nature geoscience

LETTERS

PUBLISHED ONLINE: 30 MAY 2016 | DOI: 10.1038/NGEO2724

Space-based detection of missing sulfur dioxide sources of global air pollution

Chris A. McLinden[1]*, Vitali Fioletov[1], Mark W. Shephard[1], Nick Krotkov[2], Can Li[2,3], Randall V. Martin[4,5], Michael D. Moran[1] and Joanna Joiner[2]

Sulfur dioxide is designated a criteria air contaminant (or equivalent) by virtually all developed nations. When released into the atmosphere, sulfur dioxide forms sulfuric acid and fine particulate matter, secondary pollutants[1] that have significant adverse effects on human health[2–5], the environment[1] and the economy[5]. The conventional, bottom-up emissions inventories used to assess impacts, however, are often incomplete or outdated, particularly for developing nations that lack comprehensive emission reporting requirements and infrastructure. Here we present a satellite-based, global emission inventory for SO_2 that is derived through a simultaneous detection, mapping and emission-quantifying procedure, and thereby independent of conventional information sources. We find that of the 500 or so large sources in our inventory, nearly 40 are not captured in leading conventional inventories. These missing sources are scattered throughout the developing world—over a third are clustered around the Persian Gulf—and add up to 7 to 14 Tg of SO_2 yr^{-1}, or roughly 6–12% of the global anthropogenic source. Our estimates of national total emissions are generally in line with conventional numbers, but for some regions, and for SO_2 emissions from volcanoes, discrepancies can be as large as a factor of three or more. We anticipate that our inventory will help eliminate gaps in bottom-up inventories, independent of geopolitical borders and source types.

Oxidation of sulfur dioxide (SO_2) leads to the formation of sulfuric acid and fine particulate matter[1] which, between them, are associated with negative health outcomes such as cardiovascular disease and cancer[2] and ecological impacts on soil, forests and freshwater[1]. Understanding current levels of SO_2 and its products and predicting their future state require computer models that accurately represent the chemical and physical processes in the atmosphere. However, underpinning these models and arguably the single most important factor governing their predictive capability is their description of emissions. Virtually all models rely on 'bottom-up' inventories that are based in part on indirect information such as activity data and emission factors. Despite this critical link, these bottom-up inventories are often incomplete, suffer from regional inconsistencies, and are typically outdated by the time they are compiled[6].

Satellites provide an ideal platform for closing this information gap, particularly for remote or otherwise inaccessible locations. There are a variety of methods used to derive SO_2 emission information from satellite measurements[6–11], but most rely on prior knowledge of the location of the source. As SO_2 has a short atmospheric lifetime, larger isolated sources can sometimes be identified as 'hotspots' on maps of the average distribution. However, winds can obscure the true source location and weaker sources are often masked by stronger ones. Furthermore, troublesome measurement biases can appear as false sources, and near the detection limit it becomes difficult to distinguish between a weaker source and measurement noise.

Here we present a new method for identifying and mapping SO_2 emission sources using satellite measurements of atmospheric composition that is independent of air quality models and requires no prior assumptions as to their location (Methods). We analyse measurement of SO_2 (ref. 12) from the Ozone Monitoring Instrument[13] (OMI), which provides a vertically integrated number density profile, or column. Each point on a high-resolution ($0.05° \times 0.05°$) global grid is evaluated as a potential (or 'test') source location—it is this evaluation scheme that is at the heart of our method. Satellite-measured columns, merged with coincident wind information from a meteorological reanalysis[14], are repositioned through a rotation about this test location such that, after rotation, all of the wind directions are aligned[15] (Methods and Supplementary Fig. 1). If the difference between the average downwind and average upwind column (Supplementary Fig. 2) exceeds a specified margin (Methods), this location is flagged as a source. Iterating over the entire grid generates a 'source map' in which emission sources appear as very localized hotspots. Once the locations of the sources are known their emissions are calculated[10].

Our approach is evaluated using OMI SO_2 columns over the eastern US, where the emission locations and intensities are well known. Figure 1 shows the map of average SO_2 columns and the corresponding source map. Virtually all sources greater than 40 kt yr^{-1} appear on the source map and to within about 10 km of their true location (Methods and Supplementary Figs 5 and 6). Figure 1 also shows the application of our algorithm over southeastern Europe. Here we find examples of sources not found in the widely used Hemispheric Transport of Air Pollutants[16] (HTAP) emissions inventory. There are many instances where the downwind–upwind difference is able to detect sources, locate them, and better discriminate between sources/non-sources better as compared with the average column (Supplementary Information). One example is that of Mount Stromboli, a volcanic source of SO_2 just north of Sicily, which we successfully identify from underneath the extremely large SO_2 signal from Mount Etna (Fig. 2b,c).

