



Technical Note: DACNO₂ – A Multi-Constraint Deep Learning Framework for High-Resolution 3D NO₂ Field Estimation

Wenfu Sun^{1,2}, Frederik Tack¹, Lieven Clarisse², Michel Van Roozendael¹

¹ Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium.

5 ² Université libre de Bruxelles (ULB), Spectroscopy, Quantum Chemistry and Atmospheric Remote Sensing (SQUARES), Brussels, Belgium.

Correspondence to: Wenfu Sun (wenfu.sun@aeronomie.be)

Abstract. Accurate, high-resolution 3D fields of nitrogen dioxide (NO₂) are critical for air quality management and satellite retrievals, yet traditional chemistry-transport models (CTMs) face challenges in fine-scale modeling. Machine learning (ML) alternatives often struggle with generalization and transferability, inheriting biases from CTMs or being limited by sparse surface measurements. We present the Deep Atmospheric Chemistry NO₂ model (DACNO₂), a deep learning model that generates daily 2 km 3D NO₂ fields over Western Europe. The model's three-phase and multi-constraint training strategy begins by pre-training on European Copernicus Atmosphere Monitoring Service (CAMS) reanalysis data to learn large-scale atmospheric patterns, then fine-tunes with both CAMS and in-situ European Environmental Agency (EEA) surface data to correct biases and refine local detail, and completes with an adaptive fine-tuning to capture evolving trends. An evaluation for 2023 shows that DACNO₂ reproduces broad-scale 3D CAMS fields ($R^2 = 0.90$) while improving agreement with independent EEA stations over the CAMS reanalysis (R^2 enhanced from 0.61 to 0.66; bias reduced from -1.15 to -0.38 $\mu\text{g}/\text{m}^3$). The model resolves more spatial detail and learns physically interpretable relationships. This hybrid training approach fuses the physical consistency of a process-based model with the real-world accuracy of surface measurements, overcoming the limitations of using either constraint data alone. Applying DACNO₂ a-priori profiles to TROPOMI retrievals increases tropospheric NO₂ columns by 3% on average over those using European CAMS profiles, with larger enhancements over emission hotspots. These results demonstrate the framework's potential to advance air quality monitoring and satellite remote sensing.



25 1. Introduction

Nitrogen dioxide (NO₂) is a key atmospheric pollutant with significant impacts on air quality, human health, ecosystems, and atmospheric chemistry. Primary sources include traffic, industrial activities, and energy production, with additional contributions from natural emissions (Crippa et al., 2018). Accurate characterization of the spatiotemporal distribution of NO₂ is critical for both air pollution management and atmospheric chemistry
30 research.

Chemistry Transport Models (CTMs) such as GEOS-Chem (Bey et al., 2001), TM5-MP (Krol et al., 2005; Williams et al., 2017; Huijnen et al., 2010), WRF-Chem (Grell et al., 2005), and the Copernicus Atmosphere Monitoring Service (CAMS) (Peuch et al., 2022; Inness et al., 2019) are widely used to simulate atmospheric NO₂
35 based on physical and chemical processes. However, most CTMs operate at coarse spatial resolution due to computational constraints and the limited availability of high-resolution emission inventories. This restricts their ability to represent fine-scale NO₂ variability and often results in spatial smoothing and underestimation, particularly in urban environments. Emission inventories are usually outdated and may omit localized and small-scale sources (Lu et al., 2025), contributing to uncertainties and discrepancies between bottom-up and top-down
40 emission estimates (Kuik et al., 2018; Yang et al., 2021). While regional high-resolution CTMs are available, such as CAMS at 10 km (Douros et al., 2023; Ialongo et al., 2020) and regional WRF-Chem at 3 km (Kuhn et al., 2024b), challenges remain in accurately capturing urban and fine-scale NO₂ patterns (Meleux et al., 2024), and model optimization is often resource-intensive (Kuhn et al., 2024a; Kuhn et al., 2024b).

