

Estimating surface sulfur dioxide concentrations from satellite data over eastern China: Using chemical transport models vs. machine learning

Zachary Watson¹, Can Li², Fei Liu^{2,3}, Sean W. Freeman¹, Huanxin Zhang⁴, Jun Wang⁴, Shan-Hu Lee¹

¹Department of Atmospheric and Earth Science, University of Alabama in Huntsville, Huntsville, AL 35758
²NASA Goddard Space Flight Center, Greenbelt, MD 20771
³Goddard Earth Sciences Technology and Research (GESTAR) II, Morgan State University, Baltimore, MD 21251
⁴Department of Chemical and Biochemical Engineering, Center for Global & Regional Environmental Research, and Iowa Technology Institute, The University of Iowa, Iowa City, Iowa 52240

Correspondence to: Zachary Watson (zw0033@uah.edu) and Shan-Hu Lee (shanhu.lee@uah.edu)

Abstract. Sulfur dioxide (SO₂) is an important air pollutant that contributes to negative health effects, acid rain, and aerosol formation and growth. SO₂ has been measured using ground-based air quality monitoring networks, but routine monitoring sites are predominantly placed in urban areas, leaving large gaps in the network in less populated locations. Previous studies have used chemical transport models (CTMs) or machine learning (ML) techniques to estimate surface SO₂ concentrations from satellite vertical column densities, but their performance has never been directly compared. In this study, we estimated surface SO₂ concentrations using Ozone Monitoring Instrument (OMI) retrievals over eastern China from 2015-2018 utilizing GEOS-Chem CTM simulations and an extreme gradient boosting ML model. For the first time, we quantified methodological uncertainties for both methods, directly compared their performance on the same truth dataset, and validated the CTM-based method on a sub-annual timescale. The surface concentrations estimated from the CTM-based method had similar spatial distributions (r = 0.58) and temporal variations compared to the in situ measurements but were underestimated (slope = 0.24; RPE = 75%) and had worsening performance over time. The ML-based method produced more accurate spatial distributions (r = 0.77) and temporal variations with a smaller discrepancy (slope = 0.69; RPE = 30%) and stable performance over time. Despite the higher accuracy of the ML-based method at the monitoring sites, the CTM-based method produced more reasonable gridded spatial distributions over areas without monitoring data, such as over the oceans, since its estimations are independent from the in situ measurements.

Sulfur dioxide (SO₂) is an important air pollutant that contributes to negative health effects, acid rain, and aerosol formation and growth. SO₂ has been measured using ground-based air quality monitoring networks, but the routine monitoring sites are predominantly placed in urban areas, leaving large gaps in the network in less populated locations. Previous studies have used chemical transport models (CTMs) or machine learning techniques to estimate surface SO₂ concentrations from satellite vertical column densities, but no direct comparisons between the methods have been made. In this study, we estimated surface SO₂ concentrations using Ozone Monitoring Instrument (OMI) retrievals over eastern China from 2015-2018 utilizing GEOS-Chem simulations and an extreme gradient boosting machine learning model. Compared to the in situ measurements, the SO₂

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

concentrations estimated from the CTM method had similar spatial distributions ($r=0.58$) and intra- and interannual variations but were underestimated (slope = 0.24) with a relative percent error of ~75% and had worsening performance over time. The machine-learning method produced more accurate spatial distributions ($r=0.77$) and temporal variations, a smaller discrepancy and bias (~30%; slope = 0.69) and relatively stable performance over time. The machine-learning method performed better than the GEOS-Chem method on smaller datasets and timescales with shorter temporal averaging periods. Ultimately, both methods were useful for estimating surface SO_2 concentrations since the CTM-based method does not rely on in-situ monitoring and produced more reasonable spatial distributions than the machine-learning method over areas without surface monitoring data.

1 Introduction

Sulfur dioxide (SO_2) is an important air pollutant due to its effects on human health, air quality, weather, and climate. SO_2 has many anthropogenic sources such as fossil fuel combustion in power plants and ore smelters, as well as natural sources from volcanoes (Engdahl, 1973). Surface SO_2 concentrations are mainly driven by anthropogenic activity in urban areas and are known to ~~causenegatively impact~~ cardiovascular and respiratory health ~~impacts~~ (Engdahl, 1973; Krzyzanowski & Wojtyniak, 1982). SO_2 ~~also~~ readily undergoes oxidation reactions in the atmosphere to form sulfuric acid, which ~~further~~ contributes to acid rain (Seinfeld and Pandis, 2016) and participates in aerosol formation and growth (Lee et al., 2019), ~~leading to further~~ These aerosols can then additionally ~~affects-on~~ weather and the global energy budget (NASEM, 2016).

Concentrations of SO_2 at the surface have been ~~regularly~~ measured using ground-based air quality monitoring networks. Surface concentrations are ~~typically~~ measured on hourly to daily time intervals, but the sites are predominantly located in urban areas, leaving large gaps in the network elsewhere. ~~In addition to surface-based air quality monitors,~~ ~~S~~satellite-based instruments can measure total-column concentrations of SO_2 globally ~~from space~~. These SO_2 vertical column densities (VCDs) are retrieved using the absorption of backscattered solar radiation in the ultraviolet wavelengths measured by a spectrometer (e.g., Krotkov et al., 2008; Levelt et al., 2006; Li et al., 2013; Li et al., 2020a; Nowlan et al., 2011; Theys et al., 2015). The VCDs are typically available ~~overfor cloud-free locations~~ over large areas ~~in cloud-free locations~~ on a daily basis but do not directly provide the ~~surface~~ concentrations ~~at the surface~~. Additional ~~toolsanalysis are is~~ required to estimate the surface concentrations from the satellite-retrieved VCDs ~~as discussed below~~.

~~The first method is to use a~~Chemical transport models (CTMs) ~~can be used~~ to convert satellite VCDs into surface concentrations using simulated surface-to-VCD ratios (SVRs). This method was initially developed for estimating surface ~~$\text{PM}_{2.5}$ concentrations of particulate matter~~ from satellite-based aerosol optical depth retrievals (Liu et al., 2004) and was later applied to nitrogen dioxide (NO_2 ; Lamsal et al., 2008) and SO_2 (Lee et al., 2011). Lee et al. (2011) and Zhang et al. (2021) ~~each~~ used coarse-resolution CTMs (grid spacings on the order of 100 km) to convert SO_2 VCDs from the Ozone Monitoring Instrument (OMI) into surface concentrations over North America for 2006, and China for 2005-2018, respectively. McLinden et al. (2014) and Kharol et al. (2017) used higher-resolution CTMs (grid spacing on the order of 10 km) and OMI ~~SO_2~~ VCDs

Formatted: Subscript

65 to estimate the surface concentrations ~~with focuses on~~ the Canadian oil sands from 2005-2011, and ~~the~~ North American
continent from 2005-2015, respectively. These ~~four~~ studies ~~each~~ demonstrated that annual mean satellite-derived surface SO₂
concentrations ~~can~~ accurately capture the spatial distribution ~~offrom the~~ ground-based air quality monitoring networks,
despite~~although~~ the estimated surface concentrations ~~werebeing~~ generally underestimated. An advantage of the CTM-based
method is that it is based on fundamental principles of atmospheric dynamics and chemistry and can produce results that are
70 independent of observed surface concentrations. The main limitations of CTMs are the computational expense of running the
simulations (Fan et al., 2022) and ~~relatively coarse-resolution, simulations may have~~ which may lead to large biases ~~due toin~~
~~the representation of~~ emissions, meteorology, and chemical processes (Wang et al., 2020b; Wang et al., 2020c).

More recently, machine learning (ML) techniques have been used to estimate surface SO₂ concentrations from
satellite retrievals, meteorology, and other ~~geographic~~ variables such as emission inventories ~~and land use types~~. Zhang et al.
75 (2022) used a Light Gradient Boosting Machine (LightGBM) to estimate surface SO₂ concentrations over northern China using
OMI SO₂ VCDs, meteorological variables, emissions, land use classifications, population density, and others. Yang et al.
(2023a) used radiances from the Geostationary Environment Monitoring Spectrometer (GEMS) satellite to estimate the surface
concentrations of SO₂ and other criteria air pollutants in a multi-output random forest model. Both studies showed that ML
techniques can accurately capture the spatial distribution and magnitude of the surface concentrations but ~~had may have~~
80 artificial biases due to nonphysical links between ~~variables-predictors~~ and the ~~observed~~ surface concentrations, such as
interactions between ~~certain~~ land use ~~typesclassifications~~ and skin temperature as shown by Zhang et al. (2022). In these
studies, the ML models ~~also~~ incorporated spatial (e.g., longitude, latitude, population density) and/or temporal (e.g., numeric
day of year, hour of day) proxies to improve performance ~~rather than depending only on measurable quantities~~, but this ~~can~~
~~lead the model to learn the locations of cities and introduce an artificial seasonality rather than depending on measurable~~
95 ~~quantities, limiting~~ the physical usefulness, ~~and interpretability, and applicability~~ of the model (Zhang et al., 2022; Yang et
al., 2023a; Yang et al., 2023b). ~~An advantage of the ML-based method is that the~~ ML models are typically much faster to train
and run than a full CTM simulation and can ~~often~~ utilize higher ~~spatial and temporal~~ resolution data (Fan et al., 2022);
~~however, s~~Since ML models can only use statistical relationships to make predictions, they are often limited in their physical
interpretability ~~and may make their predictions based on predictors that have no physical relevance to the estimated surface~~
90 ~~concentrations~~.

Although the CTM- and ML-based methods have ~~both each~~ been ~~employed in estimating used to estimate~~ surface SO₂
concentrations from satellite retrievals, ~~so far there has is been~~ a lack of direct comparisons between the ~~m two~~ methods. Here,
we estimated surface SO₂ concentrations using OMI ~~SO₂~~ VCDs over eastern China (105-125°E, 25-45°N) from 2015-2018 to
directly compare the two methods. ~~First, we quantified methodological uncertainties for each method for the first time.~~
95 ~~FirstNext~~, we used ~~the relationship between the surface and total column concentrations-simulated SVRs fromby the~~ GEOS-
Chem ~~model~~ to estimate ~~the~~ surface SO₂ concentrations from ~~the satellite data~~ OMI using the CTM-based method. Then, we
used a ML model to predict surface SO₂ concentrations from OMI VCDs, meteorological variables, and an emission inventory,
which are all physically relevant to the spatial distribution or lifetime of SO₂. The results from each method were validated

Formatted: Subscript

Formatted: Subscript

against ground-based in situ measurements from the China National Environmental Monitoring Centre (CNEMC) air quality monitoring network ~~on annual and seasonal mean timescales, the latter of which has never been done for the CTM-based method~~. Finally, we compared the performance of each method on ~~an identical-the same truth~~ dataset ~~over the same times and locations~~ for the first time to gain insights on their abilities and limitations to accurately estimate the surface SO₂ concentrations ~~from satellite data~~.

2 Data and methods

2.1 Study region

Eastern China has abundant anthropogenic SO₂ emissions and thus is a region with elevated surface concentrations. ~~Satellite SO₂ retrievals typically have a low signal-to-noise ratio due to interfering absorbers (Li et al., 2020), so regions with large SO₂ emissions and pollution, such as eastern China, are required to obtain sufficient signals from the spectrometer and provide more reliable retrievals compared to less polluted regions.~~ A map of our study region ~~with-including~~ the locations of OMI-derived SO₂ emission sources (Fioletov et al., 2022; ~~Fioletov et al., 2023~~) and CNEMC monitoring sites ~~-in-the-study region-are~~ shown in Fig. 1. The ~~largestmain~~ sources of SO₂ in the study region ~~come-from~~ ~~include around~~ 70 power plants, ~~as-well-as~~ five ore smelters, and one area of oil and gas production (Fig. 1b). There are also approximately 1000 ~~air quality monitoring~~ stations located ~~across-in~~ the ~~study~~ region that ~~can-be~~ ~~were~~ used to validate the estimated surface concentrations ~~from-the-satellite-data~~ (Fig. 1c). Our analysis covers the period from 2015 (the first full year of in situ measurements) to 2018 (to avoid the impacts of the COVID-19 lockdowns).

