

Article

Not peer-reviewed version

Surface Ozone Estimation over the Beijing–Tianjin–Hebei Region Using EMI-II Total Ozone Observations and Machine Learning Integration

Hua Cheng , [Jian Chen](#) ^{*} , Zhiyi Zhang , Yihui Huang , [Keke Zhu](#)

Posted Date: 6 March 2026

doi: 10.20944/preprints202603.0441.v1

Keywords: Gaofen-5; Environmental Trace Gases Monitoring Instrument 2 (EMI-II); total ozone column; surface ozone; machine learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Surface Ozone Estimation over the Beijing–Tianjin–Hebei Region Using EMI-II Total Ozone Observations and Machine Learning Integration

Hua Cheng, Jian Chen *, Zhiyi Zhang, Yihui Huang and Keke Zhu

School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China

* Correspondence: chjnju@163.com

Highlights

What are the main findings?

- EMI-II total ozone observations contain independent and usable information related to surface ozone variability, as demonstrated by machine learning models that integrate multi-source predictors.
- The applicability of EMI-II for surface ozone estimation is strongly regime-dependent: the information can be effectively exploited under strong photochemical activity and deep boundary-layer mixing, while reduced skill occurs under weak radiation and stable boundary-layer conditions.

What are the implications of the main findings?

- This study clarifies the potential role of China's UV–VIS satellite measurements in regional air quality analysis, demonstrating that EMI-II products can be systematically integrated into surface pollution studies despite intrinsic vertical sensitivity limitations.
- A transferable methodological framework is established for incorporating validated national satellite products into multi-source surface ozone estimation, with applicability to other regions and atmospheric composition sensors.

Abstract

Surface ozone monitoring remains challenging due to sparse ground networks and limited satellite boundary-layer sensitivity. This study evaluates, for the first time, China's Environmental Trace Gases Monitoring Instrument II (EMI-II) for estimating surface ozone over the Beijing–Tianjin–Hebei (BTH) region. EMI-II total ozone columns (TOCs) are retrieved using the differential optical absorption spectroscopy (DOAS) algorithm and validated against the TROPOspheric Monitoring Instrument (TROPOMI) ($R = 0.96$), Geostationary Environment Monitoring Spectrometer (GEMS) ($R = 0.97$), and the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) ground measurements ($R > 0.92$, bias $< 4\%$). TOCs are then combined with ERA5 meteorology, satellite NO_2/HCHO , and surface observations within machine learning models, achieving cross-validated R^2 of 0.94 and RMSE of $12.29 \mu\text{g}/\text{m}^3$ for surface ozone estimation. EMI-II estimates show strong agreement with independent observations ($R = 0.97$, RMSE = $10.05 \mu\text{g}/\text{m}^3$) and reproduce seasonal gradients, with summer concentrations ($130 \mu\text{g}/\text{m}^3$) more than double winter levels ($59 \mu\text{g}/\text{m}^3$). Estimation skill is regime-dependent: performance comparable to TROPOMI occurs under strong photochemical activity, while reduced sensitivity occurs under weak radiation and stable boundary layers—consistent with averaging kernel diagnostics. This first comprehensive validation demonstrates that EMI-II, despite vertical sensitivity limitations, provides meaningful surface ozone

constraints under favorable atmospheric conditions. The framework is transferable to other regions and sensors, supporting broader applications of national satellite assets in air pollution monitoring.

Keywords: Gaofen-5; Environmental Trace Gases Monitoring Instrument 2 (EMI-II); total ozone column; surface ozone; machine learning

1. Introduction

Surface ozone is a secondary photochemical pollutant formed through complex reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs) [1]. In China, ozone pollution has intensified during recent decades and has become a dominant warm-season air quality issue in many urban and industrial regions. Elevated ozone concentrations are associated with adverse health outcomes and ecosystem stress, including reduced photosynthetic activity and crop productivity [2,3]. Accurate characterization of surface ozone variability is therefore important for air quality assessment and for understanding regional atmospheric chemistry dynamics.

Ground-based monitoring networks provide high-accuracy observations but remain spatially uneven and limited in regional representativeness [4]. Satellite remote sensing offers broader spatial coverage and long-term continuity; however, direct retrieval of surface ozone from nadir-viewing ultraviolet–visible (UV–VIS) instruments remains fundamentally challenging. Total ozone columns (TOCs) are dominated by stratospheric contributions, and the vertical sensitivity of UV–VIS sensors within the boundary layer is limited. As a result, most satellite-based surface ozone studies rely on indirect approaches that combine column observations with meteorological variables, chemical proxies, and statistical or machine learning models to estimate surface concentrations.

Over the past two decades, instruments such as Global Ozone Monitoring Experiment (GOME) [5], Ozone Monitoring Instrument (OMI) [6], TROPOspheric Monitoring Instrument (TROPOMI) [7], and Geostationary Environment Monitoring Spectrometer (GEMS) [8] have significantly advanced satellite monitoring of atmospheric ozone. These missions have enabled analyses of long-term trends, photochemical regimes, and regional transport processes. Nevertheless, most surface ozone estimation studies have relied on international satellite datasets, and the capability of China's national atmospheric composition sensors for surface air quality applications remains insufficiently evaluated.

The Environmental Trace Gases Monitoring Instrument II (EMI-II), launched onboard the Gaofen-5B satellite in 2021, is China's new-generation UV–VIS spectrometer with improved spectral performance and a spatial resolution of approximately 13 km × 24 km [9,10]. Recent studies have demonstrated the reliability of EMI-II for retrieving ozone and other trace-gas columns. However, it remains unclear to what extent EMI-II total ozone observations contain independent and usable information related to surface ozone variability, and under which atmospheric conditions this information can be effectively exploited.

Addressing this question requires both rigorous retrieval validation and quantitative evaluation of information transfer from column measurements to surface concentrations. Satellite-derived TOCs reflect integrated atmospheric loading and photochemical accumulation. When combined with meteorological drivers and photochemical regime indicators (e.g., NO₂ and HCHO), TOCs may provide indirect but physically meaningful constraints on surface ozone variability. Machine learning approaches offer a flexible framework to capture nonlinear relationships among multi-source predictors and have shown strong performance in previous OMI and TROPOMI-based studies. However, their applicability to EMI-II products has not yet been systematically assessed [11,12].

In this study, we conduct a comprehensive evaluation of EMI-II for surface ozone estimation over the Beijing–Tianjin–Hebei (BTH) region. First, EMI-II TOCs are retrieved using the differential optical absorption spectroscopy (DOAS) algorithm coupled with SCIATRAN radiative transfer simulations and validated against TROPOMI, GEMS, and the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) ground-based observations to ensure retrieval consistency and accuracy.

Second, EMI-II TOCs are integrated with ERA5 meteorological variables, satellite-derived NO₂ and HCHO, and surface ozone measurements within machine learning frameworks, including Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), to quantify the statistical sensitivity and seasonal dependence of EMI-II-derived surface ozone estimates. Finally, the resulting products are evaluated against independent surface datasets and used to characterize spatial and seasonal ozone variability across the BTH region in 2022.

By systematically assessing the information content and applicability of EMI-II observations for surface ozone estimation, this study clarifies the potential role of China's UV-VIS satellite measurements in regional air quality analysis and provides a methodological framework for integrating national satellite products into surface pollution studies.

2. Materials and Methods

2.1. Study Area and Datasets

2.1.1. Study Area

The Beijing–Tianjin–Hebei (BTH) region (36–42°N, 113–119°E) is located in northern China and comprises Beijing, Tianjin, and eleven prefecture-level cities in Hebei Province. The region covers approximately 216,000 km², with a population of over 110 million, making it one of China's most densely populated and industrialized megaregions. Topographically, BTH slopes from the mountainous and high-elevation areas in the northwest toward low-lying plains in the southeast, resulting in pronounced spatial gradients in meteorology, emissions, and boundary-layer dynamics.

Climatologically, BTH is characterized by a temperate monsoon climate, featuring hot, humid summers and cold, dry winters. These conditions, combined with persistently high emissions of ozone precursors (NO_x and VOCs), contribute to frequent and severe ozone pollution episodes, particularly during warm seasons. The region is covered by comprehensive ground-based air quality monitoring networks, offering a valuable means to evaluate the information content of satellite ozone observations for constraining surface ozone variability. These characteristics make BTH a representative region for assessing the capability and limitations of EMI-II observations in surface ozone applications.

2.1.2. EMI-II Ozone Measurements

The Environmental Trace Gases Monitoring Instrument II (EMI-II) onboard the Gaofen-5B satellite is a nadir-viewing UV–VIS imaging spectrometer with four spectral channels: UV1 (234–311 nm), UV2 (306–401 nm), VIS1 (400–552 nm), and VIS2 (544–714 nm). This study uses measurements from the UV2 channel, which offers a nadir spatial resolution of approximately 13 km (along-track) × 24 km (across-track) and a spectral resolution of 0.3–0.6 nm. EMI-II operates in a sun-synchronous orbit at an altitude of approximately 705 km, with a local overpass time of 10:30 LT and a field of view of 114°.

EMI-II TOC products retrieved from these measurements constitute the core observational dataset whose physical reliability and surface relevance are evaluated in this study. Rather than treating EMI-II products as direct proxies for surface ozone, they are examined as column-integrated variables whose information content for surface ozone variability is quantified through independent validation and statistical analysis.

2.1.3. TROPOMI and GEMS Observations

TROPOMI: The TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor satellite provides daily global observations of atmospheric composition across the UV–VIS, NIR, and SWIR spectral ranges, with a nadir spatial resolution of up to 3.5 × 5.5 km² and a local overpass time of approximately 13:30 LT [13]. In this study, TROPOMI TOC products are used as an independent satellite reference to evaluate the consistency and accuracy of EMI-II ozone retrievals.