CTM outputs also serve as a-priori NO₂ profiles for satellite retrievals (Palmer et al., 2001; Douros et al., 2023; Yang et al., 2023), supporting large-scale NO₂ monitoring. Over the past three decades, satellite NO₂ observations have been advancing toward higher spatiotemporal resolution. Satellite instruments such as the TROPOspheric Monitoring Instrument (TROPOMI, $7 \times 3.5 \text{ km}^2$, $5.5 \times 3.5 \text{ km}^2$ since August 2019) on Sentinel-5P (Veefkind et al., 2012), the Geostationary Environment Monitoring Spectrometer (GEMS, $3.5 \times 8 \text{ km}^2$) (Kim et al., 2020),
50 Tropospheric emissions: Monitoring of pollution (TEMPO, $2 \times 4.75 \text{ km}^2$) (Zoogman et al., 2017), Sentinel-4 ($8 \times 8 \text{ km}^2$) (Gulde et al., 2017), Sentinel-5 ($7.5 \times 7.5 \text{ km}^2$) (Bézy et al., 2014), Twin ANthropogenic Greenhouse Gas Observers (TANGO, $300 \times 300 \text{ m}^2$) (Landgraf et al., 2020), and the Copernicus Anthropogenic Carbon Dioxide Monitoring constellation (CO2M, $2 \times 2 \text{ km}^2$) (Bernd et al., 2021) are advancing spaceborne NO₂ observations to kilometer-scale resolution. This progress has increased demand for high-resolution a priori profiles,



55 which can improve the sensitivity of satellite NO₂ products to near-surface concentrations and emission hotspots. However, CTM-based profiles remain constrained by the limitations mentioned above, highlighting the need for alternative modeling approaches.

Machine learning (ML) provides an efficient alternative for high-resolution NO₂ estimation. ML techniques have
60 been widely applied for surface NO₂ mapping (Sun et al., 2024; Kim et al., 2021; Wei et al., 2022), and recent studies have extended these approaches for 3D NO₂ modeling above the surface. These studies have trained models on process-based 3D NO₂ fields generated by CTMs (Bodnar et al., 2024; Kuhn et al., 2024a), on vertical profiles from MAX-DOAS observations (Zhang et al., 2025; Zhang et al., 2022b; Jiang et al., 2025), and on a combination of process-based CTM outputs with satellite observations (Li and Xing, 2024). While these studies
65 demonstrate the feasibility of ML-based 3D NO₂ modeling, challenges remain in achieving high spatial resolution, robust generalization, and transferability. Process-based data carries inherent biases and has relatively coarse resolution. Ground-based observations are sparse and unevenly distributed, limiting the model's spatial generalization. While Li and Xing (2024) combine process-based NO₂ fields with satellite NO₂ observations from the Ozone Monitoring Instrument (OMI) to train the ML model, the resulting product is still limited to a coarse
70 resolution (27 km).

In this study, we present the Deep Atmospheric Chemistry NO₂ model (DACNO₂), a deep learning model developed to produce daily, high-resolution (2 km) 3D NO₂ fields with high accuracy, robust generalization, and transferability. DACNO₂ integrates multi-source inputs, including emissions, geography, meteorology, and
75 temporal indicators. The model is trained using a phased, multi-constraint approach that combines process-based CAMS fields with ground-based EEA measurements. This method enables the model to reproduce broad-scale, process-based NO₂ patterns and capture local NO₂ gradients. The training strategy consists of three phases: pre-training, multi-constraint fine-tuning, and adaptive fine-tuning. Western Europe (5°W–9°E, 42°N–54°N) is chosen as the study region, given its diverse topography, high urbanization, and substantial industrial activity.

80 This study addresses two key research questions: (1) Can a deep learning framework combining multi-constraint and phased training overcome the resolution, bias, and generalization limitations of current CTM and ML approaches for 3D NO₂ modeling? (2) Does the DACNO₂ 2 km product improve fine-scale NO₂ representation to support applications in regional air quality management and satellite retrievals?



The remainder of this paper is organized as follows. Section 2 describes the DACNO₂ development framework, including dataset preparation, model architecture, and training strategy. Section 3 evaluates model performance. Section 4 discusses broader insights and implications. Conclusions and outlook are provided in Section 5.