We applied our algorithm globally to identify SO_2 sources not included in leading bottom-up emission inventories. We use the HTAP inventory[16] for the year 2008 as our primary reference but

[1]Air Quality Research Division, Environment and Climate Change Canada, Toronto M3H 5T4, Canada. [2]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742, USA. [4]Dalhousie University, Department of Physics and Atmospheric Science, Halifax, Nova Scotia B3H 4R2, Canada. [5]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. *e-mail: chris.mclinden@canada.ca

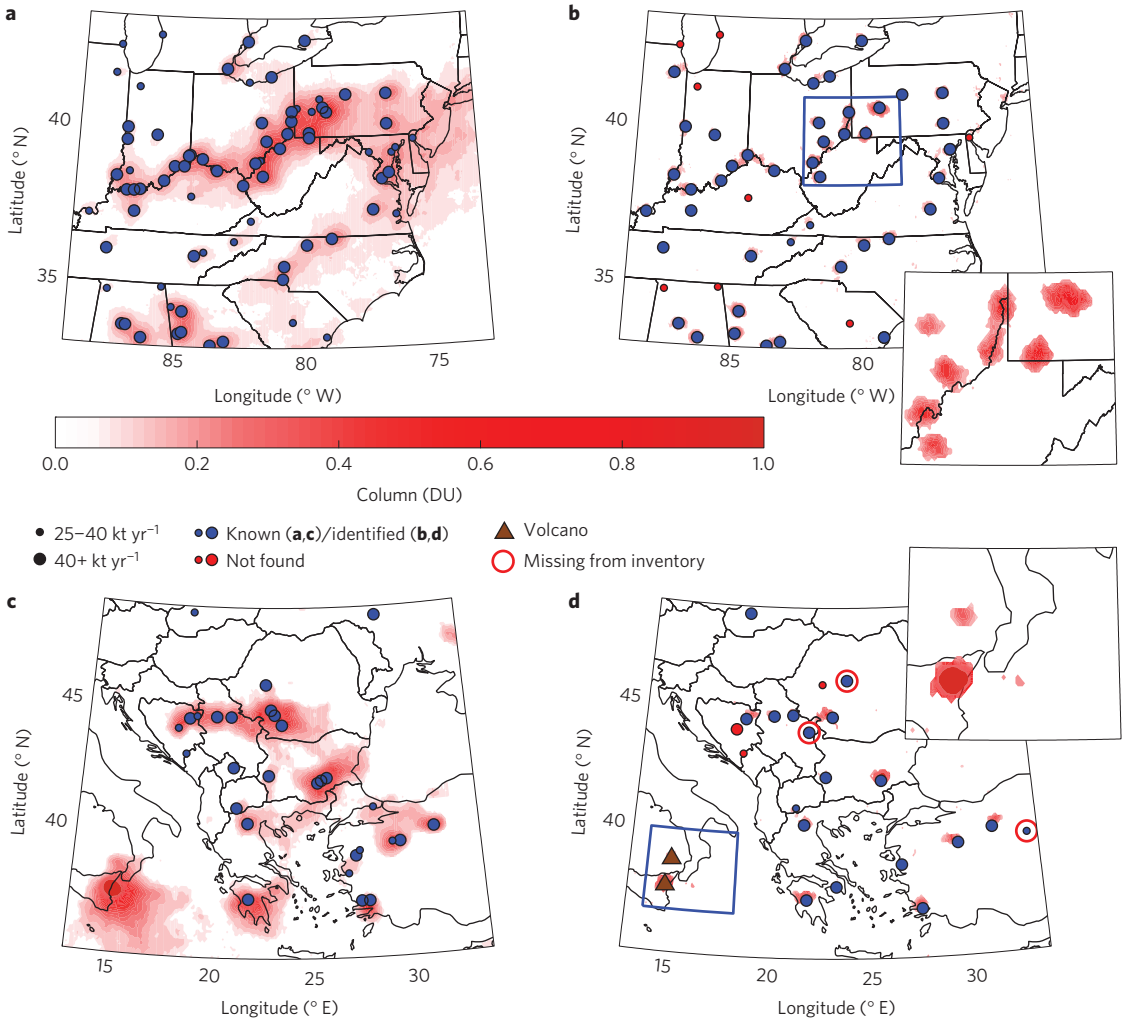

Figure 1 | Demonstration of source-detection method over selected regions. a, Average 2007–2009 SO_2 columns from the OMI satellite instrument over eastern USA expressed in Dobson units (Methods). Also shown are the locations of large power plants. **b**, Eastern US source map (or the downwind–upwind difference), also in Dobson units, and the locations of sources from **a** identified by OMI and those not found. **c**, Average SO_2 column over eastern Europe and sources from the HTAP[16] inventory. **d**, Eastern European source map, emission sites identified and not found, and identified sources that are missing (Methods) from HTAP.

