Improving Sentinel-5P Imagery Usability Through Machine Learning Gap-Filling

Zhanbin Wu[1], Vasil Yordanov[1,*], Maria Antonia Brovelli[1]

[1] Politecnico di Milano, Department of Civil and Environmental Engineering, Piazza Leonardo da Vinci, 32, Milan, Italy - zhanbin.wu@mail.polimi.it, (vasil.yordanov, maria.brovelli)@polimi.it

Keywords: Air Pollution, Sentinel 5P, Gap-filling, Trace Gases, Machine Learning, Po Valley.

Abstract

Satellite-based air quality monitoring, particularly using Sentinel-5P (S5P) TROPOMI observations of NO_2 and SO_2, provides high-resolution global coverage but is substantially affected by spatio-temporal data gaps. These discontinuities, caused by cloud cover, surface reflectance, viewing geometry constraints, and strict quality assurance filtering, limit the reliability of downstream applications such as emission inversion, exposure assessment, and time-series analysis. This study investigates missing data patterns and develops a robust gap-filling framework for the Po Valley (Northern Italy) over 2019–2023, a region characterized by complex basin topography and frequent winter inversions that exacerbate both pollution accumulation and data loss. A statistical assessment revealed average missing rates of 45.4% for NO_2 and 77.4% for SO_2, with strong seasonality and extreme winter sparsity (up to 93.1% missing for SO_2). Missingness was strongly correlated with elevation (r $\approx 0.77 for NO_2$, r $\approx 0.92 for SO_2$), highlighting terrain-related influences on retrieval quality. To reconstruct missing observations, we implemented two complementary machine learning models: a LightGBM baseline and a 3D Convolutional Neural Network (3D CNN). Both models were trained on a harmonized multi-source feature stack including lagged pollutant fields, spatial neighborhood statistics, ERA5 meteorology, static geographical variables, and cyclic temporal encodings. Synthetic masking simulated realistic gap scenarios during training (2019–2022), with evaluation on 2023 data. The 3D CNN outperformed LightGBM, achieving R^2 values of 0.9469 (NO_2) and 0.7387 (SO_2), indicating strong spatio-temporal modeling capability. This framework enhances the continuity of S5P observations and supports their use for air quality analysis in data-sparse conditions.

1. Introduction

Accurate air quality monitoring increasingly relies on satellite-based observations of atmospheric trace gases. In particular, nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) products from the Sentinel-5 Precursor (S5P) mission provide daily global measurements with high spatial resolution. The TROPOMI instrument onboard S5P enables detailed monitoring of atmospheric pollutants at regional and urban scales, supporting air quality assessment, emission monitoring, and environmental policy analysis (Veefkind et al., 2012).

However, the operational use of Sentinel-5P observations is often limited by substantial spatio-temporal data gaps. Missing observations mainly arise from cloud contamination, surface reflectance issues, high solar zenith angles, instrument retrieval limitations, and missing data due to clouds, geometry, QA filtering (Veefkind et al., 2012; Loyola et al., 2018). This effect is particularly pronounced for SO_2, where low solar zenith angles lead to stricter quality filtering and increased data loss during winter months in the Northern Hemisphere, resulting in total data absence in the months between mid-November and mid-February. These gaps reduce the reliability of satellite-based products for downstream applications such as emission inversion, exposure assessment, and long-term time-series analysis.

Various gap-filling methods have been proposed for remote sensing time series. Early approaches relied on deterministic interpolation or spatio-temporal neighborhood methods, which perform well for variables with smooth temporal dynamics but struggle with highly variable atmospheric pollutants (Siabi et al., 2022). More recent studies employ machine learning models that integrate auxiliary predictors such as meteorological variables to reconstruct missing observations (Liu et al., 2023). Deep learning methods, particularly convolutional neural networks, have further improved reconstruction performance by learning complex spatio-temporal patterns directly from satellite image sequences (Fu et al., 2022; Appel, 2024). However, these approaches remain challenging when missing observations are extensive or systematically related to observation conditions.

This study investigates sporadic missing value patterns (Figure 1) in Sentinel-5P NO_2 and SO_2 observations over the Po Valley in Northern Italy during 2019–2023. The Po Valley represents one of the most polluted regions in Europe, where basin topography and frequent winter temperature inversions limit pollutant dispersion and exacerbate data gaps. To address these challenges, this research proposes a gap-filling pipeline based on two complementary machine learning approaches: a gradient boosting model implemented with LightGBM (Ke et al., 2017) and a spatio-temporal deep learning model based on a three-dimensional convolutional neural network (3D CNN).

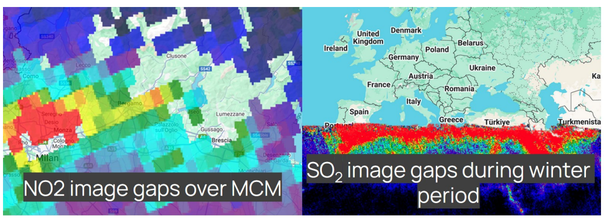

Figure 1. Sporadic gaps in NO_2 images and season (winter) data absence in SO_2.

* Corresponding author

2. Data and Methodology

2.1 Study Area

The Po Valley in Northern Italy (Figure 2) represents one of the most polluted regions in Europe and provides an ideal testbed for satellite-based air quality analysis. The region is bounded by the Alps to the north and west and the Apennines to the south, forming a basin-like topography that restricts atmospheric circulation. Frequent winter temperature inversions further limit pollutant dispersion, leading to persistent accumulation of atmospheric contaminants (Pirovano et al., 2015). These conditions make the Po Valley a suitable study area for investigating data gaps in satellite observations and evaluating gap-filling approaches.

