nature geoscience
SUPPLEMENTARY INFORMATION
DOI: 10.1038/NGEO2724

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

Chris McLinden, Vitali Fioletov, Mark W. Shephard, Nick Krotkov, Can Li,
Randall V. Martin, Michael D. Moran, and Joanna Joiner

Supplementary Methods – Additional material on methods, sensitivities and error budgets, comparisons, volcanic emissions. This also includes related references.

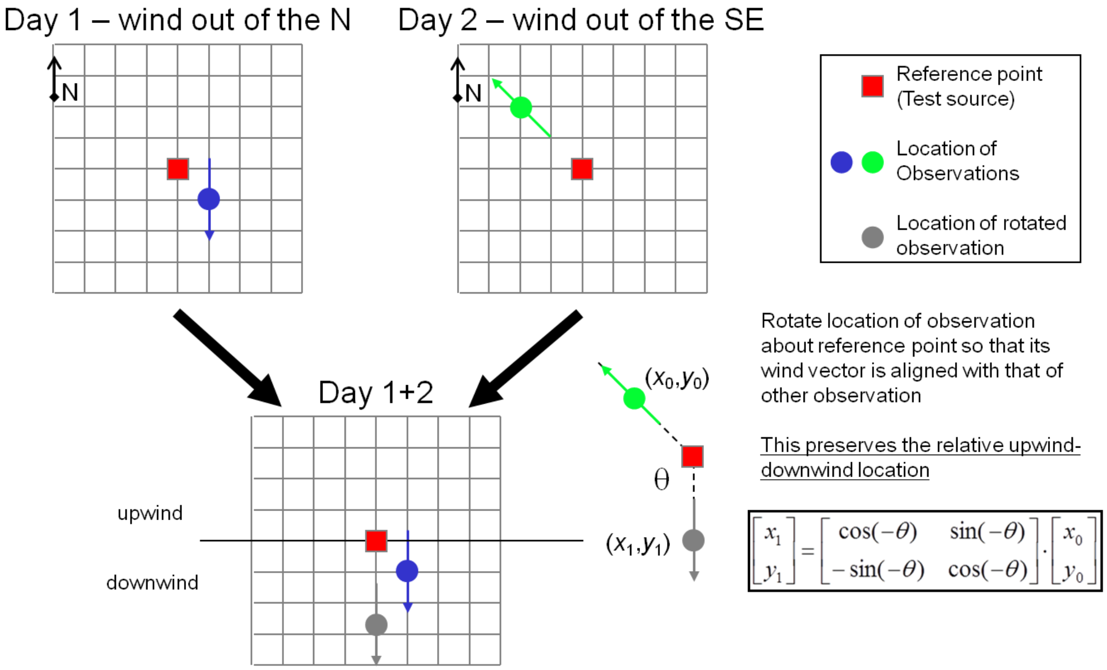
Supplementary Information file 1 – Excel spreadsheet on details (locations, source type, annual emissions, uncertainties) of missing and volcanic sources of SO_2

Supplementary Information file 2 – kml (Google-Earth) overlay showing locations and emissions of missing and volcanic SO_2 sources

Supplementary Information file 3 – kml (Google-Earth) overlays showing 2005-2007 OMI (i) average SO_2 columns and (ii) downwind-upwind differences

Supplementary Methods

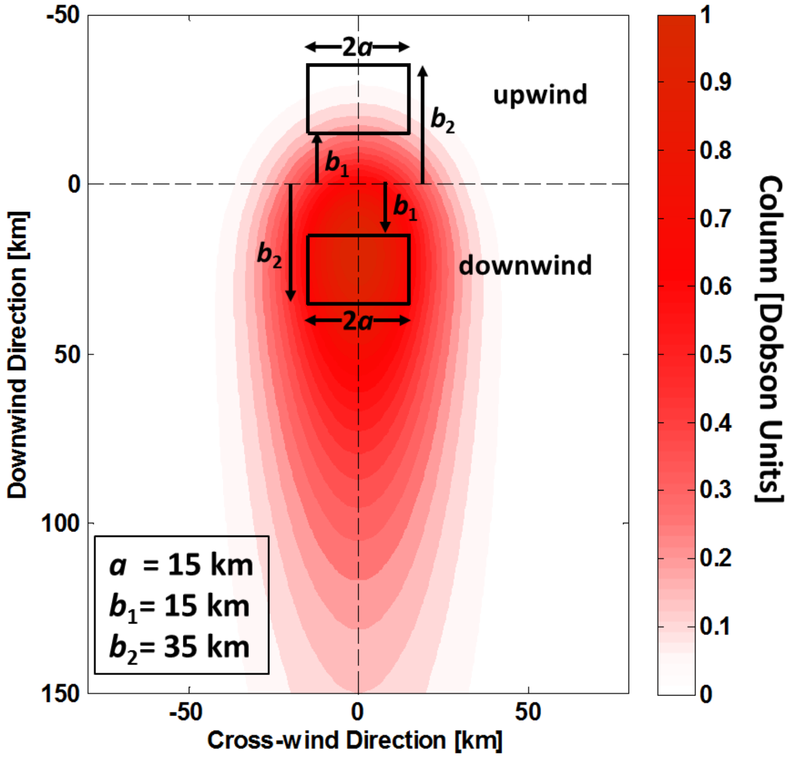
Illustration of the wind rotation scheme.