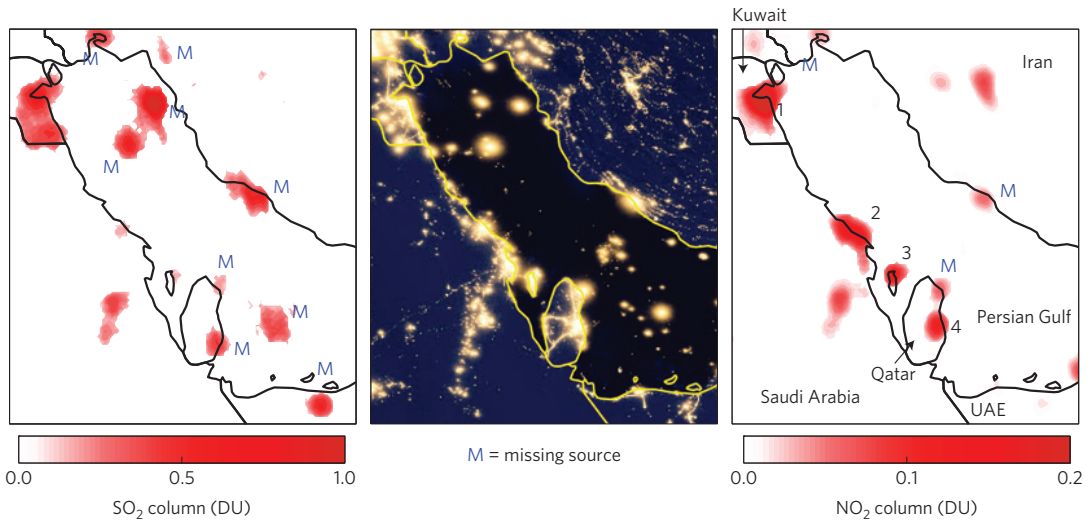

Figure 2 | Satellite detection of emission sources in the Persian Gulf. Shown are the SO_2 and NO_2 (or NO_x) emission source maps from 2007–2009 OMI data along with satellite imagery that shows night lights from cities and gas flaring sites[19]. M indicates a source that is missing from the three reference emission inventories. The numbers represent cities: 1, Kuwait City; 2, Al Jubail; 3, Manama; 4, Mesaieed. UAE, United Arab Emirates.

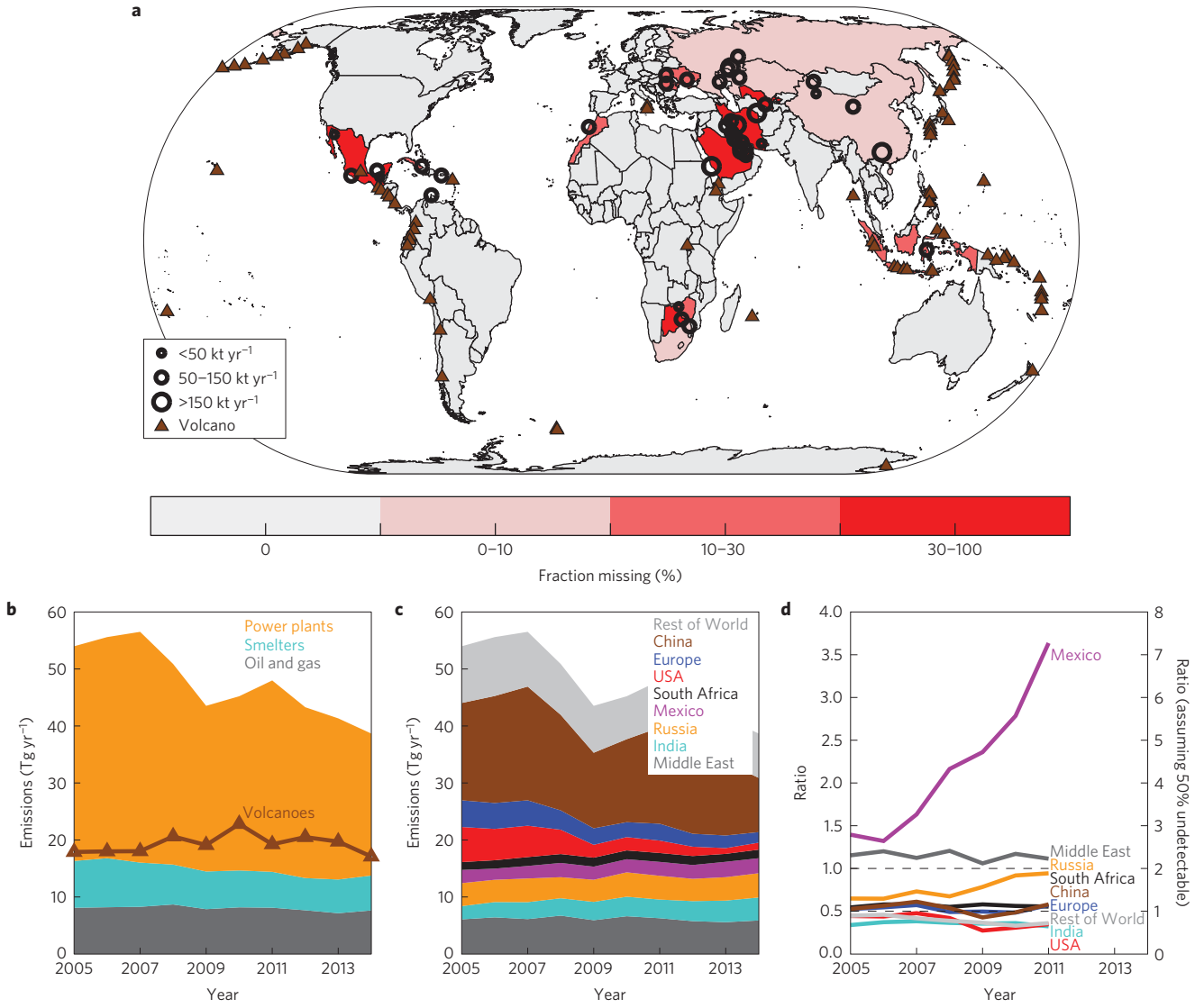

Figure 3 | Summary of satellite-derived missing sources and inventory totals. a, SO_2 sources identified that were found to be missing from three leading bottom-up inventories[16–18]. Each nation is colour-coded according to its total fraction of SO_2 emissions that are missing, relative to its total national emissions (the sum of HTAP and missing). **b**, Time series of the total satellite-derived emissions by source type. **c**, Time series of satellite-derived anthropogenic emissions for selected regions. **d**, Time series of the ratio of satellite-measured anthropogenic emissions to those from an annual SO_2 inventory[25].