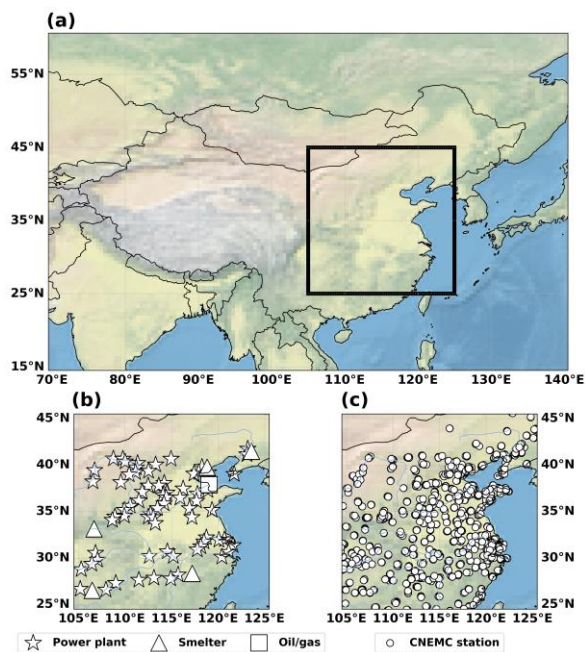


Figure 1: Maps showing the (a) study region (solid box; 105°E--125°E, 25°N--45°N) relative to the rest of the Asian continent, (b) locations of large SO₂ sources from the 2015 OMI emission catalogue during 2015 (Fioletov et al., 2022) including 70 power plants (stars), five ore smelters (triangles), and one area of oil and gas production (square), and (c) locations of the CNEMC monitoring stations (circles).

2.2 OMI satellite data

We employed data from the Ozone Monitoring Instrument (OMI; Levelt et al., 2006), a hyperspectral ultraviolet/visible nadir solar backscatter spectrometer launched onboard the Aura satellite in 2004. Aura flies in a sun-synchronous polar orbit, and OMI is used to retrieve SO₂ VCDs with daily global coverage and a spatial resolution of 13 km x 24 km at nadir, a significant improvement from previous satellite-based instruments. The VCDs were gridded to a horizontal resolution of 0.25° x 0.25° to decrease noise in the SO₂ retrieval without significantly coarsening it from the native measurement resolution. The OMI overpass time of our study region ranges from approximately 12:15 pm to 2:45 pm local time. For both the CTM- and ML-based methods, we used the OMI Planetary Boundary Layer (PBL) SO₂ product to estimate the surface concentrations due to its main application for anthropogenic, near-surface SO₂ (Krotkov et al., 2014; Li et al.,

Formatted: Subscript

2020b). The OMI retrievals use a principal component analysis- (PCA) based algorithm for spectral fitting based on the radiances ~~of wavelengths between 310.5–340 nm~~ for each row in the measurement swath with wavelengths between 310.5–340 nm (Li et al., 2013; Li et al., 2020a). This version of the PCA retrievals include pixel-specific air mass factor calculations to convert slant column densities (SCDs) to VCDs rather than using a fixed value worldwide (Li et al., 2020a). The VCDs express the number of SO₂ molecules in the column and are reported in Dobson Units (DU; 1 DU = 2.69 x 10¹⁶ molecules cm⁻²). To ensure good data quality, we ~~gridded the data to 0.25° x 0.25° resolution and~~ screened out measurements with cloud fractions greater than 0.3, solar zenith angles greater than 65°, located in the outer ten cross-track positions, or affected by the row anomaly (NASA, 2020). We also excluded extreme outliers that fell outside of five standard deviations from the mean as thresholds less than this appeared to remove legitimate data.

2.3 CNEMC ground-based monitoring data

Ground-based SO₂ concentrations from the China National Environmental Monitoring Centre (CNEMC) air quality monitoring network were used to validate the performance of ~~each both the CTM- and ML-based~~ methods. The concentrations were converted from µg m⁻³ to parts per billion (ppbv) following the procedure outlined in Wei et al. (2023). To ensure the ground-based measurements were temporally aligned with the OMI overpass, we averaged the hourly concentrations from 12:00 pm to 3:00 pm local time on days where there was at least one OMI observation within 40 km of the station. Like the OMI data, we also removed data that fell more than five standard deviations outside of the mean.

2.4 ~~GEOS-Chem~~CTM-based technique

We used simulated SVRs from the GEOS-Chem model (~~version 14.2.2; Bey et al., 2004~~The International GEOS-Chem User Community, 2023) to convert the OMI VCDs into surface concentrations for the CTM-based method. We ~~conducted ran~~ simulations for January, April, July, and October 2015, ~~each with a one-month spin-up to represent the SO₂ profiles in different seasons. Each simulation was conducted with a one-month spin-up following Kharol et al. (2015).~~ To reduce the computational expense, we used the monthly average SVR from each simulation to estimate the daily surface concentrations within the corresponding winter (DJF), spring (MAM), summer (JJA), and autumn (SON) months (referred hereafter as quasi-seasonal temporal sampling) for all years of the study period. The model was run at a horizontal resolution of 2.5° (longitude) x 2.0° (latitude) with 47 vertical layers and was driven by assimilated GEOS-FP meteorology (Lucchesi, 2018) and the Community Emissions Data System (CEDS) anthropogenic emission inventory (Hoesly et al., 2018). The internal time steps for the chemistry and advection calculations in the model were lengthened by 50% from the default values to reduce simulation times while minimizing errors ~~following (Philip et al., (2016)). Despite the longer internal timesteps, the~~ Courant-Friedrichs-Lewy condition is maintained with a Courant number of 0.041, indicating numerical stability of the simulations. We used model output at the lowest model level. ~~The surface concentrations were assumed to be equal to the concentrations at the lowest model level (~60 m above ground level). At the model output data output timestep of GEOS-Chem was every three hours, so it was sampled at 2:00 pm local time, which is the only output timestep inside the OMI overpass~~

window. We only included GEOS-Chem data in the analysis if there was at least one valid OMI observation within the model grid cell on a given day.

The approach from Lee et al. (2011) was used to infer surface SO₂ concentrations from OMI VCDs using the GEOS-Chem (GC) and simulated vertical SO₂ profiles from GEOS-Chem (GC). Lee et al. (2011) showed that the CTM-based method provided accurate results even with CTM resolutions that are much coarser than the satellite data. The monthly averaged profiles and SVRs from GEOS-Chem are shown in Fig. S1. The profiles indicate that most of the SO₂ within the vertical column is located near the surface and within the boundary layer (Fig. S1). The concentrations then drop to near zero in the free troposphere and have small variations, indicating a lack of elevated SO₂ plumes (Fig. S1). The profiles from the GEOS-Chem simulations are similar to those from aircraft observations (e.g., Li et al., 2012; Norman et al., 2025; Shan et al., 2025; Xue et al., 2010) and higher resolution simulations (Norman et al., 2025) over China. The daily surface SO₂ concentrations for the CTM-based method (S_{OMI}) were calculated on a daily basis at 0.25° x 0.25° resolution using the daily OMI VCDs and averaged GEOS-Chem SVRs from the model grid cell that the OMI measurement lies within using Eqn. 1. The conversion was done using the following relationship:

$$S_{OMI} = \frac{v S_{GC}}{v \Omega_{GC,PBL} + \Omega_{GC,FT}} \times \Omega_{OMI}, \quad (1)$$

where S is the surface SO₂ concentration in ppbv and Ω is the SO₂ VCD in DU. The FT and PBL subscripts are the free-tropospheric and boundary layer VCDs, respectively, which were calculated relative to the GEOS-FP PBL height. Since there is a significant difference in horizontal resolution between the satellite and model data, OMI VCDs were used to provide sub-model grid variability (v) using Eqn. 2:

$$v = \frac{\Omega_{OMI}}{\Omega'_{OMI}}, \quad (2)$$

where Ω_{OMI} is the OMI VCD at 0.25° x 0.25° resolution and Ω'_{OMI} is the average OMI VCD over the 2.5° x 2.0° GEOS-Chem grid cell. To compare the estimated surface concentrations to the in situ surface monitoring data, we used a 40 km averaging radius around each station to increase the amount of usable data and further reduce the noise in the OMI data. This is similar to previous studies (i.e., Kharol et al., 2017) and maximizes both the slope and correlation compared to other radii, as shown in Fig. S2. Since this method does not require prior knowledge of in situ measurements, the analysis in Sect. 3.1 will be performed over the full dataset.

Since only simulations for January, April, July, and October 2015 were available to provide SVRs, there are two inherent assumptions regarding the temporal representativeness of the SVRs. The first assumption was using quasi-seasonal temporal sampling for the SVRs and resultant calculating the estimated surface concentrations. To test the impact of temporal representativeness on the estimated surface concentrations, we ran an additional GEOS-Chem simulation to cover all of spring (MAM) 2015. We also employed a full year of archived 2018 GEOS-CF data (NASA GMAO, 2023), which has improved temporal (hourly) and spatial (0.25° x 0.25°) resolution compared to GEOS-Chem and uses the same chemistry module, so they tend to produce similar results (Keller et al., 2021). We found that the intraseasonal variability in the SVR was 0.6 ppbv DU⁻¹ for MAM in both GEOS-Chem and GEOS-CF, as shown by Fig. S3. Therefore, we used the GEOS-CF data to estimate

Formatted: Not Highlight

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

Formatted: Not Highlight

200 this uncertainty for the entire year. We found that the average intraseasonal variability in the SVRs for the full year was 0.8
ppbv DU⁻¹ (Fig. S3). We also used the full year of GEOS-CF data to test the impact of temporal representativeness on the
annual mean surface SO₂ concentrations. Figures S4b-f show that there was no significant difference in the accuracy of the
annual average surface SO₂ concentrations among the different temporal sampling techniques ranging from daily to annual
mean SVRs. The slopes and correlations of the surface concentrations were consistent only ranging between 0.23 – 0.29 and
0.33 – 0.40, respectively (Fig. S4b-f). The sensitivity analysis with GEOS-CF also suggested that improving the spatial
resolution of the CTM while maintaining the same temporal sampling of the SVRs did not have a large impact on the accuracy
of the estimated surface concentrations despite an improvement in spatial resolution by nearly a factor a 10, as indicated by
205 Figs. S4a and S4d for 2018 data, as well as Fig. S5 for all years of the study period.

The other assumption was only using a single year of simulations to convert four years of OMI data into surface
concentrations. Kharol et al. (2017) did not have simulations that spanned their entire analysis period, but the implications of
this were never discussed. To address this, we first compared the monthly averaged SVRs from observations (calculated using
CNEMC surface concentrations and OMI VCDs) for each year in the study period to the 2015 GEOS-Chem simulations to
210 ensure there is no significant changes over time. Figure S6 shows boxplots of the observed and GEOS-Chem SVRs with the
percent difference between them. In general, the differences between the observed and GEOS-Chem SVRs were consistent
across all years of the study period, typically ranging from 73 – 89% (Fig. S6). We also ran additional GEOS-Chem simulations
for January, April, July, and October 2018 to assess if the simulated SVRs change over time. Boxplots for these two sets of
simulations can be seen in Fig. S7 and indicate that the GEOS-Chem SVRs only changed by 0.8 ppbv DU⁻¹, or 9%, from 2015
215 to 2018. The implications of these uncertainties on the resultant concentrations are discussed further in Sect. 2.6.

One simplification of our approach is to use January, April, July, and October simulations for a single year (2015) to
estimate the surface SO₂ concentrations over the entire study period. To evaluate this approach, we first compared the GEOS-
Chem and OMI SO₂ VCDs. We found that there was no significant change in the correlation between them from 2015-2018
(Fig. S1). This indicates the spatial distributions remained similar, and the model can distinguish between relatively polluted
220 and unpolluted areas, and thus, the SVRs in those environments. We also ran four additional GEOS-Chem simulations for
January, April, July, and October 2018 to assess the year-to-year changes in the SVR. The slopes in Fig. S2 indicate that the
monthly average SVR does not have a systematic change from 2015-2018 and has a maximum discrepancy of 13%. Since the
spatial distribution of SO₂ and simulated SVRs remained relatively constant over time, we believe this simplification made to
reduce computational expense will not have a significant impact on the results.

