Systematic Review of Sentinel-5P Nitrogen Dioxide (NO_2) Validation and Surface Estimation Methods (2018–2025): Implications and Perspectives for Indonesia

Setiowati, A. O.,[1,2] Pok. S.[3] and Jaelani, L. M.[1*]
[1]Department of Geomatics Engineering, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, 60111, Indonesia, E-mail: ana.setiowati@bmkg.go.id, lmjaelani@its.ac.id*
[2]Regional Center II of BMKG, Banten 15221, Indonesia
[3]Faculty of Land Management and Land Administration, Royal University of Agriculture, Phnom Penh, Cambodia, E-mail: psophak@rua.edu.kh
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v22i10.5235

Abstract
This review examines validation and conversion of Sentinel-5P TROPOMI column data to near-surface concentrations, focusing on NO_2 and contextualizing CO, SO_2, O_3, and HCHO from 2018–2025. Following PRISMA guidelines, a three-stage Scopus search identified 1,022 studies, narrowed to 150 validation-focused studies. Studies were included if they (i) converted TROPOMI vertical column density to surface concentration and (ii) reported at least one quantitative metric (R^2, RMSE, MAE, or correlation) against ground observations. After full-text assessment, 15 met both criteria. A single reviewer conducted screening and extraction, a limitation. Extracted fields included conversion method, algorithm, reference type, region, and performance metrics. Bibliometric mapping showed a shift toward machine learning from 2021 onward, alongside a Global North–Global South disparity and limited research in the Maritime Continent. In this predominantly mid-latitude corpus, Random Forest, XGBoost, neural networks, and deep learning were most common and generally agreed more strongly with ground observations than direct column–surface correlations or chemical-transport-model scaling. However, no head-to-head comparisons on identical datasets existed, so this advantage is context-dependent. On held-out data, deep learning achieved validation R^2 of 0.914–0.919 across three annual UK test sets; Random Forest achieved $R^2 = 0.875$ (training $R^2 = 0.949$) in China's Fenwei Plain; and XGBoost achieved pan-European test-set MAE of 7.77 $\mu g/m^3$. Metrics represent validation/test performance unless noted otherwise. NO_2 dominates global validation, while evidence for CO, SO_2, and O_3 remains limited, particularly in tropical, cloudy regions with sparse monitoring. None of the 15 studies addressed the tropics or provides direct evidence for the Maritime Continent. In Indonesia, the critical bottleneck is the absence of Pandora or MAX-DOAS reference spectrometers. Installation in Jakarta and Surabaya should be prioritised, followed by staged NO_2, CO, and volcanic SO_2 estimation workflows adapted to tropical monsoon conditions.

Keywords: Indonesia, Machine Learning, Nitrogen Dioxide (NO_2), Sentinel-5P TROPOMI, Surface Concentration Estimation

1. Introduction

Satellite observations now provide the only spatially continuous view of air pollution available at global scale. The TROPOspheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 Precursor platform, operational since 2018, retrieves daily global columns of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), carbon monoxide (CO), and formaldehyde at a nadir resolution of 3.5 × 7 km² (5.5 × 3.5 km² after the 2019 orbit change), fine enough to resolve intra-urban gradients and individual emission sources that earlier instruments such as GOME, SCIAMACHY, and OMI could not [1][2][3][4] and [5]. Satellite data therefore reach the spatial scale at which air quality policy is made.

What satellite instruments measure, however, is not what determines human exposure. TROPOMI retrieves the vertical column density (VCD), the concentration integrated from the surface to the top of the atmosphere, whereas air quality standards, health assessments, and emission inventories refer to

near-surface concentrations. The conversion between the two is governed by planetary boundary layer height, the vertical sensitivity described by the averaging kernel, cloud and aerosol loading, surface reflectance, and the temporal sampling of a single daily overpass; these factors differ by region, so conversion relationships are region-dependent and no transferable formulation has been established [6][7] and [8]. Validation practice compounds the difficulty, because studies report accuracy against different reference networks, temporal aggregations, and data partitions, which prevents direct comparison between methods [9] and [10]. The physical mechanisms behind this disconnect are set out in Section 5.2.

Tropical maritime regions are almost absent from this evidence base, and Indonesia illustrates the gap. The country combines dense urban and industrial emissions in Jakarta and Surabaya, coal-fired power generation, seasonal peatland and forest fires in Sumatra and Kalimantan, and volcanic SO_2 along the Sunda Arc [11][12][13][14] and [15]. At the same time, persistent deep convection with annual mean cloud fractions above 70% over much of the landmass removes a large share of potential TROPOMI observations [16] and [17], while the ground-based network of approximately 50 BMKG monitoring stations, concentrated on Java, is sparse by international standards and includes no operational Pandora or MAX-DOAS reference instrument. The conditions governing column-to-surface conversion here differ fundamentally from those of the mid-latitude, reference-dense settings where current methods were developed, so methods validated there cannot be assumed to transfer.

Three gaps motivate this review. Scientifically, column-to-surface conversion remains region-dependent and physically incompletely constrained. Methodologically, heterogeneous validation practice prevents comparison between reported accuracies. Regionally, tropical maritime environments are under-represented in the validation record [18]. This review therefore pursues three objectives: (1) to identify and classify the column-to-surface conversion methods applied to TROPOMI data for NO_2, CO, SO_2, and O_3; (2) to assess their reported performance and the rigour of their validation protocols; and (3) to translate the global evidence into a staged roadmap for operationalising surface estimation over Indonesia. The roadmap is developed in Section 5.6, where the atmospheric, infrastructural, and methodological conditions specific to the Maritime Continent are examined in detail.

2. Methodology

To document the selection process transparently and to substantiate the PRISMA claim, a PRISMA 2020 flow diagram is provided as Figure 1, summarising the progression from identification to inclusion (1,022 records identified → 150 screened → 15 included) with the number of records excluded at each stage and the reasons for exclusion. This systematic review uses a rigorous Three-Phase Search and Selection Protocol (PRISMA-compliant) to isolate studies that specifically address the challenge of converting satellite Vertical Column Densities (VCD) into Surface Concentrations. The literature search was conducted on the Scopus database on January 5, 2026, covering the publication-year range from the launch of Sentinel-5P (2018) through the end of 2025. The title range "2018–2025" therefore refers to the publication-year window, not the search execution date. Scopus was selected because it indexes over 25,000 active peer-reviewed journals across environmental science, earth science, and public health, with complete coverage of Medline and comprehensive indexing of the remote sensing literature relevant to TROPOMI validation studies. Scopus also provides the RIS export format required for the dual-stream bibliometric workflow (Section 2.4.1). This is therefore a Scopus-based review; reliance on a single bibliographic database is acknowledged as a limitation: studies indexed exclusively in Web of Science, PubMed, Google Scholar, or regional databases may be underrepresented.

2.1 Phase 1: Broad Identification (The Global Dataset)

Studies were identified through a three-phase protocol applied to the Scopus database. Phase 1 (broad identification) retrieved 1,022 records covering standard TROPOMI trace gas products (NO_2, SO_2, O_3, and CO). Phase 2 (targeted validation screening) retained the 150 records that linked satellite-derived columns to ground-level measurements through quantitative comparison, excluding 872 records that did not combine satellite-geometry, ground-truth, and statistical terms. Phase 3 (manual eligibility check) applied the conversion criterion, separating direct comparisons from studies reporting a column-to-surface conversion method; 135 records were excluded because they reported no conversion method or no quantitative performance metric, leaving 15 studies. The eligibility criteria and the record counts at each stage are given in Table 1; the corresponding flow is shown in Figure 1.

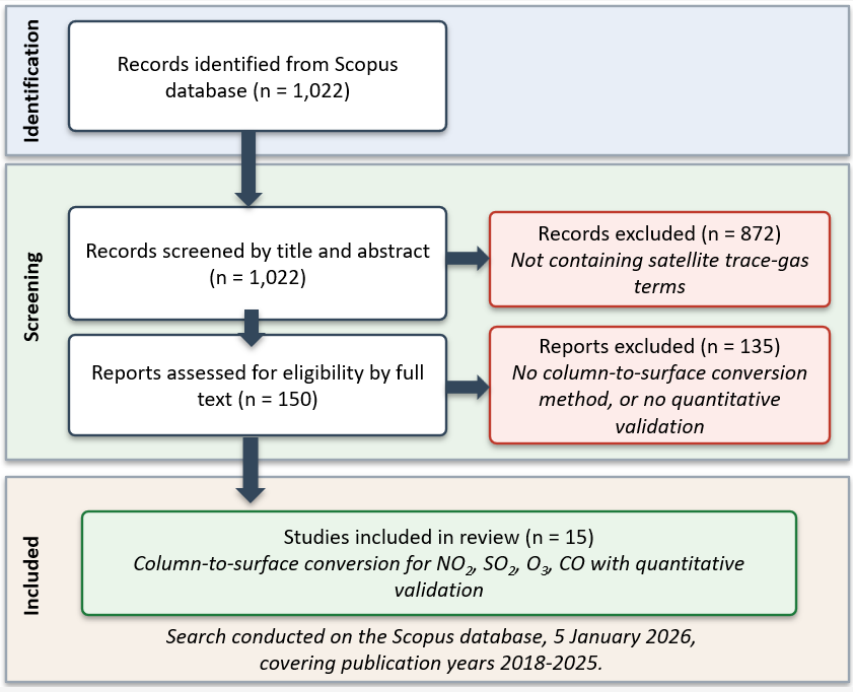

Figure 1: PRISMA 2020 flow diagram of study selection process

Table 1: Eligibility criteria and study attrition across the three-phase screening protocol

Phase	Criterion	Operational Definition	Records Remaining
Phase 1 (Identification)	TROPOMI + trace gas keyword match	Scopus search: TITLE-ABS-KEY contains any of {NO_2, NOx, SO_2, O_3, CO} AND any of {TROPOMI, Sentinel-5P, S-5P}. DOCTYPE = article or review. Language = English. Subject area = ENVI or EART.	n = 1,022
Phase 2 (Screening)	Satellite geometry + ground reference + validation terms	Script S2: TITLE-ABS-KEY must contain at least one term from EACH of three categories: (a) Column terms: column density, vertical column, total column, VCD, mixing ratio. (b) Ground terms: ground-level, surface concentration, in-situ, ground station, monitoring station. (c) Validation terms: validation, evaluation, accuracy, comparison, performance, agreement, correlation. Same DOCTYPE , language, and subject-area filters as Phase 1.	n = 150 (−872)
Phase 3 (Eligibility)	VCD-to-surface conversion method + quantitative metric	Manual full-text review. Both criteria must be met: (i) Applies a method to convert TROPOMI VCD → surface concentration (CTM scaling, ML model, statistical regression, or hybrid). (ii) Reports ≥1 quantitative metric (R^2, RMSE, MAE, r) against reference-grade or regulatory ground observations. Studies with only direct VCD–surface correlation excluded.	n = 15 (−135)
Final	All criteria satisfied	15 studies with VCD-to-surface conversion + quantitative validation against reference-grade networks.	n = 15