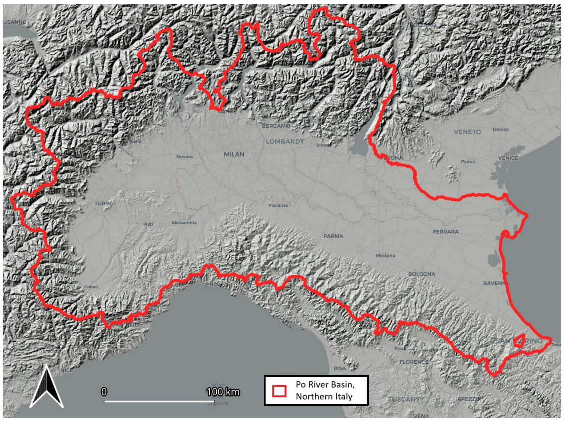

Figure 2. Study area: Po River Plain; outline = Po River Basin. Basemap: © CartoDB. Map created by the author using QGIS.

2.2 Data Sources

This study integrates satellite observations, meteorological reanalysis data, geographical information, and derived contextual predictors to support the gap analysis and reconstruction of atmospheric pollutants.

The primary datasets are Sentinel-5P TROPOMI Level-3 products for NO_2 and SO_2, obtained from the Google Earth Engine cloud-based geospatial analysis platform for the period 2019–2023. Daily composites were generated by aggregating all available satellite observations within each day over the study area. Each year was exported as a multiband GeoTIFF containing 365 or 366 daily bands in order to preserve the temporal structure of the time series.

Meteorological predictors were derived from the ERA5 reanalysis dataset provided by the Copernicus Climate Data Store. These variables include boundary layer height, 2 m air temperature, 10 m wind components, precipitation, surface pressure, and radiation variables, which characterize atmospheric conditions influencing pollutant transport, dispersion, and chemical transformation.

To represent underlying geographical conditions, static layers including elevation and slope derived from the Shuttle Radar Topography Mission (SRTM) digital elevation model and land use/land cover information from the ESA WorldCover 2020 dataset were incorporated. Population density from the WorldPop UN-Adjusted 2020 dataset was also included as an anthropogenic indicator representing potential emission intensity.

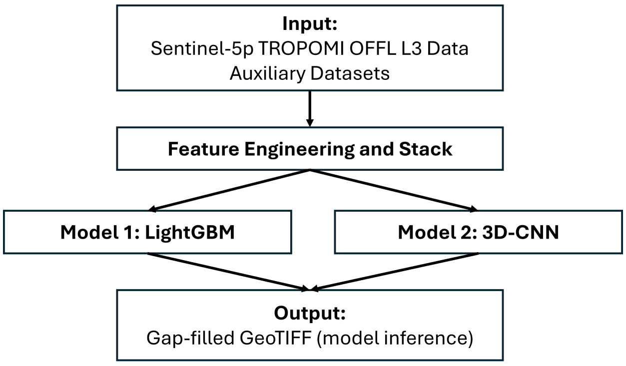

Figure 3. Overall workflow from inputs (S5P NO_2/SO_2 and auxiliary datasets) through feature engineering and two complementary models (LightGBM and 3D-CNN) to model-inferred, gap-filled GeoTIFF outputs.

All datasets were harmonized to the Sentinel-5P spatial grid and aggregated to a daily temporal resolution. In addition to these base datasets, several derived predictors were constructed to capture temporal and spatial dependencies in pollutant dynamics. Temporal descriptors include weekly activity indicators (day-of-week and weekend-to-weekday emission ratios) as well as annual cyclic features derived from day-of-year encoding. Spatial context features were also introduced, including one-day lagged concentration fields from the Copernicus Atmosphere Monitoring Service (CAMS) and local neighbourhood statistics computed from the Sentinel-5P grid. These contextual predictors provide information on temporal persistence and spatial correlation, thereby improving the robustness of the subsequent gap-filling models.

2.3 Overall workflow

We turn raw observations into daily gap-filled maps in four steps. The Figure 3 summarises the pipeline: we start from S5P TROPOMI OFFL L3 NO_2/SO_2 (2019–2023) together with auxiliary sources, harmonise all inputs to the S5P grid (CRS, resolution, AOI clip, pixel alignment), then build a feature stack combining meteorology, static geography, population, temporal signals (day-of-week weights, day-of-year sine/cosine) and spatio-temporal context (CAMS lag-1d, 3×3 neighbour mean). A LightGBM model and a 3D-CNN model are trained and hyperparameter tuning, after which model inference produces gap-filled GeoTIFF outputs that are assessed by masked validation check.

2.4 Gap Characterization Analysis

To understand the spatial and temporal distribution of missing observations in the Sentinel-5P products, a gap characterization analysis was conducted for both NO_2 and SO_2 over the study period (2019–2023). A data gap was defined as any pixel-day with missing or invalid observations after quality filtering, including pixels with non-positive values or NaN. Based on the daily multi-band GeoTIFF stacks, a binary gap mask $G(t, \mathbf{x})$ was constructed to indicate whether a valid observation exists at location $\mathbf{x}$ and day t.

Using this mask, data gaps were summarized across time and space to examine their magnitude and distribution. Annual gap ratios were calculated by counting the number of missing days per pixel and dividing by the total number of days in each year. Seasonal patterns were further analysed by aggregating daily observations into four climatological seasons: DJF (December–February), MAM (March–May), JJA (June–August), and SON (September– November). And compute seasonal gap fractions within the area of interest.

To compare the behaviour of the two pollutants, AOI-wide statistics were computed for both NO_2 and SO_2, including annual averages and seasonal variability. Interannual trends in data gaps were evaluated using the Mann–Kendall trend test and Sen’s slope estimator applied to the yearly gap-ratio maps, allowing identification of areas with increasing or decreasing data availability.