2. Development Framework for DACNO₂

90 2.1 Framework Overview

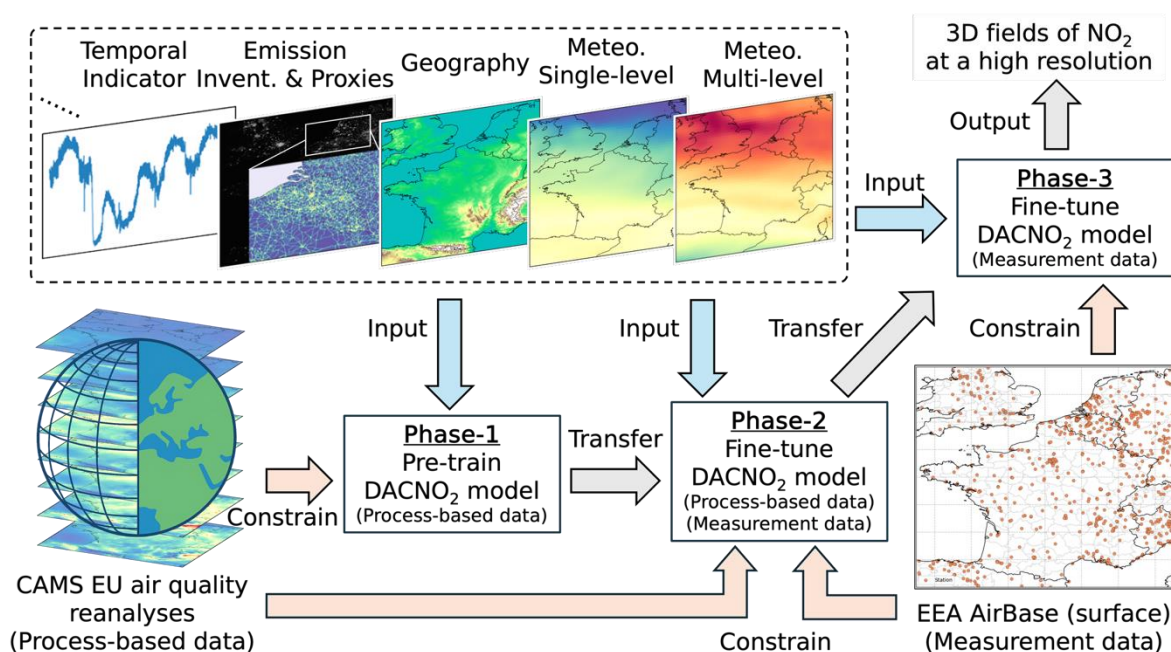


Figure 1. Overview of the DACNO₂ model development framework. The framework integrates multiple input data streams: temporal indicators, emission inventories and proxies, geography, and ERA5 meteorological data, with two target datasets: process-based NO₂ from CAMS European air quality reanalysis and ground-based in-situ EEA NO₂ measurements. The training is organized in three sequential phases: pre-training on process-based CAMS NO₂, multi-constraint fine-tuning with both CAMS and EEA data, and adaptive fine-tuning to recent NO₂ trends. The resulting model generates daily, high-resolution (2 km) 3D NO₂ fields. Arrows indicate the data flow and phased training process.

DACNO₂ is developed to provide daily 3D NO₂ fields at high spatial resolution (2 km) with improved accuracy and generalizability by integrating multi-source data, physically consistent process-based datasets, and real-world measurements. The overall framework, illustrated in Fig. 1, combines diverse data streams with a phased training strategy.

DACNO₂ uses five groups of input features: temporal indicators, emission inventories and proxies, geographic data, ERA5 single-level meteorological variables, and ERA5 multi-level meteorological variables. Together, they provide complementary information on spatial and temporal NO₂ variability. For model training, the targets are process-based NO₂ fields from the CAMS European air quality reanalyses (Inness et al., 2019; Peuch et al., 2022)



and real-world surface NO₂ measurements from the EEA AirBase network (European Environment Agency, 2024). CAMS supplies physically consistent large-scale 3D NO₂ distributions, while EEA data constrain the model to match local concentration patterns. Details on data preparation are provided in Section 2.2.

110

To effectively learn NO₂ patterns from diverse datasets, DACNO₂ employs an encoder–decoder architecture with five dedicated encoder branches, each tailored to a specific group of input features. The model structure is described in Section 2.3.

115

Model training is organized into three sequential phases. In Phase 1, a baseline model is pre-trained on process-based CAMS data. In Phase 2, the model is further trained with both process-based and measurement data, improving its ability to represent local NO₂ gradients. In Phase 3, the model is fine-tuned using recent measurements to reflect current NO₂ trends and support real-world applications. Details of the training approach are provided in Section 2.4.



2.2 Dataset Preparation

2.2.1 Input Features

DACNO₂ utilizes 38 input datasets, organized into five groups: temporal indicators, emission inventories and proxies, geography, single-level meteorology, and multi-level meteorology. Details of all input features and their sources are provided in Table 1.

