674 (2011).
44. Crippa, M. et al. Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth Syst. Sci. Data* **10**, 1987–2013 (2018).
45. Eastham, S. et al. GEOS-Chem High Performance (GCHP v11-02c): a next-generation implementation of the GEOS-Chem chemical transport model for massively parallel applications. *Geosci. Model Dev.* **11**, 2941–2953 (2018).
46. Zhu, L. et al. Long-term (2005–2014) trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing emissions of volatile organic compounds. *Geophys. Res. Lett.* **44**, 7079–7086 (2017).
47. Sun, W. et al. Global Significant Changes in Formaldehyde (HCHO) Columns Observed From Space at the Early Stage of the COVID–19 Pandemic. *Geophys. Res. Lett.* **48**, n/a–n/a (2021).
48. Ma, T., Zhou, C., Pei, T., Haynie, S. & Fan, J. Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sens. Environ.* **124**, 99–107 (2012).
49. Guo, W., Zhang, Y. & Gao, L. Using VIIRS-DNB and landsat data for impervious surface area mapping in an arid/semiarid region. *Remote Sens. Lett.* **9**, 587–596 (2018).
50. Small, C., Pozzi, F. & Elvidge, C. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sens. Environ.* **96**, 277–291 (2005).
51. Zhou, Y., Li, X., Asrar, G. & Smith, S. A global record of annual urban dynamics (1992–2013) from nighttime lights. *Remote Sens. Environ.* **219**, 206–220 (2018). Imhoff MJRSOE.
52. Chen, X. & Nordhaus, W. Using luminosity data as a proxy for economic statistics. *Proc. Natl Acad. Sci. USA* **108**, 8589–8594 (2011).
53. Li, X., Xu, H., Chen, X. & Li, C. Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China. *Remote Sens* **5**, 3057–3081 (2013).
54. Fendorf, S., Michael, H. & van Geen, A. Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science* **328**, 1123–1127 (2010).
55. Gorelick, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
56. Zhang, Z. et al. Assessment of Annual Composite Images Obtained by Google Earth Engine for Urban Areas Mapping Using Random Forest. *Remote Sens* **13**, 2695–2717 (2021).
57. Marais, E. et al. Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns. *Atmos. Chem. Phys.* **12**, 6219–6235 (2012).
58. Wang, T. & Sun, F. Global gridded GDP data set consistent with the shared socioeconomic pathways. *Sci. Data* **9**, 221 (2022).
59. Dinerstein, E. et al. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *Bioscience* **67**, 534–545 (2017).
60. Miller, C. et al. Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data. *Atmos. Chem. Phys.* **17**, 8725–8738 (2017).
61. Wolfe, G. et al. Formaldehyde production from isoprene oxidation across NO<sub>x</sub> regimes. *Atmos. Chem. Phys.* **16**, 2597–2610 (2016).
62. Griffin, D. et al. High resolution mapping of nitrogen dioxide with TROPOMI: First results and validation over the Canadian oil sands. *Geophys Res Lett.* **46**, 1049–1060 (2019).
63. Chan, K. et al. MAX-DOAS measurements of tropospheric NO<sub>2</sub> and HCHO in Munich and the comparison to OMI and TROPOMI satellite observations. *Atmos. Meas. Tech.* **13**, 4499–4520 (2020).
64. De Smedt, I. et al. Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements. *Atmos. Chem. Phys.* **21**, 12561–12593 (2021).
65. van Geffen, J. et al. S5P TROPOMI NO<sub>2</sub> slant column retrieval: method, stability, uncertainties and comparisons with OMI. *Atmos. Meas. Tech.* **13**, 1315–1335 (2020).
66. Duncan, B., Yoshida, Y., Damon, M., Douglass, A. & Witte, J. Temperature dependence of factors controlling isoprene emissions. *Geophys. Res. Lett.* **36**, n/a–n/a (2009).
67. Zhu, L. et al. Formaldehyde (HCHO) As a Hazardous Air Pollutant: Mapping Surface Air Concentrations from Satellite and Inferring Cancer Risks in the United States. *Environ. Sci. Technol.* **51**, 5650–5657 (2017).
68. van der Werf, G. et al. Global fire emissions estimates during 1997–2016. *Earth Syst. Sci. Data* **9**, 697–720 (2017).
69. Wang, Z. et al. Quantifying uncertainties in nighttime light retrievals from Suomi-NPP and NOAA–20 VIIRS Day/Night Band data. *Remote Sens. Environ.* **263**, n/a–n/a (2021).
70. Gelaro, R. et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
71. Guenther, A. B. et al. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci. Model Dev.* **5**, 1471–1492 (2012).

## Acknowledgements

This work is funded by the National Natural Science Foundation of China (42375090), the Shenzhen Key Laboratory of Precision Measurement and Early Warning Technology for Urban Environmental Health Risks (ZDSYS20220606100604008), Guangdong Basic and Applied Basic Research Foundation (2021A1515110713), Guangdong University Research Project Science Team (2021KCXTD004), Major Talent Project of Guangdong Province (2021QN020924), Shenzhen Science and Technology Program (KQTD20210811090048025, JCYJ20210324104604012, JCYJ20220530115404009), and High level of special funds (G030290001). This work is supported by the Center for Computational Science and Engineering at Southern University of Science and Technology.

## Author contributions

D.P. and L.Z. designed the work; I.D., J.Y., H.S., L.S., D.W., X.L., X.Z., X.Y., and T.F. performed the work; L.Z., J.L., J.Y., and H.S., contributed analytic tools; D.P., J.L., L.S., D.W., X.L., X.Z., X.Y., and T.F. analyzed data; D.P., L.Z., and H.S. wrote the manuscript.

## Competing interests

The authors declare no competing interests.



## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41612-024-00683-5>.

**Correspondence** and requests for materials should be addressed to Lei Zhu.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024