we also consider two other state-of-the-science global, gridded, bottom-up inventories[17,18]. Our search yielded 481 locations that we were able to authenticate as sources through a two-step process (Methods): a statistically significant emissions retrieval; and an identification of the emitting facility using a combination of satellite imagery, external databases, and/or other resources[19,20]. From this process, we determined that 75 are volcanic SO_2 sources that passively outgas SO_2 (eruptive episodes were excluded from our analysis; see Methods). Of the remaining 406 sources, 340 had significant emissions in the year 2008 and 301 of these correspond to sources in at least one of the three emission inventories. This leaves 39 sources that do not correspond to any of the inventories we considered, which we refer to as 'missing' (that is, a real source that is unknown to the emissions community, known but unquantified or significantly under-reported, or incorrectly located). We also find several weak, false positives over locations at very high elevation or that are ice-covered. In addition there are 20 weak sources that could not be authenticated; however, considering that only one false positive was found in the eastern US, we suggest some may be real.

Of the 39 verified missing sources, 14 are located in the Middle East, and 12 of these are oil and gas-related. The SO_2 source map in Fig. 2 shows hotspots corresponding to several large oil refinery complexes in and around the Persian Gulf, most of which are not accounted for in any of the emission inventories considered. The analogous OMI nitrogen dioxide[21] (NO_2, a proxy for nitrogen oxides, or NO_x) source map (Fig. 2c) shows that these locations are modest NO_x sources that we have determined are also missing from NO_x emission inventories. Night-light satellite imagery[19] (Fig. 2b) reveals that most are gas flaring locations. In contrast, the largest sources of NO_x correspond to urban areas where lights are brightest and SO_2 emissions are low. Overall, the three types of data (SO_2, NO_2, lights) portray a consistent picture. These NO_x sources are also expected to be sources of CO_2, an important greenhouse gas, as both are by-products of combustion[22], yet we find some of these missing NO_x sources are also missing from CO_2 inventories[18,23].

Globally in 2008, emissions from the 39 missing SO_2 sources total 7 $Tg[SO_2]\,yr^{-1}$, or roughly 6% of the anthropogenic total[16–18]. Figure 3a shows the location and size of each missing source

as well as their national totals expressed as a fraction of their (HTAP + missing) national total. Total OMI-derived anthropogenic emissions (Fig. 3b) show a decline from a peak in 2006–2007 that is driven primarily by a reduction in power plant emissions. Regionally (Fig. 3c), US, European and Chinese emissions show declines, with emissions from India increasing. The 2011–2014 declines in Chinese emissions are more consistent with projections based on fully implemented emission controls[24]. Figure 3 also suggests that if emissions from China, India, Russia and the Middle East continue along their current trajectories, they will all be of comparable magnitude by the end of the decade.

OMI-derived regional emission totals, shown as fractions of an annually varying bottom-up inventory[25] in Fig. 3d, are typically near 0.5 for most regions and fairly constant with time. This indicates that despite capturing only about half of the anthropogenic total, a consequence of its 40 kt yr^{-1} detection limit, our method can reliably track changes in regional emissions. The bottom-up inventories similarly indicate that roughly half of their totals are from sources that exceed 40 kt yr^{-1} globally (and 30–70% regionally; Supplementary Fig. 7), providing additional quantitative verification to our OMI-derived regional totals. Extending this reasoning further, we estimate the upper limit to the total (including an unmeasured 50%) from missing sources to be 12% of global anthropogenic emissions. Significant departures from this in which OMI is as large or larger are seen for Russia, the Middle East and Mexico, which points to discrepancies between OMI-derived emissions and the bottom-up inventories that should be investigated. Mexico is particularly striking in that we find emissions are increasing slightly as compared with a 60% decline in reported emissions between 2005 and 2011[25].

The total emissions from the 75 continuously emitting volcanoes detected by OMI vary between 20 and 25 Tg[SO_2] yr^{-1} (or up to half the OMI anthropogenic total), with many volcanoes exhibiting significant inter-annual variability. As roughly 90% of active volcanoes have rudimentary or no monitoring[26], this natural SO_2 source represents a significant gap in our knowledge of its global source. We find good agreement between OMI-derived emission estimates and those measured from Japanese volcanoes (Methods and Supplementary Fig. 9), one of the few areas actively monitored. However, a comparison with a global volcanic emissions database[27] often used in air quality models[28] shows a weak correlation, with emissions from 31 volcanoes differing by a factor of 3 of more (Supplementary Fig. 10). Using analogous criteria as for anthropogenic sources, 17 OMI-identified volcanoes qualify as missing sources, with emissions totalling 8 Tg[SO_2] yr^{-1} (Methods).

Emissions from the anthropogenic and volcanic missing sources, taken together, total between 15 and 30 Tg[SO_2] yr^{-1} (again assuming only 50% can be detected), representing roughly 10–20% of the total, global SO_2 input into the atmosphere. Their omission from model-based mortality and environmental impact studies can give rise to biased findings at the local, regional and global[29] scales, potentially leading to misinformed policy and mitigation decisions. Gaps in pollutant inventories are likely to extend beyond SO_2 as our analysis of NO_x in the Persian Gulf suggests. The utility of satellite data for further closing these gaps now seems clear, be it through the application of algorithms such as this one to other pollutants, or improved detection limits from new satellite instruments, including the upcoming geostationary air quality constellation[30].