2.2 Phase 2: Targeted Validation Screening (The Validation Subset)

The second phase refined the initial set of 1,022 studies to identify those that attempted to link satellite-derived data with ground-level measurements through quantitative comparison. The search was refined using Script S2, which requires the simultaneous presence of terms from three predefined categories: satellite geometry (e.g., "column density," "vertical column"), ground reference (e.g., "in-situ," "ground station"), and validation or correlation (e.g., "validation," "comparison"). Only peer-reviewed articles and reviews (Scopus DOCTYPE "ar" and "re") in English were included; conference proceedings, book chapters, editorials, and preprints were excluded by the Scopus search filters. Title and abstract screening were conducted by a single reviewer (AS), which is acknowledged as a limitation. To reduce selection bias, the screening criteria were defined a priori as the simultaneous presence of at least one term from each of the three categories in the title, abstract, or keywords; studies meeting this condition were advanced to full-text review without discretionary exclusion. This phase narrowed the list to 150 records; because Phase 2 screening requires only the co-occurrence of column, ground-reference, and validation terms (not a conversion method), the subset may contain studies whose validation is limited to direct column surface comparisons; such studies are subsequently removed at Phase 3 unless they apply a VCD-to-surface conversion method (Section 2.3). The query strings used in Phase 1 and Phase 2 are given in full as Script S1 and Script S2 in Appendix A.

2.3 Phase 3: Manual Eligibility Check (The Conversion Criterion)

The 150 studies from Phase 2 were reviewed in detail to separate simple direct comparisons from more advanced conversion methods. Studies were classified into three tiers. Tier 1 (core conversion studies) met both criteria: (i) applied a method to convert TROPOMI vertical column density (VCD) into surface concentrations, such as chemical transport models (CTMs), machine learning approaches, or linear scaling factors; and (ii) reported at least one quantitative performance metric, coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), or Pearson/Spearman correlation coefficient (r), against reference-grade ground-based observations. Tier 2 (contextual validation studies) met Criterion 2 but not Criterion 1: they reported quantitative correlation or agreement metrics between VCD and surface data without employing a conversion model. These studies are retained as contextual evidence because the reliability of any column-to-surface conversion depends on the quality of the underlying retrieval; they are flagged as "Direct Validation" in Table 5 and are excluded from the performance synthesis. Tier 3 (excluded) failed one or both criteria. Screening was conducted by a single reviewer (AS); no second reviewer or inter-rater reliability assessment was performed, which is a recognised limitation. All 15 included studies (Tiers 1 and 2) validated against reference-grade or regulatory networks, not low-cost sensors.

A quantified breakdown of study attrition across the three phases is as follows. Phase 1 (broad identification) retrieved 1,022 records. Phase 2 (targeted validation screening) applied the refined query shown in Script S2, which required the simultaneous presence of satellite geometry terms (e.g., "column density," "vertical column"), ground-reference terms (e.g., "in-situ," "ground station"), and validation terms (e.g., "validation," "comparison"). This eliminated 872 records that did not contain all three term categories, for example, studies focused on retrieval algorithm development without ground comparison, or emission inventory studies that used TROPOMI data without ground-based validation. The remaining 150 records entered Phase 3 (manual eligibility check), where each full-text report was assessed against two criteria: (i) application of a method to convert TROPOMI VCD to surface concentrations, and (ii) reporting of at least one quantitative performance metric (R^2, RMSE, MAE, or correlation coefficient) against ground-based observations. Of the 150 reports, 135 were excluded for three reasons. First, 104 did not employ any column-to-surface conversion method: they reported only direct correlations between VCD and surface data, or focused exclusively on retrieval algorithm validation, and this group includes the COVID-19 lockdown studies that compared TROPOMI columns with surface data without applying a conversion method. Second, 18 applied a conversion method but reported only qualitative or visual comparisons without quantitative metrics. Third, 13 used low-cost sensors without regulatory certification as the sole ground reference; these were excluded on the basis of the reference-grade instrument criterion, though they were retained as contextual evidence. The final corpus comprised 15 studies meeting both inclusion criteria.

2.4 Data Cleaning and Preparation Protocol

A strict data-cleaning process was applied to the final set of 150 studies to ensure that bibliometric data were accurate and that citations could be managed easily.

2.4.1 Dual-stream data export

Two distinct datasets were generated from Scopus:

1. Analytical Dataset (CSV): Included full metadata, affiliations, references, and abstracts for VOSviewer analysis and manual screening.
2. Citation Dataset (RIS): a lightweight version (citations only) used solely for reference management in Mendeley. The full metadata (including affiliations and cited references) were retained in the CSV analytical dataset; the RIS file omits these fields only because they are not required for citation insertion, not as a data-quality judgement.

2.4.2 Standardisation and thesaurus application

A thesaurus was applied during pre-processing to standardise geographic labels and to harmonise author keywords. Administrative labels were merged to a single national label, while mainland China, Hong Kong, and Taiwan were retained as separate entities following standard bibliometric classification. Keyword synonyms were consolidated at the method-category level, so that algorithm names map to the machine-learning category while the algorithm identities remain distinct in Section 4 and in the merged table. The full mapping tables are provided in Tables 8 and 9.

2.4.3 Manual screening and classification

The Analytical Dataset was checked manually, row by row in Excel, to add information that was not included in the metadata:

- Conversion Method: Classified into Direct Correlation, Geophysical/CTM, Machine Learning, or Hybrid.
- Validation Region: Categorised by Köppen–Geiger climate zone (Tropical Af/Am/Aw, Temperate Cfa/Cfb/Csa/Csb, Continental Dfa/Dfb/Dfc, Arid BWh/BWk/BSk/BSh) based on the validation site coordinates. Cloud fraction statistics from ERA5 reported where available in the source studies [19] and [20].
- Ground Instrumentation: Identified as Reference Grade or Low-Cost Sensors. Low-cost sensors are defined here as devices without independent regulatory certification (e.g., US EPA FEM or EU equivalent) or relying solely on manufacturer-stated accuracy without peer-reviewed validation. Reference-grade denotes regulatory or research-network instruments (e.g., US EPA FRM, EU reference methods, Pandora, MAX-DOAS, NDACC FTIR). Performance validated against low-cost sensors is not equivalent to validation against regulatory networks; all 15 studies meeting the eligibility criteria in fact validate against reference-grade networks. Studies validated exclusively against low-cost sensors were excluded from the final corpus (13 studies at Phase 3).
- Data availability: The full cleaned dataset (CSV) and thesaurus files used for bibliometric analysis are available from the corresponding author upon reasonable request. Because the retained studies report performance metrics on different data partitions (training, validation, or test), temporal aggregations (hourly to annual), and accuracy measures (R^2, MAE, RMSE, correlation) that are not statistically comparable (Section 5.1), a meta-analysis was not performed; the evidence is synthesised narratively in accordance with PRISMA 2020 guidance, and pooled statistics (e.g., weighted averages, I^2 statistics, funnel plots) are therefore not reported.

3. Bibliometric Results

3.1 Thematic Evolution: From "Validation" to "Machine Learning"

The overlay visualisation of author keywords (Figure 2) shows a shift in the research focus of the field [21], from the assessment of the retrieved products themselves to the modelling of surface concentrations. Figure 2 is coloured by mean publication year, so the node colour of any individual term reflects the mean year of all records carrying that term rather than the date at which the term first appeared; the per-period frequencies on which the statements below rest are given in Table 8 in Appendix A. The first period, 2018–2020, is dominated by the terms that define the retrieval itself: the instrument name and the species retrieved. The concern of this period is the consistency of the column product, both against earlier instruments and against the ground networks then available. The terms naming those networks, including the ground-based spectrometers used as reference instruments, persist into the following period, which reflects the continuity of the validation effort rather than a change of subject. The third period, 2023–2025, is defined by two developments, and the year-by-year distribution of the 15 synthesis studies by method category is given in Table 2. The term for machine learning grows monotonically across the three periods (0, 3, and 6 occurrences) [22][23] and [24].

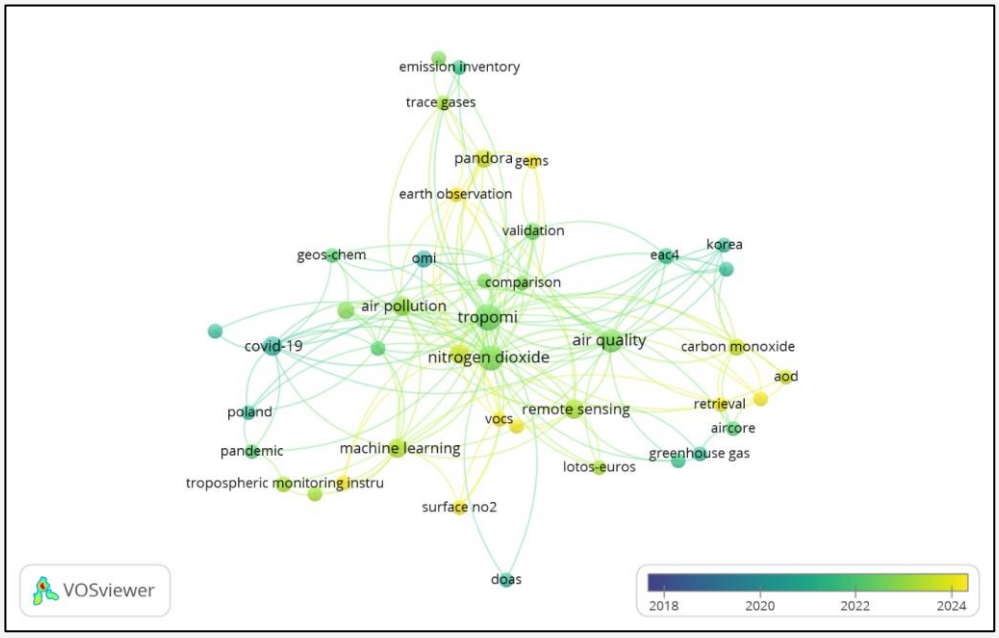

Figure 2: VOSviewer keyword co-occurrence map of author keywords (n = 150), coloured by mean publication year (2018–2025); node size reflects occurrence frequency

Table 2: Year-by-year distribution of the 15 synthesis studies by method category (publication year of each study as indexed in the original records)

Year	ML	CTM/Geophysical Scaling	Direct Validation	Statistical (GWR)	Total
2021	3	1	0	0	4
2022	0	0	2	0	2
2023	2	2	0	0	4
2024	1	0	1	0	2
2025	2	0	0	1	3
Total	**8**	**3**	**3**	**1**	**15**

The third period is also where terms naming the surface quantity itself appear for the first time, together with terms for multi-sensor and multi-parameter frameworks. The growth of the machine-learning term is a property of the wider Phase-2 corpus of 150 records rather than of the 15 synthesis studies: machine-learning conversion is the most frequent method category among the synthesis studies from the earliest years, not only after 2023 (Table 2) [18] and [22].