Finally, potential drivers of spatial gap patterns were explored by examining correlations between the gap ratios and environmental variables. Elevation from the SRTM digital elevation model was used to assess topographic influences, while cloud fraction derived from Sentinel-5P observations was used to evaluate the effect of cloudiness. Pearson and Spearman correlation coefficients were computed across all valid pixels within the study area to quantify these relationships.

Together, these analyses provide a comprehensive overview of the frequency, spatial distribution, and potential drivers of data gaps in the Sentinel-5P NO_2 and SO_2 observations, forming the basis for the subsequent gap-filling modelling.

2.5 Auxiliary Datasets Preprocess

To reconstruct missing observations, both models were trained using a harmonized multi-source feature dataset combining spatio-temporal, meteorological, and geographical predictors:

- **Spatio-temporal predictors:** Lagged pollutant concentration $(t-1)$ from CAMS data and local 3×3 spatial neighborhood means derived from Sentinel-5P observations.
- **Meteorological variables:** ERA5 atmospheric parameters including boundary layer height (BLH), 2 m air temperature, 10 m wind components ($u10$, $v10$), total precipitation, and surface pressure.
- **Static geographical factors:** Elevation and slope derived from SRTM DEM, land cover from ESA WorldCover 2020, and population density from WorldPop.
- **Temporal features:** Day-of-year encoded as sine and cosine components to represent seasonal cycles, together with weekday/weekend indicators derived from historical satellite observations.

These features serve as input variables for the gap-filling models described in the following section.

2.6 Feature Stack Construction

To provide the models with consistent spatial and temporal context, all predictor layers were combined into daily feature stacks aligned to the Sentinel-5P grid. A feature stack is defined as a set of co-registered raster layers sharing the same spatial grid, coordinate reference system, and resolution. For each date t, the input data are organised as a multi-channel array

$$X_t \in R^{H \times W \times C},$$

where H and W denote the spatial dimensions of the grid and C represents the number of feature channels.

The stacks integrate both dynamic and static predictors. Dynamic variables include meteorological fields from ERA5, temporal descriptors (day-of-year sine/cosine and weekday activity weight), spatial context features (neighbourhood mean and one-day lagged concentration), and pollutant-specific signals. Static predictors include elevation, slope, land cover classes, and population density. All layers were reprojected to a common template grid (EPSG:4326; 300×621 cells) using bilinear interpolation for continuous variables and nearest-neighbour resampling for categorical layers. Invalid values and product NoData were mapped to `NaN`.

Separate stacks were produced for NO_2 and SO_2 (Dataset: 10.5281/zenodo.17226552). While the overall feature design is consistent, the SO_2 stack includes an additional monthly climatology prior to stabilise winter retrievals affected by low solar irradiance. Across both pollutants, a fixed channel ordering and shared spatial grid ensure consistent preprocessing and facilitate model training.

2.7 Feature Standardization

To ensure numerical stability and consistent feature scales across space and time, continuous predictors were standardized before model training. Global per-channel means and standard deviations were estimated using the training dataset (2019–2021) and then reused for validation and testing to avoid data leakage.

For each feature channel c, normalization was performed using standard z-score transformation,

$$z = \frac{x - \mu_c}{\sigma_c},$$

where μ_c and σ_c denote the mean and standard deviation computed over all valid training pixels. Non-finite values and masked pixels were excluded from the estimation. Categorical variables such as land-cover one-hot channels were not scaled.

The resulting normalization parameters were stored and consistently applied to all feature stacks during both training and inference, ensuring that the models operate on a stable and comparable feature space.

2.8 Machine Learning Model: LightGBM

LightGBM is a gradient-boosted decision tree (GBDT) algorithm designed for efficient learning on large tabular datasets (Ke et al., 2017). By constructing an ensemble of decision trees, it captures non-linear relationships and interactions among predictors while maintaining high computational efficiency. The method is well suited to the feature stacks used in this study, which contain a mixture of continuous variables, binary indicators, and one-hot encoded categorical features. In addition, LightGBM handles missing values natively by learning optimal split directions for incomplete inputs.

In this work, LightGBM is used as a baseline model to evaluate the effectiveness of the constructed feature stack and data preprocessing pipeline before applying more complex deep learning approaches. For each day, the spatial feature stacks are flattened into per-pixel feature vectors (N_{pixels}, C), while target values are taken from observed satellite pixels only. The model is trained on data from 2019–2021, validated on 2022 using early stopping, and tested on 2023 to preserve temporal independence.

This setup allows LightGBM to learn relationships between pollutant concentrations and the combined meteorological, geographical, temporal, and spatial-context predictors, providing a robust benchmark for the subsequent gap-filling models.

2.9 Deep Learning Model: 3D CNN

To capture spatio-temporal dependencies in atmospheric pollutant dynamics, a three-dimensional convolutional neural network (3D CNN) was adopted. Unlike conventional 2D CNNs, 3D convolutions operate simultaneously across time, height, and width, allowing the model to learn joint temporal evolution and spatial context of pollutant fields. This property is particularly suitable for Sentinel-5P gap filling, where missing observations and pollutant plumes exhibit strong spatio-temporal correlations across consecutive days (Appel, 2024).

Input data are organized as spatio-temporal tensors

$$\mathcal{X} \in R^{T\times C\times H\times W},$$

where T denotes the temporal window, C the number of feature channels, and H, W the spatial dimensions of the grid. Sliding windows of $T = 7$ days are extracted from the daily feature stacks, and training samples are generated as spatial patches of 64×64 pixels.

The network architecture consists of three convolutional blocks (Conv3D–normalization–ReLU) with increasing channel depth (32, 64, and 128), followed by a $1 \times 1 \times 1$ convolution layer that produces the final prediction. Training follows a temporal split (2019–2021 for training, 2022 for validation, and 2023 for testing) to preserve chronological independence.