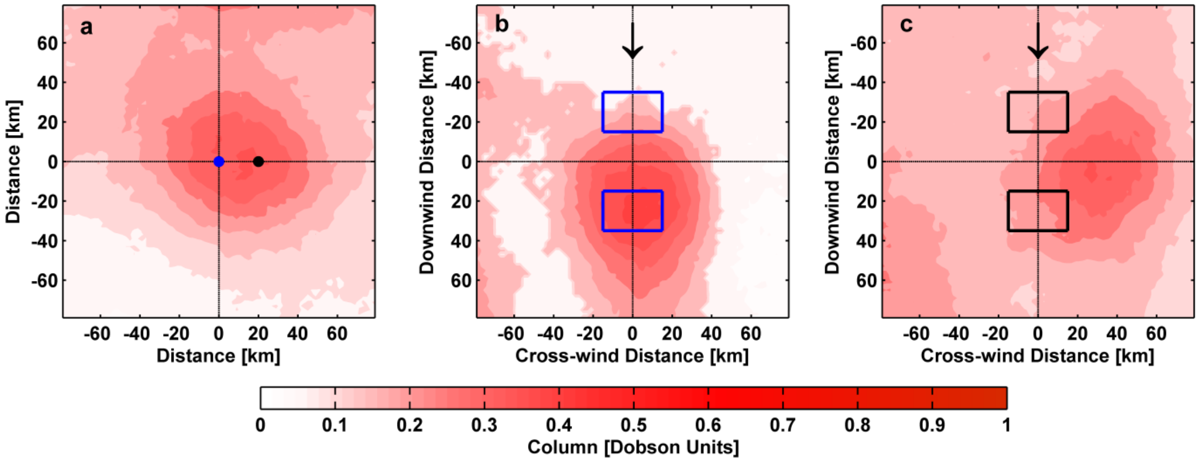
$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & \sin(-\theta) \\ -\sin(-\theta) & \cos(-\theta) \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

Supplemental Information Figure 1. Schematic of the rotation procedure, shown considering two days: Day 1 in which the wind is out of the north, and Day 2 in which the wind is out of the southeast. The location of the observation on Day 2 (green) is adjusted through a rotation about the Test source so that the wind is effectively out of the north (gray). In this way measurements from many days can be analyzed together[1,2].

Addition Information on Emissions Mapping.



Supplemental Information Figure 2. Idealized rotated SO_2 distribution indicating the downwind and upwind averaging regions, shown as the the two rectangles defined by a, b_1, and b_2. The SO_2 source is located at (0,0). The downwind average column is 0.85 DU and the upwind average column is 0.08 DU, giving an downwind-upwind difference of 0.77 DU.



Supplemental Information Figure 3. (a) Average SO_2 column amount over the Montour power plant in central Pennsylvania, U.S. (41.07°N; 76.67°W), with the power plant located here at (0,0); (b) The average SO_2 column after a rotation about the blue dot (the location of the power plant) in panel (a); (c) the average column after a rotation about the black dot, 20 km east of the location of the power plant, in panel (a). The rotated distributions in panels (b) and (c) are such that the wind is from the top of the plot and the boxes in panels (b) and (c) show the regions averaged to calculate the downwind and upwind average columns.

There are several potential sources of error that need to be considered when determining the overall uncertainty of the emissions. These are summarized in Supplemental Information Table 1, and are grouped into three categories: in the determination of the vertical column density (VCD), or column, those that come from the fit of the column distributions, and those related to the ECMWF wind information and its usage.

The OMI SO_2 VCDs are subject to uncertainties. Each VCD measurement possesses a random error on the order of 0.5 DU (ref. 3). We assume here that these can be neglected since we are averaging over a large number of individual measurements. However, there is also an uncertainty from the air mass factors (AMFs) used to convert slant column density (SCD) into VCD. Based on a sensitivity study performed in McLinden et al.[4], uncertainties in surface reflectivity, surface pressure, ozone column, and cloud fraction and pressure all combine for an uncertainty of 18%. The uncertainty from profile shape is more difficult to evaluate. For this, the sensitivity in AMFs was assessed by recalculating for different SO_2 profile assumptions including: exponentially decreasing number densities to the top of the boundary layer[5], and fixed SO_2 layer thicknesses of 1, 1.5, and 2 km. The standard deviation of these variations, 18%, was used to quantify this uncertainty. The AMF calculations assumed a Lambertian surface and we estimate this approximation could lead to a 10% uncertainty. We estimate the uncertainty due to aerosols by adjusting the optical depth of the aerosol layer[6] by ±0.25 about its assumed mean value (to a minimum of 0 and a maximum of 1) and recalculating AMFs. This changes the AMF by 10%. Based on these, the overall AMF uncertainty (which translates directly into emission uncertainty) was found to be 27%.

There is a statistical uncertainty from the fit related to noise in the OMI observations and the representativeness of the two-dimensional functional form[2]. We determined here that this uncertainty can be well quantified at 15 kt/yr + 5%. To determine the uncertainty arising from the use of an incorrect lifetime or width parameter, we recomputed emissions using the 10th and 90th percentile values from the lifetime (shown in Supplemental Information Figure 4) and width parameter distributions for a subset of the emissions sources analysed. Combined, these led to an uncertainty of 35%. We also examined the impact of modifying the domain over which the VCDs are fit. Different combinations, varied by ±20 km, were attempted and changes no larger than 13% were found. The data filtering choices, cloud fraction and snow/snow-free pixels, were found each have a small impact on the emissions.

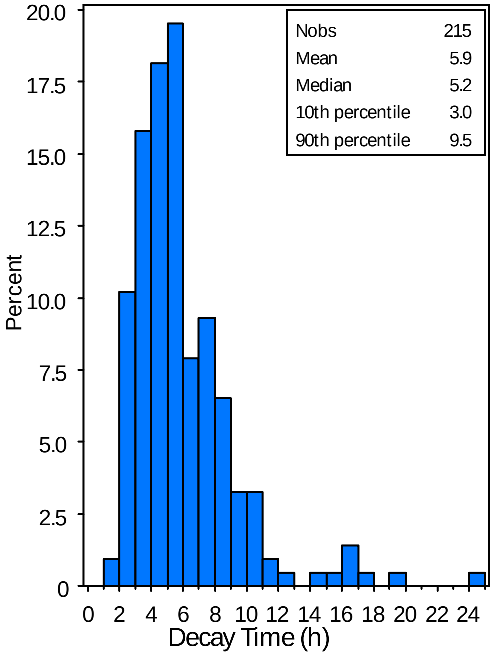
Supplemental Information Table 1. Uncertainty budget for annual emission rates from a single source.