In addition, TROPOMI tropospheric NO₂ and HCHO column products are employed as chemical proxy variables representing precursor concentrations and photochemical regime conditions in the surface ozone estimation framework. To ensure data reliability, only pixels with quality assurance values (qa_value) greater than 0.5 are retained.

GEMS: The Geostationary Environment Monitoring Spectrometer (GEMS) onboard the GEO-KOMPSAT-2B satellite operates in geostationary orbit and provides UV–VIS measurements of trace gases over East Asia, with a spatial resolution of approximately 7×8 km² and a spectral resolution of 0.6 nm [14,15]. GEMS TOC products acquired at 10:45 LT are employed in this study for inter-sensor comparison with EMI-II to further assess spatial consistency and systematic biases under similar illumination conditions.

2.1.4. Ancillary and Validation Datasets

Absorption cross sections: High-resolution absorption cross sections of O₃ (223 K, 243 K), NO₂ (298 K), BrO (223 K), and HCHO (297 K) were obtained from the MPI-Mainz UV/VIS Spectral Atlas of Gaseous Molecules of Atmospheric Interest [16]. These cross sections are widely used in atmospheric radiative transfer modelling and DOAS retrieval algorithms and provide a physically consistent basis for ozone column estimation.

Surface reflectance: Surface reflectance was characterized using the Lambertian Equivalent Reflectivity (LER) climatology derived from GOME-2 observations, which provides global coverage across 27 spectral bands between 328 and 772 nm [17]. The LER value at 328 nm was adopted to represent surface reflectivity in the EMI-II ozone fitting window.

Ground-based ozone observations (CNEMC): Surface ozone measurements were obtained from the China National Environmental Monitoring Center (CNEMC) network [18], which provides hourly observations of major air pollutants at nationwide monitoring sites. Maximum daily 8-h average ozone concentrations (MDA8 O₃) from March 2022 to February 2023 were used as the target variable for model training and evaluation. These observations serve as the reference for assessing the ability of models using EMI-II observations, in combination with meteorological and chemical predictors, to estimate surface ozone variability.

Meteorological variables (ERA5): Meteorological drivers play a dominant role in ozone formation, accumulation, and dispersion through their influence on photochemical reaction rates, boundary-layer dynamics, and transport processes. Meteorological variables from the ERA5 reanalysis were therefore incorporated, including 2-m air temperature (T2M), relative humidity (RH), surface downward shortwave radiation (SSRD), surface downward longwave radiation (STR), zonal and meridional wind components at 10 m (U10, V10), surface pressure (SP), total precipitation (PRE), evaporation (E), total cloud cover (TCC), planetary boundary layer height (BLH), and leaf area index of high and low vegetation (LAI_hv and LAI_lv). These predictors provide physically interpretable information on ozone photochemistry and mixing processes and serve as baseline drivers in the surface ozone estimation framework.

WOUDC ground-based ozone measurements: Ground-based TOC observations from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) [19] were used as an independent reference dataset to validate EMI-II ozone retrievals. These measurements, derived from high-precision Dobson and Brewer spectrophotometers, were collocated to quantify retrieval accuracy, systematic bias, and stability, thereby establishing the physical reliability of EMI-II ozone columns prior to surface ozone estimation.

ChinaHighO₃ dataset: The China High-Resolution Air Pollutants (CHAP) dataset provides nationwide daily surface ozone concentrations at 1 km spatial resolution by integrating satellite observations, ground measurements, reanalysis meteorology, and emission inventories within a machine learning framework [20]. The ChinaHighO₃ product achieves an R² of 0.89 and an RMSE of 15.77 µg/m³ under 10-fold cross-validation. In this study, ChinaHighO₃ is used as an independent benchmark for evaluating the spatial consistency of EMI-II-based surface ozone estimates at the regional scale.

2.2. EMI-II Ozone Column Retrieval and Physical Validation

Accurate retrieval of TOCs is a prerequisite for assessing the information content of EMI-II observations for surface ozone applications. We therefore implemented a physically based DOAS retrieval framework, followed by independent validation against ground-based and satellite reference datasets.

2.2.1. Spectral Calibration and DOAS Fitting

The DOAS technique is susceptible to wavelength shifts. EMI-II irradiance spectra exhibit pronounced across-track variations in wavelength shift and slit function width, likely associated with vibrations and thermal effects during launch and in-orbit operations. If not properly corrected, these variations can lead to striping artifacts and systematic biases in retrieved slant column densities (SCDs) [21].

To address this issue, pixel-level spectral calibration was performed based on Fraunhofer line fitting through a three-step procedure. First, a high-resolution solar reference spectrum was convolved with a representative slit function to approximate EMI-II spectral resolution. Second, for each irradiance spectrum, wavelength shift, spectral stretch, and slit function full width at half maximum (FWHM) were optimized using a Levenberg–Marquardt least-squares algorithm by minimizing residuals between measured and simulated spectra. Third, wavelength correction polynomials were constructed for each detector row and applied to the corresponding irradiance and radiance spectra. This calibration significantly improved the alignment of Fraunhofer lines and reduced fitting residuals (Figure 1 and Figure A1), ensuring spectral consistency across the swath.

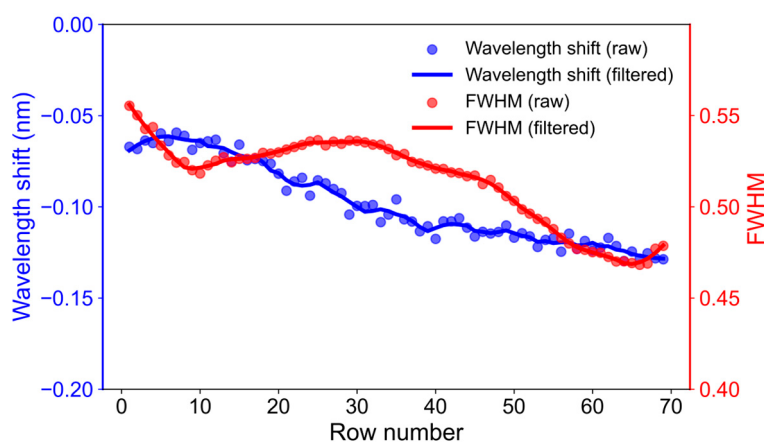


Figure 1. Across-track variations of the wavelength shift (left y -axis, blue) and slit function width (right y -axis, red) in the EMI-II irradiance spectrum on 18 October 2022, both showing pronounced row-dependent variations across the detector.

Ozone SCDs were retrieved using a standard DOAS approach [22] based on optical density fitting over the 325–335 nm window. The fitting equation is expressed as:

$$-\ln\left(\frac{I(\lambda - \Delta(\lambda)) - \text{offset}(\lambda)}{I_0(\lambda)}\right) = \sum_i \sigma'_i(\lambda) \text{SCD}_i + P(\lambda) \quad (1)$$

where, $I(\lambda)$ is the observed radiance spectrum and I_0 is the irradiance spectrum. The term $\Delta\lambda$ refers to the wavelength shift of the radiance spectrum, and the offset, expressed as a first-order polynomial, accounts for intensity variations. The absorption cross section of gas i is denoted by $\sigma'_i(\lambda)$, and its SCD is represented by SCD_i . A low-order polynomial $P(\lambda)$ is included to account for slowly varying atmospheric extinction and scattering effects (e.g., Rayleigh and Mie scattering). The trace gases

included in the fitting process are O₃, NO₂, BrO, and HCHO. The key fitting parameters are summarized in Table 1.

Table 1. DOAS fitting parameters for EMI-II slant column density retrieval.

Parameters	Parameter Settings	Reference
Fitting window	325nm–335nm	
Polynomial	4th order	
Cross-section	O ₃ (223 K,243 K)	Bogumil et al., 2003 [23]
	NO ₂ (298 K)	Vandaele et al., 1996 [24]
	BrO(223 K)	Fleischmann et al., 2004 [25]
	HCHO(297 K)	Meller and Moortgat, 2000 [26]
	Raman spectrum	Chance and Spurr, 1997 [27]

To mitigate residual wavelength shift and stretch effects in radiance spectra, first-order Taylor expansion terms were introduced as pseudo-absorption cross sections [28]. This correction substantially reduced root-mean-square (RMS) fitting residuals and eliminated the systematic overestimation of ozone SCDs (Figure 2), thereby improving the physical reliability of the retrieved columns.

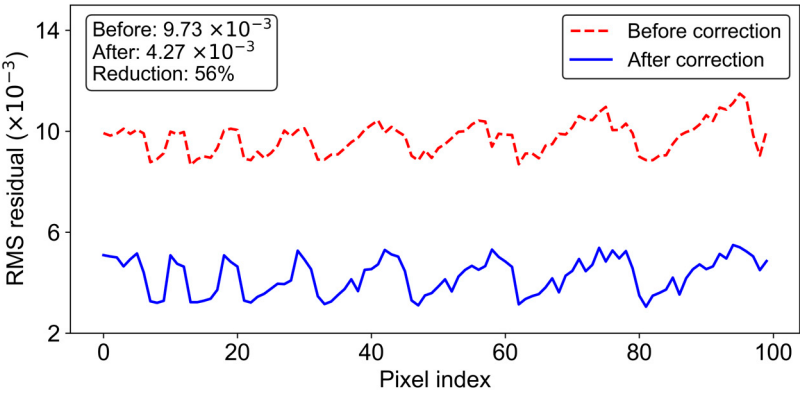


Figure 2. RMS fitting residuals in EMI-II retrievals on 3 April 2022 before and after spectral shift correction. The figure shows that spectral shift correction of the irradiance data effectively reduces the fitting residuals.