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

Received 24 January 2016; accepted 27 April 2016; published online 30 May 2016

References

1. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics* 2nd edn (Wiley, 2006).
2. Lelieveld, J. *et al*. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525,** 367–371 (2015).
3. Lim, S. S. *et al*. 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. *Lancet* **380,** 2224–2260 (2012); correction **381,** 628 (2013).
4. Prüss-Ustün, A. *et al*. *Preventing Disease through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks* (World Health Organization, 2016); http://www.who.int/quantifying_ehimpacts/publications/preventing-disease/en
5. *The Cost of Air Pollution: Health Impacts of Road Transport* (OECD, 2014).
6. Lamsal, L. N. *et al*. Application of satellite observations for timely updates to global anthropogenic NO_x emission inventories. *Geophys. Res. Lett.* **38,** L05810 (2011).
7. Martin, R. V. Satellite remote sensing of surface air quality. *Atmos. Environ.* **42,** 7823–7843 (2008).
8. Streets, D. G. *et al*. Emissions estimation from satellite retrievals: a review of current capability. *Atmos. Environ.* **77,** 1011–1042 (2013).
9. Beirle, S. *et al*. Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science* **333,** 1737–1739 (2011).
10. Fioletov, V. E., McLinden, C. A., Krotkov, N. & Li, C. Lifetimes and emissions of SO_2 from point sources estimated from OMI. *Geophys. Res. Lett.* **42,** 1969–1976 (2015).
11. de Foy, B. *et al*. Estimates of power plant NO_x emissions and lifetimes from OMI NO_2 satellite retrievals. *Atmos. Environ.* **116,** 1–11 (2015).
12. Li, C., Joiner, J., Krotkov, N. A. & Bhartia, P. K. A fast and sensitive new satellite SO_2 retrieval algorithm based on principal component analysis: application to the ozone monitoring instrument. *Geophys. Res. Lett.* **40,** 6314–6318 (2013).
13. Levelt, P. F. *et al*. The ozone monitoring instrument. *IEEE Trans. Geosci. Remote* **44,** 1093–1101 (2006).
14. Dee, D. P. *et al*. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137,** 553–597 (2011).
15. Pommier, M., McLinden, C. A. & Deeter, M. Relative changes in CO emissions over megacities based on observations from space. *Geophys. Res. Lett.* **40,** 3766–3771 (2013).
16. Janssens-Maenhout, G. *et al*. HTAP_v2: a mosaic of regional and global emission gridmaps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmos. Chem. Phys.* **15,** 11411–11432 (2015).
17. Granier, C. *et al*. Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic Change* **109,** 163–190 (2011).
18. Janssens-Maenhout, G., Pagliari, V., Guizzardi, D. & Muntean, M. *Global Emission Inventories in the Emission Database for Global Atmospheric Research (EDGAR) Manual (I): Gridding: EDGAR Emissions Distribution on Global Gridmaps* JRC Report EUR 25785 EN (2013); http://dx.doi.org/10.2788/81454
19. *Night Lights 2012—The Black Marble* (NASA Earth Observatory, 2012); http://earthobservatory.nasa.gov/IOTD/view.php?id=79803
20. Davis, C. B. *et al*. *Enipedia & Energy Industry Group, Faculty of Technology, Policy and Management* (TU Delft, 2015); http://enipedia.tudelft.nl
21. Bucsela, E. J. *et al*. A new stratospheric and tropospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI. *Atmos. Meas. Tech.* **6,** 2607–2626 (2013).
22. Reuter, M. *et al*. Decreasing emissions of NO_x relative to CO_2 in East Asia inferred from satellite observations. *Nature Geosci.* **7,** 792–795 (2013).
23. Oda, T. & Maksyutov, S. A very high-resolution (1 km × 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos. Chem. Phys.* **11,** 543–556 (2011).
24. Zhao, Y., Zhang, J. & Nielse, C. P. The effects of energy paths and emission controls and standards on future trends in China's emissions of primary air pollutants. *Atmos. Chem. Phys.* **14,** 8849–8868 (2014).
25. Klimont, Z., Smith, S. J. & Cofala, J. The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. *Environ. Res. Lett.* **8,** 014003 (2013).
26. Sparks, R. S. J., Biggs, J. & Neuberg, J. W. Monitoring volcanoes. *Science* **335,** 1310–1311 (2012).
27. Diehl, T. *et al*. Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO_2 from 1980 to 2010 for hindcast model experiments. *Atmos. Chem. Phys. Discuss.* **12,** 24895–24954 (2012).
28. Keller, C. A. *et al*. HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models. *Geosci. Model Dev.* **7,** 1409–1417 (2014).
29. *Global Sources of Local Pollution: An Assessment of Long-Range Transport of Key Air Pollutants to and from the United States* (National Academies, 2010).
30. Chance, K. *et al*. in *Proc. SPIE 8866, Earth Observing Systems XVIII* (eds Butler, J. J., Xiong, X. & Gu, X.) (SPIE, 2013).