Error source	Source	Category	Magnitude	Note
Air mass factor	Variability	VCD	18%	Considers random errors in cloud fraction, cloud pressure, surface albedo, surface pressure, column ozone, and profile shape following McLinden et al.[4]
	Variability	VCD	18%	Profile shape
	Uncertainty	VCD	14%	Bidirectional reflection distribution function (10%) and aerosol (10%)
Mass	Linear fit	Fit	5%	Statistical errors from the regression model.
	Uncertainties in OMI retrieved values	Fit	10-20 kt/yr	
Lifetime and Width	Uncertainty	Fit	35%	Based on using lifetimes of 3 and 10 hours and widths 12 and 31 km; these represent the 10th and 90th percentiles
Two-dimensional fit parameters	Uncertainty	Fit	13%	Fitting limits; assessed by adjusting up/down wind and cross-wind fitting limits by ±20 km
Cloud fraction	Uncertainty	Fit	1%	Changed maximum cloud fraction to 0.1, 0.3
Snow on ground	Uncertainty	Fit	2%	Impact of including measurements of snow-covered pixels
Wind-speed and direction	Variability	Winds	4%	Determined by applying random errors in wind speed (2 m/s) and direction (15°)
	Uncertainty	Winds	6%	Determined by applying systematic errors in wind speed (2 m/s) and direction (10°)
Wind height	Uncertainty	Winds	20%	Systematic effect from taking the winds at a different height; wind height adjusted by ±500 m
Source of wind information	Uncertainty	Winds	7%	Change in emissions using winds from NCEP[7] reanalysis for one test source
Total			50% > 60%	For sources above 100 kt For sources under 50 kt

Another source of error related to emissions retrieval are errors in the ECMWF ERA-interim reanalysis[8] wind-speed and direction, as well as an error related to the height of the wind. The impact of uncertainties in wind speed and direction were estimated by artificially adding random and/or systematic errors in speed and direction (see Supplemental Information Table 1) and recalculating emissions. The use of winds at an inappropriate altitude was quantified by changing the height of the winds used by ±500 m and then recalculating. Wind height appears to be one of the larger sources of error (20%). The use of an alternative source of wind information, NCEP[7], was also examined for a small subset of sources, with a modest (7%) impact on emissions.

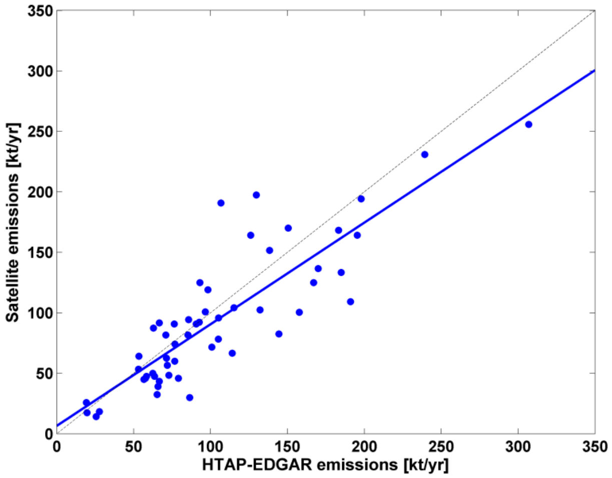
Combining these components, we estimate that the overall single-source uncertainty in annual emissions to be 50% for a larger source and about 60% for a smaller source. These values appear to represent more of an upper limit as comparisons with reference data made here (see below) indicate differences in the range of 20-30% are more typical.

Parts of China have SO_2 emissions dense enough that they can no longer be considered as (multiple) point sources. Here we use maxima in the source map and perform an emissions calculation that roughly represents a regional emission rate. While these estimates are expected to have a larger uncertainty, we note that national OMI-emission totals are still roughly 50% of the bottom-up inventory values.



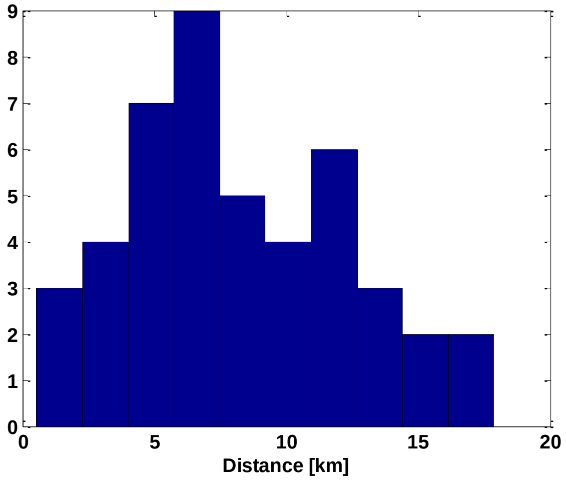
Supplemental Information Figure 4. The distribution of the estimated decay time (or, effective lifetime) estimated from the fit of rotated OMI data for the 2005-2007 period using the methodology of Fioletov et al.[10] Data from 215 emission sources that produced estimates with small uncertainties were used for the plot. The main statistical characteristics of the distribution are also shown.