2.2.2. Air Mass Factor Calculation and VCD Retrieval

Vertical column densities (VCDs) were obtained from SCDs using air mass factors (AMFs), which account for the influence of observation geometry and atmospheric conditions on the light path:

$$VCD = \frac{SCD(\lambda, \theta, \alpha, \phi)}{AMF(\lambda, \theta, \alpha, \phi)} \tag{2}$$

AMFs were simulated using the SCIATRAN radiative transfer model and stored in a multidimensional look-up table (LUT) covering solar zenith angle (SZA), viewing zenith angle (VZA), relative azimuth angle (RAA), surface albedo, and surface pressure (Table 2). Sensitivity experiments under representative conditions revealed that AMFs are primarily controlled by SZA and VZA, while the impacts of RAA within the fitting window is comparatively minor (Figure 3). Furthermore, surface albedo exerts a stronger influence than wavelength on AMFs under fixed geometry. A priori ozone profiles from the SCIATRAN climatology were adopted, with seasonal and latitudinal dependencies incorporated.

For partially cloudy pixels, the independent pixel approximation (IPA) was applied by combining clear-sky and cloudy-sky AMFs weighted by cloud fraction derived from GOME-2C cloud products:

$$AMF = wM_{cloudy} + (1 - w)M_{clear} \tag{3}$$

Final pixel-level VCDs were obtained through multidimensional linear interpolation within the LUT.

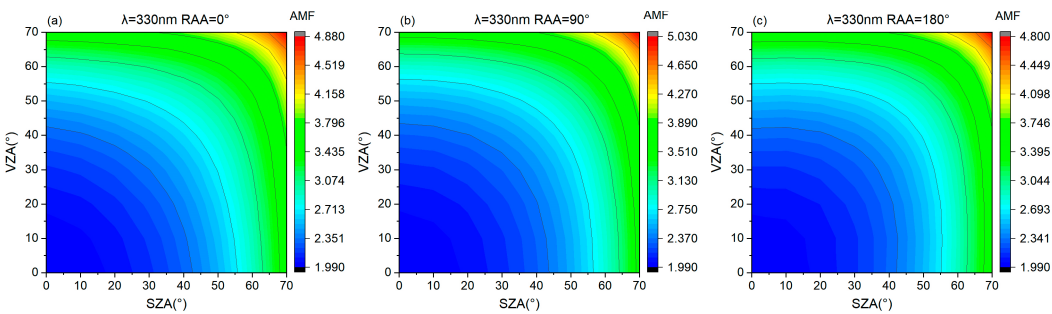


Figure 3. Variation of AMF with observation geometry at 330 nm: (a) RAA = 0°; (b) RAA = 90°; (c) RAA = 180°. The figure shows that SZA has the most significant impact on ozone AMF, followed by VZA, while RAA have relatively minor effects.

Table 2. The parameter node settings of the AMF LUT simulated using SCIATRAN radiative transfer model.

Parameter	Number of Nodes	Values
Solar zenith angle (°)	12	0, 10, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70
Viewing zenith angle(°)	12	0, 10, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70
Relative azimuth angle(°)	5	0, 45, 90, 135, 180
Surface albedo	9	0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.45
Surface pressure (hPa)	9	1013, 795, 701, 616, 472, 356, 264, 164, 96

2.2.3. Stripe Artifact Correction

Retrieved EMI-II VCD fields exhibit systematic along-track striping patterns that can obscure spatial gradients and introduce biases in regional analyses. To suppress this artifact, we applied a Fourier-based destriping technique [29]. First, analysis windows (70 pixels across-track × 100 pixels along-track) were identified along each orbit using covariance minimization. Second, cross-track mean profiles were computed within the selected windows and transformed into the frequency domain. High-frequency components associated with striping were isolated using a low-pass filter and subsequently removed from the original data. This procedure effectively suppressed striping artifacts (Figure A2), yielding spatially coherent ozone column fields suitable for regional-scale applications.

2.2.4. Independent Validation and Satellite Consistency

To ensure the physical reliability of EMI-II ozone columns prior to surface ozone estimation, we conducted independent validation against ground-based observations from the WOUDC network and cross-comparisons with TROPOMI and GEMS satellite products. Validation statistics, including correlation coefficients (R), mean relative bias (Bias), and standard deviation (SD), were calculated across multiple sites and spatial scales.

2.3. Surface Ozone Estimation Framework

Rather than treating satellite columns as direct substitutes for surface ozone observations, we constructed a statistical modeling framework that explicitly integrates EMI-II TOCs with

meteorological drivers and chemical proxy variables to quantify their contribution to surface ozone variability.

2.3.1. Multi-Source Data Integration and Harmonization

Due to differences in spatial resolution, sampling frequency, and coverage among datasets, all variables were harmonized to a uniform daily temporal resolution and $0.25^{\circ} \times 0.25^{\circ}$ spatial grid. ERA5 meteorological fields were aggregated from hourly to daily means. TROPOMI Level-2 products from multiple orbits were mosaicked and averaged to daily grids. Surface ozone observations from monitoring stations were spatially aggregated by averaging observations from multiple stations falling within the same grid cell. For each grid cell and day, EMI-II TOCs, meteorological variables, precursor columns, and surface ozone concentrations were collocated to form the integrated analysis dataset.

2.3.2. Feature Selection and Feature Importance Evaluation

Predictor variables were selected based on both physical relevance to ozone formation processes and statistical associations with observed surface ozone concentrations. Meteorological variables included T2M, SSRD, BLH, RH, SP, wind components (U10, V10), and other related parameters. Chemical variables included TROPOMI NO₂ and HCHO columns as proxies for photochemical precursor availability, together with EMI-II TOCs.

Initial screening was conducted using Pearson correlation analysis to assess monotonic associations with surface ozone (Figure 4a). Recursive feature elimination (RFE) was subsequently applied within both RF and XGBoost frameworks to evaluate the contribution of each input variable to model performance. Although features with low linear correlation exhibited limited individual explanatory power, their removal resulted in a consistent decline in nonlinear model performance (Figure 4b). To maximize information utilization and preserve potential nonlinear interactions, all physically relevant predictors were retained in the final models.

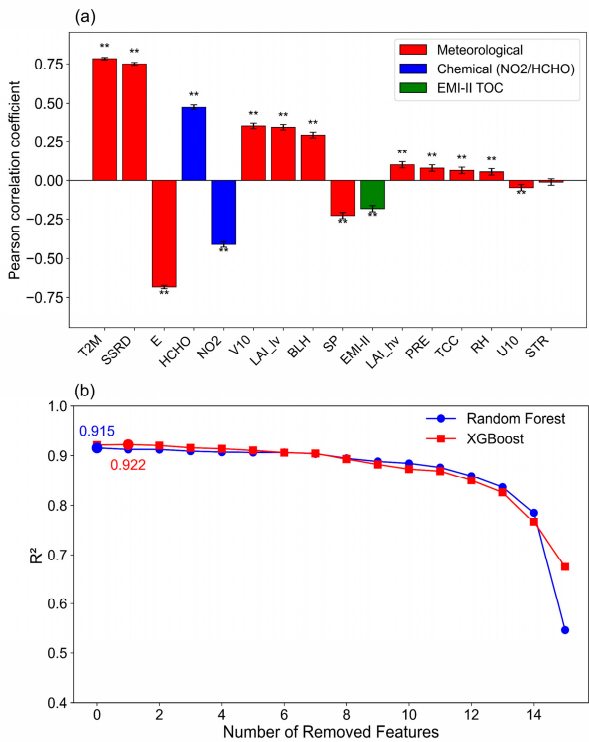


Figure 4. (a) Ranking of feature variables by correlations with surface ozone concentrations (* $p < 0.05$; ** $p < 0.01$); (b) Model performance during recursive feature elimination (RFE) using the Random Forest (blue) and XGBoost (red) algorithm.

2.3.3. Machine Learning Models and Experimental Design

Two tree-based ensemble learning algorithms—RF and XGBoost—were employed to model the nonlinear relationships between multi-source variables and surface ozone concentrations. These algorithms were selected due to their robustness to multicollinearity, ability to capture nonlinear interactions, and widespread application in atmospheric composition studies.

Model hyperparameters were optimized using grid search combined with ten-fold cross-validation. Key parameters for RF included the number of trees ($n_estimators$), maximum tree depth (max_depth), and minimum samples per leaf ($min_samples_leaf$) and split ($min_samples_split$). For XGBoost, tuning focused on the number of boosting rounds ($n_estimators$), learning rate (eta), maximum depth (max_depth), and minimum child weight (min_child_weight). All remaining parameters were set to default values.

To quantify the independent contribution of EMI-II TOCs, we designed parallel baseline experiments excluding EMI-II variables while retaining all other variables. Performance differences between models with and without EMI-II variables were used to assess the incremental contribution of EMI-II observations under varying meteorological and seasonal conditions.

2.3.4. Model Performance Evaluation

Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

In these equations, y_i denotes the observed ozone concentration for the i th sample, $\hat{y}_i$ represents the corresponding model-predicted ozone concentration, n is the total number of samples, and $\bar{y}_i$ indicates the mean of the observed values.

2.4. Attribution Analysis Using Geographical Detector

To further investigate the spatial drivers and interaction mechanisms governing surface ozone variability, we applied the geographical detector method [30,31]. This approach quantifies the explanatory power of individual factors and their interactions by comparing within-stratum variance with total variance across the study domain. The factor detector and interaction detector modules were used to evaluate the seasonal influence of meteorological conditions, chemical precursors, and EMI-II TOCs on the spatial heterogeneity of surface ozone.

The explanatory power of each factor was quantified using the q -statistic:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (7)$$

Here, h denotes the category of the stratified factor, and L is the total number of strata. N_h and σ_h^2 represent the sample size and the variance of the dependent variable within the h -th stratum, respectively; While N and σ^2 denote the total sample size and the overall variance of the dependent variable in the study area. The q -statistic ranges from 0 to 1, with larger values indicating stronger explanatory power of the factor for the spatial distribution of surface ozone.

The interaction detector evaluates whether the combined explanatory power of two factors exceeds that of each factor individually. By comparing $q(X_1 \cap X_2)$ with $q(X_1)$ and $q(X_2)$, the method identifies independence, enhancement, or weakening effects. In this study, interaction analysis was conducted separately for each season to diagnose nonlinear coupling between meteorological and chemical drivers.