To improve robustness to missing observations, artificial gaps are introduced during training by randomly masking earlier frames within each temporal window, while the final prediction frame remains unchanged. Model parameters are optimized using the Adam optimizer with a masked mean squared error loss computed on valid pixels of the target frame.

2.10 Model Evaluation Strategy

2.10.1 Masked Validation for LightGBM We evaluate gap filling on the 2023 test split using a masked (counterfactual) validation that injects missing pixels only at inference time. For each test day, a fraction $r \in \{0.1, 0.2, 0.3\}$ of valid pixels in the last frame is randomly selected and hidden while preserving their true targets. At masked locations, all feature channels are removed so that the model must reconstruct the values using the remaining spatial and temporal context.

To further test robustness and prevent potential information leakage, we introduce a sensitive-features mode where lag and neighbour predictors (`lag1`, `neighbor`) are also set to NaN at masked pixels. LightGBM handles these missing values through its native missing-value mechanism.

Predictions are evaluated only on masked pixels with valid ground truth. We report RMSE, MAE, R^2, Pearson correlation, bias, and normalized errors $NRMSE_{std}$ (normalized by the target standard deviation) and $NRMSE_{range}$ (normalized by the target range). Each configuration is repeated with different random seeds and summarised as mean $\pm$ standard deviation.

2.10.2 Masked Validation for 3D CNN The trained 3D-CNN model is evaluated using a masked validation designed to simulate missing pixels in the last frame of a temporal window ($T = 7$ days). For each day in 2023, a spatio-temporal feature stack is constructed and the model predicts the pollutant concentration of the final frame.

Masking is applied only to valid pixels of the last frame. Several masking strategies are considered: (1) **Random Pixel**, where valid pixels are randomly sampled across the spatial domain; (2) **Block32**, where contiguous square regions (e.g., 32×32 pixels) are masked to simulate realistic spatial gaps; (3) **Sensitive-channel masking**, where all channels, including neighbour-based predictors, are masked at held-out locations to prevent information leakage; and (4) **Monthly Block32**, where evaluation is restricted to selected months to investigate seasonal effects.

The model predictions are compared with ground truth only at the masked pixels. Performance is assessed using RMSE, MAE, R^2, Pearson correlation, and $NRMSE_{std}$. Each configuration is repeated multiple times with different random seeds and summarised as mean $\pm$ standard deviation.

2.11 Gap-filling Inference and Output Generation

The trained models are applied to generate gap-filled pollutant maps for the year 2023. For each day, the feature stack is constructed using the same variables and standardization parameters as in training. Missing inputs are retained as NaN and handled by the model during inference.

Predictions are generated for all pixels and reconstructed on the native Sentinel-5P raster grid. The resulting daily maps are exported as compressed Float32 GeoTIFF files aligned with the original spatial reference system.

For deep learning models, predictions are de-standardized to recover pollutant concentrations in physical units before export.

3. Results

3.1 Results of NO_2 and SO_2 Annual Exploratory Analysis

The gap maps in Figures 4 and 5 are consistent with the statistics in Table 1.

Across the AOI, NO_2 exhibits moderate data gaps, with a five-year mean of 45.4% (median 41.2%) and a 90th percentile of 60.7%. Most pixels fall within the 20–50% category. In contrast, SO_2 shows substantially poorer coverage, with a mean gap fraction of 77.4% (median 74.9%) and a 90th percentile of 88.9%. The majority of pixels lie within the 50–80% range or above 80%.

These results indicate that SO_2 observations are considerably sparser than NO_2 across the study region.

3.2 Results of NO_2 and SO_2 Seasonal Data Gaps Pattern

Seasonal patterns are pronounced and consistent across years: NO_2 shows moderate gaps that are higher in December–January–February (DJF) and lower in June–July–August (JJA), while SO_2 exhibits persistently high data gaps, with the strongest deficits in DJF.

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-195-2026 |

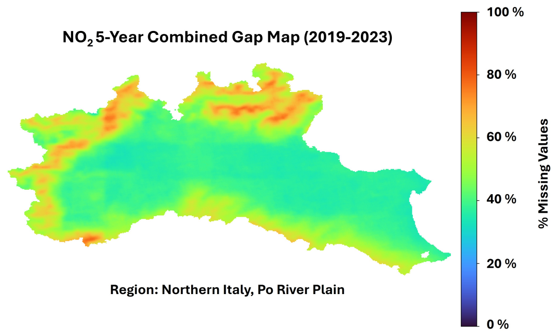

Figure 4. NO_2 5-Year Combined Gap Map (2019–2023).

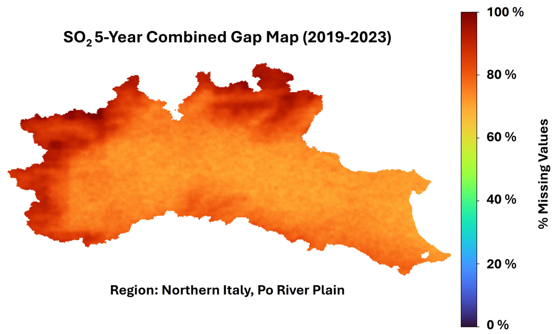

Figure 5. SO_2 5-Year Combined Gap Map (2019–2023).

Species	Mean	Median	Std. dev.	90th pct.
NO_2 (2019–2023)	45.4%	41.2%	0.099	60.7%
SO_2 (2019–2023)	77.4%	74.9%	0.068	88.9%

Species	$< 20\%$	20–50%	50–80%	$\geq 80\%$
NO_2	0.0%	70.4%	29.6%	0.0%
SO_2	0.0%	0.0%	73.1%	26.9%

Table 1. Five-year summary by species over the AOI (2019–2023). Ratios are computed per pixel from daily coverage, treating values ≤ 0 or `NaN` as missing. Std. dev. = standard deviation of pixel-wise five-year data gaps ratios; 90th pct. = value below which 90% of pixels fall.