225 **2.5 Machine learning ML-based technique**

To estimate the surface SO₂ concentrations using a ML model, we used an eXtreme Gradient Boosting regression
model (XGBoost; Chen & Guestrin, 2016) to statistically relate satellite-based SO₂ VCDs, meteorological variables, and
emissions data to the in situ measurements. XGBoost models use a scalable tree boosting system to efficiently train an
ensemble of decision trees by adding a new tree with each training epoch and learning with each iteration (Chen & Guestrin,

Formatted: Subscript

Formatted: Subscript

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Superscript

Formatted: Subscript

230 2016; Friedman, 2001). Previous studies have showed that XGBoost and LightGBM models are able to estimate surface concentrations from satellite data more effectively than other ML architectures as shown by Kang et al. (2021) and Zhang et al. (2022). We trained ~~the our~~ XGBoost model with an ensemble of 500 trees, ~~with~~ a maximum tree depth of 15 splits, and a learning rate of 0.15 on a mean squared error loss function. ~~Neither Using a largern ensemble with more nor deeper trees did not improved the performance of the model, as shown by Fig. S8 and Fig. S9, respectively, and using a depth of 15 splits was found to be the best balance between overfitting and underfitting during training.~~

235 Our ML model was trained on a small number of variables (five) that each have known physical relationships to the spatial distribution or lifetime of atmospheric SO₂. By using a small number of variables, it is easier to derive physical meaning from the ML predictions without sacrificing accuracy since the input variables are already known to affect surface SO₂ concentrations. ~~First, w~~We used daily OMI SO₂ VCDs to estimate the spatial distribution of SO₂. ~~Next, we used hourly European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5; Hersbach et al., 2020; ECMWF, 2019) 100 m u-winds, 100 m v-winds, and boundary layer PBL heights (PBLHs) averaged over the OMI overpass window were used to from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5; Hersbach et al., 2020; ECMWF, 2019) to account for the meteorological mixing and dispersion of SO₂. Finally, we used and monthly SO₂ emissions from the CEDS inventory to capture the known locations of SO₂ sources. The ERA5 meteorological variables were provided at 0.25° x 0.25° horizontal resolution, and the CEDS emissions were provided at 0.5° x 0.5° horizontal resolution.~~ We trained the model on ~~logarithmic emissions since the values ranged several orders of magnitude, and~~ logarithmic boundary layer heights to get better sensitivity to variations in low boundary layers, ~~and logarithmic emissions since the values ranged several orders of magnitude.~~ The model ~~can be~~ summarized in Eqn. 3 as:

$$S_{ML} = XGBoost(\Omega_{OMI}, U_{ERA5}, V_{ERA5}, \log_{10}[PBLH_{ERA5}], \log_{10}[E_{CEDS}]), \quad (3)$$

250 where S_{ML} is the predicted surface concentrations from the XGBoost ML model, Ω_{OMI} is the satellite SO₂ VCD, U_{ERA5} is the u-wind, V_{ERA5} is the v-wind, PBLH_{ERA5} is the boundary layer height, and E_{CEDS} is the SO₂ emissions. ~~Earlier versions of the model were trained on 11 predictors, but the predicted surface concentrations produced an unrealistic spatial distribution of SO₂, as shown in Fig. S10. Additionally, some of the predictors were shown to be relatively unimportant to the model output, as indicated by the permutation importance in Fig. S11. The reduction of predictors from 11 down to five led to an improvement in the statistical performance and spatial distribution of the estimated surface concentrations, suggesting that utilizing known physical relationships between variables is more beneficial than the number of predictors in a ML model.~~

255 We trained the model on 90% of the daily data (N = 137630) from 2015 ~~to~~ 2018 with ~~meteorology ERA5 and emission CEDS variables predictors~~ sampled to match the valid OMI observations. The input variables were ~~sampled and averaged within 40 km of the CNEMC sites for training, as done in the GEOS-ChemCTM-based method, and the predicted surface concentrations from the XGBoost model are provided at each CNEMC site in the dataset.~~ The remaining 10% of the data (N = 15292) was reserved for a sample-based independent validation. ~~This split of the training and independent testing datasets, as was done used in by previous studies (e.g., Zhang et al., 2022; Yang et al., 2023a; Yang et al., 2023b), and was shown to have the best performance for the independent testing dataset for our model as shown in Fig. S12. each~~

Formatted: Subscript

CNEMCgridded the Figure 2 shows that the model had noticeably better performance with the training data (slope = 0.89; r = 0.95) compared to the testing data (slope = 0.67; r = 0.76), indicating that the model has good performance, but is slightly overfitting, a common artifact of complex machine learning models such as XGBoost. While the model was trained to estimate the surface SO₂ concentration at each CNEMC station, the trained model can then be used to make predictions on gridded input data to obtain estimates of the surface SO₂ concentrations on a continuous domain at the same horizontal resolution as the inputs.

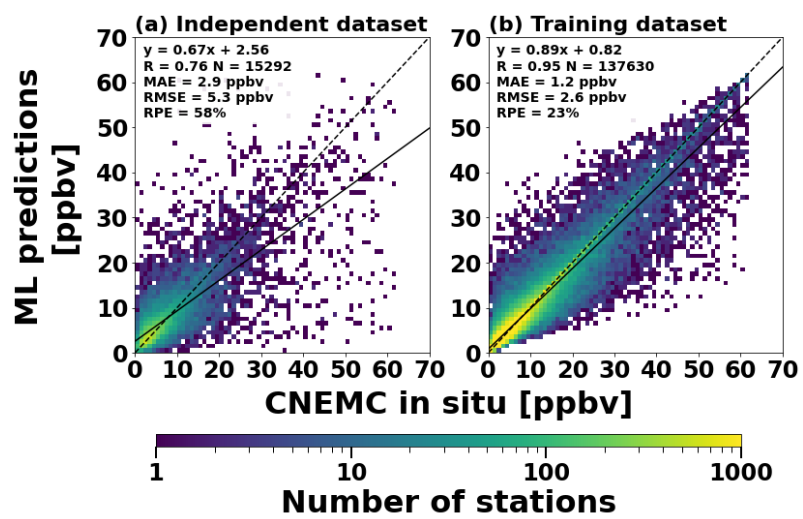


Figure 2: Scatterplots between the daily ML model predictions and CNEMC in situ measurements for the (a) independent dataset and (b) training dataset. Each panel includes a linear regression analysis with best fit line (solid line) and discrepancy statistics for the estimated surface SO₂ concentrations compared to in-situ measurements. Scatterplots are binned every 1 ppbv. Each scatterplot is colored by the number of stations in each bin and includes a linear regression analysis with the best fit line (solid line), best-fit equation, correlation coefficient, total number of stations, 1:1 line (black dashed line), MAE, RMSE, and RPE. The scatterplots are binned every 1 ppbv. The dashed line indicates the 1:1 line.

2.6 Methodological Uncertainties

This study provides the first detailed discussion of the individual sources and summation of uncertainties for either methodology. To estimate the uncertainty of the input variables for both methods, we performed moving-block bootstrapping with 10000 iterations on the daily gridded data. For each bootstrap, a horizontal coordinate and date was randomly sampled with replacement. For each random sample, a temporal block of five days in each direction from the randomly sampled day was used to calculate the standard deviation. After all bootstraps were completed, the uncertainty was defined as the average

of the standard deviations calculated from each iteration. This was done for the OMI SO₂ VCDs, GEOS-Chem SVRs, and ERA5 meteorology. The uncertainty in the CEDS emission inventory was not included due to the monthly temporal resolution, and a lack of uncertainty quantification in previous literature (e.g., Hoesly et al., 2018; McDuffie et al., 2020).

For the CTM-based method, the summation of error was determined using error propagation. For the error propagation, Eq. 1 was simplified such that:

$$\sigma_{\text{OMI}} = (\text{SVR}_{\text{GC}}) \times \rho_{\text{OMI}} \tag{4}$$

where SVR_{GC} is the monthly averaged SVR from the GEOS-Chem simulations. Equation 4 was used in the error propagation formula to obtain Eq. 5:

$$\sigma_{\text{OMI}} = \sqrt{\sigma_{\text{SVR}_{\text{GC}}}^2 (\rho_{\text{OMI}})^2 + \sigma_{\rho_{\text{OMI}}}^2 (\text{SVR}_{\text{GC}})^2} \tag{5}$$

where σ_{OMI} is the propagated error of the CTM-derived concentration, $\sigma_{\text{SVR}_{\text{GC}}}$ is the uncertainty in the GEOS-Chem SVR, and $\sigma_{\rho_{\text{OMI}}}$ is the uncertainty in the OMI SO₂ VCD. The uncertainty in the GEOS-Chem SVR was initially calculated with bootstrapping, but also needs to account for the uncertainties of the quasi-seasonal and single-year assumptions in the CTM-based methodology. The quasi-seasonal and single-year assumptions were defined and quantified in Section 2.4. These three sources of GEOS-Chem SVR uncertainty were assumed to be independent of each other and- were combined using the sum of the squares of each term. The results of the bootstrapping and error propagation are shown in Table 1. Ultimately, the methodological uncertainty of the CTM-based method is ±4.9 ppbv when considering the uncertainties of the OMI SO₂ VCDs (±0.67 ppbv DU⁻¹) and GEOS-Chem SVRs (±1.7 ppbv DU⁻¹). The OMI SO₂ VCD uncertainty has a relative standard deviation of 136%, which is comparable to the reported uncertainty of 60 – 120% for moderately polluted areas from Li et al. (2020).

Table 13: Sources and magnitudes of uncertainty for the CTM-based method. Uncertainties for the OMI SO₂ VCDs and GEOS-Chem SO₂ SVRs were determined using moving-block bootstrapping. The uncertainty for the quasi-seasonal SVR assumption was determined using GEOS-CF data. The single-year SVR assumption was determined using the 2015 and 2018 GEOS-Chem simulations. The overall uncertainty for the CTM-based method was determined using error propagation.

Variable	Uncertainty
OMI SO ₂ VCDs	± 0.67 DU
GEOS-Chem SO ₂ SVR	± 1.4 ppbv DU ⁻¹
Quasi-Seasonal SVR Assumption	± 0.8 ppbv DU ⁻¹
Single-Year SVR Assumption	± 0.6 ppbv DU ⁻¹
Overall Uncertainty	± 4.9 ppbv

It is much less straightforward to propagate error through a ML model since it effectively acts as a “black box,” so analytical error propagation methods cannot be used. First, uncertainties of the ERA5 meteorological fields were calculated using the moving-block bootstrapping approach. To obtain the overall uncertainty, we used traditional bootstrapping techniques to resample the training dataset with replacement and train an ensemble of XGBoost models to obtain an uncertainty

in the model output based on changes in the training data given to the model. To maintain consistency, the same independent testing dataset was used to make the model predictions for each bootstrap. The standard deviation was calculated for each station and day across the different models and was then averaged over space and time to obtain the overall uncertainty. The uncertainties of the ML inputs and overall uncertainty from the retraining analysis are shown in Table 2. For the ML-based method, the overall uncertainty was estimated to be ± 2.0 ppbv, which is lower than the propagated error for the CTM-based method. The overall uncertainty for the ML-based method does not directly account for uncertainty in the model inputs, but since traditional error propagation and summation of uncertainties are not possible for ML, this is our best estimate at how the training data can impact the predictions from the model.

Table 2: Sources and magnitudes of uncertainty for the ML-based method. Uncertainties for the OMI SO₂ VCDs and ERA5 meteorology were determined using moving-block bootstrapping. The overall uncertainty for the ML-based method was determined using bootstrapping on the training dataset and retraining multiple XGBoost models to estimate the uncertainty in the model training. The uncertainty for the CEDS inventory was not able to be quantified (NQ).

Variable	Uncertainty
OMI SO ₂ VCDs	± 0.67 DU
ERA5 U-Wind	± 1.9 m s ⁻¹
ERA5 V-Wind	± 1.9 m s ⁻¹
ERA5 Boundary Layer Height	± 326 m
CEDS Emissions	NQ
Overall Uncertainty	± 2.0 ppbv

Formatted: Subscript

Formatted Table

Formatted: Normal

2.7 Evaluation metrics

To quantify the discrepancies between the ~~estimated surface SO₂ concentrations from the CTM-based method, ML-based method, and the CNEMC~~ in situ ~~measurements, surface SO₂ concentrations and the estimates using the GEOS-ChemCTM and ML-based methods;~~ we used several different metrics ~~that were utilized in from~~ previous studies (e.g., Yang et al., 2023b; Zhang et al., 2021; Zhang et al., 2022) including the mean absolute error (MAE; Eq. 64), root mean squared error (RMSE; Eq. 75), and relative percent error (RPE; Eq. 86),

$$MAE = \frac{1}{N} \sum_i^N |S_{est,i} - S_{CNEMC,i}|, \tag{64}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (S_{est,i} - S_{CNEMC,i})^2}, \tag{75}$$

$$RPE = \frac{1}{N} \left(\sum_i^N \left| \frac{S_{est,i} - S_{CNEMC,i}}{S_{CNEMC,i}} \right| \right) \times 100\%, \tag{86}$$

where N is the number of stations, S_{est} is the estimated surface concentration from the ~~GEOS-ChemCTM~~ or ML-based method, and S_{CNEMC} is the ~~in situ~~ surface concentration ~~from the in situ measurements~~. Previous studies have also used slopes and

correlations from linear regression analyses between the estimated and in situ concentrations to assess the relative magnitudes and spatial distributions, respectively (e.g., Kharol et al., 2017; Lee et al., 2011; McLinden et al., 2014). In Sect. 3.1, results from the CTM-based method were validated and compared to previous studies using the full dataset since they are independent on in situ measurements. In Sect. 3.2, The GEOS-Chem and ML results from the ML-based method were validated and compared to previous studies using only the independent testing dataset, as well as to each other. Finally, in Sect. 4, The comparison between the two methods in our study both methods were directly compared were made using an the identical, independent testing dataset (i.e., retained from ML training) dataset such that the comparisons are made on an identical truth dataset for the first time.