3. Results

3.1. Physical Consistency and Information Quality of EMI-II Ozone Columns

Before evaluating the capability of EMI-II observations to inform surface ozone variability, it is essential to establish the physical reliability of the retrieved TOCs, which constitute the only satellite-derived ozone constraint in the present framework.

EMI-II TOCs were first evaluated against ground-based Dobson and Brewer spectrophotometer measurements archived by WOUDC at three long-term East Asian stations: Xianghe (China), Seoul (South Korea), and Tateno (Japan). During 2022, correlation coefficients exceeded 0.90 at all sites (Table 3), indicating strong temporal consistency. Mean relative biases ranged from 3.29% to 3.63%, with standard deviations between 2.29% and 2.66% (Figure 5). The observed positive offsets are small and spatially consistent across stations, suggesting the presence of a systematic retrieval bias rather than site-specific effects. The magnitude of these biases falls within the typical uncertainty range reported for contemporary UV satellite ozone products.

Table 3. Validation of EMI-II TOCs against WOUDC ground-based measurements at three East Asian stations during 2022.

Station	Latitude, Longitude	Method	N	R	Bias (%)	SD%
Xianghe	39.75°N,116.96°E	Dobson	160	0.93	3.29	2.29
Seoul	37.57°N,126.95°E	Dobson	188	0.92	3.53	2.66
Tateno	36.06°N,140.13°E	Brewer	126	0.93	3.63	2.51

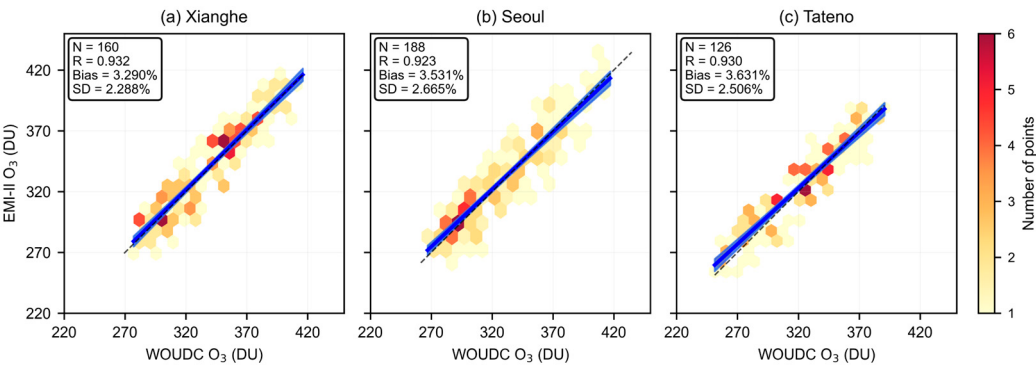


Figure 5. Linear fitting between EMI-II TOCs and ground-based measurements at the (a) Xianghe, (b) Seoul, and (c) Tateno stations.

In addition to ground validation, spatial consistency was examined through comparison with independent satellite TOC products from TROPOMI and GEMS. All datasets were aggregated to a

common 0.25° grid to minimize resolution-induced discrepancies. Figure 6 presents a representative case over the BTH region on 28 November 2022. EMI-II reproduces the large-scale ozone gradient and enhanced values over northern Hebei, exhibiting high spatial correlations with both GEMS ($r = 0.965$) and TROPOMI ($r = 0.963$). Mean relative differences remain below 5%, with slightly larger deviations in regions of elevated ozone columns. Considering differences in overpass time, viewing geometry, cloud screening procedures, and native spatial resolution, these discrepancies are expected and do not indicate structural inconsistencies in the EMI-II retrievals.

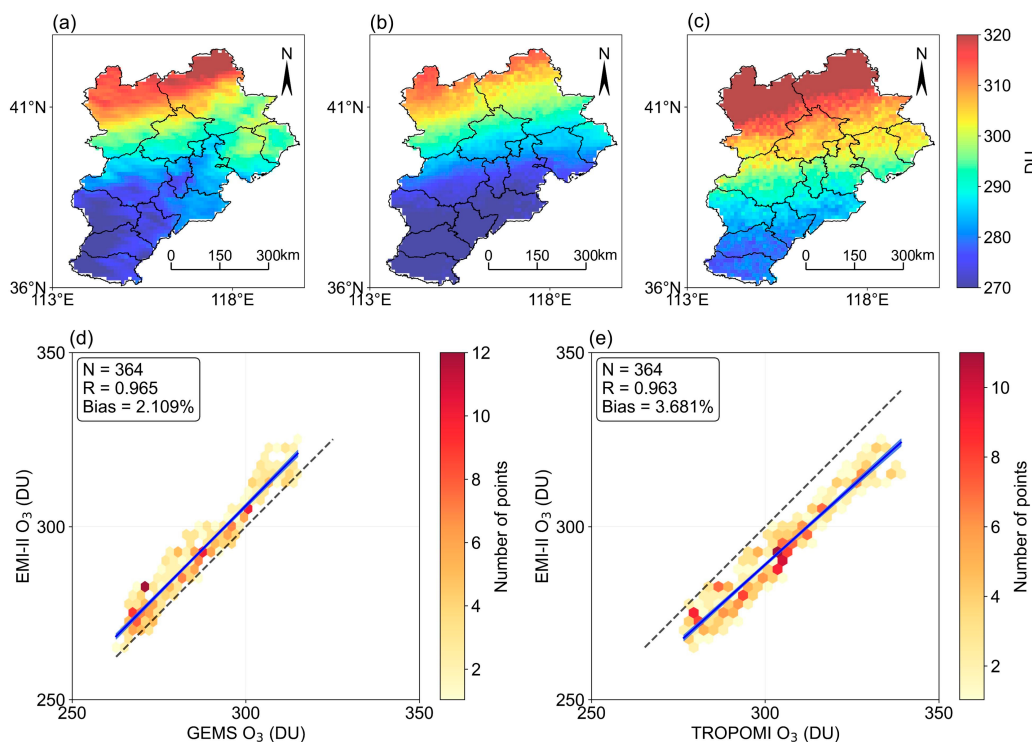


Figure 6. Comparison of TOCs over the BTH region on 28 November 2022: (a) EMI-II TOCs; (b) GEMS TOCs; (c) TROPOMI TOCs; (d) linear fitting between the EMI-II TOCs and GEMS TOCs; (e) linear fitting between the EMI-II TOCs and TROPOMI TOCs. The results demonstrate the consistency and accuracy of the satellite retrievals, with r exceeding 0.9 for TOC.

Together, the ground-based and inter-satellite comparisons demonstrate that EMI-II preserves both temporal variability and regional spatial structures of total ozone. The consistency across independent observing systems indicates that EMI-II TOCs provide a physically credible representation of column ozone variability at regional scales.

Although total column retrievals are primarily sensitive to the upper troposphere and lower stratosphere, the reliability of the integrated column amount is a prerequisite for any downstream application that seeks to extract tropospheric or surface-related information. The validation results presented here therefore establish a necessary physical foundation for subsequent analyses of information content and surface ozone estimation.

In addition to external validation, we further examined the sensitivity of the EMI-II retrieval to key parameters influencing the AMF, which constitutes a major source of uncertainty in TOC retrievals. Since AMF errors cannot be directly observed, a series of controlled radiative transfer simulations were conducted using the SCIATRAN model to evaluate the relative sensitivity of AMF to representative perturbations in input parameters across typical observation geometries (Table 4).

The relative uncertainty associated with the DOAS-derived SCD was estimated to be less than 2% based on fitting residuals. For AMF-related factors, sensitivity tests were performed by perturbing individual parameters while keeping others fixed. The AMF LUT was constructed using a standard

LOWTRAN aerosol model. The aerosol-related sensitivity (approximately 1.3%) was estimated by comparing simulations with and without aerosol loading. Surface albedo sensitivity (approximately 0.3%) was assessed using monthly mean and minimum albedo values derived from the GOME-2C dataset. The relatively small sensitivity reflects the limited dependence of ozone TOC retrievals on surface reflectance under typical UV observing conditions. However, this estimate assumes that the albedo product is unbiased and does not explicitly account for sub-pixel cloud contamination.

Cloud-related sensitivity (below 2.3%) was evaluated by comparing clear-sky and cloudy scenarios in the radiative transfer simulations. As the present retrieval does not implement an explicit cloud correction scheme, this estimate reflects the potential impact of unresolved cloud effects under simplified assumptions. The influence of the a priori ozone profile (below 3%) was quantified by comparing AMFs derived from the SCIATRAN climatological profile with those calculated using ozonesonde measurements from Hong Kong. While the resulting differences remain modest across representative solar and viewing geometries, we acknowledge that using a single sonde location to represent a geographically diverse region introduces additional structural uncertainty.

Assuming independence among these perturbations, the combined sensitivity under the tested conditions is estimated to be on the order of 4–5%. This value should be interpreted as a lower-bound estimate under controlled modelling assumptions rather than a fully propagated retrieval uncertainty budget. Nevertheless, the results indicate that AMF-related sensitivities are of comparable magnitude to the SCD fitting uncertainty and do not dominate the overall error structure of the EMI-II TOC retrieval. These analyses further support the physical credibility of the EMI-II ozone columns as a basis for subsequent information-content assessment and surface ozone estimation.

Table 4. Sensitivity of EMI-II AMF to key input parameters under controlled perturbations.

Error Sources	Relative error (%)
SCD (DOAS fitting)	<2
Aerosols	1.3
Surface Albedo	0.3
Clouds	<2.3
A Priori Ozone profile	<3

3.2. Averaging Kernel and Surface Sensitivity Diagnostics

To assess the intrinsic vertical information content of EMI-II ozone retrievals, averaging kernels (AVKs) were analyzed under representative seasonal conditions over the BTH region. The AVK characterizes the sensitivity of the retrieved total column to perturbations in the vertical ozone distribution and reflects the effective vertical weighting of the retrieval.