Acknowledgements

R.V.M. was supported by Environment and Climate Change Canada. N.K., C.L. and J.J. acknowledge NASA funding through the Aura science team programme for OMI SO_2 product development and analysis. The Dutch–Finnish-built OMI instrument is part of the NASA's EOS Aura satellite payload. The OMI project is managed by KNMI and the Netherlands Space Office (NSO). The authors acknowledge ECMWF for the provision of their ERA-interim reanalysis data. C.A.M. thanks H. Morrison for commenting on earlier versions of the manuscript.

Author contributions

C.A.M., V.F. and M.W.S. planned the research and developed the algorithm, C.A.M. and V.F. performed the computations and analysis, N.K., C.L. and J.J. provided the satellite data, C.A.M. wrote the paper, and all authors discussed the results and commented on the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.A.M.

Competing financial interests

The authors declare no competing financial interests.

Methods

Satellite observations and meteorological reanalysis. Observations of SO_2 and NO_2 from the Ozone Monitoring Instrument[13,31] (OMI, 2004–present), on-board the NASA Aura satellite, are used in this study. These gases are quantified in terms of a vertical column density (expressed in Dobson units, where 1 DU $= 2.69 \times 10^{16}$ molecules cm^{-2}), or simply column, and represent the vertically integrated number density between the surface and the tropopause. Specifically, orbit-based (Level 2) SO_2 columns from the Principal Component Analysis data product[12] and NO_2 columns from the NASA Standard Product[21] version 2.1 are used. Data are filtered using a maximum radiative cloud fraction of 0.2 and maximum solar zenith angles of 70°/75° for SO_2/NO_2. Data from OMI track positions affected by the row anomaly[32] were excluded, and of the 60 across-track positions measured by OMI, only positions 11–50 were used, which correspond to those with a finer spatial resolution.

Filtering of SO_2 from explosive volcanic eruptions is performed using a simple screening algorithm[31]. Daily 99th percentiles of the SO_2 columns within 300 km of a source are calculated. All data from a day in which this exceeds a threshold are excluded. We used a threshold of 15 DU for sources larger than 500 kt yr^{-1}, 10 DU for sources between 100 and 500 kt yr^{-1}, and 5 DU for smaller sources. Most days removed coincide with a period immediately following known eruptions.

The sensitivity of OMI to SO_2 in a particular scene is quantified through an air mass factor (AMF)[31]. AMFs are calculated using radiative transfer models that simulate the absorption and multiple-scattering in the atmosphere and reflection off the surface. To improve the accuracy of the SO_2 emissions, we have recalculated AMFs for each OMI observation largely following the approach of ref. 33, accounting for parameters such as surface reflectivity, solar zenith angle, viewing geometry, surface pressure, and cloud fraction and pressure. The SO_2 profile shape is estimated on the basis of the elevation of the source and the climatological boundary layer height (specific to the month, source location, and time of day)[34]. Between these two heights the profile is assumed to have a constant mixing ratio and above and below these heights it is assumed to be zero. We also included an aerosol layer between the surface and the top of the boundary layer, scaled so that its vertical optical depth matches that of a satellite-derived annual mean climatology at 0.5° × 0.5° resolution[35]. It is worth noting that one common source of profile information used in the calculation of AMFs, output from air quality models, cannot readily be used to determine emissions from missing sources as the simulated profiles will resemble background conditions owing to the lack of nearby sources.

Wind information is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-interim)[14]. This is a four-dimensional variational analysis system with a spatial resolution of about 80 km with output available every 6 h. We interpolate the vertical profiles of the zonal and meridional wind components in time and space to the location of the each OMI pixel. We then average each wind component between the elevation of the source and 1 km above. These mean wind components are then converted to a wind speed and direction.

Bottom-up SO_2 emission inventories. Three gridded emission inventories are considered. The primary reference is the HTAP (Hemispheric Transport of Air Pollution) - EDGAR (Emission Database for Global Atmospheric Research) version 2 high-resolution gridded emission inventory[16] for the determination of missing sources. HTAP is available for 2008 and 2010 on a 0.1° × 0.1° latitude–longitude grid. We also use the EDGAR v4.2, also available on a 0.1° × 0.1° grid, for 2008[18], and the MACCity (Monitoring Atmospheric Composition and Climate—CityZen) inventory, on a 0.5° × 0.5° grid, for 2008[17]. In addition, for the SO_2 emission time series analysis in Fig. 3d we used an annually (2005–2011) varying inventory at the national or regional level[25]. All give generally consistent national and global totals for the year 2008.

Emission mapping methodology. A global, 0.05° × 0.05° grid is defined where each grid point is a potential source location, referred to as a test source. All filtered satellite column observations over a specified time frame—we considered multiple three-year blocks—that are within a distance of 100 km of the test source are considered. These are rotated about the location of the test source such that, after rotation, the wind vectors of all observations are aligned[10,11,16]. This process ensures that the relative upwind–downwind/cross-wind distance from the observation to the test source is preserved (Supplementary Fig. 1).