3.2 Geographic Distribution: The "North-South Divide"

The co-authorship network is dominated by Europe, China, and the United States. European countries show the highest collaboration density, followed by East Asia and the Americas. The entry of these blocs into the network is chronological: European countries form the earliest and densest cluster, consistent with the institutional collaboration established under the Copernicus programme at the start of the Sentinel-5P mission; the United States and China become prominent in the middle period; and the most recent expansion is into South and Southeast Asia. One limitation of the map must be stated with the result. A country appears as a node only if it meets the minimum-two-document threshold applied in VOSviewer. Absence from the network therefore indicates fewer than two publications meeting that threshold, and does not by itself establish an absence of research activity: the research may be published in non-English national journals, in venues not indexed in Scopus, or by authors affiliated with institutions abroad. The quantitative record counts reported in the following paragraph are therefore the basis for the claim about

geographic coverage, and the map is used only to describe the structure of collaboration.

Despite the recent expansion of the co-authorship network into South and Southeast Asia, the Maritime Continent is absent from the validation record. None of the 15 synthesis studies is sited in the tropics (Table 3). Of the 150 Phase-2 validation studies, 3 (2.0%) were conducted in tropical regions, in Mexico City, Jinja, and Kinshasa, and none in Southeast Asia (Table 4) [25][26][27] and [28]. None of the 15 synthesis studies is sited in the tropics (Table 3). A targeted search of the 1,022-record corpus identified four Indonesia-related studies, all of which were excluded at Phase 2 because none reported a quantitative comparison against in-situ or reference-grade ground observations, and a cross-check of affiliations in the full Scopus export returned no Indonesian, Malaysian, Singaporean, or Philippine records. Two distinct claims should not be conflated here. Absence from the co-authorship map follows from the minimum-two-document threshold applied in VOSviewer, and is not evidence about research activity. Absence from the review corpus follows from the eligibility criteria, which require a quantitative ground-reference comparison. The tropical gap documented here is the second kind: it is a property of the published ground-validated TROPOMI literature, not an artefact of the visualisation threshold.

3.3 Methodological Implications

The juxtaposition of these two network maps creates a compelling narrative of scientific progress contrasted with systemic geographic exclusion. While the keyword map demonstrates a significant technical leap toward "machine learning" and interpretive frameworks (Figure 2), the country co-authorship map (Figure 3) reveals that this computational innovation remains highly confined to specific affluent clusters in the Global North and advanced East Asian economies [21].

Table 3: Country/region of the 15 synthesis studies (shares rounded to one decimal place; the shares sum to 100.3 because of rounding)

Country / Region	Studies	Share (%)
China	4	26.7
East Asia (multi-country)	2	13.3
Germany	1	6.7
Switzerland	1	6.7
South Korea	1	6.7
Finland	1	6.7
Poland	1	6.7
Spain	1	6.7
Europe (multi-country)	1	6.7
Lithuania	1	6.7
United Kingdom	1	6.7
Total	**15**	**100**

Table 4: Geographic classification of the 150 Phase-2 validation studies (by validation site). The Köppen class A and Tropical highland rows are subsets of the Tropical row and are not additive

Class	Definition	Count	% of 150
Tropical (latitudinal)	Validation site 23.5° N–23.5° S	3	2.0
Köppen class A (Af/Am/Aw)	Tropical rainforest / monsoon / savanna	2	1.3
Tropical highland (Cwb)	Mexico City, 19.4° N	1	0.7
Southeast Asia	Indonesia, Malaysia, Singapore, Thailand, Viet Nam, Philippines, Brunei, Cambodia, Laos, Myanmar	0	0.0
All other studies	Temperate, subtropical, arid, and global/multi-site networks	147	98.0
Total	Total validation sample size	**150**	**100.0**

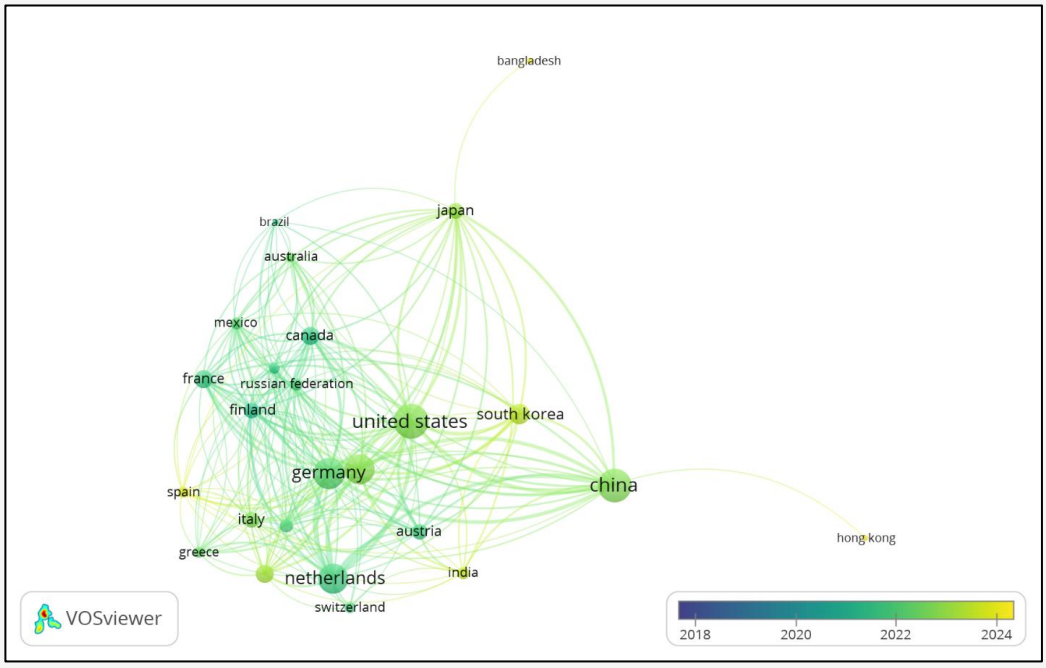

Figure 3: Origin of co-authorship map. Country nodes are sized by document count and connected by co-authorship link strength (VOSviewer; association-strength normalisation; minimum two documents per country)

The Ground-Reference Bottleneck:
This spatial disparity raises important questions about model transferability. The co-authorship map shows that the countries producing the most ML-based validation studies are concentrated in temperate regions. This pattern does not prove that ML models are trained exclusively on temperate-zone data, that depends on the specific training data used in each study, but it indicates that the geographic distribution of ML validation expertise follows the broader Global North research infrastructure pattern documented in Section 3.2. Direct evidence of training data origin would require per-study data provenance reporting, which is not currently standard in the reviewed literature. The transferability of these models to tropical conditions should therefore be treated as an empirical question requiring dedicated validation. Machine learning models require training datasets that adequately sample the variability of the target variable; in the reviewed studies, ground station counts ranged from 13 to 3,044 and temporal coverage from 3 months to 5 years. Tropical regions, where ground monitoring infrastructure is sparse, cannot provide training data of comparable density. Three mechanisms prevent direct transferability of mid-latitude-trained ML models to tropical conditions: (i) the vertical NO_2 profile differs systematically between shallow mid-latitude winter boundary layers and deep tropical convective boundary layers, altering the column-to-surface relationship [6][8] and [29]; (ii) tropical cloud cover reduces the number of valid satellite observations by 60–80% compared to mid-latitudes, biasing the training sample toward clear-sky conditions [9][30] and [31]; and (iii) the dominant aerosol types, biomass burning in the tropics versus urban-industrial in mid-latitudes, affect the air mass factor differently [6] and [32]. These mechanisms, combined with the spatial cross-validation literature [33] and [34], indicate that model performance under domain shift should be treated as an empirical question rather than assumed. Spatial cross-validation studies document RMSE increases of 30–60% when air quality models trained in temperate regions are applied to tropical environments [8] and [33].

The Risk of Out-of-Distribution (OOD) Degradation:
Because the compiled data contains zero empirical studies trained, tuned, or evaluated under tropical maritime conditions, the transferability of these AI-driven surface concentration models to the cloud-prone, high-humidity, and complex aerosol environments of the Maritime Continent (such as Indonesia) remains entirely untested. Applying regression, ensemble, or deep learning architectures trained on mid-latitude data to equatorial contexts constitutes an out-of-distribution (OOD) deployment. This is a well-recognised source of degraded predictive performance in spatial data

science. Consequently, the performance of current state-of-the-art retrieval frameworks over Indonesia must be treated as an unverified hypothesis rather than an established operational outcome, underscoring the urgent need for localised reference infrastructure and regional collaborative inclusion [35].

4. Methodological Approaches to Surface Estimation

4.1 Overview of Classification

The 15 selected studies fall into three methodological approaches, which differ in how they represent the column-to-surface relationship: machine learning, which learns the mapping empirically from predictors; geophysical scaling, which derives it from modelled vertical profiles; and statistical or direct validation, which characterises the relationship without converting it. The corpus is small by the standards of adjacent review fields, and the small size follows from the two eligibility criteria rather than from the state of the literature: a study must apply a conversion method and must report a quantitative metric against a reference-grade network. The narrow evidence base limits the generalisability of the performance comparisons reported below, and the findings are therefore treated as indicative rather than conclusive. The studies, their categories, and their ground references are listed in Table 5.

4.2 Machine Learning Approaches

Eight of the 15 studies use machine learning [36][37][38][39][40][41][42] and [43]. Tree-based ensembles predominate, accounting for six of the eight, while a neural network and a deep-learning model account for the remaining two. No single architecture consistently outperformed the others across the regions and pollutants represented in the corpus. The predictor sets used by these eight studies are compared in Table 6.

