Supplementary Information: Improved Satellite Retrieval of Tropospheric NO_2 Column Density via Updating of Air Mass Factor (AMF), Part I: Case Study of Southern China

Hugo Wai Leung Mak [1], Joshua L. Laughner [2], Jimmy Chi Hung Fung [1,3], Qindan Zhu[4], and Ronald C. Cohen [2,4]

[1] Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China
[2] Department of Chemistry, University of California, Berkeley, USA
[3] Division of Environment and Sustainability, Hong Kong University of Science and Technology, Hong Kong, China
[4] Department of Earth and Planetary Sciences, University of California, Berkeley, USA
* Correspondence: hwlmak@ust.hk; Tel.: +852-9432 7394

S1. Air Mass Factor (AMF) based upon Traditional OMI-NASA Retrieval

In this Section, the Air Mass Factor (AMF) is calculated based on Equations (2) and (3) on Page 3. Figure S1 below shows the spatial distributions of AMF within four seasons of 2015, based on OMI-NASA retrieval. The AMF values of most places in southern China are between 0.5 and 0.8, with some pixels having extraordinarily high AMF (>0.85), which happens more frequently in April (spring) and July (summer) 2015, especially in Hainan, Zhanjiang and Yulin. None of the pixels have AMF values lower than 0.25 in all four seasons, indicating that the ratio of apparent column densities (ACD) to VCD is high.

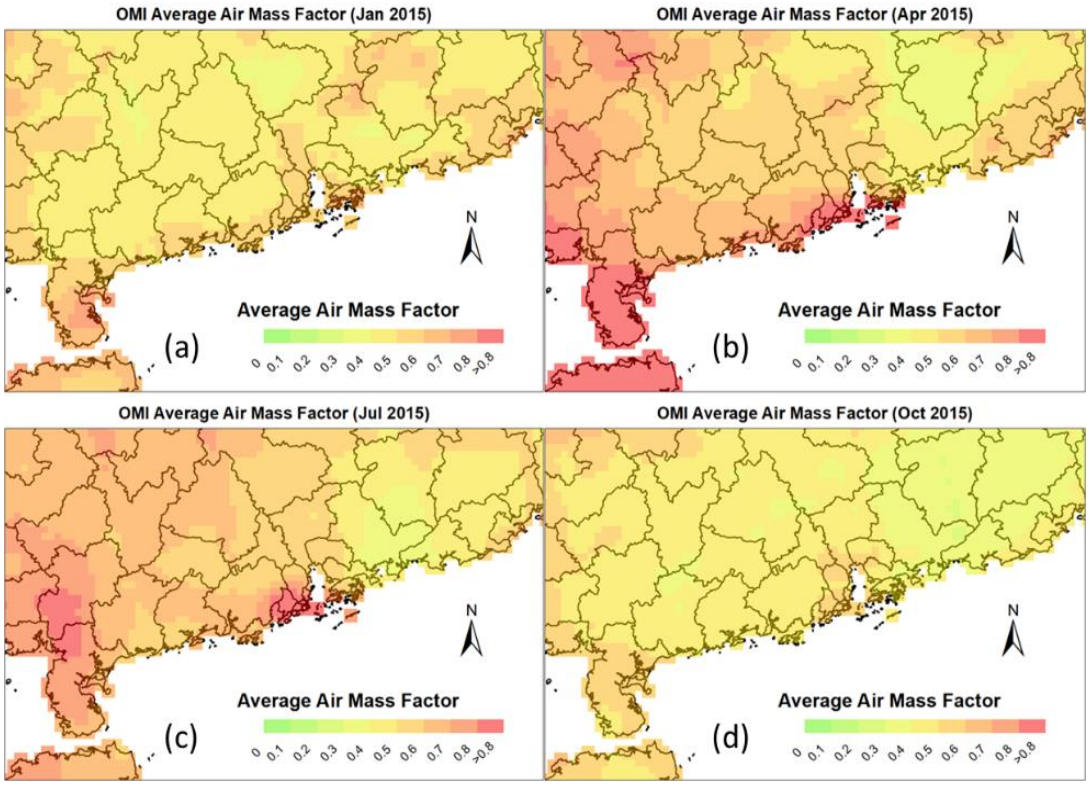

Figure S1. OMI-NASA average AMF in southern China. (**a**) Jan 2015; (**b**) Apr 2015; **(c)** Jul 2015; **(d)** Oct 2015. The units of the figures are dimensionless, and AMF ranges from 0 to 1 in most circumstances.