After rotation the upwind and downwind averages are calculated. The upwind average column (Ω_u) is computed over all columns that are within a rectangle defined by $\pm a$ (km) about the test source in the cross-wind direction and between b_1 and b_2 (km) upwind of the test source (Supplementary Fig. 2). The downwind average column (Ω_d) is likewise computed for a rectangle $\pm a$ and between b_1 and b_2 downwind of the test source. The choice of a, b_1 and b_2 are based on multiple factors including the effective 'size' of a typical source as seen by OMI, the lifetime of the gas (several hours for NO_2 and SO_2), and a trade-off between signal and resolution. For example, b_1 needs to be of the order of the spatial extent of the source so that there will be a real difference in the upwind and downwind direction whereas a and b_2 need to be large enough to allow for a sufficiently large number of data points but small enough that a reasonable resolution is maintained. An ideal choice for OMI SO_2 and NO_2 was found to be $a = 15$ km, $b_1 = 15$ km and $b_2 = 35$ km.

Also calculated are the standard deviations of the upwind and downwind averages and from these the signal-to-noise ratio (SNR) is calculated using equation (1):

$$\mathrm{SNR} = \frac{\Omega_d - \Omega_u}{\frac{\sigma_{\Omega u}}{\sqrt{N_u}} + \frac{\sigma_{\Omega d}}{\sqrt{N_d}}} \quad (1)$$

where $\sigma_{\Omega d}$ and $\sigma_{\Omega u}$ are the downwind and upwind standard deviations, and N_d and N_u are the number of observations. Note that alternative SNR definitions were tested but the above definition was found to produce the most consistent results. If the SNR exceeded a threshold value then the test source was taken as a real source (subject to confirmation). For the analysis performed over the eastern US where source locations are precisely known, a SNR value of 2 led to identification of sources larger than about 30–40 kt yr^{-1}, but with little or no 'false positives'. Further discussion on the method and its sensitivities is provided in the Supplementary Information.

Determination of emissions and their uncertainties. Once the location of the source has been determined, the algorithm of ref. 10 is used to quantify its emissions. This method is consistent in that it uses the same basic approach—a rotation about the reference location and then an examination of the downwind behaviour. In this algorithm, a specialized three-dimensional (downwind–upwind distance, cross-wind distance, and wind speed) function is fitted to the rotated column data. The function is based on three parameters: an effective lifetime, a width parameter, and a parameter that represents the total mass of the SO_2, with the emissions calculated assuming a mass balance; that is, as the ratio of mass to lifetime. To stabilize the fit against noise, the lifetime and width parameters are specified using the mean values derived from a subset of all emission sources identified. Here, lifetime is set to 6 h and the width parameter is set to 20 km. These may vary from location to location, but there is a cancellation of errors when an inaccurate lifetime or width parameter is used due to the nature of the fitting procedure. For example, when lifetime is in error by a factor of 2, the impact on the emissions is less than 30%. The 10th and 90th percentiles of the distribution of lifetime are 3 and 9.5 h, respectively (Supplementary Fig. 4).

There are several potential sources of error that need to be considered when determining the overall uncertainty of the emissions (Supplementary Table 1). Uncertainties in the SO_2 columns themselves are estimated at 29%, uncertainties in the emissions fitting are 37% + 10–20 kt yr^{-1}, and the uncertainties from the wind fields are 22%. Combining these gives an uncertainty of 50% for larger sources (>100 kt yr^{-1}) and about 60% for smaller ones (<50 kt yr^{-1}). Additional details are provided in the Supplementary Information.

Methodology for authenticating a source or identifying a missing source. For a peak on the source map to be verified as a real source it must meet the following two conditions:

Possess a statistically significant emission rate[10]. We require that an annual emission rate exceed its 3-sigma uncertainty, where sigma is the uncertainty from the fit. This need only be the case for one of the 10 years of OMI measurements.

A facility (or volcano) must be identified within a radius of 20 km of the test source location of a size and type (for example, coal-burning power plant) that could be emitting appreciable SO_2. This is determined using a variety of external data sources. Some of the more useful ones are listings of coal-burning power plants[21], the black marble, or night-lights, satellite imagery[20], which shows locations where gas flaring is occurring, Google Earth terrain images and associated (linked) photos, as well as general Internet searches. In some cases more than one candidate facility was identified and here we made a judgement call.

For a source to be identified as a so-called missing source, it must have statistically significant emissions (see above) for 2008 and be absent from each of the three 2008 gridded, bottom-up emission inventories considered[16–18]. We determined this by integrating each one over a 50 km radius from the location of the source under consideration. For each inventory we determine whether their integrated emissions are less than 8 kt yr^{-1} (our detection limit is about 40 kt yr^{-1}) or a factor of five smaller than our emissions estimate. If either of these conditions is true for all three inventories, we flag this location as a missing source. The value of 50 km is based on an analysis of several case studies that suggest that for the OMI pixel size sources located within about 50 km of one another are interpreted as one with total emissions close to the sum of their emissions. In addition, the use of a 50 km radius leads to the best correlation in a comparison with US sources.

These criteria lead to the identification of 39 missing SO_2 sources whose emissions total 6.8 $Tg[SO_2]$ yr^{-1}. As a sensitivity test we integrate the bottom-up inventories over a radius of 80 km (instead of 50 km) and require them to be at least a factor of 8 (instead of 5) smaller than OMI. There are still 21 sources that meet

these more stringent criteria, totalling 4.5 Tg[SO_2] yr^{-1}. Again, considering an undetected 50%, this amounts to 4–8% of the global anthropogenic total.