Table 5: Selected studies by methodological category

ID	Region	Pollutant	Method Category	Specific Method	Ground Reference	References
1	Germany	NO_2	Machine Learning	Neural Network	UBA Stations	[36]
2	East Asia	NO_2, O_3	Machine Learning	RF, XGBoost, LGBM, SVR	Ground Stations	[37]
3	Switzerland	NO_2	Machine Learning	XGBoost	NABEL Stations	[38]
4	South Korea	CO	Geophysical Scaling	CAMS EAC4 Scaling	AirKorea	[44]
5	Jiangsu, China	NO_2	Direct Validation	Correlation Analysis	CNEMC & MAX-DOAS	[45]
6	Xinjiang, China	NO_2	Direct Validation	Correlation & GRA	Ground Stations	[46]
7	Finland	NO_2	Geophysical Scaling	GEOS-Chem CTM Scaling	FMI Stations	[47]
8	Poland	NO_2	Machine Learning	Random Forest	GIOŚ Stations	[39]
9	East Asia	NO_2	Geophysical Scaling	GEOS-CF Retrieval	Pandora	[48]
10	Fenwei Plain, China	NO_2	Machine Learning	Random Forest	Ground Stations	[40]
11	Madrid, Spain	NO_2	Direct Validation	Correlation Analysis	CNEMC Stations	[49]
12	Europe	NO_2	Machine Learning	XGBoost	EEA Stations	[41]
13	Lithuania	NO_2	Machine Learning	Bagged Trees, Coarse Tree	EPA Stations	[42]
14	England	NO_2	Machine Learning	Deep Learning (SLNet)	AURN Stations	[43]
15	Shandong, China	NO_2	Statistical Model	GWR	Ground Stations	[50]

Table 6: Comparative summary of feature inputs used across machine learning studies for TROPOMI column-to-surface NO_2 estimation

ID	Region	Algorithm	VCD	MET	POP	ROD	NTL	LULC	ELE	TMP	CTM	Ref.
1	Germany	Neural Network	✓	✓	—	—	—	—	—	✓	—	[36]
2	East Asia	RF / XGBoost / LightGBM / SVR	✓	✓	✓	—	—	✓	✓	✓	—	[37]
3	Switzerland	XGBoost	✓	✓	—	—	—	✓	✓	✓	—	[38]
8	Poland	Random Forest	✓	✓	—	—	—	—	—	✓	—	[39]
10	Fenwei Plain, China	Random Forest	✓	✓	—	—	—	✓	✓	—	—	[40]
12	Europe	XGBoost (S-MESH)	✓	✓	—	—	✓	✓	—	✓	—	[41]
13	Lithuania	Bagged Trees / Coarse Tree	✓	✓	—	—	—	—	—	—	—	[42]
14	England	Deep Learning (SLNet)	✓	✓	—	✓	—	✓	—	✓	✓	[43]
		Total	**8**	**8**	**1**	**1**	**1**	**5**	**3**	**6**	**1**	

Algorithm choice appears to have been less influential than the choice of auxiliary predictors and the representativeness of the training data. The specific methods used by each study are listed in Table 5. Four cautions apply across these studies and limit the generalisability of their findings. Feature sets differ substantially between studies, so reported performance gains are not always attributable to the same predictors [36][39] and [41] (Table 6). Uncertainty quantification is largely absent, with only one study reporting prediction or credible intervals [41]. The training corpora are small and unevenly sampled in space and time, which raises the risk of overfitting where spatial autocorrelation is not controlled in the train-test split. Validation strategies also differ across the corpus, which prevents direct comparison of the reported metrics. Machine-learning studies for TROPOMI NO_2 column-to-surface conversion included in the synthesis are summarised in Table 9 (Appendix A).

4.3 Geophysical Scaling and Chemical Transport Models (CTM)

Three of the 15 studies convert columns using vertical profiles from a chemical transport model or reanalysis: GEOS-Chem for surface NO_2 in Finland, CAMS EAC4 for CO in South Korea, and GEOS-CF in a GEMS retrieval application [44][47] and [48]. The physical interpretability of this approach is its principal advantage, since the vertical profile represents transport and mixing explicitly; its principal limitation is dependence on the fidelity of that profile, which is least well constrained in data-sparse tropical regions. One of the three raises a specific circularity: CAMS EAC4 assimilates satellite CO retrievals, including TROPOMI, so the profiles used to convert the column are not independent of the column being converted, and the reported agreement is not an independent validation [44]. The corpus contains three CTM studies rather than more because 12 further Phase-2 studies used CTM or reanalysis profiles for scaling but reported only qualitative or model-to-model comparisons [47] and [48].

A hybrid approach may combine the two paradigms: the CTM supplies a physically grounded estimate of the vertical profile and boundary layer structure, and a machine-learning model learns the residual bias against observed surface concentrations. None of the 15 reviewed studies adopts this architecture, but the approach is attractive for Indonesia precisely because the constraint there is the scarcity of ground data for training rather than the availability of modelled profiles. The computational cost differs by orders of magnitude, since CTM simulations are impractical for operational nowcasting while machine-learning inference is near-instantaneous once trained. A hybrid design that uses modelled profiles as physical priors is therefore the most plausible pathway to operational surface estimation over the Maritime Continent.

4.4 Statistical and Direct Validation Studies

Four studies (4 of 15) are classified under this category, comprising one statistical modelling study

(GWR) and three direct validation studies. Direct validation studies play a foundational role in the conversion evidence chain: they assess whether the satellite column carries sufficient surface-level information to warrant conversion, and they provide the baseline correlation against which more complex methods can be evaluated. The three direct validation studies in this corpus [45][46] and [49] reported VCD–surface correlation coefficients of r = 0.78–0.90 at monthly scales, confirming that TROPOMI NO_2 columns contain meaningful surface information even without a dedicated conversion model. Geographically Weighted Regression (GWR): one study applied GWR to model the spatial heterogeneity of the VCD–surface NO_2 relationship in Shandong, China [50] GWR characterises how the correlation between satellite and ground data varies across space, it does not perform a physical conversion of column density to surface concentration. Consistent with the Tier 2 classification used for direct validation studies, the Shandong study [50] is included as a contextual study that provides insight into the spatial structure of the satellite-surface relationship rather than a conversion method per se. Direct Comparison and Correlation: Three studies focused primarily on validating the satellite product itself rather than developing conversion models. TROPOMI and OMI products were validated against MAX-DOAS and CNEMC stations in Jiangsu [45]. A spatiotemporal characterisation of NO_2 in Xinjiang used correlation analysis and Grey Relational Analysis (GRA) [46]. A direct validation of TROPOMI NO_2 against ground stations in Madrid [49].

5. Discussion

5.1 Performance of Machine Learning and Deep Learning Models

Across the 15 studies, reported Pearson correlations range from 0.51 to 0.96 (median 0.80) and coefficients of determination from 0.59 to 0.92 (median 0.84) [36][37][38][39][40][41][42] and [43]. These figures are not directly comparable, because they refer to different temporal aggregations, spatial resolutions, and evaluation protocols: a value may describe training fit, cross-validation, or held-out test performance, and the reported metric may be a correlation coefficient, a coefficient of determination, or an absolute error. Read collectively, they point to region and reference-data quality rather than method choice as the dominant source of between-study variation: a neural network in Germany and a GEOS-Chem scaling in Finland achieved the same correlation through different paradigms [36] and [47], whereas the highest and lowest validation values in the corpus differ by region. No study reported confidence intervals, and on the corpus median of 87 training stations an interval around the highest validation value spans roughly 0.87 to 0.94, which is wide enough that differences of about 0.03 between studies are not statistically distinguishable. The individual metrics reported by each study are listed in Table 7 (Appendix A).

Placed against regulatory benchmarks, the reported errors are small relative to the 40 µg/m³ annual NO_2 standard applied in the European Union and Indonesia: absolute errors of 3.1 to 3.7 µg/m³ correspond to roughly 8–9% of that threshold, 5.12 µg/m³ to about 13%, and 7.7 to 7.8 µg/m³ to about 19%. These levels are adequate for hotspot identification, exposure ranking, and trend detection. They are marginal against the WHO 2021 annual guideline of 10 µg/m³, where the same errors represent roughly 30–80% of the guideline value, and are therefore insufficient for near-threshold compliance decisions. The reported values support regional exposure assessment rather than station-level regulatory inference [8] and [41]. Training fit and validation performance answer different questions, since the first quantifies reproduction of the training data and the second approximates out-of-sample accuracy for a particular partition.

5.2 Importance of TROPOMI and Auxiliary Predictors

The TROPOMI column is the most important single predictor of surface NO_2 in every study that quantified predictor importance, which is expected, since predicting the surface concentration from the column is the task the studies were designed for. The informative quantity is therefore not the ranking of the column but the incremental contribution of the predictors added to it. Two studies quantify that contribution directly: a leave-one-out experiment on a Random Forest model for Poland, and a Shapley-value analysis of the European S-MESH model [39] and [41]. Meteorological predictors play a direct physical role in the column-to-surface relationship, and they are not ancillary to it. Planetary boundary layer height controls vertical dilution: for a fixed emission flux, a shallow boundary layer concentrates the pollutant near the surface, while a deep, well-mixed boundary layer distributes the same mass over a larger volume and changes the ratio of column to surface concentration. The leave-one-out experiment supports this physically: removing boundary layer height from the Polish model reduced hourly accuracy by 16%, the largest decrease among the meteorological predictors [39]. Cloud fraction appears in no study as an explicit predictor, but it governs data availability, since quality filtering

removes cloudy scenes and thereby calibrates models implicitly on clear-sky conditions.

Anthropogenic proxies behave differently. Night-time lights, road density, population density, and land cover do not modulate the column-to-surface relationship; they constrain the spatial distribution of emissions at sub-pixel scales, which the satellite column cannot resolve. Their contribution is substantial where measured: night-time lights ranked second to the TROPOMI column in the European model, with an average contribution of about 18%, and removing them from the Polish model reduced hourly accuracy by 19%, second only to removing the column itself [39][41] and [51]. Because these two families of predictors act through different mechanisms, treating them as a single undifferentiated set of auxiliary variables obscures which is which, particularly when a model is transferred to a region with a different climatology. Transferability to Indonesia differs between the two families. Meteorological predictors are available globally from the ERA5 reanalysis [19][20] and [52] and transfer in availability, but not in meaning: the seasonal cycle of performance in the mid-latitude studies is driven by winter heating and shallow boundary layers, whereas in Indonesia it is driven by the wet-dry monsoon. Anthropogenic proxies transfer in neither respect, since the local emission maps and land-cover products on which they depend are less well constrained for the Maritime Continent [53].