3 Estimations of surface SO₂ cConcentrations from OMI satellite data

3.1 Evaluation of the ~~GEOS-Chem~~CTM-based method

Maps, histograms, and scatterplots of the annual mean surface SO₂ concentrations from the ~~GEOS-Chem~~CTM-based method and CNEMC in situ measurements are shown in Fig. 3. Both datasets have a similar spatial distribution with the highest concentrations in the North China Plain (Fig. 3), a highly industrialized region with many anthropogenic sources of SO₂ (Fig. 1b). The average correlation between the estimated and in situ concentrations is 0.58, indicating that the ~~GEOS-Chem~~CTM-based method can accurately distinguish between polluted and clean areas (Fig. 3). The ~~GEOS-Chem~~CTM-based method also captures a 45% decrease in the concentrations from 2015-2018, matching the change also seen in the data from the monitoring network (Fig. 3). ~~The decrease in SO₂ is due to the regulation of emissions, which has been previously reported in previous studies using satellite VCDs (Li et al., 2017; Wang et al., 2020a) and surface concentrations (Wei et al., 2023; Zhang et al., 2021).~~ Despite the similarities in the spatial distribution and temporal trends, the surface concentrations obtained from the ~~GEOS-Chem~~CTM-based method are significantly underestimated. The slope between the estimated and in situ concentrations is 0.24 with an RPE around 75% (Fig. 3). The discrepancy in the estimated concentrations is also apparent in the frequency distributions with a-peaks and mean values at lower concentrations around 1-3 ppbv and a smaller range compared to than compared to around 5-10 ppbv from the in situ measurements. The surface concentrations from the CTM-based method concentrations from the GEOS-Chem method and CNEMC measurements were also separated by season, and averaged from 2015-2018, from and validated against in situ measurements for the first time 2015-2018. As shown in Fig. S13, the ~~GEOS-Chem~~CTM-based method was able to accurately capture the spatial distribution (r = 0.56) and seasonality of the in situ measurements with higher concentrations in the winter and lower concentrations in the summer but still suffered from underestimation (slope = 0.24; RPE = 76%).

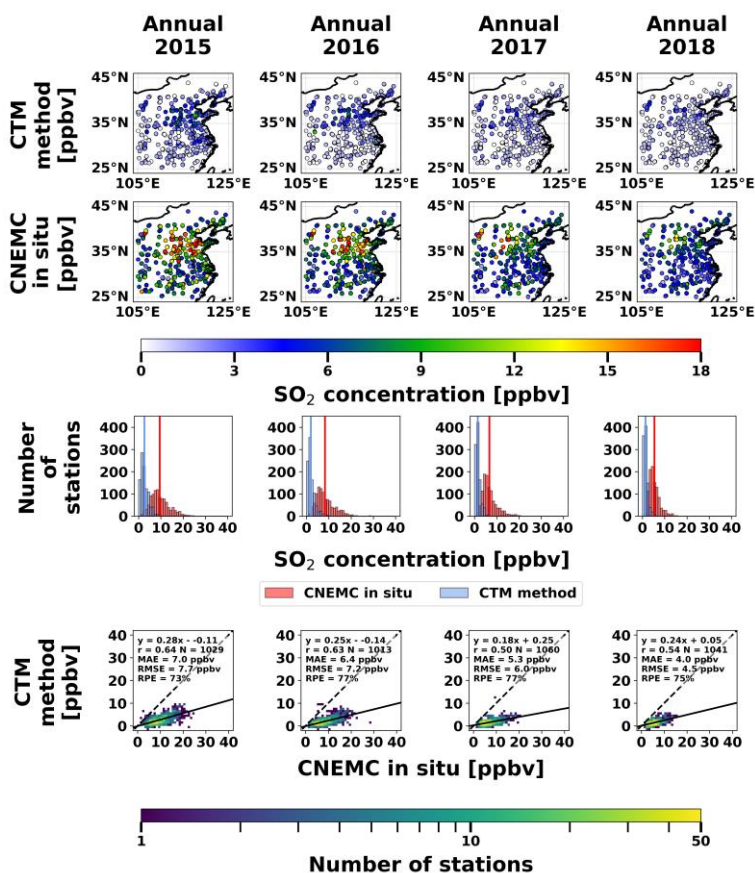


Figure 3: Spatial distributions of the annual average surface SO_2 concentrations from the CTM-based method (top row) and CNEMC in situ measurements (second row), histograms of the surface concentrations from each dataset with vertical bars representing the means (third row), and scatterplots between the two datasets (bottom row). Each column represents a different year in the study period. Histograms and scatterplots are binned every 1 ppbv. Each scatterplot is colored by the number of stations in each bin and includes a linear regression analysis with the best fit line (solid lines), best-fit equation, correlation coefficient, total number of stations, 1:1 line (black dashed line), MAE, RMSE, and RPE.

Table 34 summarizes the results from the validation of annual mean concentrations from our study compared to and previous studies using the CTM-based method. These previous studies were primarily focused on estimating annual mean

~~surface SO₂ concentrations using OMI VCDs and CTMs of varying resolution.~~ The studies by Lee et al. (2011) and McLinden et al. (2014) each utilized the OMI band residual difference (BRD) SO₂ product and used SVRs from coarse-~~resolution~~ and high-resolution CTMs, respectively. McLinden et al. (2014) outperformed Lee et al. (2011) with slopes of 0.88 and 0.79, respectively, and correlations of 0.91 and 0.81, respectively. Similarly, our study and Kharol et al. (2017) both use the OMI PCA SO₂ product and used SVRs from coarse-resolution and high-resolution CTMs, respectively. Our study had slightly worse performance than Kharol et al. (2017) with slopes of 0.24 and 0.39, respectively, and correlations of 0.58 and 0.61, respectively. These two sets of studies suggest that given the same OMI ~~dataprodu~~ct, the model resolution ~~may plays an important role in accurately estimating~~ affect the accuracy of the ~~estimated~~ surface concentrations compared to the in situ observations, assuming that the surface monitoring data are accurate. Our sensitivity tests comparing the impact of spatial resolution on the accuracy of the CTM-based method (Fig. S5) showed a discrepancy in the correlations of 0.05 between GEOS-Chem and GEOS-CF compared to a difference of 0.03 between our study and Kharol et al. (2017); however, the sensitivity tests only showed a discrepancy in the slopes of 0.02 between GEOS-Chem and GEOS-CF, which is much smaller than the 0.15 difference in slopes between our study and Kharol et al. (2017). This suggests that the difference in spatial resolution of our CTM simulations may account for the discrepancy in the correlations between our study and Kharol et al. (2017), but not the slopes, indicating that there may be another factor contributing to the underestimation of the CTM-based method. Additionally, ~~Pp~~previous studies have also shown that there are differences in SO₂ ~~VCDs as a result of across~~ different retrieval algorithms and sensors (Wang et al., 2020a). The higher slopes from the BRD product may be due to a high bias in the retrievals in polluted areas whereas the PCA product is thought to be more accurate (Li et al., 2013). Additionally, the slope of 0.75 from Kharol et al. (2017) and the results from Zhang et al. (2021) ~~are based on applying~~ use a scaling factor ~~onto~~ the in situ measurements to eliminate some of the bias, so these results are not directly comparable to our study.

Table 34: Comparison of study design (satellite data, model name and resolution, study location and study period) and performance metrics (mean absolute error, root mean square error, relative percent error, slope, and correlation) between our study and previous studies that utilized the CTM-based method for annual mean surface SO₂ concentrations. NR indicates that the value was not reported, and asterisks (*) indicate a scaling factor applied to the in situ surface concentrations.

Formatted: Subscript

Study	Satellite data	CTM (resolution)	Study location (time period)	MAE [ppbv]	RMSE [ppbv]	RPE [%]	Slope [-]	Correlation [-]
This Our study	OMI SO ₂ PCA	GEOS-Chem, (2.5° x 2.0°)	Eastern China (2015-2018)	5.7	6.3	74	0.23	0.58
Lee et al. (2011)	OMI SO ₂ BRD	GEOS-Chem, (2.5° x 2.0°)	North America (2006)	NR	NR	NR	0.79	0.81
McLinden et al. (2014)	OMI SO ₂ BRD	GEM-MACH, (15 km)	Canadian oil sands (2005-2011)	NR	NR	NR	0.88	0.91
Kharol et al. (2017)	OMI SO ₂ PCA	GEM-MACH, (15 km)	North America (2005-2015)	NR	NR	NR	0.39/0.75*	0.61
Zhang et al. (2021)	OMI SO ₂ PCA	MOZART, 1.9° x 2.5° Resolution	China (2014)	NR	3.9	19	0.83*	0.86

Inaccuracies in the CTM-based method can be partially attributed to noise in the satellite data. Individual VCD retrievals have ~~very~~ large uncertainties estimated from 60 – 120% from Li et al. (2020a) and 136% from our bootstrapping analysis (Table 1)~~(60-130%; Li et al., 2020a)~~, making it difficult to compare to the ground-based measurements on short

timescales. However, the noise in the data can decrease with temporal averaging by a factor of $n^{1/2}$ where n is the number of measurements being averaged (Krotkov et al., 2008). As a result, longer averaging periods (i.e., annual means) tend to have better performance than shorter timescales (i.e., seasonal means). The CTM resolution is also important for obtaining accurate surface concentrations. Coarse grid-cells may smooth-out SO_2 hotspots potentially resulting in an SVR that is too small. This may account for the consistent underestimation observed from this method and the relatively better performance in Kharol et al. (2017) with a higher resolution CTM. Additionally, the consistent underestimation of the CTM-based method may be a result of either underestimated SVRs from the GEOS-Chem simulations or a low bias in the OMI PCA product. Figure S6 suggests that the SVRs from GEOS-Chem are typically between 70-90% lower than observed SVRs calculated from CNEMC in situ surface concentrations and OMI VCDs, which is similar to the discrepancy in the surface concentrations from the CTM-based method compared with the in situ measurements. Based on the error propagation from Sect. 2.6, the combined methodological uncertainty of the CTM-based method considering both the OMI retrievals and GEOS-Chem SVRs is ± 4.9 ppbv (Table 1). This is larger than most of the estimated surface concentrations shown in the histograms from Fig. 3, indicating the methodological uncertainty associated with this method is large and is highly affected by the accuracy of the input data.

3.2 Evaluation of the machine learning ML-based method

The spatial distribution, frequency distribution, and validation scatterplots of the ML and CNEMC annual mean surface SO_2 concentrations from the independent testing dataset are shown in Fig. 4. The ML model estimated the surface concentrations more accurately than the GEOS-Chem CTM-based method. The with an improved average spatial correlation was of 0.77, and the ML predictions also matched the 45% decline from 2015 to 2018 observed from the CNEMC network. The most significant improvement compared to the GEOS-Chem method is the lower RPE of the ML method is much smaller at 33%, and the average slope of 0.69. These improvements indicate that the ML-based method has better accuracy in the spatial distributions and magnitudes compared to the CTM-based method indicating both less discrepancy and underestimation, respectively. The ML estimated concentrations also have a 45% decline from 2015-2018, which is the same as the CNEMC in situ measurements (Fig. 4). The shapes of the ML-based frequency distributions also agree well with the CNEMC observations with peaks at the same concentrations (5-10 ppbv) and similar ranges (Fig. 4). The ML-derived and in situ concentrations were also assessed using the seasonal concentrations averaged from 2015-2018. As shown in Fig. S14, the ML-based method was able to capture the spatial distribution ($r = 0.72$), seasonality, and magnitudes (slope = 0.64; RPE = 36%) of the seasonal mean surface concentrations on the seasonal data more accurately than the GEOS-Chem CTM-based method. Additionally, the overall uncertainty of the ML-based method is much lower than the CTM-based method at around ± 2 ppbv (Table 2). Since the ML predictions have a much larger magnitude, the methodological uncertainty for the ML-based method appears to be more reasonable compared to the CTM-based method.

