## **Supplemental Information.**

## A New Machine Learning based Analysis for Improving Satellite Retrieved Atmospheric Composition Data: OMI SO<sub>2</sub> as an Example

Can Li<sup>1,2</sup>, Joanna Joiner<sup>2</sup>, Fei Liu<sup>2,3</sup>, Nickolay A. Krotkov<sup>2</sup>, Vitali Fioletov<sup>4</sup>, Chris McLinden<sup>4</sup>

<sup>1</sup>Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

<sup>2</sup>Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA <sup>3</sup>Goddard Earth Sciences Technology and Research (GESTAR) II, Morgan State University, Baltimore, MD 21251, USA

<sup>4</sup>Environment and Climate Change Canada, Toronto, Ontario, Canada

Correspondence to: Can Li (can.li@nasa.gov)



Figure S1: Monthly mean SO<sub>2</sub> SCDs for March 2005 from (first column) the original PCA retrievals and (second column) the neural network based analysis, and (third column) their differences for (a-c) January, (d-f) April, (g-i) July and (j-l) October 2005.



Figure S2: Same as Figure 6 but for India. For polluted areas ( $SCD_{NN} > 0.15 \text{ DU}$ ), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is 9%, with the original being greater.



Figure S3: Same as Figure 6 but for the Middle East. For polluted areas ( $SCD_{NN} > 0.15 \text{ DU}$ ), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is 14%, with the original being greater.



Figure S4: Same as Figure 6 but for South Africa. For polluted areas ( $SCD_{NN} > 0.15$  DU), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is -1%, with the original being smaller.



Figure S5: Same as Figure 6 but for the eastern U.S. For polluted areas (SCD<sub>NN</sub> > 0.15 DU), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is -11%, with the original being smaller.



Figure S6: Same as Figure 6 but for Norilsk, Russia. For polluted areas (SCD<sub>NN</sub> > 0.5 DU), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is -4%, with the original being smaller.



Figure S7: Same as Figure 6 but for the southeast Pacific for October 2005. The large SO<sub>2</sub> plume is from the 2005 Sierra Negra eruption that started on October 22. For pixels with relatively large SO<sub>2</sub> loading (SCD<sub>NN</sub> > 1 DU), the mean relative difference between the original and analyzed SO<sub>2</sub> SCDs is 20%, with the original being greater.