5.3 Spatiotemporal Variations and Model Validity

Every study in the corpus was conducted in a mid-latitude setting, and none in the tropics. The relationships they report are therefore hypotheses about tropical behaviour rather than results that transfer to it. The physical reason matters more than the geography: in the reviewed studies the seasonal cycle of model performance is driven by winter heating emissions and shallow boundary layers, which raise near-surface concentrations and strengthen the surface imprint in the column. In the Maritime Continent the seasonal cycle is instead driven by the wet-dry monsoon, so the mechanism that produces the mid-latitude pattern does not operate in the same way. The reviewed methods succeed where the column carries a strong surface imprint, which is precisely the condition that tropical convection and deep boundary layers weaken. Tropical performance must therefore be established empirically rather than assumed from the mid-latitude results.

5.4 The Geostationary Inflection Point: Implications for Indonesia and Southeast Asia

TROPOMI's dominance of the validation literature is unlikely to persist. GEMS, launched in February 2020, was the first geostationary instrument able to resolve the diurnal variation of tropospheric NO_2, and its observation domain already covers the northern Maritime Continent, including northern Sumatra, Kalimantan, Sulawesi, and much of Maluku [54][55][56] and [57]. TEMPO, launched in August 2023, provides the equivalent capability over North America [58][59] and [60]. Both instruments sample hourly, which means that the diurnal cycle, invisible to a polar-orbiting instrument, becomes observable [61]; validation work in East Asia has shown that the geostationary products require independent regional validation rather than being assumed consistent with the polar-orbiting record [56] and [57]. Three implications follow for Indonesia. Validation infrastructure should be planned for a multi-satellite environment in which GEMS, TROPOMI, and Sentinel-4 operate concurrently, rather than treating TROPOMI as the sole reference. The Maritime Continental portion of the GEMS footprint is under-explored, which is an opportunity for original regional contributions, but exploiting it requires reference-grade instruments on the ground. Hourly observations also enable a class of models that a single daily overpass cannot support, in which observations from preceding hours are used as predictors; this is particularly suited to the tropical diurnal cycle, which is modulated by deep convective mixing and afternoon thunderstorms.

5.5 Methodological Frontiers Beyond NO_2

Methodological maturity is not distributed evenly across pollutants, and the difference in readiness is large enough to determine the order in which species should be attempted in Indonesia. Carbon monoxide is the most developed of the four non-NO_2 species. A simple linear relationship between the TROPOMI column and the surface concentration has been shown to hold [44][62] and [63], and the species is therefore ready for adaptation to Indonesian conditions, particularly for the peatland-fire season. Sulfur dioxide is the most constrained. Only four SO_2-relevant studies appear in the Phase 2 subset [64][65][66] and [67], which reflects the difficulty of retrieving a species with a short atmospheric lifetime and spatially concentrated sources; the volcanic plumes over the Sunda Arc are a case in point [66][68] and [69]. Formaldehyde has been validated extensively, but against ground-based Fourier-transform spectrometry rather than against surface concentrations [70] and [71].

The validation question for HCHO is therefore about the column rather than about the conversion, and the surface-conversion problem has not been posed for this species. Ozone is the least advanced for surface estimation. The tropospheric ozone product carries a substantial stratospheric contribution, and the studies in the corpus do not attempt a column-to-surface conversion comparable to the NO_2 workflows [68][72][73] and [74]. The order of readiness for Indonesia is therefore CO first, then NO_2, with SO_2, HCHO, and O_3 requiring dedicated retrieval and validation work before surface estimation is attempted.

5.6 The Indonesian Context: Atmospheric Conditions, Monitoring Infrastructure, and a Roadmap for Local Validation

Indonesia's air quality is shaped by three interacting source categories: urban and industrial emissions, concentrated in Jakarta, Surabaya, and other metropolitan areas; seasonal peatland and forest fires in Sumatra and Kalimantan during the dry season, which emit large quantities of carbon monoxide and aerosol; and volcanic degassing along the Sunda Arc, which produces sulfur dioxide plumes. The country lies within the Maritime Continent, where persistent deep convection produces annual mean cloud fractions exceeding 70% over much of the landmass and reduces the number of valid satellite observations. This matters for retrieval rather than only for sampling: air-mass-factor structural uncertainty is largest precisely under the cloud- and aerosol-dominated, deep-convective conditions that prevail over the region [16] and [17]. The retrieval settings applied over Indonesia were developed and validated against mid-latitude reference data, and their behaviour under tropical conditions is therefore not established.

The ground-based monitoring network is sparse by international standards, consisting of approximately 50 air quality monitoring stations operated by BMKG and concentrated on Java, with no reference-grade spectrometer of the Pandora or MAX-DOAS class operating anywhere in the country. This constrains the problem in two ways. It limits the training data available for data-driven methods, since models require ground observations that sample the variability they are asked to predict. It also changes what the predictors mean: meteorological predictors are available globally from reanalysis products and therefore transfer in availability, but not in meaning, because the seasonal cycle of model performance in the reviewed studies is driven by winter heating emissions and shallow boundary layers, whereas the Indonesian cycle is driven by the wet-dry monsoon. Four actions follow, staged by readiness. Deploy at least four Pandora or MAX-DOAS reference spectrometers at representative sites covering urban, fire-prone, and volcanic environments, beginning with Jakarta and Surabaya, to provide the reference data the country currently lacks. Target nitrogen dioxide surface estimation as the near-term priority, given the methodological maturity documented for that species [47]. Operationalise carbon monoxide surface estimation for the peatland-fire season using the linear scaling approach established in the reviewed literature, which is the only species for which a conversion method is already demonstrated [44]. Establish collaborative validation programmes with the neighbouring GEMS consortium countries, since the Maritime Continental portion of the geostationary footprint is under-explored and shared infrastructure is more efficient than duplicated national efforts.

6. Conclusion

This review set out to determine what the global TROPOMI literature can establish about converting satellite column measurements to the near-surface concentrations that determine human exposure, and what that evidence implies for Indonesia. The principal finding is that data availability, not methodological sophistication, is the primary constraint on conversion accuracy: the highest reported performance comes from the region with the densest reference network, not from the most advanced algorithm. Multi-method comparisons within the corpus show that relative method performance is region- and pollutant-dependent, and no head-to-head comparison of machine learning against chemical transport modelling on identical data exists in the reviewed literature. The practical consequence is that the accuracy reported for mid-latitude applications supports hotspot identification, exposure ranking, and trend detection, but not near-threshold regulatory compliance decisions. The apparent advantage of machine learning is therefore context-dependent rather than general, and it is not a substitute for reference infrastructure.

The tropical validation gap is the central limitation of the current evidence base. None of the studies meeting the inclusion criteria was conducted in a tropical maritime setting, and the absence is structural rather than incidental: it follows from the global distribution of reference-grade monitoring infrastructure and is reinforced by eligibility criteria that require quantitative validation against such networks. A further limitation of this review is that its screening was performed by a single reviewer, and that the Scopus-only, English-language search may under-represent research published in national

journals or grey literature. Because the deep convective boundary layer, persistent cloud cover, and biomass-burning aerosol of the Maritime Continent produce error structures absent from the mid-latitude corpora, the transferability of existing methods to Indonesia should be treated as an open empirical question rather than an established result.

Four actions follow for research and method development. First, reference infrastructure should be established: at least four Pandora or MAX-DOAS spectrometers at representative sites, covering urban, fire-prone, and volcanic environments, are required before the most basic validation step can be performed over Indonesian territory. Second, NO_2 surface estimation should be pursued as the near-term priority, using a hybrid framework that combines physical priors with machine learning and is evaluated by spatial hold-out against national monitoring stations. Third, CO surface estimation should be operationalised for the peatland-fire season, with local calibration for tropical conditions. Fourth, collaborative validation programmes with neighbouring geostationary-constellation countries would generate regionally representative statistics while domestic reference infrastructure is developed. None of the reviewed methods has been validated for regulatory or policy application in Indonesia; operational use would require dedicated uncertainty quantification and independent campaign validation. The transition to geostationary sensors offers the most promising longer-term pathway, because the diurnal sampling density they provide is what physically meaningful conversion models require in regions where ground networks remain sparse.

References

[1] Burrows, J. P., Weber, M. and Buchwitz, M., (1999). The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results. *Journal of the Atmospheric Sciences,* Vol. 56(2), 151–175. https://doi.org/10.1175/1520-0469(1999)056%3C0151:TGOMEG%3E2.0.CO;2.

[2] Bovensmann, H., Burrows, J. P. and Buchwitz, M., (1999). SCIAMACHY: Mission Objectives and Measurement Modes. *Journal of the Atmospheric Sciences,* Vol. 56(2), 127–150. https://doi.org/10.1175/1520-0469(1999)056<0127:smoamm>2.0.co;2.

[3] Levelt, P. F., van den Oord, G. H. J. and Dobber, M. R., (2006). The Ozone Monitoring Instrument. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 44(5), 1093–1101. https://doi.org/10.1109/tgrs.2006.872333.

[4] Veefkind, J. P., Aben, I. and McMullan, K., (2012). 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 Sensing of Environment,* Vol. 120, 70–83. https://doi.org/10.1016/j.rse.2011.09.027.

[5] van Geffen, J., Boersma, K. F. and Eskes, H., (2020). S5P TROPOMI NO_2 Slant Column Retrieval: Method, Stability, Uncertainties and Comparisons with OMI. *Atmospheric Measurement Techniques,* Vol. 13(3), 1315–1335. https://doi.org/10.5194/amt-13-1315-2020.

[6] Boersma, K. F., Eskes, H. J. and Dirksen, R. J., (2011). An Improved Tropospheric NO_2 Column Retrieval Algorithm for the Ozone Monitoring Instrument,. *Atmospheric Measurement Techniques,* Vol. 4(9), 1905–1928. https://doi.org/10.5194/amt-4-1905-2011.

[7] Lamsal, L. N., Martin, R. V. and van Donkelaar, A., (2008). Ground-Level Nitrogen Dioxide Concentrations Inferred from the Satellite-Borne Ozone Monitoring Instrument. *Journal of Geophysical Research,* Vol. 113(D16). https://doi.org/10.1029/2007jd009235.

[8] Chang, B., Liu, H., Zhang, C., Xing, C., Tan, W. and Liu, C., (2025). Relating Satellite NO_2 Tropospheric Columns to Near-Surface Concentrations: Implications from Ground-Based MAX-DOAS NO_2 Vertical Profile Observations. *npj Climate and Atmospheric Science,* Vol. 8(1), 1. https://doi.org/10.1038/s41612-024-00891-z.