Formatted: Font: Bold

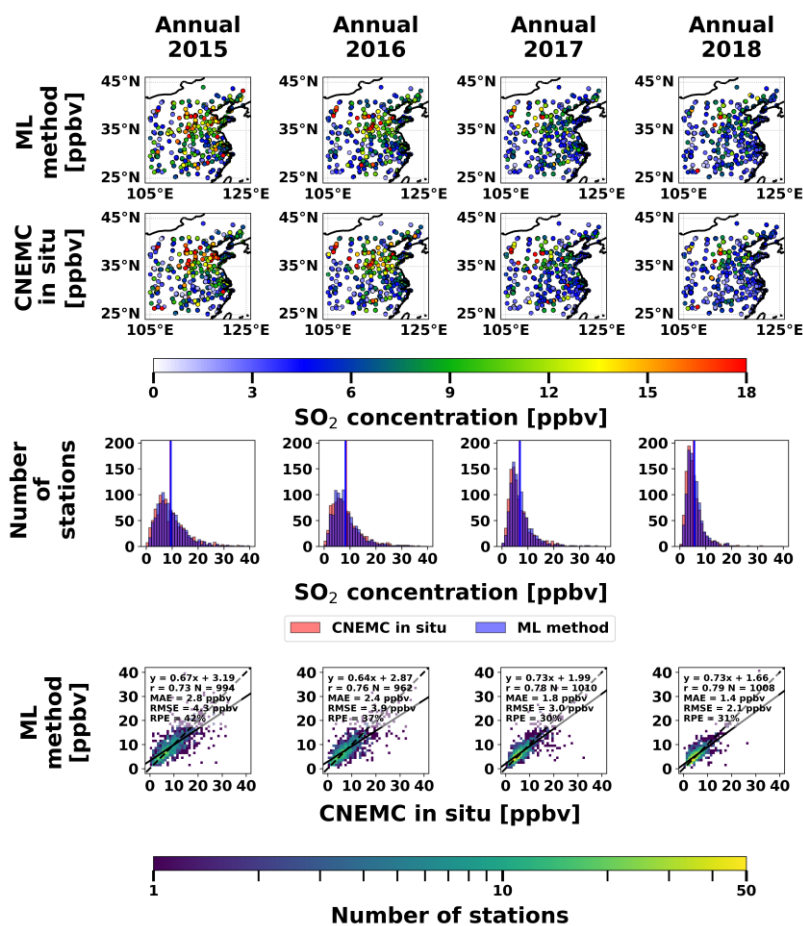


Figure 4: Spatial distributions of the annual average surface SO₂ concentrations from the ML-based method (top row) and CNEMC in situ measurements (second row), histograms of the surface concentrations from each dataset with vertical bars representing the means (third row), and scatterplots between the two datasets (bottom row). Each column represents a different year in the study period. Histograms and scatterplots are binned every 1 ppbv. Each scatterplot is colored by the number of stations in each bin and includes a linear regression analysis with the best fit line (solid lines), best-fit equation, correlation coefficient, total number of stations, 1:1 line (black dashed line), MAE, RMSE, and RPE.

445 Previous studies have shown that ML models can skillfully capture day-to-day variations in surface SO₂
concentrations in addition to the annual and seasonal means as summarized in Table 42 (e.g., Zhang et al., 2022; Yang et al.,
2023b). The estimated daily surface concentrations from our independent testing dataset had a slope of 0.67, correlation of
0.76, and RPE of 58% compared to the in situ measurements, indicating accuracy-good performance on short timescales (Fig.
2; Table 24). The performance of our model was comparable to previous studies but had a slightly larger discrepancy (Table
450 42). Our ML model only used five predictors compared to nine in Yang et al. (2023b) and 66 in Zhang et al. (2022), which
may partially account for the increased discrepancy. Additionally, our study did not use any spatial or temporal proxies, which
could also explain the slight reduction in performance compared to other studies that have used them.

Formatted: Indent: First line: 0"

Table 42: Comparison of study design (satellite data, machine learning model type and number of predictors, study location and study period) and performance metrics (mean absolute error, root mean square error, relative percent error, slope, and correlation) between our study and previous studies that utilized a ML-based method for daily surface SO₂ concentrations. NR indicates that the value was not reported.

Study	Satellite data	Machine learning model (number of predictors)	Study location (time period)	MAE [ppbv]	RMSE [ppbv]	RPE [%]	Slope [-]	Correlation [-]
This Our study	OMI SO ₂ PCA	XGBoost (5)	Eastern China (2015-2018)	3.0	5.2	59	0.67	0.75
Zhang et al. (2022)	OMI SO ₂ PCA	LightGBM (66)	Northern China (2013-2019)	NR	4.0	39	NR	0.94
Yang et al. (2023b)	Landsat-8 visible and infrared reflectance	Deep neural network multi-task learning (9)	China (2019)	3.5	5.7	47	0.76	0.85

We performed a permutation importance analysis to assess how each predictor impacted the model predictions. Figure 5a indicates that the PBLH-boundary layer heights and OMI SO₂ VCDs are the two most influential predictors followed by emissions and wind speeds. It is also worth noting that all of the predictors contribute toward the estimated surface concentrations with all permutation importance scores falling between 0.2 and 0.5 with none being unused. The boundary layer heights have a much smaller variation on short timescales compared to the OMI SO₂ VCDs. Based on the bootstrapped uncertainties from Table 2, the relative standard deviations are 20% for boundary layer heights, and 136% for OMI VCDs. As a result, the ML model is likely able to learn the relationship between boundary layer height and surface SO₂ concentrations

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

more easily than the OMI SO₂ VCDs. Scatterplots between each ML predictor variable and the ML estimated surface SO₂ concentrations with Spearman rank coefficients (r_s) are shown in Figs. 5b-f. The ML-derived SO₂ concentrations increase with larger SO₂ VCDs and emissions, as well as decrease with increasing PBLH-boundary layer heights and wind speeds (Figs. 5b-f). The surface concentrations and boundary layer heights also each have a strong, inverse seasonality, as shown in Fig. S14 and Fig. S15, respectively, so the strong temporal correlations between them also likely lead to a high permutation importance in the model. These trendsThe behavior of the ML predictions is-are consistent with the expected physical relationships between each variable-predictor and the surface SO₂ concentrations in the real atmosphere. Large OMI VCDs and emissions indicateing areas of high SO₂ loading, and elevated-large PBLHs-boundary layer heights and wind speeds lead to mixing and the dilution of SO₂. The magnitudes of the r_s values are small, indicating that the model may be making predictions based on the interactions between variables rather than any individual predictor. The small number of predictors used in our model allows us to link the ML predictions to known atmospheric processes, adding confidence to the model in its ability to accurately estimate the surface concentrations.

Formatted: Subscript

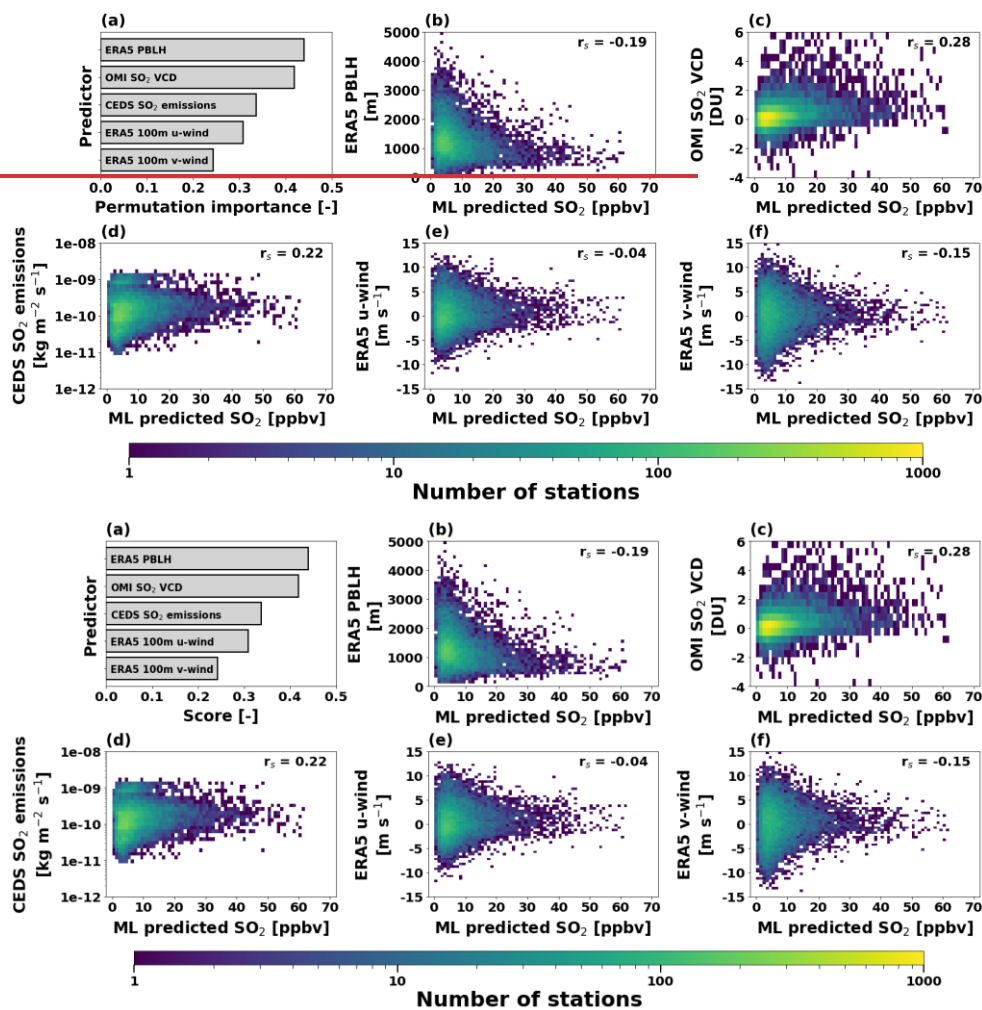


Figure 5: Evaluation of the daily ML-predicted surface concentrations using (a) permutation importance analysis, and scatterplots showing the ML predictor variables against the ML estimated surface SO₂ concentrations for (b) ERA5 PBLH, (c) OMI SO₂ VCDs, (d) CEDS SO₂ emissions, (e) ERA5 U-wind speeds, and (f) ERA5 V-wind speeds. Each scatterplot is colored by the number of stations in each bin and includes the Spearman rank coefficient (r_s).

4 Comparing results from the GEOS-Chem and machine learning methods Direct comparison of the CTM- and ML-based methods

The validation results from the GEOS-ChemCTM-based method in Sect. 3.1 were based on the full dataset since the methodology produces results that estimated surface concentrations are independent of the in situ monitoring data. However, the validation results from the ML-based method in Sect. 3.2 were only based on 10% of the data that was not used for training and reserved for an independent validation and not used for training. The comparison of these results using different datasets is still important but does not provide a direct comparison of their performance. Here, the GEOS-ChemCTM- and ML-based methods will be were directly compared using the same truth dataset over the same locations and study period for the first time, using Each method was resampled to match the independent testing dataset (i.e., data retained from ML training) was used to assess and the relative performance of each method was assessed given identically sampled data. First, each technique was will be validated at the CNEMC measurement sites in Sect. 4.1, similar to the analyses in Sect. 3. Then, both methods will be used to create gridded surface SO₂ concentrations in Sect. 4.2 to assess how effective both methods are for filling in the gaps of the CNEMC monitoring network, one of the main motivations for estimating surface concentrations from satellite data VCDs.

4.1 Performance on independent data

Scatterplots between the in situ concentrations and estimates of the surface concentrations from both the GEOS-ChemCTM- and ML-based methods for the identical testing dataset are shown in Fig. 6. The surface concentrations estimated by the ML model are much closer to the in situ measurements (i.e., the 1:1 line) than the GEOS-ChemCTM-based method, which is consistent with the previous results in Figs. 3-4. For the annual mean concentrations, the ML-based method had an average slope of 0.69 and correlation of 0.77, compared to values of 0.18 and 0.30, respectively, from the GEOS-ChemCTM-based method, respectively (Figs. 6a-d). The ML model also outperforms the GEOS-ChemCTM-based method on the seasonal data averaged over 2015-2018. The ML-based method had an average slope and correlation of 0.64 and 0.73, compared to 0.19 and 0.31, respectively, from the CTM-based method (Figs. 6e-h). The GEOS-ChemCTM-based method performed worse on this smaller dataset compared to the full dataset in Sect. 3.1 due to less temporal averaging, leading to larger discrepancies with the in situ measurements. As shown in Fig. S5, there is a smaller decrease in the performance of the ML-based method compared to the GEOS-ChemCTM-based method when assessing the performance for individual seasons rather than the 2015-2018 average for each season, as shown by Fig. S16. The slope and correlation for the ML-based method each decreased by around 0.1 to 0.59 and 0.67, compared to a decrease of 0.05 and 0.1 to 0.15 and 0.22, respectively, for the GEOS-ChemCTM-based method, respectively (Fig. S516). Despite the smaller amounts of data in the independent dataset and for individual seasons On both annual and seasonal timescales, the ML-based method still more accurately captured the spatial distribution and magnitudes of the surface SO₂ concentrations, indicating better consistency with the CNEMC measurements than the compared to the GEOS-ChemCTM-based method.