Inventory evaluation. A comparison of emissions derived using this approach to those from the HTAP inventory in the eastern U.S. is shown in Supplemental Information Figure 5. Here we integrate the HTAP emissions over all grid-boxes within a radius of 50 km about the source location. There is good correlation (correlation coefficient of 0.86) and the slope is 0.85. These values are well within the uncertainties from Supplemental Information Table 1, and suggest that the satellite emissions have a slight low bias. However, the slope is somewhat sensitive to the integration radius used. A more direct comparison was conducted in Fioletov et al.[2] using power plant emissions measured from Continuous Emissions Monitoring Systems and this showed even better correlation and a slope closer to unity.



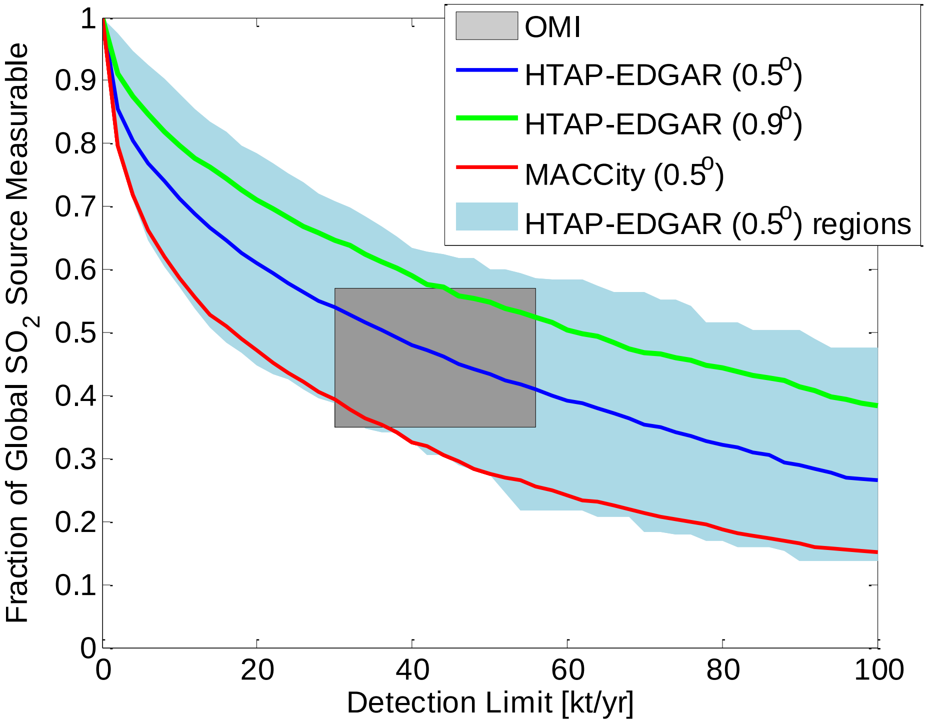
Supplemental Information Figure 5. Comparison of HTAP-EDGAR v2 2008 SO_2 emissions with those derived from OMI over the eastern U.S. HTAP emissions were calculated by summing over all grid-boxes within a 50 km radius of the emission location. The slope of the best-fit line (blue) is 0.85 and the correlation coefficient is 0.86.

To provide a quantitative measure of how well our algorithm can locate a source we assume the peak in a hot-spot on the source map corresponds to the source location, considering single sources in the eastern U.S. We employ an automated, two-dimensional peak finding algorithm to determine their co-ordinates and calculate the distance between these peak locations and the true co-ordinates of the sources. The histogram of distances is shown in Supplemental Information Figure 6. We find the average distance is 8±4 km with no values larger than 20 km.



Supplemental Information Figure 6. Distribution of the distance between the actual source location and the location assigned to that OMI-identified source using a two-dimensional local peak finding routine. The average ± 1 standard deviation, which can be interpreted as a geolocation error, is 8.2 ± 4.2 km. Only single sources in the Eastern U. S. were considered.

The uncertainty of regional emissions should be smaller than that of individual sources due to a cancellation of random errors. Totalling HTAP emissions and OMI emissions for sources captured by both inventories we find agreement to better than 5%. Also, we find on average that 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. This is consistent with the distribution of sources globally in the inventories for a comparable detection limit as demonstrated in Supplemental Information Figure 7 where we used the bottom-up inventories to assess what fraction would be measureable for a given detection limits. Globally, we find 40-55% would be measured. We also computed this for the different regions considered in Figure 3 where up to 70% is measureable for a 40 kt/yr detection limit. On this basis it is reasonable to conclude a discrepancy exists between OMI and bottom-up emissions for a region where OMI exceeds 70% of the bottom-up.