[9] Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J. C., Eskes, H., Eichmann, K. U., Fjaeraa, A. M., Granville, J., Niemeijer, S., Cede, A. M., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K. F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G. H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P. F., Liu, C., Müller, M., Navarro-Comas, M., Piters, A., Pommereau, J. P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J. P., Wagner, T., Wittrock, F., Yela, M. and Zehner, C., (2021). Ground-based Validation of the Copernicus Sentinel-5P TROPOMI NO_2 Measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia Global Networks. *Atmospheric Measurement Techniques,* Vol. 14(1), 481–510. https://doi.org/10.5194/amt-14-481-2021.

[10] Ialongo, I., Virta, H., Eskes, H., Hovila, J. and Douros, J., (2020). Comparison of TROPOMI/Sentinel-5 Precursor NO_2 Observations with Ground-Based Measurements in Helsinki. *Atmospheric Measurement Techniques,* Vol. 13(1), 205–218. https://doi.org/10.5194/amt-13-205-2020.

[11] Ciptaningayu, T. N., Syafei, A. D., Santoso, M. and Tangahu, B. V., (2021). Identification of Pollutant Sources on PM10: Case Study in West Surabaya. *IOP Conference Series: Materials Science and Engineering,* Vol. 1144(1). https://doi.org/10.1088/1757-899x/1144/1/012059.

[12] Santoso, M., Hopke, P. K., Permadi, D. A., Damastuti, E., Lestiani, D. D., Kurniawati, S., Khoerotunnisya, D. and Sukir, S. K., (2021). Multiple Air Quality Monitoring Evidence of the Impacts of Large-Scale Social Restrictions During the COVID-19 Pandemic in Jakarta, Indonesia. *Aerosol and Air Quality Research,* Vol. 21(8). https://doi.org/10.4209/aaqr.200645.

[13] Permadi, D. A., Kim Oanh, N. T. and Vautard, R., (2018). Integrated Emission Inventory and Modeling to Assess Distribution of Particulate Matter Mass and Black Carbon Composition in Southeast Asia. *Atmospheric Chemistry and Physics,* Vol. 18(4), 2725–2747. https://doi.org/10.5194/acp-18-2725-2018.

[14] Kiely, L., Spracklen, D. V., Arnold, S. R., Papargyropoulou, E., Conibear, L., Wiedinmyer, C., Knote, C. and Adrianto, H. A., (2021). Assessing Costs of Indonesian Fires and the Benefits of Restoring Peatland. *Nature Communications,* Vol. 12(1), 7044. https://doi.org/10.1038/s41467-021-27353-x.

[15] Carn, S. A., Fioletov, V. E., McLinden, C. A., Li, C. and Krotkov, N. A., (2017). A Decade of Global Volcanic SO_2 Emissions Measured from Space. *Scientific Reports,* Vol. 7(1). https://doi.org/10.1038/srep44095.

[16] Aldrian, E. and Dwi Susanto, R., (2003). Identification of Three Dominant Rainfall Regions Within Indonesia and Their Relationship to Sea Surface Temperature. *International Journal of Climatology,* Vol. 23(12), 1435–1452. https://doi.org/10.1002/joc.950.

[17] Reid, J. S., Xian, P., Hyer, E. J., Flatau, M. K., Ramirez, E. M., Turk, F. J., Sampson, C. R., Zhang, C., Fukada, E. M. and Maloney, E. D., (2012). Multi-Scale Meteorological Conceptual Analysis of Observed Active Fire Hotspot Activity and Smoke Optical Depth in the Maritime Continent. *Atmospheric Chemistry and Physics,* Vol. 12(4), 2117–2147. https://doi.org/10.5194/acp-12-2117-2012.

[18] Ghahremanloo, M., Lops, Y., Choi, Y. and Yeganeh, B., (2021). Deep Learning Estimation of Daily Ground-Level NO_2 Concentrations from Remote Sensing Data. Journal of Geophysical Research: Atmospheres, Vol. 126(21), https://doi.org/10.1029/2021jd034925.

[19] 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., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R.G., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hogan, R.J., Hólm, E.V., Janisková, M., Keeley, S.P., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P.D., Rozum, I., Vamborg, F., Villaume, S. and Thepaut, J., (2020). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society,* Vol. 146(730), 1999–2049. https://doi.org/10.1002/qj.3803.

[20] Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C. and Thépaut, J. N., (2021). ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications. Earth System Science Data, Vol. 13(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.

[21] van Eck, N. J. and Waltman, L., (2009). Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics,* Vol. 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3.

[22] Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F. and Fraundorfer, F., (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine, Vol. 5(4), 8–36. https://doi.org/10.1109/mgrs.2017.2762307.

[23] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B. A., (2019). Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

[24] Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L. and Chen, J. M., (2020). Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sensing of Environment,* Vol. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716.

[25] Taquet, N., Stremme, W., del Castillo, M. E. G., Almanza, V., Bezanilla, A., Laurent, O., Alberti, C., Hase, F., Ramonet, M., Richards, T., Che, K. and Grutter, M., (2024). CO_2 and CO Temporal Variability over Mexico City from Ground-Based Total Column and Surface Measurements. *Atmospheric Chemistry and Physics,* Vol. 24(20), 11823–11848. https://doi.org/10.5194/acp-24-11823-2024.

[26] Humpage, N., Boesch, H., Okello, W., Chen, J., Dietrich, F., Lunt, M. F., Feng, L., Palmer, P. I. and Hase, F., (2024). Greenhouse Gas Column Observations from a Portable Spectrometer in Uganda. *Atmospheric Measurement Techniques*, Vol. 17(19), 5679–5707. https://doi.org/10.5194/amt-17-5679-2024.

[27] Phaka, R. Y., Merlaud, A., Pinardi, G., Friedrich, M. M., van Roozendael, M., Müller, J. F., Stavrakou, T., de Smedt, I., Hendrick, F., Dimitropoulou, E., Lepiba, R. B. M., Phuati, E. P., Djibi, B. L., Jacobs, L., Fayt, C., Mbungu Tsumbu, J. P. and Mahieu, E., (2023). Ground-Based Multi-AXis Differential Optical Absorption Spectroscopy (MAX-DOAS) Observations of NO2 and H2CO at Kinshasa and Comparisons with TROPOMI Observations. *Atmospheric Measurement Techniques,* Vol. 16(21), 5029–5050. https://doi.org/10.5194/amt-16-5029-2023.

[28] Borsdorff, T., García-Reynoso, A., Maldonado, G., Mar-Morales, B., Stremme, W., Grutter, M. and Landgraf, J., (2020). Monitoring CO Emissions of the Metropolis Mexico City Using TROPOMI CO Observations. *Atmospheric Chemistry and Physics,* Vol. 20(24), 15761–15774. https://doi.org/10.5194/acp-20-15761-2020.

[29] Lorente, A., Boersma, K. F., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M. P., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J. T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J. and Krol, M., (2017). Structural Uncertainty in Air Mass Factor Calculation for NO_2 and HCHO Satellite Retrievals. Atmospheric Measurement Techniques, Vol. 10(3), 759–782. https://doi.org/10.5194/amt-10-759-2017.

[30] Loyola, D., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R., Pedergnana, M., Doicu, A., Molina García, V. and Schüssler, O., (2018). The Operational Cloud Retrieval Algorithms from TROPOMI on Board Sentinel-5 Precursor. Atmospheric Measurement Techniques, Vol. 11(1), 409–427. https://doi.org/10.5194/amt-11-409-2018.

[31] Gianotti, R. L. and Eltahir, E. A. B., (2014). Regional Climate Modeling over the Maritime Continent. Part I: New Parameterization for Convective Cloud Fraction. Journal of Climate, Vol. 27(4), 1488–1503. https://doi.org/10.1175/jcli-d-13-00127.1.

[32] Leitão, J., Richter, A., Vrekoussis, M., Kokhanovsky, A., Zhang, Q. J., Beekmann, M. and Burrows, J. P., (2010). On the Improvement of NO_2 Satellite Retrievals – Aerosol Impact on the Airmass Factors. Atmospheric Measurement Techniques, Vol. 3(2), 475–493. https://doi.org/10.5194/amt-3-475-2010.

[33] Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. and Nauss, T., (2018). Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation. Environmental Modelling & Software, Vol. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

[34] Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C. F., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S. and Pélissier, R., (2020). Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models. Nature Communications, Vol. 11(1), 4540. https://doi.org/10.1038/s41467-020-18321-y.

[35] Horner, R. P., Marais, E. A., Wei, N., Ryan, R. G. and Shah, V., (2024). Vertical Profiles of Global Tropospheric Nitrogen Dioxide (NO_2) Obtained by Cloud Slicing the TROPOspheric Monitoring Instrument (TROPOMI). Atmospheric Chemistry and Physics, Vol. 24(22), 13047–13064. https://doi.org/10.5194/acp-24-13047-2024.

[36] Chan, K. L., Khorsandi, E., Liu, S., Baier, F. and Valks, P., (2021). Estimation of Surface NO_2 Concentrations over Germany from Tropomi Satellite Observations Using a Machine Learning Method. Remote Sensing, Vol. 13(5), 969. https://doi.org/10.3390/rs13050969.

[37] Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C. K. and Kim, S., (2021). Estimation of Surface-Level NO_2 and O_3 Concentrations Using TROPOMI Data and Machine Learning over East Asia. Environmental Pollution, Vol. 288, 117711. https://doi.org/10.1016/j.envpol.2021.117711.

[38] Kim, M., Brunner, D. and Kuhlmann, G., (2021). Importance of Satellite Observations for High-Resolution Mapping of Near-Surface NO_2 by Machine Learning. *Remote Sensing of Environment,* Vol. 264. https://doi.org/10.1016/j.rse.2021.112573.

[39] Grzybowski, P. T., Markowicz, K. M. and Musiał, J. P., (2023). Estimations of the Ground-Level NO_2 Concentrations Based on the Sentinel-5P NO_2 Tropospheric Column Number Density Product. *Remote Sensing 2023,* Vol. 15(2). https://doi.org/10.3390/rs15020378.

[40] Wu, Y., Liu, H., Liu, S. and Lou, C., (2023). Estimate of Near-Surface NO_2 Concentrations in Fenwei Plain, China, Based on TROPOMI Data and Random Forest Model. *Environmental Monitoring and Assessment,* Vol. 195(11). https://doi.org/10.1007/s10661-023-11993-1.