Figure 7a presents the seasonal mean vertical distributions of the AVK. Maximum sensitivity occurs in the upper troposphere and lower stratosphere, with peak values near 200 hPa slightly exceeding unity. Sensitivity decreases rapidly below approximately 500 hPa and approaches values of 0.20–0.25 within the planetary boundary layer. This vertical structure reflects the dominance of Rayleigh scattering in the UV spectral range, which enhances sensitivity to higher altitudes while attenuating contributions from the lower atmosphere.

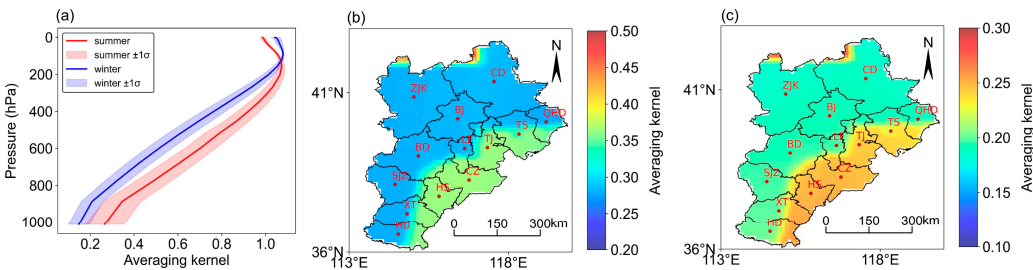


Figure 7. Example of averaging kernels: (a) Averaging kernels under summer and winter conditions; (b) Spatial distribution of 0–2 km averaged kernels over the BTH region in summer; (c) Spatial distribution of 0–2 km averaged kernels over the BTH region in winter.

Seasonal differences are evident. In winter, reduced sensitivity is observed in the 400–800 hPa layer compared with summer. The reduction is consistent with larger solar zenith angles and longer effective optical paths during winter, which increase multiple scattering and further suppress lower-tropospheric contributions. These characteristics are typical of nadir-viewing UV ozone retrievals and indicate that direct sensitivity to boundary-layer ozone is intrinsically limited.

To examine horizontal variability in near-surface sensitivity, AVKs were averaged over the 0–2 km layer and mapped spatially (Figure 7b–c). A clear gradient emerges, with lower sensitivity over the northern mountainous regions and relatively enhanced sensitivity over the southern and southeastern plains. The pattern closely follows spatial variations in surface UV reflectivity. Regions with higher albedo increase the fraction of backscattered radiation originating from the lower atmosphere, thereby enhancing near-surface sensitivity. Conversely, densely vegetated and mountainous areas with lower reflectance exhibit reduced boundary-layer sensitivity.

Overall, the AVK analysis indicates that EMI-II TOCs contain limited but physically structured information related to lower-tropospheric variability. While the retrieval is not directly sensitive to boundary-layer ozone in a strict inversion sense, the presence of systematic and spatially coherent sensitivity implies that column measurements may still carry weak constraints on surface variability. Consequently, surface ozone estimation based on EMI-II should be interpreted as a statistical inference constrained by column-integrated information and modulated by meteorological and chemical covariates, rather than as a vertically resolved physical retrieval of boundary-layer ozone.

The AVK diagnostics therefore provide a quantitative basis for understanding the strengths and limitations of EMI-II observations and for interpreting subsequent surface ozone estimation results within a physically consistent framework.

3.3. Information Content of EMI-II for Surface Ozone Estimation

Building upon the physical consistency analysis and AVK diagnostics, the contribution of EMI-II TOCs to surface ozone estimation was evaluated within a statistical learning framework.

Ten-fold cross-validation was conducted using RF and XGBoost models (Figure 8). The RF model yields a mean R^2 of 0.91 for validation samples, with RMSE and MAE of approximately $15.0 \mu\text{g}/\text{m}^3$ and $10.7 \mu\text{g}/\text{m}^3$, respectively. XGBoost provides moderately improved performance ($R^2 = 0.94$) together with reduced error metrics, reflecting its stronger capacity to represent nonlinear relationships among predictors.

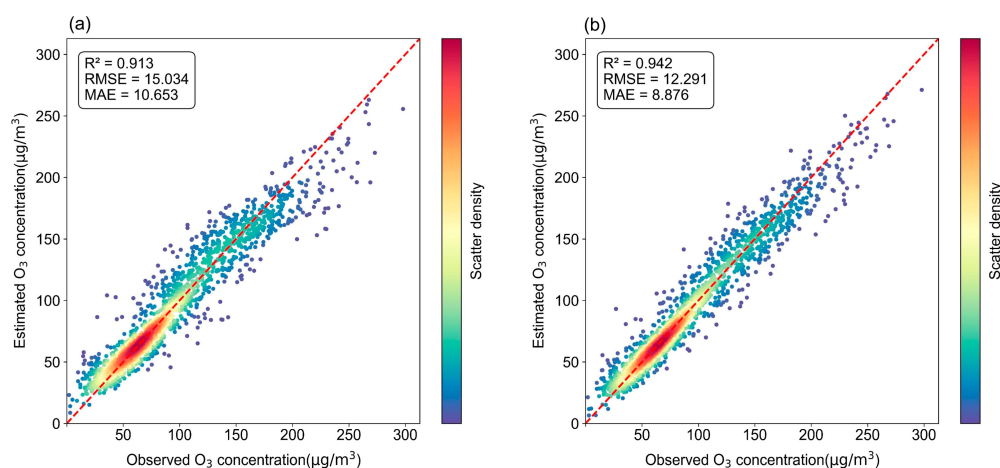


Figure 8. Scatter plots of estimated versus observed values for the test set: (a) Random Forest, (b) XGBoost.

Although these statistics indicate substantial predictive skill, model performance reflects the combined influence of meteorological variables, precursor information, and satellite-derived ozone columns. To assess the specific contribution of EMI-II, baseline models excluding TOCs were constructed using meteorological parameters and chemical proxies (NO_2 and HCHO) alone. Models including EMI-II TOCs achieved a cross-validated R^2 of 0.94 and RMSE of $12.29 \mu\text{g}/\text{m}^3$, with the inclusion of TOCs yielding consistent, albeit modest, improvements across all cross-validation folds. The stability of this incremental gain suggests that EMI-II observations provide complementary information that is systematically embedded within the multi-predictor framework, rather than being dominated by any single contributing factor.

The magnitude of improvement is modest but systematic. This suggests that EMI-II TOCs provide additional explanatory information beyond meteorology and precursor indicators. Given the limited direct boundary-layer sensitivity indicated by the AVK analysis, this incremental gain likely reflects weak yet spatially coherent column–surface coupling embedded within the total ozone signal rather than a dominant control.

To further assess spatial robustness, the derived surface ozone fields were compared with the independent CHAP dataset after resampling to a common 0.25° grid. On 3 June 2022 (Figure 9), the two products exhibit strong spatial correspondence ($r = 0.97$), with MAE and RMSE of $7.98 \mu\text{g}/\text{m}^3$ and $10.05 \mu\text{g}/\text{m}^3$, respectively. Both datasets reproduce enhanced ozone over southern Hebei and the central plains. Differences are mainly observed in urban cores and areas with complex terrain, where discrepancies may arise from differences in modeling assumptions and spatial resolution.

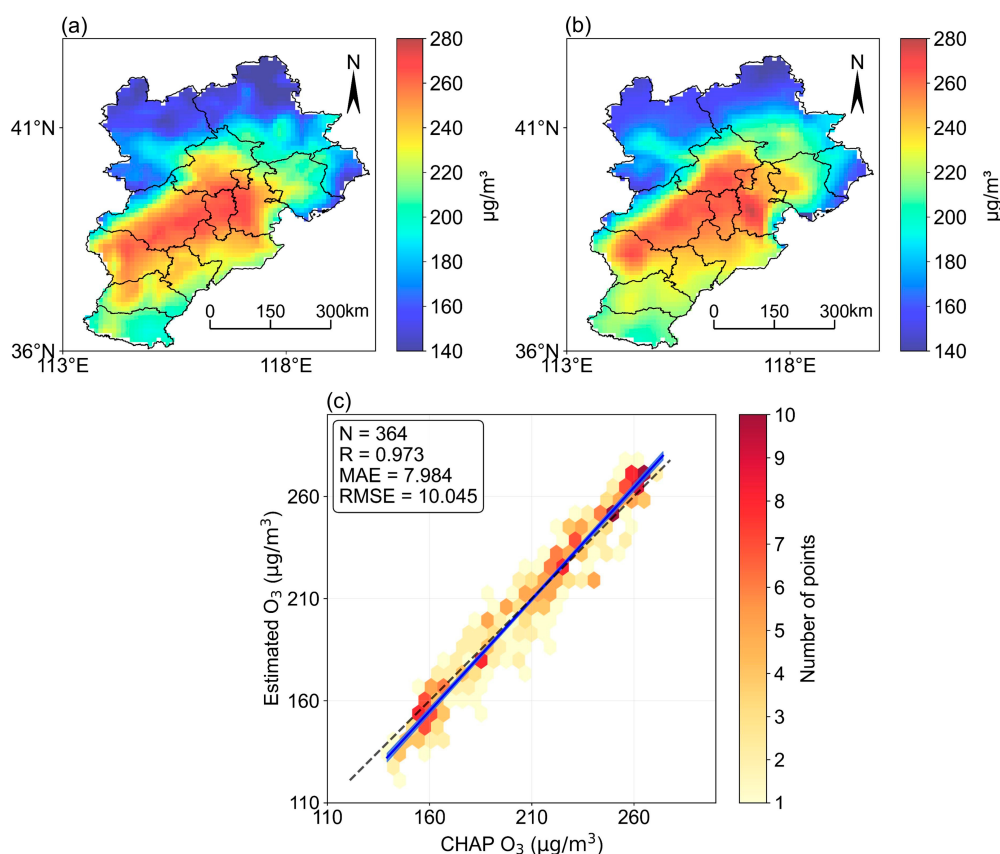


Figure 9. Comparison of surface ozone over the BTH region on 3 June 2022: (a) EMI-II; (b) CHAP; (c) linear fitting between EMI-II and CHAP.