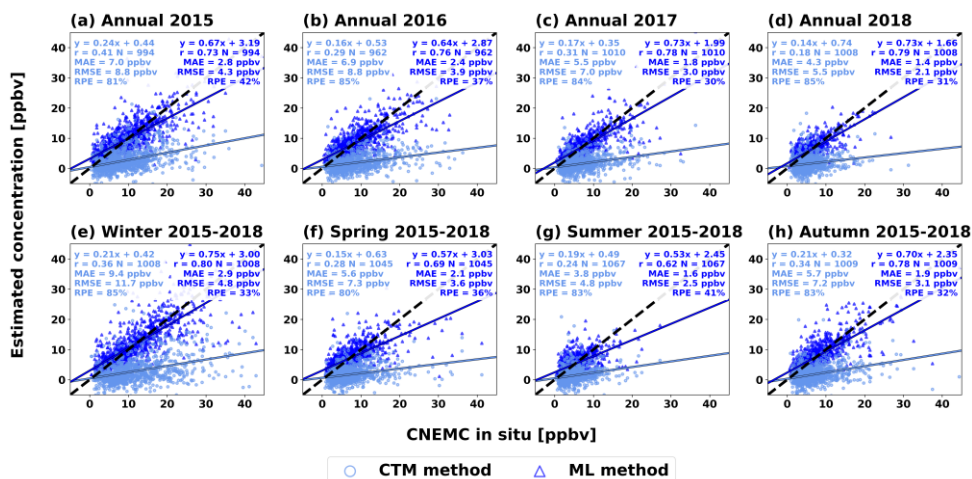


Figure 6: Scatterplots showing the estimated surface SO_2 concentrations from the GEOS-ChemCTM-based method (light blue squares) and ML-based method (dark blue triangles) against the in situ measurements from the independent dataset for (a-d) annual mean concentrations for each year in the study period, and (e-h) the 2015-2018 mean concentrations separated by season. Each scatterplot includes a linear regression analysis with the best fit line (solid lines), best-fit equation, correlation coefficient, total number of stations, 1:1 line (black dashed line), MAE, RMSE, and RPE.

Time series of the annual and seasonal mean surface SO_2 concentrations from the in situ measurements, and estimated concentrations from the GEOS-ChemCTM- and ML-based methods are shown in Figs. 7a-b. The ML estimated concentrations were much more-accurate closer to the CNEMC measurements than the GEOS-ChemCTM-based method compared to the CNEMC in situ concentrations. The overall mean ML concentrations had an average discrepancy of 5% with the in situ measurements, compared to a 58% discrepancy of 58% from the GEOS-ChemCTM-based method (Figs. 7a-b). The ML-based method also captured the same temporal variations as the in situ measurements each with a 44% decrease in concentrations from 2015-2018, and an average seasonal fluctuation by a factor of 1.9 between the winter and summer seasons (Figs. 7a-b). The GEOS-ChemCTM-based method also had good agreement in the temporal trends of the in situ measurements; but was not as good as the ML-based method with a 36% decrease from 2015-2018 and a seasonal fluctuation by a factor of 2.4 between the winter and summer, but not as good as the ML method (Figs. 7a-b). Since the CTM-based surface SO_2 concentrations were underestimated, the magnitude of the temporal trends is much smaller than the observations and ML-based method, but the relative change was similar, as shown by Table S1. The decrease in SO_2 from 2015-2018 detected by both methods is consistent with previous studies that showed a due to the regulation of reduction in emissions over China, which has been previously reported in previous studies using shown using satellite VCDs (Li et al., 2017; Wang et al., 2020a), satellite-derived emissions (Fioletov et al., 2023), and surface concentrations (Wei et al., 2023; Zhang et al., 2021). Despite the similarities in

Formatted: Subscript

the overall-year-to-year and season-to-season variations, the greatest difference between the time series of the two methods was the magnitude of the concentrations, as shown in Sect. 3.

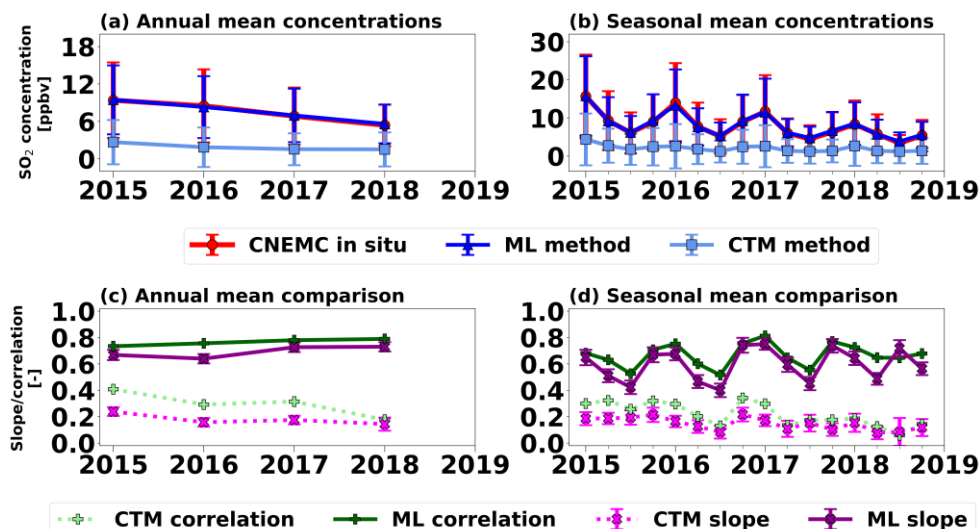


Figure 7: Time series of the surface SO₂ concentrations from the GEOS-Chem (CTM-based) method (light blue squares), ML-based method (dark blue triangles), and CNEMC in situ measurements (red circles) from the independent dataset as (a) annual and (b) seasonal means, as well as the slopes (pink x's) and correlations (green plus-sign crosses) from the (c) annual and (d) seasonal mean validations between the GEOS-Chem (CTM-based) method (dashed line) and ML-based method (solid line) with the in situ measurements. Error bars on the concentrations represent a 1 standard deviation uncertainty, and error bars on the slopes represent a 95% confidence interval based on the standard error of the linear regression fit. Data for all panels can be found in Table S1.

To assess how the accuracy of each method changed s over time, time series of the slopes and correlations from the individual annual and seasonal comparisons between the estimated and in situ surface concentrations (from Fig. S165 and Table S1) are shown in Figs. 7c-d. For the entire study period, the performance of the ML-based method was much more accurate than the GEOS-ChemCTM-based method as indicated by the higher slopes and correlations (Figs. 7c-d). Additionally, the GEOS-ChemCTM-based method suffered from a decrease in accuracy over time alongside declining SO₂ concentrations while the ML-based method remained stable from year to year (Figs. 7c-d). The accuracy of the CTM-based method is highly dependent on noise in the satellite data. Smaller datasets with less temporal averaging tend to have more noise, which leads to worse performance. Additionally, as SO₂ loading decreased s over China, it becomesbecame hardermore difficult for OMI to detect from the satellite, which may have introducing additional noise into the VCDsover time as SO₂ loading drops below

the detection limit of OMI. Comparatively, the ML-based method was more resistant to noise in the satellite data. As the SO₂ VCDs decreased, the ML predictions can become more reliant upon utilize other predictors such as meteorological and emissions predictors to estimate the surface concentrations, limiting the impact of the noisy satellite data (Figs. S176a-d). The accuracy of the ML-based method also had a distinct seasonality with generally better performance in the winter and worse in the summer (Fig. 7d). The boundary layer heights PBLH and OMI SO₂ VCDs were the dominant predictors in the winter, compared to whereas the CEDS emissions and PBLH boundary layer heights were the dominant predictors in the summer (Figs. S176e-h). The CEDS emissions are less consistent with the in situ measurements than the OMI VCDs with correlations r_c values of 0.2215 and 0.289, respectively (Fig. 5), which may account for the increased discrepancy of the ML-derived concentrations during the summer.

In summary, both methods the CTM- and ML-based methods captured the same similar temporal variations as the in situ measurements, but the ML-based method performed better was more accurate and had more stable performance over time compared to than the CTM-based method, which had decreasing performance over time, likely due to increased more noise in the satellite data from decreasing SO₂ loading.

4.2 Comparison of gridded products

Here, the GEOS-Chem CTM- and ML-based methods will be were used to create high-resolution gridded products of surface SO₂ concentrations at 0.25° x 0.25° horizontal resolution to assess how effective each technique is for filling in the gaps of the CNEMC air quality monitoring network. The gridded annual mean surface SO₂ concentrations from the CTM- and ML-based and GEOS-Chem methods are shown in Fig. 8 at 0.25° x 0.25° horizontal resolution. Both methods produced have similar spatial distributions to one another over land with the highest concentrations in the North China Plain and lower concentrations elsewhere. Over land, each method also has a spatial distribution similar to the retrieved SO₂ VCDs from OMI as shown in Fig. S187a-d, further indicating that both methods effectively utilizing the OMI data. Over the oceans, there is disagreement in the spatial distributions with the ML-based method producing high concentrations and the CTM-based method producing low concentrations. Since the ML predictions are significantly affected by boundary layer heights (Fig. 5), the model is most likely incorrectly associating the low marine boundary layer with areas elevated pollutants SO₂, as suggested by the seasonally averaged ERA5 boundary layer heights in Fig. S15, typical of low continental boundary layers, as shown in Fig. 8. Inaccuracies over the oceans have also been reported in Kang et al. (2021) where ML was used to estimate surface concentrations of NO₂ and ozone and were attributed to a lack of training data for the ML model in these locations. Since the ML model was only trained for conditions over land, it learned the relationship between high surface SO₂ concentrations and low continental boundary layer heights during the winter months, it is but was not able to make could not accurately apply this knowledge predictions over the oceans, or other areas that are different from where the model was trained. As a result, the CTM-based method may be more reliable for estimating the produce more reasonable spatial distributions of surface SO₂ concentrations in locations with a lack of surface observations where a ML model cannot be trained.

Formatted: Subscript

Formatted: Subscript

Formatted: Not Highlight

Formatted: Subscript

As shown in Fig. 8, both gridded products captured the decrease in annual mean concentrations from 2015 to 2018 observed at the CNEMC sites. Both methods were also able to capture the seasonal variations in their gridded products with the highest concentrations in the winter and lowest concentrations in the summer, as shown in Fig. S198. The seasonal gridded surface SO₂ concentrations were also still consistent with the OMI SO₂ VCDs (Fig. S187e-h). Although it is not possible to validate the gridded products, since the ML-based method had more accurate spatial distributions, temporal variations, and magnitudes than the CTM-based method when validated at the CNEMC sites, the gridded product is likely to be more accurate as well, but only over land. The unexpected area of elevated concentrations over the oceans exposed a major limitation of the ML-based method and suggests that future work in improving the CTM-based method may be worthwhile, especially for estimating surface SO₂ concentrations in locations where training data are not available.

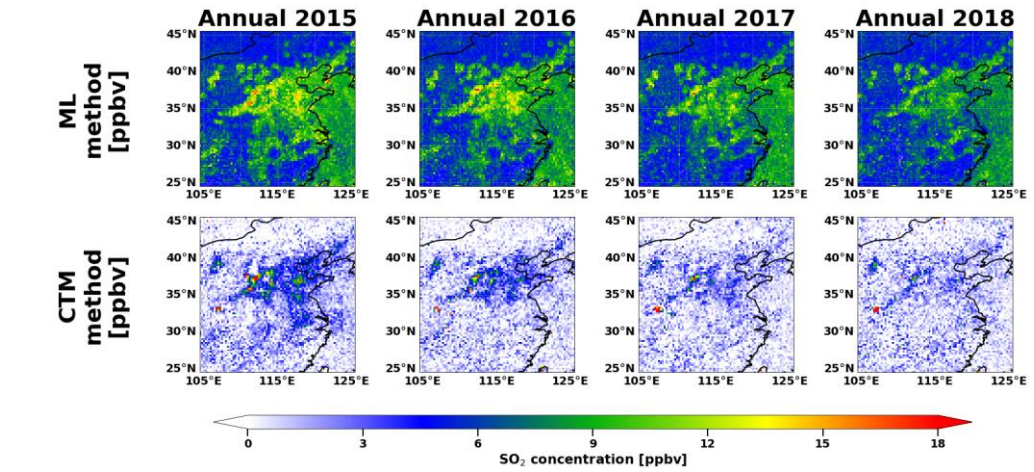


Figure 8: Maps of the annual mean surface SO₂ concentrations in ppbv from the ML-based method (top row) and GEOS-Chem CTM-based method (bottom row) over the study area at 0.25° x 0.25° horizontal resolution. Each column represents a different year of the study period (from left to right: 2015, 2016, 2017, and 2018).