[41] Shetty, S., Schneider, P., Stebel, K., David Hamer, P., Kylling, A. and Koren Berntsen, T., (2024). Estimating Surface NO_2 Concentrations over Europe Using Sentinel-5P TROPOMI Observations and Machine Learning. *Remote Sensing of Environment,* Vol. 312. https://doi.org/10.1016/j.rse.2024.114321.

[42] Fahim, M. A. S. and Sužiedelytė Visockienė, J. S., (2025). Satellite-Driven Monitoring and Prediction of Atmospheric Nitrogen Dioxide in Lithuania. *International Journal of Environmental Science and Development,* Vol. 16(4), 265–271. https://doi.org/10.18178/ijesd.2025.16.4.1533.

[43] Wang, S., Zhang, S., Wang, D. and Li, W., (2025). Spatial-and-Local-Aware Deep Learning Approach for Ground-Level NO_2 Estimation in England with Multisource Data from Satellite-Based Observations and Chemical Transport Models. *International Journal of Applied Earth Observation and Geoinformation,* Vol. 139. https://doi.org/10.1016/j.jag.2025.104506.

[44] Jeong, U. and Hong, H., (2021). Comparison of Total Column and Surface Mixing Ratio of Carbon Monoxide Derived from the TROPOMI/Sentinel-5 Precursor with In Situ Measurements from Extensive Ground-Based Network over South Korea. *Remote Sensing,* Vol. 13(19). https://doi.org/10.3390/rs13193987.

[45] Cai, K., Li, S., Lai, J., Xia, Y., Wang, Y., Hu, X. and Li, A., (2022). Evaluation of TROPOMI and OMI Tropospheric NO_2 Products Using Measurements from MAX-DOAS and State-Controlled Stations in the Jiangsu Province of China. *Atmosphere,* Vol. 13(6). https://doi.org/10.3390/atmos13060886.

[46] Yu, Z. and Li, X., (2022). The Temporal–Spatial Characteristics of Column NO_2 Concentration and Influence Factors in Xinjiang of Northwestern Arid Region in China. *Atmosphere 2022,* Vol. 13(10). https://doi.org/10.3390/atmos13101533.

[47] Virta, H., Ialongo, I., Szeląg, M. and Eskes, H., (2023). Estimating Surface-Level Nitrogen Dioxide Concentrations from Sentinel-5P/TROPOMI Observations in Finland. *Atmospheric Environment,* Vol. 312. https://doi.org/10.1016/j.atmosenv.2023.119989.

[48] Zhang, Y., Lin, J., Kim, J., Lee, H., Park, J., Hong, H., Van Roozendael, M., Hendrick, F., Wang, T., Wang, P., He, Q., Qin, K., Choi, Y., Kanaya, Y., Xu, J., Xie, P., Tian, X., Zhang, S., Wang, S., Cheng, S., Cheng, X., Ma, J., Wagner, T., Spurr, R., Chen, L., Kong, H. and Liu, M., (2023). A Research Product for Tropospheric NO_2 Columns from Geostationary Environment Monitoring Spectrometer Based on Peking University OMI NO_2 Algorithm. *Atmospheric Measurement Techniques,* Vol. 16(19), 4643–4665. https://doi.org/10.5194/amt-16-4643-2023.

[49] Morillas, C., Alvarez, S., Serio, C., Masiello, G. and Martinez, S., (2024). TROPOMI NO_2 Sentinel-5P Data in the Community of Madrid: A Detailed Consistency Analysis with In Situ Surface Observations. *Remote Sensing Applications: Society and Environment,* Vol. 33, https://doi.org/10.1016/j.rsase.2023.101083.

[50] Zeng, S., Zhou, S., Wang, F., Zhou, Y. and Zhang, Y., (2025). Spatial–Temporal Variations of NO_2 Pollution in Shandong Province Based on Sentinel-5P Satellite Data and Influencing Factors. *Open Geosciences,* Vol. 17(1), https://doi.org/10.1515/geo-2025-0833.

[51] Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., Pandey, B., Seto, K. C., Hall, D., Oda, T., Wolfe, R. E., Lin, G., Golpayegani, N., Devadiga, S., Davidson, C., Sarkar, S., Praderas, C., Schmaltz, J., Boller, R., Stevens, J., Ramos, O., Padilla, E., Alonso, J., Detrés, Y., Armstrong, R., Miranda, I., Conte, Y., Marrero, N., MacManus, K., Esch, T. and Masuoka, E. J., (2018). NASA's Black Marble Nighttime Lights Product Suite. *Remote Sensing of Environment,* Vol. 210, 113–143. https://doi.org/10.1016/j.rse.2018.03.017.

[52] Alladi, H. K., Sandeep, A., Prasad, P. and Bv, B., (2026). Evaluation of ERA5 Boundary Layer Height Against Radiosonde Measurements: Regional Drivers, Spatiotemporal Variability, and Trends over the Indian Monsoon Region. *Atmospheric Research,* Vol. 329. https://doi.org/10.1016/j.atmosres.2025.108511.

[53] Barrington-Leigh, C. and Millard-Ball, A., (2017). The World's User-Generated Road Map is more than 80% Complete. *PLoS ONE,* Vol. 12(8), https://doi.org/10.1371/journal.pone.0180698.

[54] Lange, K., Richter, A., Bösch, T., Zilker, B., Latsch, M., Behrens, L. K., Okafor, C. M., Boesch, H., Burrows, J. P., Merlaud, A., Pinardi, G., Fayt, C., Friedrich, M. M., Dimitropoulou, E., Van Roozendael, M., Ziegler, S., Ripperger-Lukosiunaite, S., Kuhn, L., Lauster, B., Wagner, T., Hong, H., Kim, D., Chang, L. S., Bae, K., Song, C. K., Park, J. U. and Lee, H., (2024). Validation of GEMS Tropospheric NO_2 Columns and Their Diurnal Variation with Ground-Based DOAS Measurements. *Atmospheric Measurement Techniques,* Vol. 17(21), 6315–6344. https://doi.org/10.5194/amt-17-6315-2024.

[55] Oak, Y. J., Jacob, D. J., Balasus, N., Yang, L. H., Chong, H., Park, J., Lee, H., Lee, G. T., Ha, E. S., Park, R. J., Kwon, H. A. and Kim, J., (2024). A Bias-Corrected GEMS Geostationary Satellite Product for Nitrogen Dioxide Using Machine Learning to Enforce Consistency with the TROPOMI Satellite Instrument. *Atmospheric Measurement Techniques,* Vol. 17(17), 5147–5159. https://doi.org/10.5194/amt-17-5147-2024.

[56] Jung, Y., Park, J., Hong, H., Lee, H. and Jeong, U., (2025). Evaluation of GEMS NO_2 Retrieval Algorithm Version 2.0 and 3.0 Using TROPOMI and Pandora Observations. *Korean Journal of Remote Sensing,* Vol. 41(5), 803–811. https://doi.org/10.7780/kjrs.2025.41.5.8.

[57] Gödeke, J., Richter, A., Lange, K., Maaß, P., Hong, H., Lee, H. and Park, J., (2025). Hourly Surface Nitrogen Dioxide Retrieval from GEMS Tropospheric Vertical Column Densities: Benefit of Using Time-Contiguous Input Features for Machine Learning Models. *Atmospheric Measurement Techniques,* Vol. 18(15), 3747–3779. https://doi.org/10.5194/amt-18-3747-2025.

[58] Yang, L. H., Jacob, D. J., Dang, R., Oak, Y. J., Lin, H. and Kim, J., (2024). Interpreting Geostationary Environment Monitoring Spectrometer (GEMS) Geostationary Satellite Observations of the Diurnal Variation of Nitrogen Dioxide. *Atmospheric Chemistry and Physics,* Vol. 24(12), 7027–7039. https://doi.org/10.5194/acp-24-7027-2024.

[59] Stratoulias, D., Nuthammachot, N., Dejchanchaiwong, R., Tekasakul, P. and Carmichael, G. R., (2024). Recent Developments in Satellite Remote Sensing for Air Pollution Surveillance in Support of Sustainable Development Goals. *Remote Sensing,* Vol. 16(16). https://doi.org/10.3390/rs16162932.

[60] Nawaz, M. O., Huber, D. E., Kerr, G. H., Judd, L. M., Acker, S. J. and Goldberg, D. L., (2025). A Comparative Analysis of TEMPO NO_2 Remote Sensing with Surface-Level Monitoring Through Diurnal and Seasonal Trends, Meteorology, and Monitor Characteristics. *Journal of Geophysical Research: Atmospheres,* Vol. 130(20), https://doi.org/10.1029/2025jd043923.

[61] Baek, K., Kim, J. H., Bak, J., Haffner, D. P., Kang, M. and Hong, H., (2023). Evaluation of Total Ozone Measurements from Geostationary Environmental Monitoring Spectrometer (GEMS). *Atmospheric Measurement Techniques,* Vol. 16(22), 5461–5478. https://doi.org/10.5194/amt-16-5461-2023.

[62] Karppinen, T., Sundström, A. M., Lindqvist, H., Hatakka, J. and Tamminen, J., (2024). Satellite-Based Assessment of National Carbon Monoxide Concentrations for Air Quality Reporting in Finland. *Remote Sensing Applications: Society and Environment,* Vol. 33. https://doi.org/10.1016/j.rsase.2023.101120.

[63] Borsdorff, T., Campos, T., Kille, N., Zarzana, K. J., Volkamer, R. and Landgraf, J., (2023). Vertical Information of CO from TROPOMI Total Column Measurements in Context of the CAMS-IFS Data Assimilation Scheme. *Atmospheric Measurement Techniques,* Vol. 16(11), 3027–3038. https://doi.org/10.5194/amt-16-3027-2023.

[64] Hedelt, P., Reichardt, J., Lauermann, F., Weiß, B., Theys, N., Redondas, A., Barreto, A., García, O. and Loyola, R. D., (2025). Analysis of the Long-Range Transport of the Volcanic Plume from the 2021 Tajogaite/Cumbre Vieja Eruption to Europe Using TROPOMI and Ground-Based Measurements. *Atmospheric Chemistry and Physics,* Vol. 25(2), 1253–1272. https://doi.org/10.5194/acp-25-1253-2025.

[65] Rawat, P., Naja, M., Rajwar, M. C., Irie, H., Lerot, C., Kumar, M. and Lal, S., (2024). Long-Term Observations of NO2, SO2, HCHO, and CHOCHO over the Himalayan Foothills: Insights from MAX-DOAS, TROPOMI, and GOME-2. *Atmospheric Environment,* Vol. 336. https://doi.org/10.1016/j.atmosenv.2024.120746.