Overall, the results indicate that EMI-II TOCs contribute measurable, though not dominant, information to surface ozone estimation. The added value is consistent with the limited but structured lower-tropospheric sensitivity identified in the AVK analysis and supports the use of EMI-II as a complementary constraint within multi-factor surface ozone estimation frameworks.

3.4. Spatial Patterns of EMI-II–Derived Surface Ozone

Seasonal mean surface ozone fields derived from EMI-II exhibit pronounced variations over the BTH region (Figure 10). Regional averages peak in summer ($130 \mu\text{g}/\text{m}^3$), followed by spring ($113 \mu\text{g}/\text{m}^3$), while substantially lower concentrations occur in autumn ($80 \mu\text{g}/\text{m}^3$) and winter ($59 \mu\text{g}/\text{m}^3$). This seasonal progression is consistent with the photochemical regime of northern China, where strong solar radiation and elevated temperatures enhance ozone production during warm seasons.

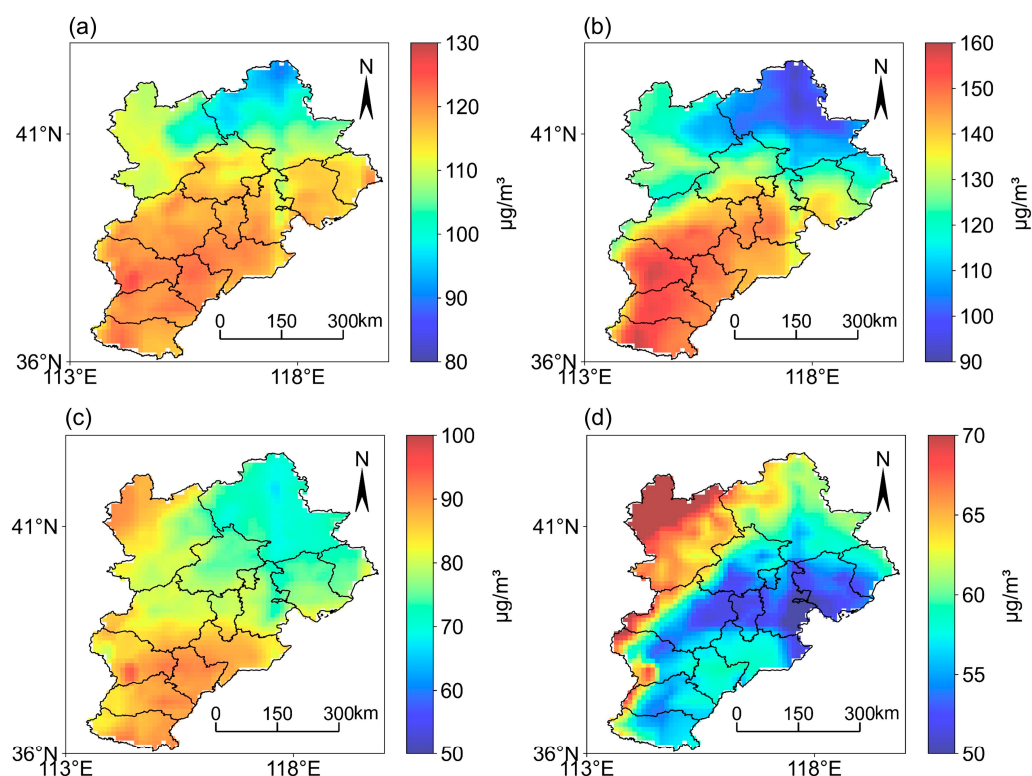


Figure 10. Seasonal mean spatial distribution of surface ozone in the BTH region: (a) spring, (b) summer, (c) autumn, (d) winter.

Spatially, enhanced ozone concentrations during spring and summer are primarily located over the southern and central plains of the BTH region. In contrast, lower values are consistently observed over the northern mountainous areas. The contrast reflects combined influences of precursor emissions, boundary-layer development, and topographic effects. The southern plains, characterized by dense anthropogenic activity and stronger summertime radiation, favor active photochemistry and regional accumulation. Mountainous northern areas, with generally lower emissions and different boundary-layer dynamics, exhibit comparatively reduced ozone levels.

From summer to winter, the spatial extent of high-ozone regions contracts and mean concentrations decline markedly. Reduced solar radiation, weaker boundary-layer mixing, and diminished photochemical activity likely account for the seasonal attenuation. Nevertheless, localized enhancements persist in certain suburban or downwind areas during winter, suggesting the influence of regional transport and complex chemical–dynamical interactions.

Overall, the derived spatial patterns align with established regional ozone climatology and reflect the interplay between photochemistry, emissions distribution, and terrain. The consistency of these patterns supports the spatial credibility of EMI-II–based surface ozone estimates at the regional scale.

3.5. Meteorological and Chemical Drivers of Surface Ozone Variability

The relative contributions of meteorological variables, precursor indicators, and EMI-II TOCs to surface ozone variability were examined using the geographical detector framework (Figure 11a). Across all seasons, T2M, SP, and BLH consistently exhibit high explanatory power, highlighting the dominant influence of thermodynamic conditions and boundary-layer dynamics on ozone variability.

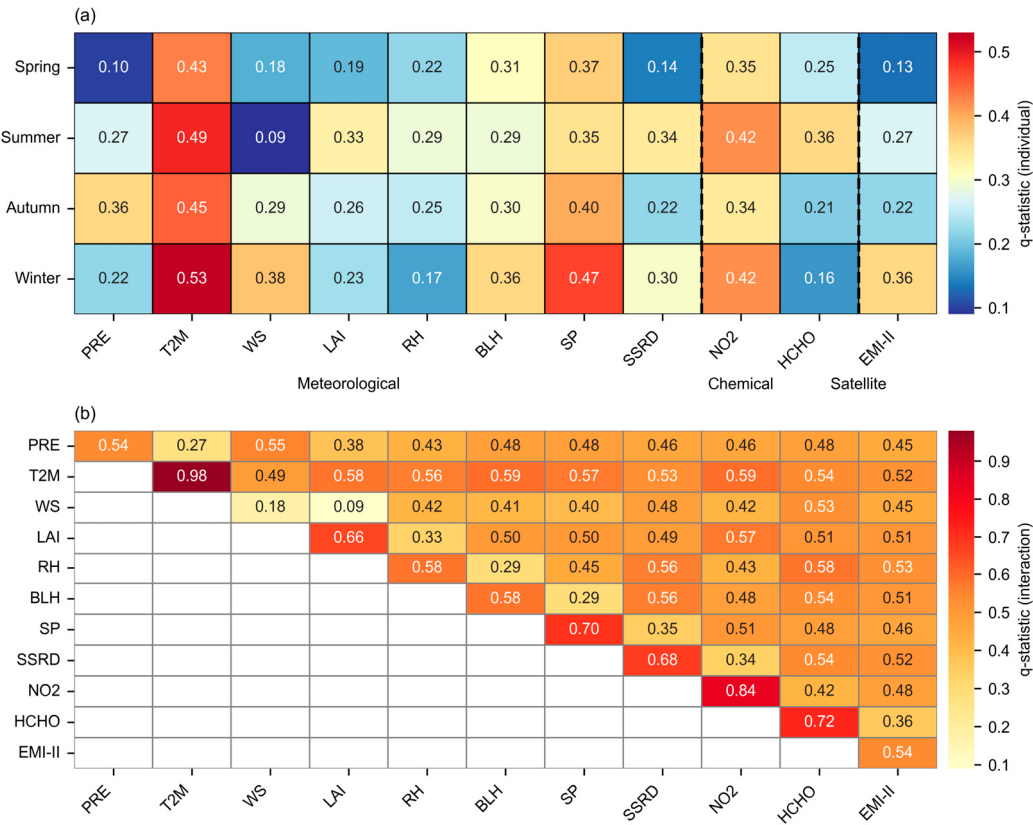


Figure 11. q values of influencing factors for surface ozone concentrations over the BTH region: (a) seasonal q values of individual factors, (b) summer q values of interactions among factors.

Temperature plays a central role by modulating photochemical reaction rates and influencing biogenic and anthropogenic precursor emissions. BLH reflects vertical mixing intensity, thereby regulating pollutant accumulation and dilution. SP, closely linked to synoptic-scale circulation patterns, likely captures large-scale transport and stagnation effects. Together, these variables characterize the dynamical and radiative environment within which ozone formation and accumulation occur.

SSRD shows enhanced importance during summer, consistent with intensified photochemical production under strong solar forcing. In contrast, the relative contribution of wind speed decreases slightly in summer, suggesting that chemical production may outweigh dispersion effects during peak ozone seasons.

Among precursor proxies, NO₂ maintains substantial explanatory power throughout the year, reflecting its role in ozone formation chemistry. HCHO exhibits stronger influence in summer, consistent with enhanced VOC-related photochemical activity under warm and high-radiation conditions.

The contribution of EMI-II TOCs displays seasonal variability, with relatively greater explanatory power in winter. This behavior is consistent with weaker boundary-layer mixing during

cold seasons, when reduced vertical dilution may strengthen the coupling between column-integrated ozone signals and near-surface concentrations.

Interaction analysis (Figure 11b) indicates that ozone variability is governed by nonlinear interactions among chemical and meteorological drivers. Temperature emerges as the dominant interacting variable, exhibiting enhanced combined effects with BLH, radiation, relative humidity, and wind speed. Strong interaction signals are also observed between precursor indicators (NO_2 and HCHO) and radiation-related variables, reflecting the radiation-dependent nature of photochemical ozone formation.