The temporal indicator group consists of the day of the week and the daily number of flights. The day of the week captures regular human activity cycles, reflecting variability between weekdays and weekends. Data on the daily number of flights, aggregated for nine countries in the study area (Eurocontrol, 2025), can indicate irregular activity such as holiday periods or major events, which may help explain the irregular changes in NO₂ emissions.

Emission inventories and proxies include anthropogenic NO_x emission inventories, road density, population density, and nighttime light. These features provide direct and indirect measures of NO_x emissions, with high-resolution proxies complementing inventories at finer spatial scales. All datasets are resampled to a 2 km grid using interpolation, averaging, or rasterization methods.

Geographic datasets include land elevation and land cover, providing terrain context to the ML model. Elevation influences atmospheric transport by creating physical barriers that can trap pollutants (Giovannini et al., 2020), while land cover serves as a proxy for the location and type of surface emissions (Beelen et al., 2013). Land cover is categorized into five classes: artificial surfaces, agricultural areas, forests and semi-natural areas, wetlands, and water bodies, aggregated from the original 44 categories by the mode method (the most frequently occurring land cover type). Both elevation and land cover data are resampled to a 2 km grid.

Meteorological features provide atmospheric information from the surface through the free troposphere, obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 hourly single-level and multi-level (pressure-level) datasets (Hersbach et al., 2020). We use 24-hour meteorological features for the target day.

Meteorological data are horizontally resampled to a 16 km grid, for three reasons: (1) the native ERA5 resolution (0.25°, approximately 25 km) is coarser than 2 km, and bilinear interpolation would mainly introduce artificial smoothness rather than genuine fine-scale gradients; (2) retaining many meteorological variables at 2 km would impose a significant computational burden; and (3) the DACNO₂ architecture uses a hierarchical encoder-decoder, where upscaling and downscaling follow a factor-of-two scaling scheme (e.g., 2 km, 4 km, 8 km, 16 km). Although



150 the ERA5-Land data can provide higher resolution (0.1° , approximately 9 km), it only covers the continental areas, which is not consistent with the model application scope.

The day-of-week feature is normalized using sine and cosine transforms to retain its cyclical nature. Land cover is one-hot encoded to convert categorical data into a numerical format. All other input features are normalized
155 with z-scores, based on the mean and standard deviation of the training set.

Notably, satellite-derived NO_2 products were deliberately excluded from the input features for two key reasons. First, frequent data gaps in satellite products, due to cloud cover and quality control, would propagate into the model's output, preventing the generation of continuous, gap-free fields. Second, this exclusion allows for an
160 independent evaluation of the model against satellite observations and preserves the potential to use satellite data as an independent constraint in future work.



Table 1. Summary of inputs and training targets for the DACNO₂ model development.

Group	Data	Spatial Resolution	Temporal Resolution	Dimension	Data Source
Inputs					
Temporal indicator	Daily number of flights	-	Daily	1D	(Eurocontrol, 2025)
	Day of week	-	-	(time)	-
Emission inventories and proxies	CAMS global emission inventories	Anthropogenic sector (0.1° × 0.1°)	-	2D (latitude × longitude)	(Soulie et al., 2024)
		Shipping sector (0.25° × 0.25°)	(average for 2018)		
	Road networks (five types)	Vector data	-		Global Roads Inventory Project (GRIP) global roads database (Meijer et al., 2018)
	Nighttime light	500 m	-		Annual global Visible Infrared Imaging Radiometer Suite (VIIRS) dataset (Elvidge et al., 2021)
	Population	1 km	(average for 2019 to 2022)		JRC-GEOSTAT 2018 gridded dataset (Silva et al., 2021)
Geography	Land elevation	90 m	-		Multi-Error-Removed Improved-Terrain digital elevation models (MERIT DEM) (Yamazaki et al., 2017)
	Land cover (resampled to 5 classes via mode aggregation)	100m	-		Coordination of Information on the Environment (CORINE) Land Cover 2018 (Feranec et al., 2016)
Meteorology	Boundary layer height	0.25°	Hourly	3D (latitude × longitude × time)	European Center for Medium-Range Weather Forecasts (ECMWF) ERA5 hourly time-series data on single levels from 1940 to the present (Hersbach et al., 2020)
	Mean boundary layer dissipation				
	Surface pressure				
	Dewpoint Temperature				
	2m Temperature				
	10 m U and V wind				
	Leaf area index (for low and high vegetation)				