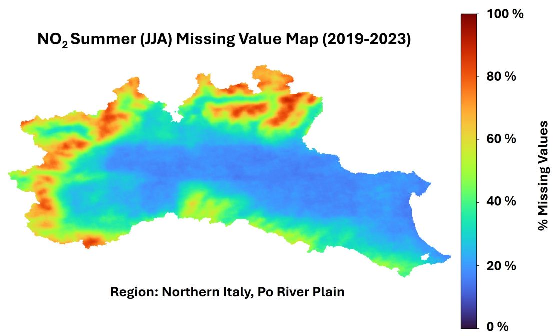

Figure 6. NO_2 Summer (JJA)

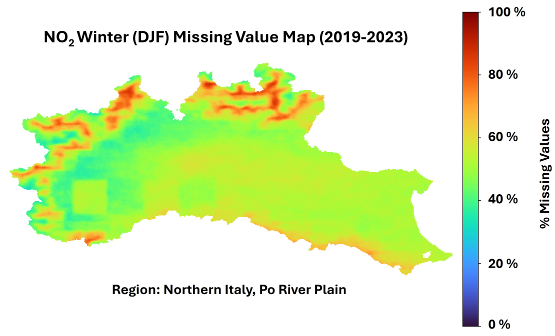

Figure 7. NO_2 Winter (DJF)

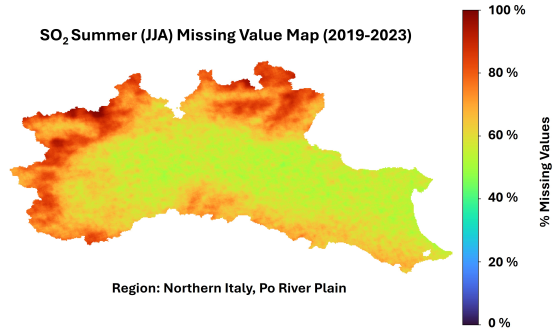

Figure 8. SO_2 Summer (JJA)

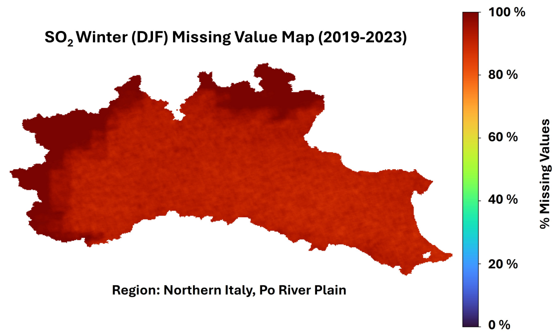

Figure 9. SO_2 Winter (DJF)

Species	Season	Mean	Median	Std. dev.	90th pct.
NO_2	Spring	46.0%	42.2%	0.119	63.7%
NO_2	Summer	35.6%	29.8%	0.180	65.4%
NO_2	Autumn	47.5%	45.5%	0.078	59.6%
NO_2	Winter	52.7%	52.3%	0.079	62.7%
SO_2	Spring	72.5%	67.6%	0.121	95.7%
SO_2	Summer	64.3%	61.7%	0.095	79.3%
SO_2	Autumn	80.0%	79.3%	0.039	84.6%
SO_2	Winter	93.1%	91.8%	0.035	100.0%

Table 2. Seasonal data gaps summary over the AOI (pooled 2019–2023). Ratios are pixel-wise fractions of missing days within each season.

Seasonal patterns are consistent across species (see Figures 4–9 and Tables 2–3). For NO_2, summer (JJA) shows the lowest gaps (mean 35.6%) and winter (DJF) the highest (52.7%); for SO_2, data gaps is high in all seasons and peaks in winter (DJF mean 93.1%, 90th pct. 100%). Category shares reinforce this contrast: in JJA, 25.4% of NO_2 pixels fall below 20% data gaps, whereas SO_2 has essentially no pixels below 50% in any season.

3.3 Results of Comparative Analysis of NO_2/SO_2

Overall, SO_2 exhibits substantially higher data gaps than NO_2 (average 77.4% vs. 45.4%). Seasonally, both species are worst in winter (DJF) and best in summer (JJA), but SO_2 remains high in every season (Table 4). The seasonal contrast between

Species	Season	$< 20\%$	20–50%	50–80%	$\geq 80\%$
NO_2	Spring	0.0%	65.9%	33.9%	0.2%
NO_2	Summer	25.4%	52.0%	21.4%	1.1%
NO_2	Autumn	0.0%	67.2%	32.8%	0.0%
NO_2	Winter	0.0%	32.2%	67.3%	0.5%
SO_2	Spring	0.0%	0.0%	77.1%	22.9%
SO_2	Summer	0.0%	0.1%	90.6%	9.3%
SO_2	Autumn	0.0%	0.0%	56.0%	44.0%
SO_2	Winter	0.0%	0.0%	0.0%	100.0%

Table 3. Share of AOI pixels by seasonal data gaps category (2019–2023).

species (SO_2 minus NO_2) is smallest in spring and largest in winter.

	NO_2 mean	SO_2 mean	Δ (SO_2–NO_2)
Overall (2019–2023)	45.4%	77.4%	+32.0 pp
Spring (MAM)	46.0%	72.5%	+26.5 pp
Summer (JJA)	35.6%	64.3%	+28.7 pp
Autumn (SON)	47.5%	80.0%	+32.5 pp
Winter (DJF)	52.7%	93.1%	**+40.4 pp**

Table 4. Average data gaps (AOI means) for NO_2 vs. SO_2 (2019–2023) and seasonal differences. Percentages denote the pixel-wise fraction of missing days; Δ reports percentage-point differences.