S2. Air Mass Factor (AMF) based upon BEHR v3.0C Retrieval

Using BEHR retrieval (with higher spatial resolution datasets), the resulting AMF becomes much lower in all four seasons in 2015, possibly due to the lack of lightning emissions within NO_2 profiles. There are three retrieval versions for BEHR, namely BEHR v3.0A, v3.0B and v3.0C respectively. All of them give similar AMF spatial distributions. Even for the most adverse circumstances, the difference in AMF between the two retrieval algorithms is less than 0.1 (i.e., a percentage difference of around 10%). We provide BEHR v3.0C AMF spatial plots in Figure S2, as a comparison with Figure S1.

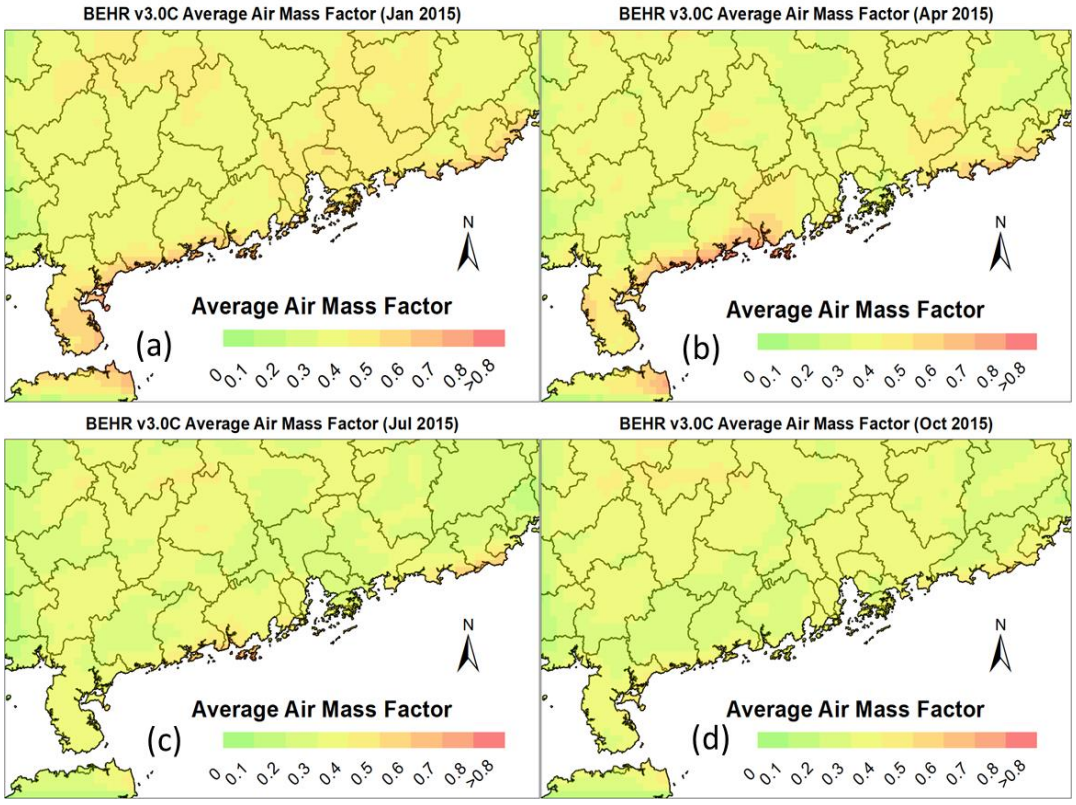

Figure S2. BEHR v3.0C average AMF in southern China (**a**) Jan 2015; (**b**) Apr 2015; **(c)** Jul 2015; **(d)** Oct 2015.; The units of the figures are dimensionless, and AMF ranges from 0 to 1 in most circumstances.