Overall, these results suggest that surface ozone variability in the BTH region arises from tightly coupled chemical–dynamical processes. EMI-II TOCs do not act as primary drivers but provide complementary information that interacts with meteorological conditions, particularly under dynamically stable environments. The observed interaction structure is physically consistent with the seasonal modulation of photochemistry and boundary-layer processes discussed in previous sections.

4. Discussion

The central objective of this study was to assess whether EMI-II TOC observations contain physically meaningful and statistically robust information relevant to surface ozone variability. The results collectively demonstrate that the skill of EMI-II-based surface ozone estimation is strongly modulated by atmospheric regime, particularly by seasonal variations in photochemical activity and boundary-layer dynamics.

The seasonal spatial patterns derived from EMI-II-based estimates reveal pronounced enhancement of surface ozone during spring and summer, consistent with intensified photochemical production over the BTH region. Geographical detector results further indicate that thermodynamic variables—especially near-surface temperature and boundary-layer height—consistently exert dominant explanatory power. These findings are physically coherent: temperature regulates photochemical reaction rates and precursor emissions, while boundary-layer development governs vertical mixing and pollutant accumulation.

The interaction analysis provides additional insight into the coupled nature of ozone formation. The observed nonlinear enhancement between temperature and radiation, boundary-layer height, and precursor indicators (NO_2 and HCHO) underscores that surface ozone variability cannot be attributed to isolated drivers. Instead, it emerges from tightly linked chemical–dynamical feedbacks. In this context, EMI-II TOC does not function as an independent predictor but rather contributes information that interacts with boundary-layer and photochemical conditions. The stronger statistical coupling observed during summer can therefore be interpreted as a regime in which enhanced vertical mixing increases the relative contribution of lower-tropospheric ozone to total column variability.

During winter and other dynamically stable periods, shallow boundary layers and weak photochemistry reduce the proportion of boundary-layer ozone within the total column. Under such conditions, stratospheric variability and large-scale transport exert comparatively greater influence on TOC signals, weakening the effective information transfer from column measurements to surface concentrations. This regime-dependent behavior aligns with theoretical expectations derived from averaging-kernel analyses of nadir-viewing UV–VIS instruments and reflects an intrinsic limitation of passive column observations rather than a deficiency specific to EMI-II. Comparable seasonal dependence reported for OMI and TROPOMI further suggests that such constraints are fundamental to column-based retrieval approaches.

Although machine learning models effectively exploit nonlinear relationships among meteorological and chemical predictors, the results indicate that statistical learning cannot fully overcome physical limitations in vertical sensitivity. Auxiliary predictors partially compensate for weak column–surface coupling by incorporating boundary-layer and photochemical indicators, yet estimation performance remains contingent upon atmospheric regime. This distinction is important:

statistical enhancement improves inference capability but does not eliminate the underlying information-content constraints.

From a broader perspective, this study suggests that the applicability of national UV–VIS satellite measurements to surface air quality monitoring should be interpreted within a regime-aware framework. Rather than assuming uniform performance across seasons, satellite-derived TOC provides meaningful surface-related information primarily when vertical coupling between the boundary layer and free troposphere is sufficiently strong. Such a regime-dependent understanding has implications for operational data assimilation, model evaluation, and satellite product development.

5. Conclusions

This study evaluated the capability of EMI-II TOC observations to support surface ozone estimation over the BTH region and developed an integrated satellite–machine learning framework for this purpose. Three main conclusions can be drawn:

First, EMI-II TOCs exhibit strong physical consistency with independent satellite and ground-based observations. Validation against TROPOMI ($R = 0.96$), GEMS ($R = 0.97$), and WOUDC ground-based measurements ($R > 0.92$, bias $< 4\%$) confirms that EMI-II reliably captures key seasonal and spatial ozone patterns, establishing a credible observational basis for subsequent surface ozone analysis.

Second, EMI-II TOCs provide complementary information for surface ozone estimation when combined with meteorological and chemical predictors. The integrated machine learning models achieve robust performance, with cross-validated R^2 of 0.94 and RMSE of $12.29 \mu\text{g}/\text{m}^3$ for surface ozone estimation over the BTH region. The inclusion of EMI-II observations yields consistent improvements across model configurations, indicating that EMI-II column data contain physically meaningful information relevant to boundary-layer ozone variability, despite inherent vertical sensitivity limitations. These findings demonstrate that EMI-II contributes as a stable and complementary constraint within multi-source estimation frameworks.

Third, the contribution of EMI-II is strongly atmosphere-regime dependent. Improved estimation performance occurs under conditions of strong photochemical activity and deep boundary-layer mixing (spring and summer), while reduced sensitivity is observed under stable stratification and weak radiative forcing (winter). This behavior reflects the dynamical and chemical modulation of column–surface ozone coupling, as supported by averaging kernel diagnostics and seasonal feature contribution analysis.

Methodologically, this study demonstrates a transferable framework for integrating validated satellite column products into surface pollutant estimation using data-driven approaches. The framework can be extended to other regions, longer time series, and additional atmospheric composition sensors.

Overall, EMI-II observations extend beyond total column monitoring and show clear potential for supporting regional surface ozone assessment when combined with auxiliary meteorological and chemical predictors. Future work should focus on: (1) quantifying vertical information content using more extensive averaging-kernel diagnostics and chemical transport model simulations; (2) extending the analysis to multi-year and multi-regional contexts to test generalizability; and (3) exploring integration with geostationary observations to capture diurnal variability and further improve estimation robustness.

Author Contributions: Conceptualization, H.C., Y.H. and J.C.; methodology, H.C., Z.Z. and Y.H.; software, H.C., Z.Z., K.Z. and Y.H.; validation, H.C. and J.C.; formal analysis, H.C.; investigation, H.C.; resources, H.C., Z.Z. and J.C.; data curation, K.Z.; writing—original draft preparation, H.C.; writing—review and editing, H.C. and J.C.; visualization, H.C.; supervision, J.C.; project administration, J.C.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by The Major Project of High Resolution Earth Observation System (Grant No. “30-Y60B01-9003-22/23”).

Data Availability Statement: The TROPOMI O₃ data are publicly available at ESA Copernicus Open Access Hub: <https://dataspace.copernicus.eu/>, accessed on 10 October 2025. The GEMS O₃ data are publicly available at <https://nesc.nier.go.kr/en/html/index.do>, accessed on 10 October 2025. The WOULD data are publicly available at <https://woudc.org/en/>, accessed on 10 October 2025. The EMI-II data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

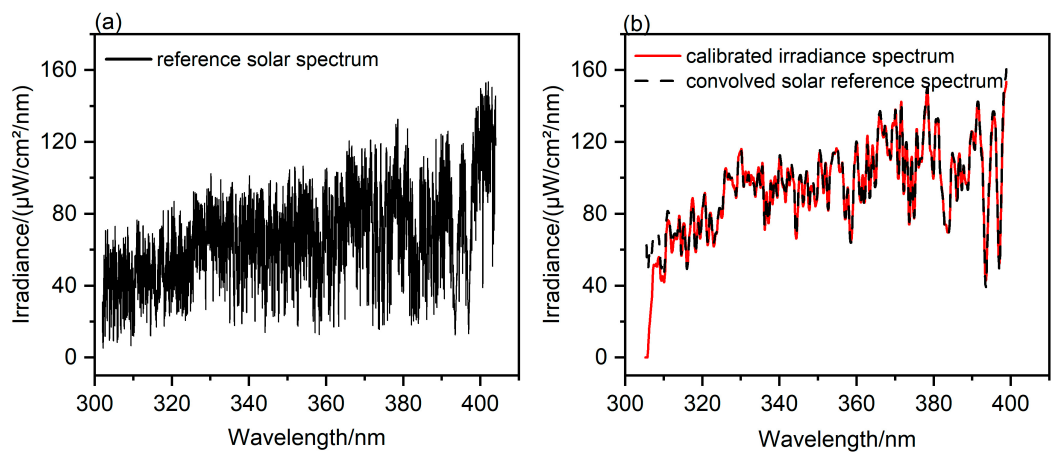


Figure A1. (a) the original reference solar spectrum; (b) spectrally calibrated irradiance from GF5B EMI-II on 3 April 2022. The figure shows that the spectral calibration significantly improves the alignment of Fraunhofer lines in the irradiance data.

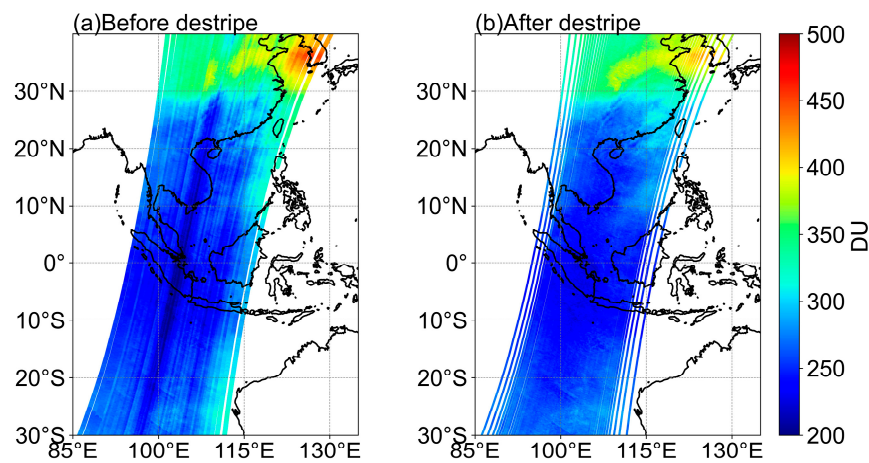


Figure A2. Comparison of TOCs along a single EMI-II orbit on 20 April 2022: (a) before and (b) after destriping correction.