The observed differences in data gaps patterns indicate that SO_2 and NO_2 may require distinct modeling considerations. For SO_2, it is plausible to emphasize temporal/context features and to adopt validation that reflects DJF scarcity; for NO_2, leveraging spatial structure may be relatively more effective. These are hypotheses to be tested rather than definitive prescriptions.

3.4 Temporal Trends and Environmental Drivers of NO_2/SO_2 Data Gaps

Across the AOI, pixel-wise temporal trends in data gaps are generally weak. Only about 3.2% of pixels show statistically significant trends ($p < 0.05$) for both NO_2 and SO_2 during 2019–2023; NO_2 trends are predominantly decreasing (2,896 pixels; median Sen's slope $\approx -0.011\,\mathrm{yr}^{-1}$), while SO_2 trends are mainly increasing (2,669 pixels; median $\approx +0.0077\,\mathrm{yr}^{-1}$). However, the corresponding significance analysis (Figure 10) indicates that the vast majority of these trends are not statistically different from zero (p $\geq$ 0.05), suggesting that the observed changes are weak and not robust. In contrast, topographic effects are pronounced: data gaps increase strongly with elevation, with Pearson correlations of $r = 0.769$ (NO_2) and $r = 0.920$ (SO_2). This strong relationship likely reflects the influence of complex terrain and higher cloud frequency in mountainous regions, which reduce satellite retrieval quality and increase the occurrence of missing data. Cloudiness also explains a substantial fraction of NO_2 missing data (Pearson $r = 0.708$), whereas SO_2 shows a strong negative correlation with cloud fraction ($r = -0.905$), reflecting that SO_2 data gaps are largely driven by additional quality screening and low signal-to-noise filtering rather than cloud cover alone.

3.5 Model Performance:LightGBM and 3D CNN

3.5.1 LightGBM In masked validation, a random $r\%$ of the observed test pixels in 2023 (pooled over all days) are hidden and predicted. A valid sample is any masked pixel with an observed target; other inputs are available because non-sensitive missing features are mean-imputed, while the sensitive channels (`lag1`, `neighbor`) are intentionally set to NaN

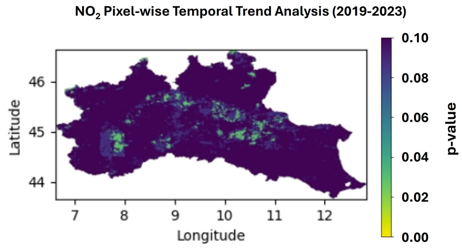

Figure 10. NO_2 pixel-wise temporal trend significance (p-values) for 2019-2023.

(which LightGBM handles natively). With the same r, NO_2 yields 531,849 / 1,063,699 / 1,595,548 valid masked pixels (10/20/30%), whereas SO_2 yields far fewer, reflecting its lower observational coverage in 2023.

Both pollutants show consistent performance across mask ratios (Tables 5–6), with SO_2 achieving slightly higher R^2 (0.1327–0.1351) compared to NO_2 (0.0879–0.0888), while NO_2 exhibits better correlation (0.4198–0.4208 vs. 0.3698–0.3724). $\mathrm{NRMSE_{std}}$ values remain near 1.0 for both species, and bias is negligible across all conditions.

The modest R^2 values indicate limited predictive power when temporal and spatial context features are unavailable, establishing a performance baseline for subsequent 3D-CNN evaluation.

Table 5. SO_2 vs NO_2 LightGBM gap-filling masked validation on 2023 (10% and 20% mask).

	10% mask		20% mask	
Metric	SO_2	NO_2	SO_2	NO_2
R^2	0.134	0.089	0.135	0.088
$\mathrm{NRMSE_{std}}$	0.930	0.955	0.930	0.955
Corr	0.372	0.421	0.372	0.420
Bias	4.2e-05	-7e-06	4.1e-05	-7e-06
Samples n	1.77e5	5.32e5	3.54e5	1.06e6

Table 6. SO_2 vs NO_2 LightGBM gap-filling masked validation on 2023 (30% mask).

	30% mask	
Metric	SO_2	NO_2
R^2	0.1327	0.0879
$\mathrm{NRMSE_{std}}$	0.9313	0.9551
Corr	0.3698	0.4198
Bias	0.000042	-0.000007
Samples n	530,908	1,595,548

3.5.2 3D CNN This study comprehensively evaluates 3D CNN models for predicting NO_2 and SO_2 concentrations using four sampling strategies, with three independent runs for statistical reliability. Table 7 summarizes the comparative performance metrics.

Random pixel sampling demonstrated higher performance under the chosen validation setup. NO_2 achieved strong results with $R^2 = 0.9469$, Corr = 0.9748, and minimal errors (RMSE = 0.000013, MAE = 0.000010), while SO_2 showed moderate performance with $R^2 = 0.7387$ and Corr = 0.8755. Block sampling strategies performed moderately for NO_2 ($R^2 = 0.0971$) but poorly for SO_2 ($R^2 = 0.0503$), with masked neighborhood processing showing negligible impact. March block sampling revealed significant differences, achieving the best block strategy

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-195-2026 |

performance for NO_2 (R^2 = 0.1259) but worst for SO_2 (R^2 = -0.1018). NO_2 models consistently outperformed SO_2 across all strategies, with the performance gap most pronounced in random pixel sampling.