[66] Theys, N., Lerot, C., Brenot, H., van Gent, J., De Smedt, I., Clarisse, L., Burton, M., Varnam, M., Hayer, C., Esse, B. and Van Roozendael, M., (2022). Improved Retrieval of SO_2 Plume Height from TROPOMI Using an Iterative Covariance-Based Retrieval Algorithm. *Atmospheric Measurement Techniques,* Vol. 15(16), 4801–4817. https://doi.org/10.5194/amt-15-4801-2022.

[67] Theys, N., Fioletov, V., Li, C., de Smedt, I., Lerot, C., McLinden, C. A., Krotkov, N., Griffin, D., Clarisse, L., Hedelt, P., Loyola, D., Wagner, T., Kumar, V., Innes, A., Ribas, R., Hendrick, F., Vlietinck, J., Brenot, H. and van Roozendael, M., (2021). A Sulfur Dioxide Covariance-Based Retrieval Algorithm (COBRA): Application to TROPOMI Reveals New Emission Sources.

Atmospheric Chemistry and Physics, Vol. 21(22), 16727–16744. https://doi.org/10.5194/acp-21-16727-2021.

[68] Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H., Romahn, F., Pedergnana, M., Loyola, D. and Van Roozendael, M., (2017). Sulfur Dioxide Retrievals from TROPOMI Onboard Sentinel-5 Precursor: Algorithm Theoretical Basis. *Atmospheric Measurement Techniques,* Vol. 10(1), 119–153. https://doi.org/10.5194/amt-10-119-2017.

[69] Savenets, M., Dvoretska, I., Nadtochii, L. and Zhemera, N., (2022). Comparison of TROPOMI NO_2, CO, HCHO, and SO_2 Data Against Ground-Level Measurements in Close Proximity to Large Anthropogenic Emission Sources in the Example of Ukraine. *Meteorological Applications,* Vol. 29(6). https://doi.org/10.1002/met.2108.

[70] Vigouroux, C., Langerock, B., Augusto Bauer Aquino, C., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J. M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A. N., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., Van Roozendael, M., Wang, P. and Winkler, H., (2020). TROPOMI-Sentinel-5 Precursor Formaldehyde Validation Using an Extensive Network of Ground-Based Fourier-Transform Infrared Stations. *Atmospheric Measurement Techniques,* Vol. 13(7), 3751–3767. https://doi.org/10.5194/amt-13-3751-2020.

[71] Ryan, R. G., Silver, J. D., Querel, R., Smale, D., Rhodes, S., Tully, M., Jones, N. and Schofield, R., (2020). Comparison of Formaldehyde Tropospheric Columns in Australia and New Zealand Using MAX-DOAS, FTIR and TROPOMI. *Atmospheric Measurement Techniques,* Vol. 13(12), 6501–6519. https://doi.org/10.5194/amt-13-6501-2020.

[72] Monks, P. S., Archibald, A. T. and Colette, A., (2015). Tropospheric Ozone and its Precursors from the Urban to the Global Scale from Air Quality to Short-Lived Climate Forcer. *Atmospheric Chemistry and Physics,* Vol. 15(15), 8889–8973. https://doi.org/10.5194/acp-15-8889-2015.

[73] Gaudel, A., Bourgeois, I., Li, M., Chang, K. L., Ziemke, J., Sauvage, B., Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Smith, N., Hubert, D., Keppens, A., Cuesta, J., Heue, K. P., Veefkind, P., Aikin, K., Peischl, J., Thompson, C. R., Ryerson, T. B., Frost, G. J., Frost, B. C. and Cooper, O. R., (2024). Tropical Tropospheric Ozone Distribution and Trends from in Situ and Satellite Data. *Atmospheric Chemistry and Physics,* Vol. 24(17), 9975–10000. https://doi.org/10.5194/acp-24-9975-2024.

[74] Zhou, M., Wang, P., Langerock, B., Vigouroux, C., Hermans, C., Kumps, N., Wang, T., Yang, Y., Ji, D., Ran, L., Zhang, J., Xuan, Y., Chen, H., Posny, F., Duflot, V., Metzger, J. M. and de Mazière, M., (2020). Ground-Based Fourier Transform Infrared (FTIR) O3 Retrievals from the 3040 cm^{-1} Spectral Range at Xianghe, China. *Atmospheric Measurement Techniques,* Vol. 13(10), 5379–5394. https://doi.org/10.5194/amt-13-5379-2020.

Appendix A: Supplementary Material

This appendix provides the supporting material referenced from the main text: the search strings (Scripts S1 and S2), the per-study performance summary (Table 7), and the keyword frequency tables (Table 8).

Table 7: Performance of selected studies

ID	Region	Performance Indicators	References
1	Germany	R = 0.80, RMSD = 6.32 μg/m³	[36]
2	East Asia	R^2 = 0.63–0.70, nRMSE = 38–42%	[37]
3	Switzerland	R^2 = 0.59, MAE = 7.69 μg/m³	[38]
4	South Korea	R ≈ 0.51	[44]
5	Jiangsu, China	r = 0.9 (monthly average)	[45]
6	Xinjiang, China	R^2 = 0.8132 (Yining), 0.7307 (Aksu)	[46]
7	Finland	R = 0.80, Slope = 1.00	[47]
8	Poland	MAE = 3.7 μg/m³, MAPE = 37.2% for hourly MAE = 3.1 μg/m³, MAPE = 30.8% for weekly	[39]
9	East Asia	r = 0.96 (diurnal), r = 0.78 (spatiotemporal)	[48]
10	Fenwei Plain, China	R^2 = 0.875 (validation), 0.949 (fit)	[40]
11	Madrid, Spain	r = 0.78 (daily), r = 0.81 (averaged)	[49]
12	Europe	MAE = 7.77 μg/m³, Median Bias = 0.6 μg/m³, Spearman r = 0.66	[41]
13	Lithuania	RMSE = 5.12 μg/m³ (validation), 4.90 μg/m³ (test)	[42]
14	England	R^2 = 0.914 (2019), 0.919 (2020), 0.887 (2021)	[43]
15	Shandong, China	R = 0.8289	[50]

Table 8: Frequency of the 20 most recurrent thematically relevant author keywords by publication period (Phase 2 subset, n = 150)

Keyword	2018–2020	2021–2022	2023–2025	Total
Tropomi	5	16	19	40
Nitrogen Dioxide	4	13	18	35
Air Quality	4	5	11	20
Max-Doas	1	2	7	10
Air Pollution	1	3	6	10
Machine Learning	0	3	6	9
Covid-19	2	6	0	8
Pandora	0	2	4	6
Validation	1	2	2	5
Carbon Monoxide	0	1	4	5
Omi	2	2	0	4
Doas	1	1	0	2
Pandemic	0	1	1	2
Lockdown	0	2	0	2
Surface NO_2	0	0	2	2
Surface Concentration	0	1	1	2
Comparison	1	0	1	2
Gems	0	0	2	2
Aod	0	0	2	2
Earth Observation	0	0	2	2
Total	22	60	88	170

Table 9: Machine-learning studies for TROPOMI NO_2 column-to-surface conversion included in the synthesis (complementing the feature-input comparison of Table 6). Abbreviations: CV = cross-validation; n/r = not reported; RFE = recursive feature elimination

Study (region)	Dataset size (stations)	Temporal coverage	Spatial resolution	Validation strategy	Best reported metric	Feature set size	Ref
Germany	455 (UBA network)	2018–2020	0.5 km	90/10 random split	R = 0.80; RMSD = 6.32 µg/m³	n/r	[36]
East Asia	729 (209 KR + 297 CN + 223 JP)	May 2018 – Nov 2020	6 km	10-fold CV (random, spatial, temporal)	R^2 = 0.63–0.70; nRMSE = 38.3–42.2%	42 → RFE	[37]
Switzerland	380 (NABEL + Airbase)	Jun 2018 – May 2020	100 m	10% test split	MAE = 7.69 µg/m³	n/r	[38]
Poland	116	Jul 2018 – Jul 2021	3.5 km grid	Temporal holdout (odd/even)	MAE = 3.1 µg/m³ (weekly); R^2 = 0.60	n/r	[39]
Fenwei Plain (CN)	52–67 (per-year subset)	Jan 2019 – Dec 2021	5 × 5 km	10-fold CV	R^2 = 0.875; RMSE = 4.61 µg/m³	n/r	[40]
Europe	3,044 (EEA; 2,435 train / 609 test)	2019–2021	1 km	Random station split 80/20, repeated	MAE = 7.77; RMSE = 13.07; ρ = 0.66	12	[41]
Lithuania	13 (EPA)	Jan 2020 – Dec 2023	Station-level	5-fold CV	RMSE = 5.12 µg/m³ (bagging)	n/r	[42]
England	207 (AURN)	Jan 2019 – Dec 2021	1 km	5-fold CV + region-/month-based CV	R^2 = 0.914 / 0.919 / 0.887	n/r	[43]

Script S1. Scopus query string, first formulation (Phase 1).
TITLE-ABS-KEY (("NO_2" OR "NOx" OR "nitrogen dioxide" OR "SO_2" OR "sulfur dioxide" OR "sulphur dioxide" OR "O_3" OR "ozone" OR "CO" OR "carbon monoxide") AND ("TROPOMI" OR "Sentinel-5P" OR "S-5P")) AND PUBYEAR > 2017 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SUBJAREA , "ENVI") OR LIMIT-TO (SUBJAREA , "EART"))

Script S2. Scopus query string, second formulation, adding the surface-concentration terms (Phase 2).
TITLE-ABS-KEY (("NO_2" OR "NOx" OR "nitrogen dioxide" OR "SO_2" OR "sulfur dioxide" OR "sulphur dioxide" OR "O_3" OR "ozone" OR "CO" OR "carbon monoxide") AND ("TROPOMI" OR "Sentinel-5P" OR "S-5P") AND ("column density" OR "vertical column" OR "total column" OR "tropospheric column" OR "VCD" OR "mixing ratio") AND ("ground-level" OR "surface concentration" OR "in-situ" OR "ground station" OR "ground-based" OR "ground truth" OR "monitoring station") AND ("validat*" OR "evaluat*" OR "accura*" OR "comparison" OR "performance" OR "agreement" OR "correlation")) AND PUBYEAR > 2017 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SUBJAREA , "ENVI") OR LIMIT-TO (SUBJAREA , "EART"))