References

1. An, T.; Li, J.; Lin, Q.; Li, G. Ozone formation potential related to the release of volatile organic compounds (VOCs) and nitrogen oxide (NO_x) from a typical industrial park in the Pearl River Delta. *Environ. Sci.: Atmos.* **2024**, *4*, 1229-1238, doi:10.1039/d4ea00091a.
2. Emberson, L. Effects of ozone on agriculture, forests and grasslands. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *378*, 20190327, doi:10.1098/rsta.2019.0327.

3. Hernandez, M.; Vadlamudi, A.; Ivins, S.; Chason, K.; Almond, M.; Peden, D. Low level ozone exposure at rest causes changes in lung function among healthy volunteers. *J. Allergy Clin. Immunol.* **2020**, *145*, AB82, doi:10.1016/j.jaci.2019.12.681.
4. Li, T.; Cheng, X. Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *101*, 102356, doi:10.1016/j.jag.2021.102356.
5. Chan, K.L.; Valks, P.; Heue, K.-P.; Lutz, R.; Hedelt, P.; Loyola, D.; Pinardi, G.; Van Roozendaal, M.; Hendrick, F.; Wagner, T.; et al. Global Ozone Monitoring Experiment-2 (GOME-2) daily and monthly level-3 products of atmospheric trace gas columns. *Earth Syst. Sci. Data* **2023**, *15*, 1831-1870, doi:10.5194/essd-15-1831-2023.
6. Levelt, P.F.; Joiner, J.; Tamminen, J.; Veefkind, J.P.; Bhartia, P.K.; Stein Zweers, D.C.; Duncan, B.N.; Streets, D.G.; Eskes, H.; van der A, R.; et al. The Ozone Monitoring Instrument: overview of 14 years in space. *Atmos. Chem. Phys.* **2018**, *18*, 5699-5745, doi:10.5194/acp-18-5699-2018.
7. Reshi, A.R.; Pichuka, S.; Tripathi, A. Applications of Sentinel-5P TROPOMI Satellite Sensor: A Review. *IEEE Sens. J.* **2024**, *24*, 20312-20321, doi:10.1109/jsen.2024.3355714.
8. Kim, J.; Jeong, U.; Ahn, M.-H.; Kim, J.H.; Park, R.J.; Lee, H.; Song, C.H.; Choi, Y.-S.; Lee, K.-H.; Yoo, J.-M.; et al. New Era of Air Quality Monitoring from Space: Geostationary Environment Monitoring Spectrometer (GEMS). *Bull. Am. Meteorol. Soc.* **2020**, *101*, E1-E22, doi:10.1175/bams-d-18-0013.1.
9. Zhao, M.; Si, F.; Zhou, H.; Jiang, Y.; Ji, C.; Wang, S.; Zhan, K.; Liu, W. Pre-Launch Radiometric Characterization of EMI-2 on the GaoFen-5 Series of Satellites. *Remote Sens.* **2021**, *13*, 2843, doi:10.3390/rs13142843.
10. Qian, Y.; Luo, Y.; Zhou, H.; Yang, T.; Xi, L.; Si, F. First Retrieval of Total Ozone Columns from EMI-2 Using the DOAS Method. *Remote Sens.* **2023**, *15*, 1665, doi:10.3390/rs15061665.
11. Wang, W.; Liu, X.; Bi, J.; Liu, Y. A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environ. Int.* **2022**, *158*, 106917, doi:10.1016/j.envint.2021.106917.
12. Jung, C.-R.; Chen, W.; Chen, W.-T.; Su, S.-H.; Chen, B.-T.; Chang, L.; Hwang, B.-F. A machine learning model for estimating daily maximum 8-h average ozone concentrations using OMI and MODIS products. *Atmos. Environ.* **2024**, *331*, 120587, doi:10.1016/j.atmosenv.2024.120587.
13. Garane, K.; Koukouli, M.-E.; Verhoelst, T.; Lerot, C.; Heue, K.-P.; Fioletov, V.; Balis, D.; Bais, A.F.; Bazureau, A.; Dehn, A.; et al. TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions. **2019**, *12*, 5263-5287, doi:10.5194/amt-12-5263-2019.
14. Baek, K.; Kim, J.H.; Bak, J.; Haffner, D.P.; Kang, M.; Hong, H. Evaluation of total ozone measurements from Geostationary Environmental Monitoring Spectrometer (GEMS). *Atmos. Meas. Tech.* **2023**, *16*, 5461-5478, doi:10.5194/amt-16-5461-2023.
15. Lange, K.; Richter, A.; Bösch, T.; Zilker, B.; Latsch, M.; Behrens, L.K.; Okafor, C.M.; Bösch, H.; Burrows, J.P.; Merlaud, A.; et al. Validation of GEMS tropospheric NO₂ columns and their diurnal variation with ground-based DOAS measurements. *Atmos. Meas. Tech.* **2024**, *17*, 6315-6344, doi:10.5194/amt-17-6315-2024.
16. Keller-Rudek, H.; Moortgat, G.K.; Sander, R.; Sörensen, R. The MPI-Mainz UV/VIS Spectral Atlas of Gaseous Molecules of Atmospheric Interest. *Earth Syst. Sci. Data* **2013**, *5*, 365-373, doi:10.5194/essd-5-365-2013.
17. Tilstra, L.G.; Tuinder, O.N.E.; Wang, P.; Stammes, P. Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth's surface measured by the GOME-2 satellite instruments. *Atmos. Meas. Tech.* **2021**, *14*, 4219-4238, doi:10.5194/amt-14-4219-2021.
18. Qiao, Z.; Liu, Y.; Cui, C.; Shan, M.; Tu, Y.; Liu, Y.; Xu, S.; Mi, K.; Chen, L.; Ma, Z.; et al. Estimation of Short-Term and Long-Term Ozone Exposure Levels in Beijing-Tianjin-Hebei Region Based on Geographically Weighted Regression Model. *Atmosphere* **2022**, *13*, 1706, doi:10.3390/atmos13101706.
19. Paschou, P.; Koukouli, M.-E.; Balis, D.; Lerot, C.; Van Roozendaal, M. The effect of considering polar vortex dynamics in the validation of satellite total ozone observations. *Atmos. Res.* **2020**, *238*, 104870, doi:10.1016/j.atmosres.2020.104870.

20. Wei, J.; Li, Z.; Li, K.; Dickerson, R.R.; Pinker, R.T.; Wang, J.; Liu, X.; Sun, L.; Xue, W.; Cribb, M. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O₃) pollution from 2013 to 2020 across China. *Remote Sens. Environ.* **2022**, *270*, 112775, doi:10.1016/j.rse.2021.112775.
21. Cheng, L.; Wang, Y.; Yan, H.; Tao, J.; Wang, H.; Lin, J.; Xu, J.; Chen, L. Preliminary Global NO₂ Retrieval from EMI-II Onboard GF5B/DQ1 and Comparison to TROPOMI. *Remote Sens.* **2024**, *16*, 4087, doi:10.3390/rs16214087.
22. Boersma, K.F.; Eskes, H.J.; Richter, A.; De Smedt, I.; Lorente, A.; Beirle, S.; van Geffen, J.H.G.M.; Zara, M.; Peters, E.; Van Roozendaal, M.; et al. Improving algorithms and uncertainty estimates for satellite NO₂ retrievals: results from the quality assurance for the essential climate variables (QA4ECV) project. *Atmos. Meas. Tech.* **2018**, *11*, 6651-6678, doi:10.5194/amt-11-6651-2018.
23. Bogumil, K.; Orphal, J.; Homann, T.M.; Voigt, S.; Spietz, P.; Fleischmann, O.C.; Vogel, A.; Hartmann, M.; Kromminga, H.; Bovensmann, H.; et al. Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230–2380 nm region. *J. Photochem. Photobiol. A Chem.* **2003**, *157*, 167-184, doi:10.1016/s1010-6030(03)00062-5.
24. Vandaele, A.C.; Hermans, C.; Simon, P.C.; Van Roozendaal, M.; Guilmot, J.M.; Carleer, M.; Colin, R. Fourier transform measurement of NO₂ absorption cross-section in the visible range at room temperature. *J. Atmos. Chem.* **1996**, *25*, 289-305, doi:10.1007/bf00053797.
25. Fleischmann, O.C.; Hartmann, M.; Burrows, J.P.; Orphal, J. New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy. *J. Photochem. Photobiol. A: Chem.* **2004**, *168*, 117-132, doi:10.1016/j.jphotochem.2004.03.026.
26. Meller, R.; Moortgat, G.K. Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 K in the wavelength range 225-375 nm. *J. Geophys. Res.: Atmos.* **2000**, *105*, 7089-7101, doi:10.1029/1999jd901074.
27. Chance, K.V.; Spurr, R.J.D. Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum. *Appl. Opt.* **1997**, *36*, 5224-5230, doi:10.1364/ao.36.005224.
28. Beirle, S.; Sihler, H.; Wagner, T. Linearisation of the effects of spectral shift and stretch in DOAS analysis. *Atmos. Meas. Tech.* **2013**, *6*, 661-675, doi:10.5194/amt-6-661-2013.
29. Boersma, K.F.; Eskes, H.J.; Veefkind, J.P.; Brinksma, E.J.; van der A, R.J.; Sneep, M.; van den Oord, G.H.J.; Levelt, P.F.; Stammes, P.; Gleason, J.F. Near-real time retrieval of tropospheric NO₂ from OMI. *Atmos. Chem. Phys.* **2007**, *7*, 2103-2118, doi:10.5194/acp-7-2103-2007.
30. Chen, Y.; Li, H.; Karimian, H.; Li, M.; Fan, Q.; Xu, Z. Spatio-temporal variation of ozone pollution risk and its influencing factors in China based on Geodetector and Geospatial models. *Chemosphere* **2022**, *302*, 134843, doi:10.1016/j.chemosphere.2022.134843.
31. Song, Y.; Wang, J.; Ge, Y.; Xu, C. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. *GISci. Remote Sens.* **2020**, *57*, 593-610, doi:10.1080/15481603.2020.1760434.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.