Table 7. Validation results of 3D CNN models for NO_2 and SO_2 prediction

Pollutant	Strategy	NRMSE_std	MAE	R^2	Corr
NO_2	Random Pixel	0.2305	0.000010	0.9469	0.9748
	Block32	0.9502	0.000037	0.0971	0.4368
	Masked Block	0.9502	0.000037	0.0971	0.4368
	March Block	0.9349	0.000028	0.1259	0.4351
SO_2	Random Pixel	0.5111	0.000220	0.7387	0.8755
	Block32	0.9745	0.000353	0.0503	0.3893
	Masked Block	0.9745	0.000353	0.0503	0.3893
	March Block	1.0497	0.000474	-0.1018	0.3618

Future work should include multiple runs and uncertainty estimates to better assess model robustness and variability.

3.6 Gap-Filling results

The qualitative results further illustrate the performance of the models in reconstructing missing satellite observations. For NO_2, the 3D-CNN successfully reproduces the main spatial structures observed in the original data, including the pronounced pollution gradients across the Po Valley. The gap-filled outputs show strong agreement with the original scenes, particularly in regions with coherent large-scale patterns, although some smoothing of fine-scale variability is visible.

For SO_2, the reconstruction is more challenging due to the higher noise levels and sparsity of the original observations. While the model is able to generate spatially coherent fields and recover the general distribution of concentrations, the predictions exhibit reduced variability compared to the original data. In particular, localized peaks and sharp gradients are less pronounced, indicating a tendency toward spatial smoothing.

Overall, the gap-filling approach demonstrates the ability to reconstruct incomplete scenes and produce spatially continuous fields, especially for NO_2. However, the results also highlight that the model performance depends on the pollutant characteristics, with noisier variables such as SO_2 showing lower fidelity in capturing fine-scale spatial variability.

4. Conclusions

This study explored machine learning approaches for gap-filling missing Sentinel-5P observations of NO_2 and SO_2 over the Po Valley. Both LightGBM and 3D-CNN models were evaluated, with the 3D-CNN showing consistently better performance across all validation settings. In particular, the model achieved a substantial improvement in predictive skill, increasing R^2 from approximately 0.09 to 0.95 for NO_2 and from 0.13 to 0.74 for SO_2 under random masking. The results indicate that incorporating spatio-temporal context is essential for improving reconstruction accuracy, especially for pollutants with strong spatial coherence such as NO_2. The 3D-CNN model was able to reconstruct missing regions and improve the spatial continuity of the observations, particularly at large scales. However, the analysis also highlights limitations in capturing fine-scale variability, especially for noisier pollutants such as SO_2, where the model tends to produce smoother fields and underrepresents localized extremes. These findings suggest that while deep learning provides a promising framework for satellite data gap-filling, further work is needed to better preserve spatial variability and quantify reconstruction uncertainty.

Future work should focus on uncertainty-aware approaches, improved validation, and the integration of additional data sources to enhance robustness, while recognizing the challenges of comparing satellite column data with surface observations.

5. Acknowledgements

This study was carried out within the Space It Up project funded by the Italian Space Agency, ASI, and the Ministry of University and Research, MUR, under contract n. 2024-5-E.0 - CUP n. I53D24000060005.

References

Appel, M., 2024. Efficient data-driven gap filling of satellite image time series using deep neural networks with partial convolutions. *CoRR*, abs/2208.08781. http://dblp.uni-trier.de/db/journals/corr/corr2208.htmlabs-2208-08781.

Fu, H., Shao, Z., Fu, P., Huang, X., Cheng, T., Fan, Y., 2022. Combining ATC and 3D-CNN for reconstructing spatially and temporally continuous land surface temperature. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102733. https://doi.org/10.1016/j.jag.2022.102733.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds), *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 3146–3154.

Liu, K., Li, X., Wang, S., Zhang, H., 2023. A robust gap-filling approach for European Space Agency Climate Change Initiative (ESA CCI) soil moisture integrating satellite observations, model-driven knowledge, and spatiotemporal machine learning. *Hydrology and Earth System Sciences*, 27(2), 577–598. https://hess.copernicus.org/articles/27/577/2023/.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., Schüssler, O., 2018. The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor. *Atmospheric Measurement Techniques*, 11(1), 409–427. https://amt.copernicus.org/articles/11/409/2018/.

Pirovano, G., Colombi, C., Balzarini, A., Riva, G., Gianelle, V., Lonati, G., 2015. PM2.5 source apportionment in Lombardy (Italy): Comparison of receptor and chemistry-transport modelling results. *Atmospheric Environment*, 106, 56-70. https://doi.org/10.1016/j.atmosenv.2015.01.073.

Siabi, N., Sanaeinejad, S. H., Ghahraman, B., 2022. Effective method for filling gaps in time series of environmental remote sensing data: An example on evapotranspiration and land surface temperature images. *Computers and Electronics in Agriculture*, 193, 106619. https://doi.org/10.1016/j.compag.2021.106619.

Veefkind, J., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H., de Haan, J., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel,