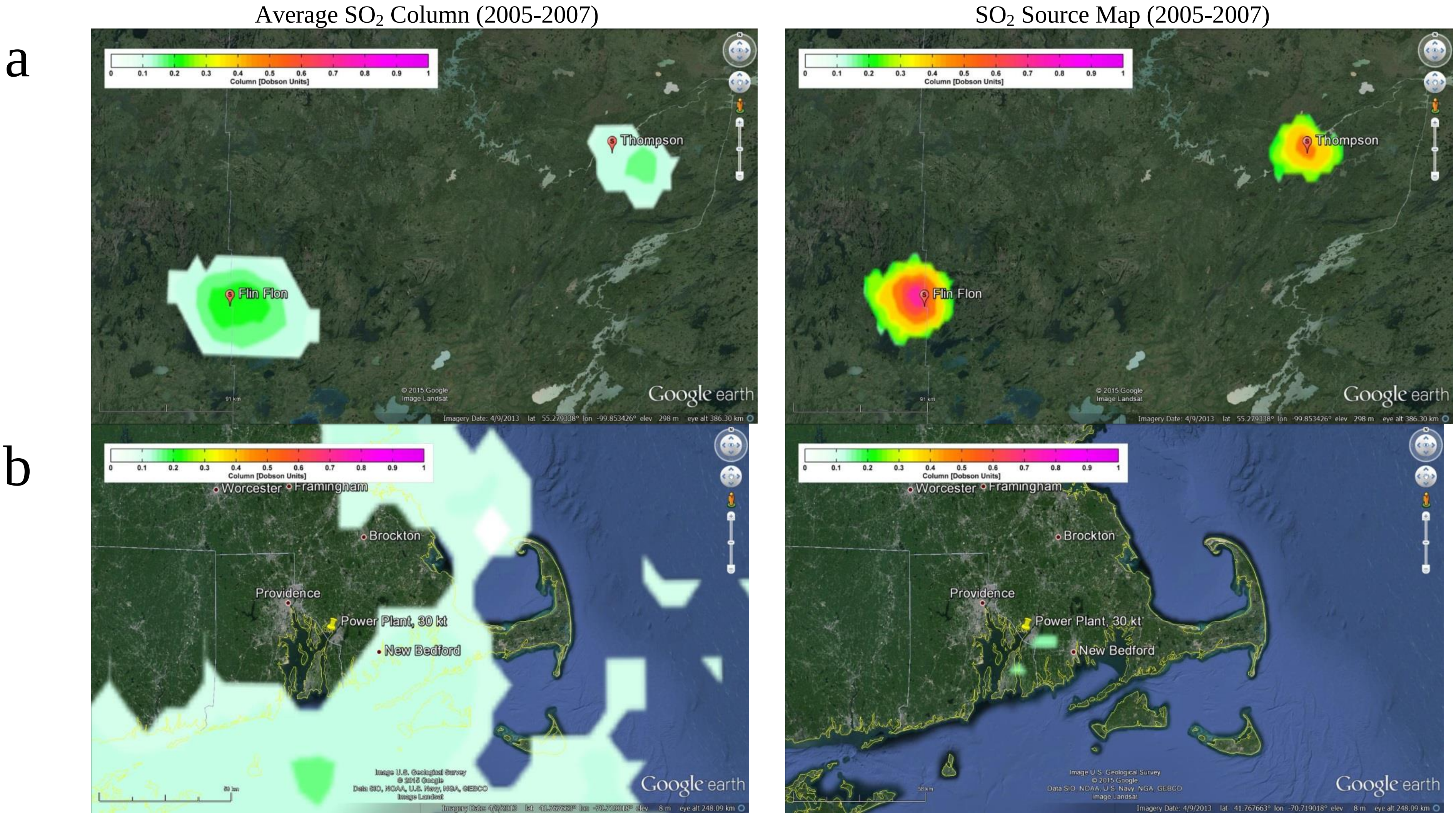


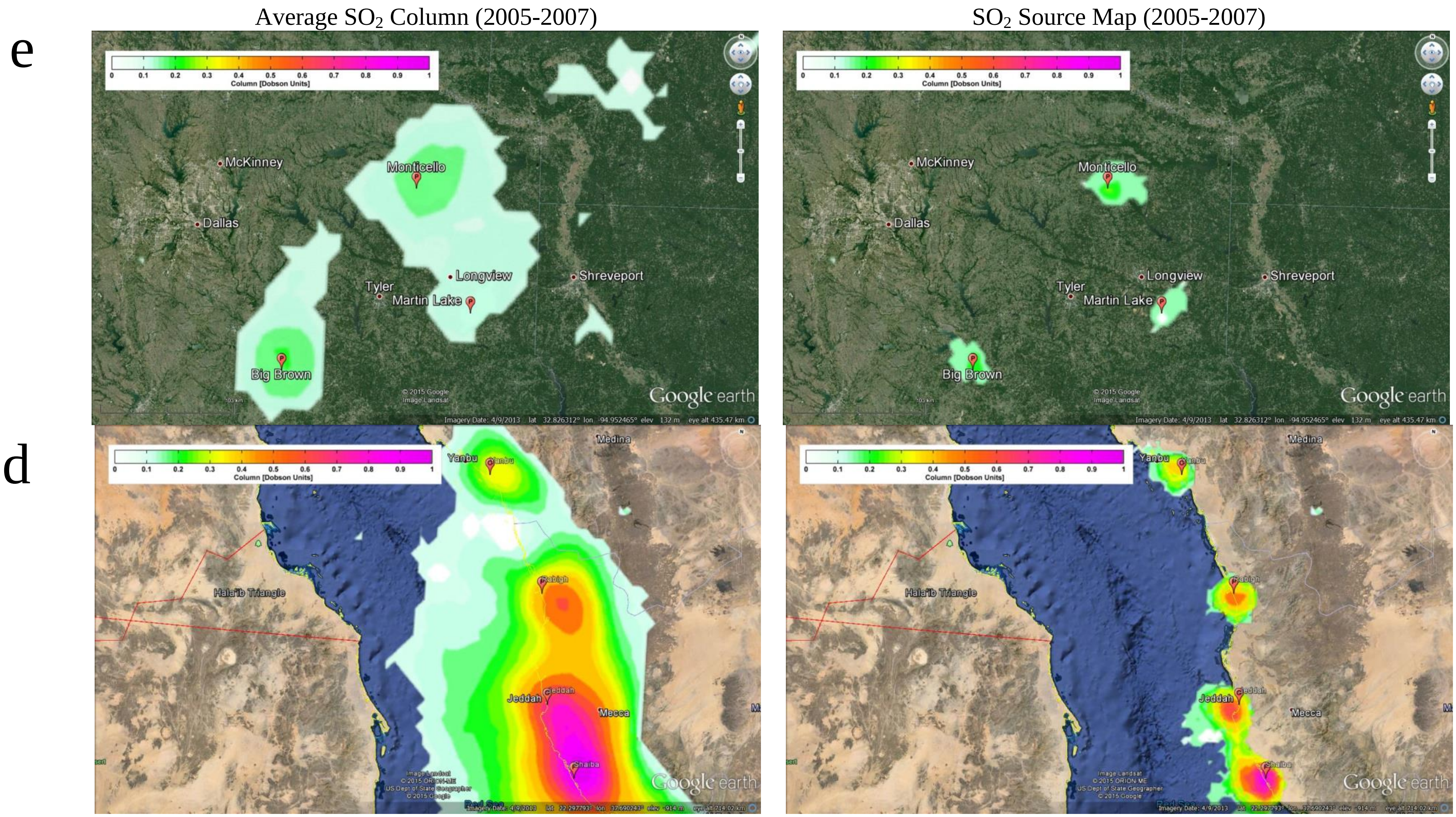
Supplemental Information Figure 7. A comparison of the fraction of total, global anthropogenic SO_2 emissions measurable for a given detection limit for two emission inventories and that found with OMI. The HTAP-EDGAR v2 inventory *(19)*, originally on a 0.1° × 0.1° grid, was summed to 0.5° × 0.5° and 0.9° × 0.9° grids to better reflect the effective resolution of OMI emissions; MACCity *(20)* was left at its native 0.5° × 0.5° resolution. The EDGAR v4.2 emission curves are very similar to the HTAP-EDGAR curves. The blue shading shows the range of values calculated (using the HTAP-EDGAR, 0.5°) over all regions considered in Figure 3c and d.