We notice that there are abrupt changes in AMF in all months compared with the corresponding spatial plot in Figure S1. Most pixels have AMF value of less than 0.5, in contrast with AMF in traditional OMI-NASA product (Figure S1). For coastal cities, average AMF is higher than inland areas in general.

S3. Statistical Indices and Methods for Evaluation

In analyzing satellite retrieval outputs, we adopt different statistical measures and quantities. In Section 5.2 (Table 3) and Section 6.2 (Table 4), we first obtain corresponding best fit line by using linear regression techniques, then base on pairwise comparison of datasets, the equation of best-fit straight line, Pearson correlation coefficient (*R*-value), *p*-value, *t*-statistics and root mean squared errors (RMSE) are deduced. Equations (S1) to (S4) below show the working formulae of these well-accepted statistical parameters used.

$$R = \begin{cases} \dfrac{\sum_{i=1}^{N}(\mathrm{VCD}_{1,i}-\overline{\mathrm{VCD}_1})(\mathrm{VCD}_{2,i}-\overline{\mathrm{VCD}_2})}{\sqrt{\sum_{i=1}^{N}(\mathrm{VCD}_{1,i}-\overline{\mathrm{VCD}_1})^2}\sqrt{\sum_{i=1}^{N}(\mathrm{VCD}_{2,i}-\overline{\mathrm{VCD}_2})^2}} & \text{(for Section 5.2)} \\ \dfrac{\sum_{i=1}^{N}(\mathrm{VCD}_{\mathrm{MAX-DOAS},i}-\overline{\mathrm{VCD}_{\mathrm{MAX-DOAS}}})(\mathrm{VCD}_{1,i}-\overline{\mathrm{VCD}_1})}{\sqrt{\sum_{i=1}^{N}(\mathrm{VCD}_{\mathrm{MAX-DOAS},i}-\overline{\mathrm{VCD}_{\mathrm{MAX-DOAS}}})^2}\sqrt{\sum_{i=1}^{N}(\mathrm{VCD}_{1,,i}-\overline{\mathrm{VCD}_1})^2}} & \text{(for Section 6.2)} \end{cases} \quad \text{(S1)}$$

$$p\text{-value} = \begin{cases} \Pr(\mathrm{VCD} \geq x|\mathrm{H}) & \text{(right tail)} \\ \Pr(\mathrm{VCD} \leq x|\mathrm{H}) & \text{(left tail)} \\ 2\min\{\Pr(\mathrm{VCD} \leq x|\mathrm{H}), \Pr(\mathrm{VCD} \geq x|\mathrm{H})\} & \text{(double tail event)} \end{cases} \quad \text{(S2)}$$

$$t\text{-statistic } (t_{\widehat{\mathrm{VCD}}}) = \frac{\widehat{\mathrm{VCD}}-\mathrm{VCD}_0}{\mathrm{s.e.}(\widehat{\mathrm{VCD}})} \quad \text{(S3)}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\widehat{\mathrm{VCD}}-\mathrm{VCD}_i)^2}{N}} \quad \text{(S4)}$$

In Equation (S1), *R*-values in Section 5.2 measure the linear correlation between VCDs retrieved by two different satellite retrieval algorithms, namely 1 and 2 respectively, while in Section 6.2, it measures the linear correlation between VCD retrieved by different algorithms and MAX-DOAS tropospheric NO_2 VCD measurements. The range of *R*-value can be from -1 to 1.

In Equation (S2), H is supposed to be the null hypothesis, *p*-value indicates the probability where statistical summary is not less than the actual observed results, if H holds. Smaller *p*-values imply higher levels of significance as the null hypothesis may not adequately explain the statistical trend.

In Equation (S3), $\widehat{\mathrm{VCD}}$ is an estimator of VCD in the linear regression model, VCD_0 is a known constant that may or may not match the actual retrieved VCD, and $\mathrm{s.e.}(\widehat{\mathrm{VCD}})$ denotes the standard error of $\widehat{\mathrm{VCD}}$ to approximate VCD.

In Equation (S4), measures the difference between predicted values (i.e., $\widehat{\mathrm{VCD}}$ projected on the best-fit line by linear regression) and true data points (VCD). A lower RMSE is desirable because it means that most data points are less deviated from the best-fit line.