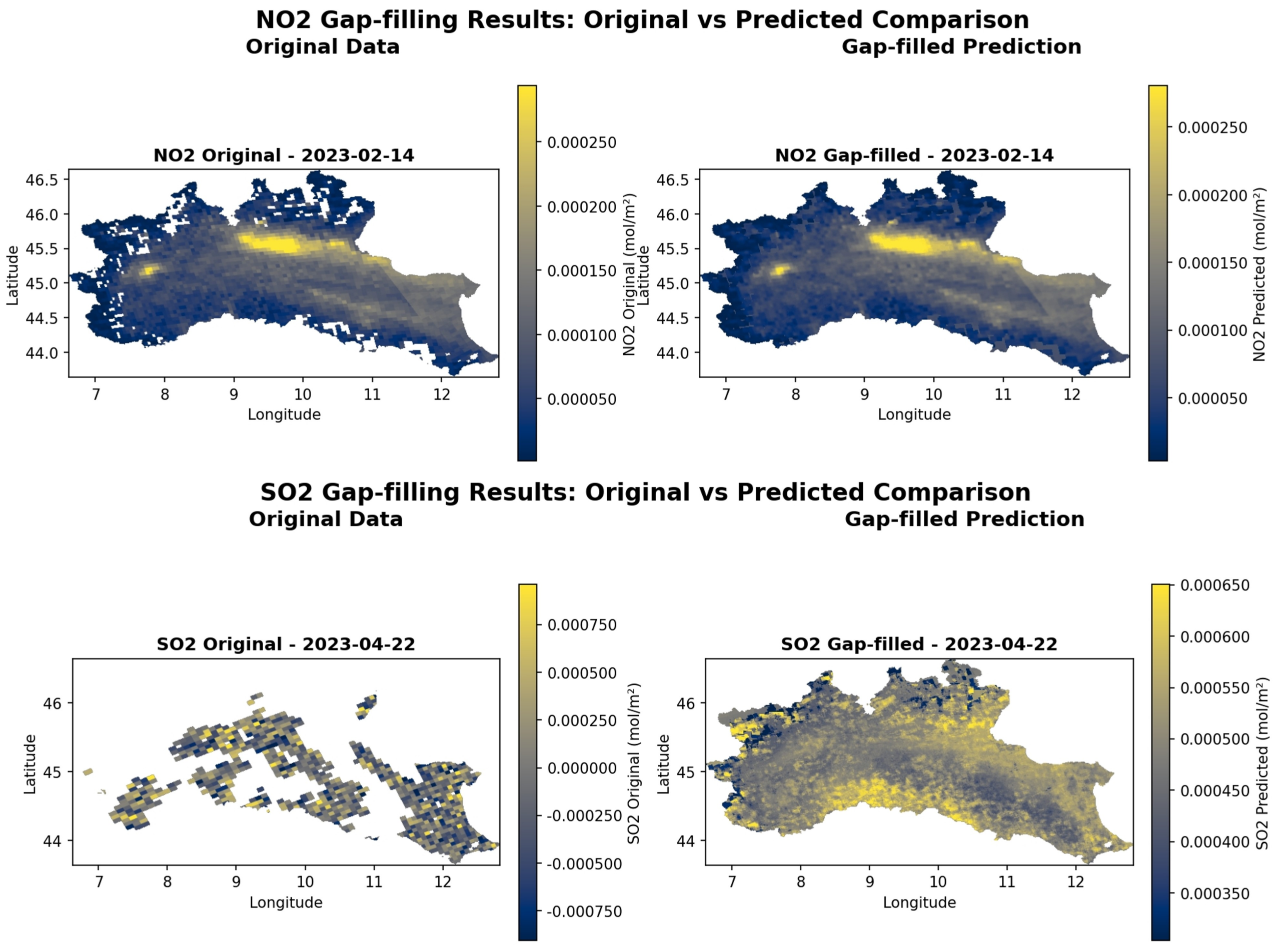

Figure 11. NO_2 and SO_2 LightGBM result examples. Left: originals; right: gap-filled. Predictions extend urban/transport plumes and fill fragmented swaths.

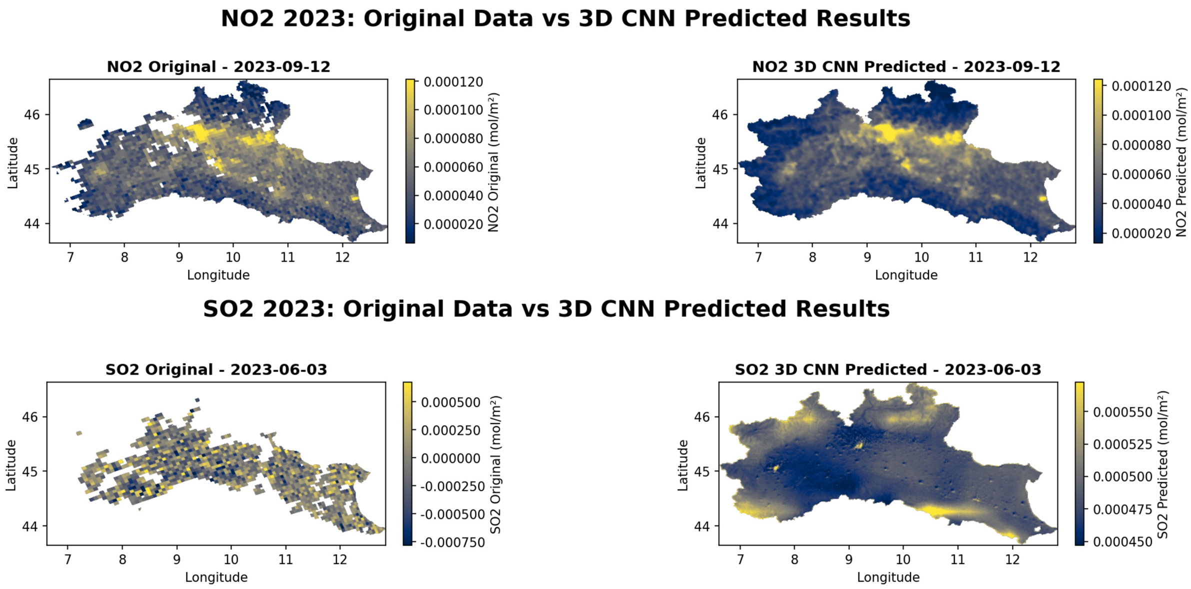

Figure 12. NO_2 and SO_2 3D-CNN result examples. Left: originals; right: gap-filled. Predictions extend urban/transport plumes and fill fragmented swaths

R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., Levelt, P., 2012. TROPOMII 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*, 120, 70-83. https://doi.org/10.1016/j.rse.2011.09.027.