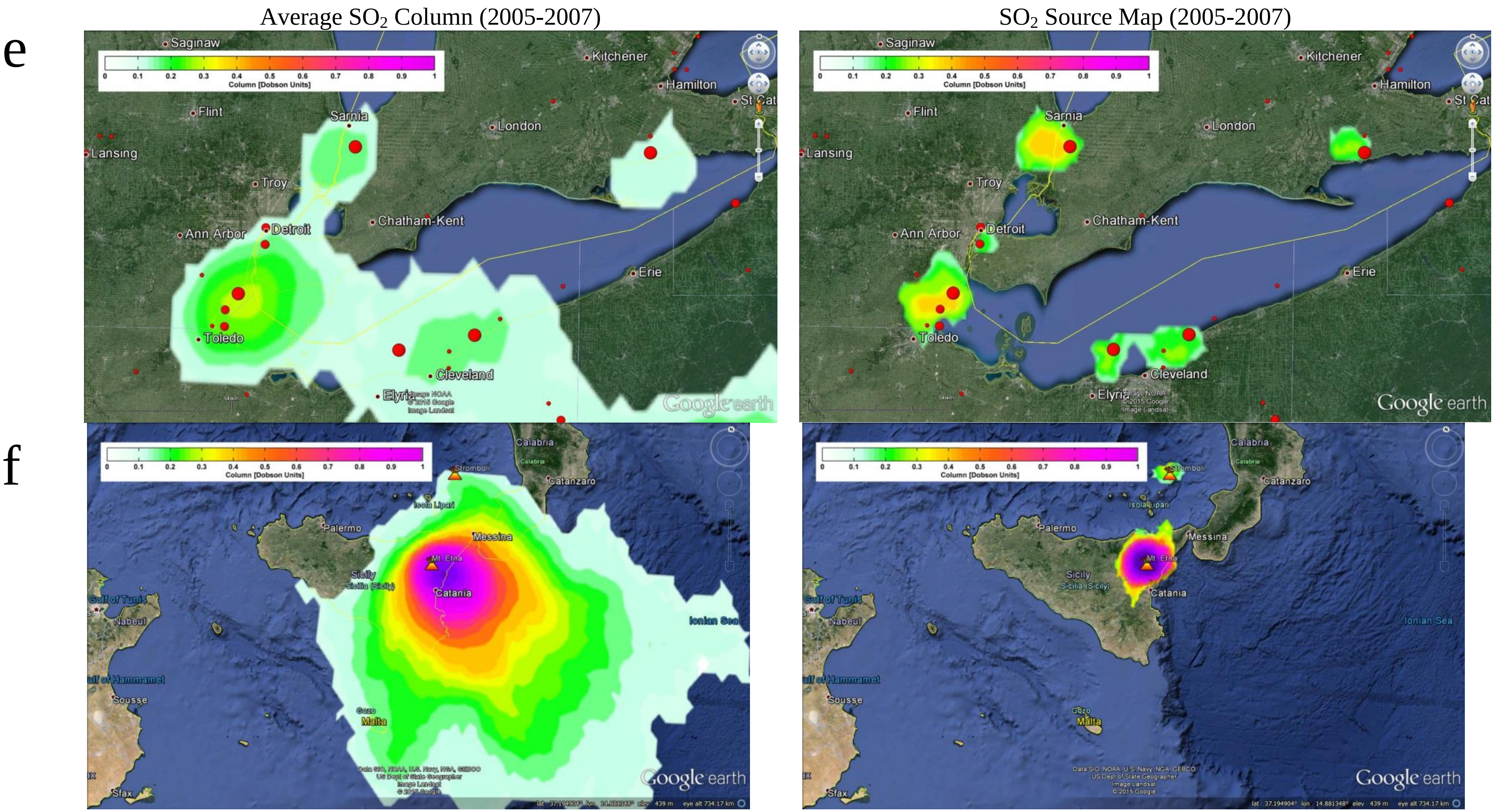
Comparisons of Average Columns and Downwind-Upwind Differences.

The OMI maps shown in this work are based on three-year averages of SO_2 columns in order to increase signal-to-noise. We also focus on years during the first half of the OMI mission (2005-2007 or 2007-2009) as this avoids the worst of the row anomaly[13] and best matches the emissions information available for comparison.

There are many examples in which the downwind-upwind column difference identified a source not apparent from the average SO_2 column map, or excluded locations that might appear to be a source. With very few exceptions it more closely corresponds to location of the source. A selection of some examples using the Google Earth overlays (Supplemental Information file 3) are shown in Supplemental Information Figure 8 and briefly discussed below. These are based on 2005-2007 OMI data.