5 Conclusion and discussion

In this study, we estimated surface SO₂ concentrations over eastern China from 2015-2018 using OMI satellite data SO₂ VCDs with two different methodologies. First, we used simulated SVRs from the GEOS-Chem model to convert the OMI SO₂ VCDs into surface concentrations using the CTM-based method. Then, we used an XGBoost model to statistically relate OMI SO₂ VCDs retrievals, ERA5 meteorology, and CEDS SO₂ emissions to in situ surface concentrations using the ML-based method. The novelty of this study includes a first time investigation of quantifying methodological

Formatted: Subscript

uncertainties for both the CTM- and ML-based techniques, a validation of seasonal mean surface concentrations from the CTM-based method, and a direct comparison between the two methods on the same truth dataset.

We found that the ML-based method ~~had better performance~~ was more accurate than the GEOS-ChemCTM-based method at estimating the surface concentrations when validated against in situ measurements from the CNEMC air quality monitoring network. The ML-based method had a discrepancy of ~30% with no significant bias (slope = 0.69), whereas the GEOS-ChemCTM-based method had a discrepancy of ~75% with a significant underestimation (slope = 0.24). Despite the underestimation, the GEOS-ChemCTM-based method also produced surface SO₂ concentrations that had similar spatial distributions (r = 0.58) and temporal patterns as the CNEMC in situ measurements, similar to previous studies. ~~using the CTM-based method. To obtain a good estimate of the spatial distribution,~~ The CTM-based method requires averaging data over ~~seasonal or annual~~ relatively long timescales to reduce the noise in the satellite retrievals and obtain more accurate estimates of the surface concentrations, and ~~the~~ The underestimation of this method is likely due to the coarse resolution of GEOS-Chem ~~smoothing out the SVR near SO₂ hotspots to a low bias in the simulated SVRs.~~ The CTM-based method also suffered from decreasing accuracy over time due to decreasing SO₂ loading over time since the retrieval that has a low signal-to-noise ratio. In addition to lower discrepancies, the ML-based method outperformed the CTM-based method in terms of the spatial distribution (r = 0.77) and temporal variations. The ~~accuracy~~ success of the ML-based method was especially apparent for smaller datasets that have limited temporal averaging and ~~thus higher noise in the noisier~~ OMI data since the model can rely on other predictors, which was also indicated by the stable accuracy over time. Even though our ML model was only based on five input variables, the results were similar to previous studies that used far more predictors. The small number of predictors also allowed us to relate the model predictions and input variables to known physical processes such as pollutant emissions and dispersion, thus lending more confidence in our ML model as compared with other “black box” ML models. Finally, both methods were used to create high-resolution gridded products to provide estimates of surface SO₂ concentrations in locations that do not have access to ground-based air quality monitoring measurements. This analysis exposed a major limitation in the ML-based method where it produced unrealistic spatial distributions of SO₂ over the ocean since it was only trained on data from over the land. Despite the underestimation of the CTM-based method, there is still value in using it to estimate surface SO₂ concentrations in locations where there is no training data available for developing ML-based techniques, but future additional steps should be taken to decrease the underestimation of this method.

In addition to using these estimated surface concentrations for filling in the gaps of air quality monitoring networks, the gridded products can be used to investigate other chemical processes in the atmosphere related to SO₂, such as estimating sulfuric acid concentrations and parameterizing aerosol nucleation and growth. New particle formation studies in China have shown that strong and frequent aerosol nucleation events occur in the presence of SO₂ in both heavily polluted urban (Dai et al., 2017; Wu et al., 2007) and relatively cleaner rural locations (Dai et al., 2017; Du et al., 2022). Estimating surface SO₂ concentrations using these satellite-based methods may be helpful for predicting the locations and intensities of new particle formation events or estimating sulfuric acid concentrations, especially for locations without in situ measurements or air quality monitoring sites.

Formatted: Subscript

In the future, ~~these performance of these methods should be applied to~~may be improved by higher-resolution satellite data, which may help to improve the results. OMI can only detect sources as small as 30 kt yr⁻¹, but newer instruments like the Tropospheric Monitoring Instrument (TROPOMI; Veefkind et al., 2012) or can detect sources as small as 8 kt yr⁻¹ (Fioletov et al., 2023). ~~Newer polar orbiting satellites like TROPOMI and~~Additionally, geostationary satellites like Tropospheric Emissions: Monitoring of Pollution (TEMPO; Zoogman et al., 2017) may offer future opportunities to estimate surface concentrations of air pollutants at even higher spatial and temporal resolution, ~~which may improve the accuracy of both methods. This also may improve the accuracy of both methods, especially if higher-resolution CTMs are also utilized.~~ Additionally, this study only focused on SO₂, but both methods can also be applied to other air pollutants such as NO₂, ozone, and particulate matter to see if the relative performance of each method is similar for other species. Since these two methods can utilize space-based measurements to fill in the gaps of ground-based air quality networks, investigating their relative performance as improvements are made to the satellite ~~retrievals~~data, CTMs, and ML models is critical for monitoring near-surface air pollution with high accuracy in locations where traditional observations are not possible.

Acknowledgements

The authors would like to thank Dr. Joanna Joiner from NASA Goddard Space Flight Center for their support and guidance and inspiration for pursuing this work.

Formatted: Normal

Code and data availability

All data used in this work are open source. The OMI PBL SO₂ VCDs are available at <https://doi.org/10.5067/Aura/OMI/DATA2023>, and the OMI emission catalogue is available at <https://so2.gsfc.nasa.gov/measures.html>. The GEOS-Chem source code is available at <https://github.com/geoschem/GCClassic>, and the GEOS-Chem input data, including the CEDS emission inventory, is available at <https://geos-chem.s3.amazonaws.com/index.html>. The ERA5 meteorology data are available at <https://nsf-ncar-era5.s3.amazonaws.com/index.html>. The CNEMC in situ measurements were obtained from <http://www.cnemc.cn>. The XGBoost model was developed using the scikit-learn (<https://scikit-learn.org/stable/>) and XGBoost (<https://xgboost.readthedocs.io/en/stable/>) Python packages. Finally, all maps were made with Natural Earth via the Cartopy Python package (<https://scitools.org.uk/cartopy>).

Supplement link

To be added.

675 **Author contributions**

ZW conducted the data analysis, prepared the paper, and created figures. ZW, CL, FL, and SL contributed to the development of the GEOS-Chem analysis in the paper. ZW, CL, SWF, and SL contributed to the development of the machine learning analysis in the paper. HZ and JW provided the CNEMC in situ data. All authors provided feedback and improvements to the paper.

680 **Competing interests**

The authors declare that they have no conflicts of interest.

Financial support

We acknowledge funding support from the National Science Foundation (Award numbers 2209772 and 2107916).

References

685 ~~Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res. Atmos.*, **106**, 23073–23095, doi:10.1029/2001JD000807, 2001.~~
~~Chen, T., and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 785–794, doi:10.1145/2939672.2939785, 2016.~~
690 China National Environmental Monitoring Centre (CNEMC): <http://www.cnemc.cn>, last access: 28 April 2024.
~~Dai, L., Wang, H., Zhou, L., An, J., Tang, L., Lu, C., Yan, W., Liu, R., Kong, S., Chen, M., Lee, S.-H., and Yu, H.: Regional and local new particle formation events observed in the Yangtze River Delta region, China, *J. Geophys. Res. Atmos.*, **122**, 2389–2402, doi:10.1002/2016JD026030, 2017.~~
~~Du, W., Cai, J., Zheng, F., Yan, C., Zhou, Y., Guo, Y., Chu, B., Yao, L., Heikkinen, L. M., Fan, X., Wang, Y., Cai, R., Hakala, S., Chan, T., Kontkanen, J., Tuovinen, S., Petäjä, T., Kangasluoma, J., Bianchi, F., Paasonen, P., Sun, Y., Kerminen, V.-M., Liu, Y., Daellenbach, K. R., Dada, L., and Kulmala, M.: Influnce of Aerosol Chemical Composition on Condensation Sink Efficiency and New Particle Formation in Beijing, *Environ. Sci. Technol. Lett.*, **9**, 375-382, doi:10.1021/acs.estlett.2c00159, 2022.~~
700 Engdahl, R. B.: A Critical Review of Regulations for the Control of Sulfur Oxide Emissions, *J. Air Pollut. Control Assoc.*, **23**, 364–375, doi:10.1080/00022470.1973.10469782, 1973.

European Centre for Medium-Range Weather Forecasts (ECMWF): ERA5 Reanalysis (0.25 Degree Latitude-Longitude Grid) (Updated monthly), Res. Data Arch. Natl. Cent. Atmos. Res., Comput. Inf. Syst. Lab., doi:10.5065/BH6N-5N20, 2019, last access: 2 Mar 2025.

Fan, K., Dhammapala, R., Harrington, K., Lamastro, R., Lamb, B., and Lee, Y.: Development of a Machine Learning Approach for Local-Scale Ozone Forecasting: Application to Kennewick, WA, Front. Big Data, 5, 781309, doi:10.3389/fdata.2022.781309, 2022.

Fioletov, V., McLinden, C. A., Griffin, D., Abboud, I., Krotkov, N., Leonard, P. J. T., Li, C., Joiner, J., Theys, N., and Carn, S.: Multi-Satellite Air Quality Sulfur Dioxide (SO₂) Database Long-Term L4 Global V2, Goddard Earth Sci. Data Inf. Serv. Cent. (GES DISC), doi:10.5067/MEASURES/SO2/DATA406, accessed: 29 October 2024.

Fioletov, V. E., McLinden, C. A., Griffin, D., Abboud, I., Krotkov, N., Leonard, P. J. T., Li, C., Joiner, J., Theys, N., and Carn, S.: Version 2 of the global catalogue of large anthropogenic and volcanic SO₂ sources and emissions derived from satellite measurements, Earth Syst. Sci. Data, 15, 75–93, doi:10.5194/essd-15-75-2023, 2023.
Flyvbjerg, H. and Petersen, H. G.: Error estimates on averages of correlated data, J. Chem. Phys., 91, 461, doi:10.1063/1.457480, 1989.

Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, Ann. Stat., 29, 1189–1232, <http://www.jstor.org/stable/2699986>, 2001.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, doi:10.1002/qj.3803, 2020.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, doi:10.5194/gmd-11-369-2018, 2018.

The International GEOS-Chem User Community.: geoschem/GCClassic: GCClassic 14.2.2 (14.2.2), Zenodo [code], doi:10.5281/zenodo.10034814, 2023.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.-K., and Kim, S.: Estimation of surface-level NO₂ and O₃ concentrations using TROPOMI data and machine learning over East Asia, Environ. Pollut., 288, 117711, doi:10.1016/j.envpol.2021.117711, 2021.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S.: Description of the NASA

Formatted: English (United Kingdom)

735 [GEOS Composition Forecast Modeling System GEOS-CF v1.0, Journal of Advances in Earth Modeling Systems, 13, e2020MS002413, doi:10.1029/2020MS002413, 2021.](#)

Kerminen, V.-M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M., and Bianchi, F.: Atmospheric new particle formation and growth: review of field observations, Environ. Res. Lett., 13, 103003, doi:10.1088/1748-9326/aadf3c, 2018.

740 [Kharol S. K., Martin, R. V., Philip, S., Boys, B., Lamsal, L. N., Jerrett, M., Brauer, M., Crouse, D. L., McLinden C., and Burnett, R. T.: Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America, Atmospheric Environment, 118, 236-245, doi:10.1016/j.atmosenv.2015.08.011, 2015,](#)

Kharol, S. K., McLinden, C. A., Sioris, C. E., Shephard, M. W., Fioletov, V., van Donkelaar, A., Philip, S., and Martin, R. V.: OMI satellite observations of decadal changes in ground-level sulfur dioxide over North America, Atmos. Chem. Phys., 17, 5921–5929, doi:10.5194/acp-17-5921-2017, 2017.