Inventory evaluation. Several comparisons between our emission inventory and the bottom-up reference inventories[16–18] were performed. This included a direct comparison of emission rates between our inventory and HTAP in the eastern US, a quantification of how well our algorithm can locate the emission sources, again using eastern US sources, and a comparison of regional totals. Additional details are available in the Supplementary Methods, but to summarize: we find biases less than 15% for individual sources and a correlation of 0.86 (Supplementary Fig. 5), differences to within 5% for regional totals considering the sources contained in both inventories, and an ability to locate an individual to 8 ± 4 km (Supplementary Fig. 6).

Another important finding was that, on average, about half of the global SO_2 source is captured with our method, with the undetected half coming from sources below the detection limit estimated at 40 kt yr^{-1} (Supplementary Fig. 7). This is consistent with the distribution of sources globally in the inventories for a comparable detection limit. We also computed this for the different regions considered in Fig. 3 where up to 70% is measurable for a 40 kt yr^{-1} detection limit. On this basis it is reasonable to conclude that a discrepancy exists between OMI and bottom-up emissions for a region where OMI exceeds 70% of the bottom-up.

Volcanic sources. Emissions of SO_2 through passive outgassing from 75 volcanoes have been detected and quantified. The locations of these volcanoes are shown in Fig. 3. We find that the emissions from some volcanoes show large inter-annual variability whereas others are nearly constant over the 10-year OMI time period (Supplementary Information). Comparisons between our SO_2 emissions estimates with those derived using *in situ* monitors by the Japan Meteorological Agency for two Japanese volcanoes shows consistency in both magnitude and seasonal variability (Supplementary Fig. 9).

On a global scale, estimates of the annual volcanic SO_2 input into the atmosphere vary widely[36]. One volcanic SO_2 emissions database, Aerocom[27], is recommended by the HTAP-EDGAR v2 authorities. Further, several leading models also use the volcanic Aerocom emissions[28,37]; the GEOS-Chem model, http://wiki.seas.harvard.edu/geos-chem/index.php/Volcanic_SO2_emissions_from_Aerocom).

The Aerocom database covers the years 1979–2010, providing a six-year overlap period with OMI (2005–2010). We compared our emissions to Aerocom by matching volcanoes considering only the passive outgassing component, and then integrating Aerocom over all other volcanic sources in a 50 km radius. There is little correlation between the two inventories (Supplementary Fig. 10) with many of the Aerocom emissions based on fill values that are constant with time. Changing the integrating radius had no impact on the correlation. If we use the same missing source criteria here; that is, requiring OMI emissions to be a factor of five larger than those from the Aerocom data, 17 volcanoes are identified whose emissions total 8.8 Tg[SO_2] yr^{-1}. OMI emissions from Ambrym, to the east of Australia, are 2,500 kt yr^{-1}, and a factor of 42 larger than Aerocom for the overlap period. Meanwhile, OMI emissions from Nevado del Ruiz in Columbia are a factor of 25 smaller than Aerocom for the overlap period, but increase rapidly after 2010.

Data availability. All data used in this work are publicly available. Level 2 Principal Component Analysis SO_2 and Standard Product version 2.1 NO_2 vertical column density data from OMI, used to identify emission source locations and derive the SO_2 emissions inventory, are available from the Goddard Earth Science and Information Service Center (http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI)[31]. Wind reanalysis data from ECMWF ERA-interim reanalysis[14] was downloaded from http://apps.ecmwf.int/datasets/data/interim-full-daily.

HTAP-EDGAR v2 emissions data[16] were downloaded from http://edgar.jrc.ec.europa.eu/htap_v2/index.php?SECURE=123, EDGAR v4.2 emissions data[18] were downloaded from http://edgar.jrc.ec.europa.eu/overview.php?v=42, MACCity emissions data[17] were downloaded from http://accent.aero.jussieu.fr/MACC_metadata.php, and ODIAC emissions data[23] were downloaded from http://db.cger.nies.go.jp/dataset/ODIAC/emission_dataset.html. Night-lights (black marble) satellite imagery[19] was obtained from NASA's Earth Observatory (http://earthobservatory.nasa.gov/IOTD/view.php?id=79803). The Aerocom volcanic SO_2 emissions database[27] was downloaded from http://aerocom.met.no/download/emissions/HTAP.

Emission summary files and kml (Google Earth) overlays created as part of this work are available as Supplementary Information.

References

31. Krotkov, N. A. *et al*. Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015. *Atmos. Chem. Phys.* **16,** 4605–4629 (2016).
32. *Background Information about the Row Anomaly in OMI* (OMI, 2012); http://projects.knmi.nl/omi/research/product/rowanomaly-background.php
33. McLinden, C. A. *et al*. Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements. *Atmos. Chem. Phys.* **14,** 3637–3656 (2014).
34. von Engeln, A. & Teixeira, J. A planetary boundary layer height climatology derived from ECMWF Re-analysis data. *J. Clim.* **26,** 6575–6590 (2013).
35. Hsu, N. C. *et al*. Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. *Atmos. Chem. Phys.* **12,** 8037–8053 (2012).
36. Halmer, M. M. *et al*. The annual volcanic gas input into the atmosphere, in particular into the stratosphere: a global data set for the past 100 years. *J. Volcanol. Geotherm. Res.* **115,** 511–528 (2002).
37. Stuefer, M. *et al*. Inclusion of ash and SO_2 emissions from volcanic eruptions in WRF-Chem: development and some applications. *Geosci. Model Dev.* **6,** 457–468 (2013).