Average SO$_2$ Column (2005-2007)
SO$_2$ Source Map (2005-2007)
a
b
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Column [Dobson Units]
Thompson
Flin Flon
Worcester
Framingham
Brockton
Providence
Power Plant, 30 kt
New Bedford
Google earth

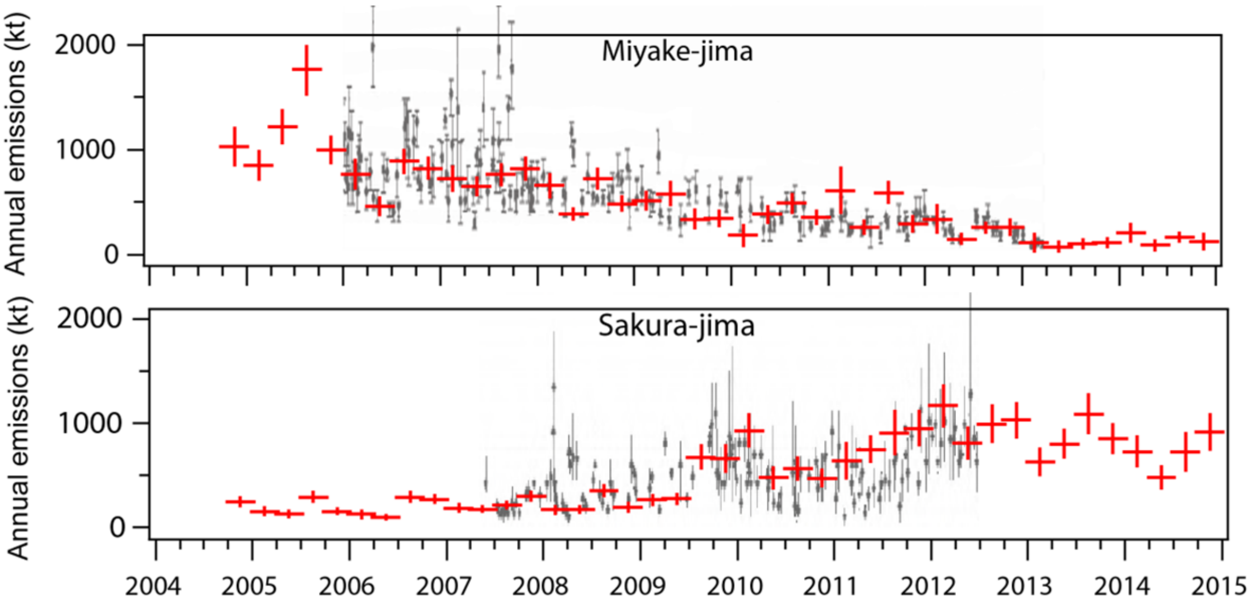

Average SO$_2$ Column (2005-2007)
SO$_2$ Source Map (2005-2007)
e
d
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Column [Dobson Units]
McKinney
Dallas
Monticello
Tyler
Longview
Martin Lake
Shreveport
Big Brown
Yanbu
Medina
Rabigh
Halaib Triangle
Jeddah
Mecca
Shaiba
Google earth


e
f
Average SO$_2$ Column (2005-2007)
SO$_2$ Source Map (2005-2007)
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Column [Dobson Units]
Saginaw
Kitchener
Hamilton
St Cat
Flint
Sarnia
London
Lansing
Troy
Detroit
Chatham-Kent
Ann Arbor
Erie
Toledo
Cleveland
Elyria
Google earth
Calabria
Stromboli
Catanzaro
Isola Lipari
Palermo
Messina
Mt. Etna
Sicily
Sicilia (Sicily)
Catania
Gulf of Tunis
Nabeul
Ionian Sea
Gulf of Hammamet
Sousse
Gozo
Malta
Sfax

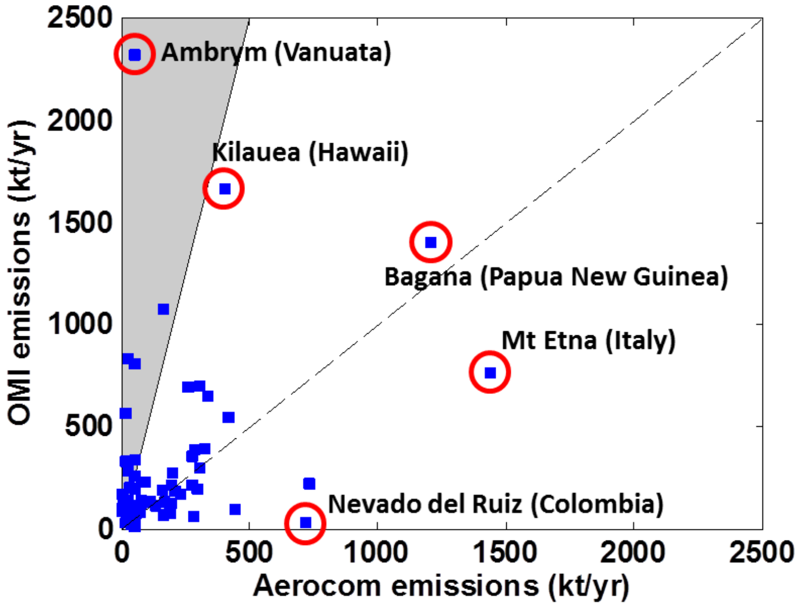
Supplemental Information Figure 8. Examples of OMI average SO_2 column and source maps from the period 2005-2007 (Supplemental Information file 3). Each row highlights a different location:

a. Smelters in northern Canada. Shown is the SO_2 distributions from two large non-ferrous smelters (emissions between 100 and 200 kt/yr) in the province of Manitoba. Due to frequent high winds their average column levels appear quite modest, about 0.2 DU. Their downwind-upwind difference, however, is much larger at about 0.8 DU, and in the case of Flin Flon, corresponds more closely to the actual location of the emission source. Map data: Google, Landsat.
b. A power plant in New England, US. The power plant (~30 kt/yr) does not appear at all in the mean column, but the downwind-upwind difference is able to detect it and locate it to within about 11 km of its true location. There is also a small local bias in the average column just offshore (to the south) that might be inferred as a source. The downwind-upwind difference was able to rule out this feature as a source. Map data: Google, U.S. Geological Survey, SIO, NOAA, U.S. Navy, NGA, GEBCO, Landsat.
c. Power plants in north Texas, U.S. This shows the SO_2 distributions from three power plants (emissions 50-100 kt/yr). Map data: Google, Landsat.
d. Power plants and oil & gas facilities along the eastern shore of the Red Sea. The source map is better able to isolate and locate sources as compared to the average distribution. Map data: Google, Landsat, ORION-ME, U.S. Dept. of State Geographer.
e. Sources around Lake Erie (U.S. and Canada). This shows the SO_2 distributions from power plants and other sources around Lake Erie. Note that the source map is able to pick out a second power plant, west of Cleveland, not evident in the average column. Sources here are from HTAP, a gridded inventory, and so may be a few km away from the true location. Map Data: Google, NOAA, Landsat.
f. Volcanoes near Sicily, Italy. In the average column map, the enormous SO_2 signal from Mt. Etna masks the smaller signal from Mt. Stromboli, whereas in the source map, Mt. Stromboli is readily apparent. Map Data: Google, SIO, NOAA, U.S. Navy, NGA, GEBCO, U.S. Dept. of State Geographer, Landsat.

Emissions from Volcanic Sources.



Supplemental Information Figure 9. Time series of seasonal OMI-estimated emission rates from the Miyake-jima and Sakura-jima volcanoes (near the southern tip of Japan) are shown in red. The error bars represent 2-σ confidence intervals. Grey dots are daily emission estimates provided by the Japan Meteorological Agency. The grey vertical bars represent minimum and maximum values.



Supplemental Information Figure 10. Comparison of OMI-derived and Aerocom[13] volcanic emissions. Scatterplot of OMI-derived annual mean volcanic emissions and the passive outgassing component from the Aerocom volcanic database emissions, both averaged over 2005-2010. The correlation coefficient is 0.36. The shaded area represents the region in which OMI emission are at least a factor of 5 larger than Aerocom emissions and which we consider as missing sources. Some noteworthy volcanoes are identified.

Relation to previous work.

Streets et al.[14] provides an excellent review on the application of satellite remote sensing for constraining or deriving emissions information. We elaborate here on some of the more relevant studies that preceded this publication. Leue et al.[15] were the first to examine the downwind decay of pollution derived from satellite observations to infer emissions, in this case NO_2 from the Global Ozone Monitoring Instrument. Beirle et al.[16] exploited the higher spatial resolution of OMI and applied a more advanced fitting procedure to derive NO_x (where NO_x=NO+NO_2 is largely emitted in the form of NO but rapidly converted to NO_2) emissions from several point sources. Fioletov et al.[2] and de Foy et al.[17] advanced this even further employing so-called exponential-modified Gaussian distributions to parameterize the downwind distributions of some U.S. sources. However, all these studies required that the location of the emissions be known beforehand and were of limited geographical scope.

Supplementary References

1. Pommier, M., C. A. McLinden, and M. Deeter, Relative changes in CO emissions over megacities based on observations from space, *Geophys. Res. Lett.*, **40**, doi:10.1002/grl.50704 (2013).
2. Fioletov, V. E., C. A. McLinden, N. Krotkov, and C. Li, Lifetimes and emissions of SO_2 from point sources estimated from OMI, *Geophys. Res. Lett.*, **42**, doi:10.1002/2015GL063148 (2015).
3. Li, C., J. Joiner, N. A. Krotkov, and P. K. Bhartia, 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, doi:10.1002/2013GL058134 (2013).
4. 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, doi:10.5194/acp-14-3637-2014 (2014)
5. von Engeln, A. and J. Teixeira, A Planetary Boundary Layer Height Climatology derived from ECMWF Re-analysis Data, J. Clim., 26, 6575-6590, doi: 10.1175/JCLI-D-12-00385.1 (2013)
6. 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, doi:10.5194/acp-12-8037-2012 (2012).
7. Suranjana Saha, A., et al., The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., 91, 1015–1057, doi: http://dx.doi.org/10.1175/2010BAMS3001.1 (2010).
8. 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).
9. 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).
10. 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).
11. G. Janssens-Maenhout, *et al.*, Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, Greenhouse Gas Measurement and Management 1: 124 pp (2010).
12. http://projects.knmi.nl/omi/research/product/rowanomaly-background.php (last accessed: April 16, 2016)
13. 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, doi:10.5194/acpd-12-24895-2012 (2012).

14. Streets, D. G., *et al.*, Emissions estimation from satellite retrievals: A review of current capability, *Atmos. Environ.*, **77**, 1011–1042, doi:10.1016/j.atmosenv.2013.05.051. (2013).

15. Leue, C., *et al.*, Quantitative analysis of NO_2 emissions from GOME satellite image sequences, *J. Geophys. Res.*, **106**, 5493-5505 (2001)

16. Beirle, S., *et al.*, Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, *Science*, **333**, 1737–1739 (2011).

17. de Foy, B., *et al.*, Estimates of power plant NO_x emissions and lifetimes from OMI NO_2 satellite retrievals, *Atmos. Env.*, 116, 1-11 (2015).