Krotkov, N. A., McClure, B., Dickerson, R. R., Carn, S. A., Li, C., Bhartia, P. K., Yang, K., Krueger, A. J., Li, Z., Levelt, P. F., Chen, H., Wang, P., and Lu, D.: Validation of SO₂ retrievals from the Ozone Monitoring Instrument over NE China, J. Geophys. Res. Atmos., 113, D16S40, doi:10.1029/2007JD008818, 2008.

745 Krotkov, N. A., Li, C., and Leonard, P.: OMI/Aura Sulphur Dioxide (SO₂) Total Column Daily L2 Global Gridded 0.125 degree x 0.125 degree V3, Goddard Earth Sci. Data Inf. Serv. Cent. (GES DISC), doi:10.5067/Aura/OMI/DATA2023, accessed: 03/10/2024, 2014.

750 Krzyzanowski, M., and Wojtyniak, B.: Ten-Year Mortality in a Sample of an Adult Population in Relation to Air Pollution, J. Epidemiol. Community Health, 36, 262–268, <http://www.jstor.org/stable/25566349>, 1992.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., Dunlea, E. J., and Pinto, J. P.: Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument, J. Geophys. Res., 113, D16308, doi:10.1029/2007JD009235, 2008.

755 Lee, C., Martin, R. V., van Donkelaar, A., Lee, H., Dickerson, R. R., Hains, J. C., Krotkov, N., Richter, A., Vinnikov, K., and Schwab, J. J.: SO₂ emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, J. Geophys. Res., 116, D06304, doi:10.1029/2010JD014758, 2011.

Lee, S.-H., Gordon, H., Yu, H., Lehtipalo, K., Haley, R., Li, Y., and Zhang, R.: New particle formation in the atmosphere: From molecular clusters to global climate, J. Geophys. Res. Atmos., 124, doi:10.1029/2018JD029356, 2019.

760 Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mäkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE Trans. Geosci. Remote Sens., 44, 1093–1101, doi:10.1109/TGRS.2006.872333, 2006.

Li, C., Joiner, J., Krotkov, N. A., and Bhartia, P. K.: A fast and sensitive new satellite SO₂ retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument, Geophys. Res. Lett., 40, 6314–6318, doi:10.1002/2013GL058134, 2013.

765

Formatted: English (United Kingdom)

Li, C., Stehr, J. W., Marufu, L. T., Li, Z., and Dickerson, R. R.: Aircraft measurements of SO₂ and aerosols over northeastern China: Vertical profiles and the influence of weather on air quality, *Atmospheric Environment*, 62, 492-501, doi:10.1016/j.atmosenv.2012.07.076, 2012.

Formatted: English (United Kingdom)

Li, C., McLinden, C., Fioletov, V., Krotkov, N., Carn, S., Joiner, J., Streets, D., He, H., Ren, X., Li, Z., and Dickerson, R. R.: India Is Overtaking China as the World's Largest Emitter of Anthropogenic Sulfur Dioxide, *Sci. Rep.*, 7, 14304, doi:10.1038/s41598-017-14639-8, 2017.

Li, C., Krotkov, N. A., Leonard, P. J. T., Carn, S., Joiner, J., Spurr, R. J. D., and Vasilkov, A.: Version 2 Ozone Monitoring Instrument SO₂ product (OMSO₂ V2): New anthropogenic SO₂ vertical column density dataset, *Atmos. Meas. Tech.*, 13, 6175–6191, doi:10.5194/amt-13-6175-2020, 2020a.

Li, C., Krotkov, N. A., Leonard, P., and Joiner, J.: OMI/Aura Sulphur Dioxide (SO₂) Total Column 1-orbit L2 Swath 13x24 km V003, Goddard Earth Sci. Data Inf. Serv. Cent. (GES DISC), doi:10.5067/Aura/OMI/DATA2022, accessed: 29 October 2024, 2020b.

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., and Sarnat, J. A.: Mapping annual mean ground-level PM_{2.5} concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States, *J. Geophys. Res.*, 109, D22206, doi:10.1029/2004JD005025, 2004.

Lucchini, R.: File Specification for GEOS FP, GMAO Office Note No. 4, Version 1.2, 61 pp., available at http://gmao.gsfc.nasa.gov/pubs/office_notes, 2018.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: 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, doi:10.5194/essd-12-3413-2020, 2020.

Formatted: English (United Kingdom)

McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., Makar, P. A., Martin, R. V., Veefkind, J. P., and Yang, K.: Improved satellite retrievals of NO₂ and SO₂ over the Canadian oil sands and comparisons with surface measurements, *Atmos. Chem. Phys.*, 14, 3637–3656, doi:10.5194/acp-14-3637-2014, 2014.

National Academy of Sciences, Engineering, and Medicine (NASEM): The Future of Atmospheric Chemistry Research: Remembering Yesterday, Understanding Today, Anticipating Tomorrow, Washington, DC, The National Academies Press, doi:10.17226/23573, 2016.

National Aeronautics and Space Administration (NASA): README Document for OMSO₂: Aura/OMI Sulfur Dioxide Level 2 Product, Goddard Earth Sciences Data and Information Services Center (GES DISC), available at https://aura.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level2/OMSO2.003/doc/OMSO2Readme_V2.pdf, 2020.

National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA GMAO): GEOS Composition Forecast (GEOS-CF), NASA Center for Climate Simulation, NASA Goddard Space Flight Center [Dataset], available at <https://portal.nccs.nasa.gov/datashare/gmao/geos-cf/v1/ana/>.

Formatted: English (United Kingdom)

800 [Norman, O. G., Heald, C. L., Bililign, S., Campuzano-Jost, P., Coe, H., Fiddler, M. N., Green, J. R., Jimenez, J. L., Kaiser, K., Liao, J., Middlebrook, A. M., Nault, B. A., Nowak, J. B., Schneider, J., and Welti, A.: Exploring the processes controlling secondary inorganic aerosol: evaluating the global GEOS-Chem simulation using a suite of aircraft campaigns, *Atmos. Chem. Phys.*, **25**, 771-795, doi:10.5194/acp-25-771-2015, 2025.](#)

Nowlan, C. R., Liu, X., Chance, K., Cai, Z., Kurosu, T. P., Lee, C., and Martin, R. V.: Retrievals of sulfur dioxide from the Global Ozone Monitoring Experiment 2 (GOME-2) using an optimal estimation approach: Algorithm and initial validation, *J. Geophys. Res.*, **116**, D18301, doi:10.1029/2011JD015808, 2011.

805 Philip, S., Martin, R. V., and Keller, C. A.: Sensitivity of chemistry-transport model simulations to the duration of chemical and transport operators: A case study with GEOS-Chem v10-01, *Geosci. Model Dev.*, **9**, 1683–1695, doi:10.5194/gmd-9-1683-2016, 2016.

Seinfeld, J. H., and Pandis, S. N. (3rd Ed.): *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Wiley, Hoboken, New Jersey, United States, 874-876, ISBN 9781118947401, 2016.

810 [Shan, Y., Zhu, Y., Sui, H., Zhao, N., Li, H., Wen, L., Chen, T., Qi, Y., Qi, W., Wang, X., Zhang, Y., Xue, L., and Wang, W.: Vertical distribution and regional transport of air pollution over Northeast China: Insights from an intensive aircraft study, *Atmospheric Environment*, **358**, 121327, doi:10.1016/j.atmosenv.2025.121327, 2025.](#)

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., 815 van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: 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, doi:10.1016/j.rse.2011.09.027, 2012.

Theys, N., De Smelt, I., van Gent, J., Danckaert, T., Wang, T., Hendrick, F., Stavrakou, T., Bauduin, S., Clarisse, L., Li, C., 820 Krotkov, N., Yu, H., Brenot, H., and Van Roozendael, M.: Sulfur dioxide vertical column DOAS retrievals from the Ozone Monitoring Instrument: Global observations and comparison to ground-based and satellite data, *J. Geophys. Res. Atmos.*, **120**, 2470–2491, doi:10.1002/2014JD022657, 2015.

Wang, Y. and Wang, J.: Tropospheric SO₂ and NO₂ in 2012-2018: Contrasting views of two sensors (OMI and OMPS) from space, *Atmospheric Environment*, **223**, 117214, doi:10.1016/j.atmosenv.2019.117214, 2020a.

825 Wang, Y., Wang, J., Xu, X., Henze, D. K., Qu, Z., and Yang, K.: Inverse modeling of SO₂ and NO_x emissions over China using multisensory satellite data – Part 1: Formulation and sensitivity analysis, *Atmos. Chem. Phys.*, **20**, 6631-6650, doi:10.5194/acp-20-6631-2020, 2020b

Wang, Y., Wang, J., Zhou, M., Henze, D. K., Ge, C., and Wang, W.: Inverse modeling of SO₂ and NO_x emissions over China using multisensory satellite data – Part 2: Downscaling techniques for air quality analysis and forecasts, *Atmos. Chem. Phys.*, 830 **20**, 6651–6670, doi:10.5194/acp-20-6651-2020, 2020c.

Formatted: English (United Kingdom)

Formatted: English (United Kingdom)

Wei, J., Li, Z., Wang, J., Li, C., Gupta, P., and Cribb, M.: Ground-level gaseous pollutants (NO₂, SO₂, and CO) in China: Daily seamless mapping and spatiotemporal variations, *Atmos. Chem. Phys.*, 23, 1511–1532, doi:10.5194/acp-23-1511-2023, 2023.

835 [Wu, Z., Hu, M., Liu, S., Wehner, B., Bauer, S., Bling, A. M., Wiedensohler, A., Petäjä, T., Maso, M. D., and Kulmala, M.: New particle formation in Beijing, China: Statistical analysis of a 1-year data set, *J. Geophys. Res.*, 112, D09209, doi: 10.1029/2006JD007406, 2007.](#)

[Xue, L., Ding, A., Gao, J., Wang, T., Wang, W., Wang, X., Lei, H., Jin, D., and Qi, Y.: Aircraft measurements of the vertical distribution of sulfur dioxide and aerosol scattering coefficient in China, *Atmospheric Environment*, 44, 278-282, doi:10.1016/j.atmosenv.2009.10.026, 2010.](#)

840 Yang, Q., Kim, J., Cho, Y., Lee, W.-J., Lee, D.-W., Yuan, Q., Wang, F., Zhou, C., Zhang, X., Xiao, X., Guo, M., Guo, Y., Carmichael, G. R., and Gao, M.: A synchronized estimation of hourly surface concentrations of six criteria air pollutants with GEMS data, *npj Clim. Atmos. Sci.*, 6, 94, doi:10.1038/s41612-023-00407-1, 2023a.

Yang, Q., Yuan, Q., Gao, M., and Li, T.: A new perspective to satellite-based retrieval of ground-level air pollution: Simultaneous estimation of multiple pollutants based on physics-informed multi-task learning, *Sci. Total Environ.*, 857, 159542, doi:10.1016/j.scitotenv.2022.159542, 2023b.

845 Zhang, S., Mi, T., Wu, Q., Luo, Y., Grieneisen, M. L., Shi, G., Yang, F., and Zhan, Y.: A data-augmentation approach to deriving long-term surface SO₂ across Northern China: Implications for interpretable machine learning, *Sci. Total Environ.*, 827, 154278, doi:10.1016/j.scitotenv.2022.154278, 2022.

Zhang, X., Wang, Z., Cheng, M., Wu, X., Zhan, N., and Xu, J.: Long-term ambient SO₂ concentration and its exposure risk across China inferred from OMI observations from 2005 to 2018, *Atmos. Res.*, 247, 105150, doi:10.1016/j.atmosres.2020.105150, 2021.

850 Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., Hilton, B. B., Nicks, D. K., Newchurch, M. J., Carr, J. L., Janz, S. J., Andraschko, M. R., Arola, A., Baker, B. D., Canova, B. P., Chan Miller, C., Cohen, R. C., Davis, J. E., Dussault, M. E., Edwards, D. P., Fishman, J., Ghulam, A., González Abad, G., Grutter, M., Herman, J. R., Houck, J., Jacob, D. J., Joiner, J., Kerridge, B. J., Kim, J., Krotkov, N. A., Lamsal, L., Li, C., Lindfors, A., Martin, R. V., McElroy, C. T., McLinden, C., Natraj, V., Neil, D. O., Nowlan, C. R., O’Sullivan, E. J., Palmer, P. I., Pierce, R. B., Pippin, M. R., Saiz-Lopez, A., Spurr, R. J. D., Szykman, J. J., Torres, O., Veefkind, J. P., Veihelmann, B., Wang, H., Wang, J., and Chance, K.: Tropospheric emissions: Monitoring of pollution (TEMPO), *J. Quant. Spectrosc. Radiat. Transf.*, 186, 17–39, doi:10.1016/j.jqsrt.2016.05.008, 2017.

Formatted: English (United Kingdom)