	Mean evaporation rate				
	Mean snow rate				
	Mean surface net radiation flux (short and long wave)				
	Mean total precipitation rate				
	Geopotential			4D	
	Vorticity	0.25°		(latitude × longitude × height × time)	ECMWF ERA5 hourly data on pressure levels from 1940 to the present
	Relative humidity	8 layers from 1000 hPa to 550 hPa			
	Temperature				
	U and V wind				
	Vertical velocity				
Training targets					
Process-based data	CAMS European air quality reanalysis NO ₂	0.1° 8 layers from the surface to 5000 m	Daily (average from hourly)	3D (latitude × longitude × height)	(Peuch et al., 2022; Inness et al., 2019)
Surface measurements	EEA AirBase surface NO ₂	-	Hourly	2D (latitude × longitude)	EEA AirBase network (European Environment Agency, 2024)

2.2.2 Training Targets

The training targets include CAMS European air quality reanalysis profile data (CAMS NO₂) and in-situ measurements from the EEA AirBase network (EEA NO₂). The datasets are both listed in Table 1. CAMS NO₂ offers extensive and continuous 3D NO₂ data aligned with physical and chemical processes, while EEA NO₂ provides ground-based in-situ measurements from sparsely distributed monitoring stations.

CAMS NO₂ is the median ensemble of 11 different regional models (Inness et al., 2019; Peuch et al., 2022). The dataset provides hourly NO₂ distributions at eight vertical heights above the surface (surface, 50m, 100m, 250m, 500m, 750m, 1000m, 2000m, 3000m, and 5000m) and has a horizontal resolution of 0.1° × 0.1° (10 km × 10 km). CAMS NO₂ has assimilated EEA observations and includes both interim and validated reanalyses. Interim data relies on near-real-time observations without full validation, whereas validated data undergo rigorous quality control with an additional delay. In this study, we used CAMS NO₂ data from 2019 to 2023, where the 2019–2021 data are validated reanalysis data and the 2022–2023 data are interim reanalysis data, based on data availability.



CAMS NO₂ was processed by averaging hourly data to daily values and by bilinearly interpolating its horizontal resolution from 10 km to 8 km to match the model's scaling scheme. In addition, CAMS NO₂ concentrations at each vertical level were rescaled by multiplying them by the inverse of the ratio between the mean NO₂ concentration at that level and the mean surface-level NO₂ concentration, where the ratio was calculated from the training dataset. This adjustment ensures that the model gives adequate attention to higher-altitude NO₂ concentrations, which are otherwise much lower than surface values and could be neglected during training (Li and Xing, 2024; Kuhn et al., 2024a). During model inference, the predicted NO₂ concentrations at each level were multiplied by the corresponding ratio to restore the original vertical profile.

EEA NO₂ was collected from background and industrial monitoring stations (European Environment Agency, 2024) and mapped onto a 2 km grid. Traffic stations were excluded because their measurements represent a very local area, significantly smaller than the 2 km grid cells of our study. If multiple stations were located within the same grid cell, their values were averaged. When both background and industrial stations existed in a grid cell, the cell was classified as background. In total, 773 grid cells with measurements were identified, with 575 assigned for training and 198 for final evaluation. Because background EEA NO₂ is assimilated into CAMS, the split of background stations followed the CAMS model assimilation system (Copernicus Atmosphere Monitoring Service, 2024) to prevent data leakage, while industrial stations were randomly split. All EEA NO₂ data were converted from hourly to daily averages.

2.2.3 Patch-Based Data Processing and Reconstruction

To balance the model's receptive field and computational efficiency, we used a patching method. Specifically, all datasets except the temporal indicator were divided into patches of 512 km × 512 km with partial overlap. This produced grid sizes of 32 × 32 for ERA5 meteorological data, 64 × 64 for CAMS NO₂ data, and 256 × 256 for emission inventories and proxies, geographic data, and EEA NO₂ data. In this study, each patch was treated as a single input sample for the model, and 12 samples were generated for one day. More samples can be generated as needed by reducing the stride of the sliding window. During model inference, the output patches were merged using a weighted averaging scheme based on a 2D Hann window (Oppenheim, 1999), which assigns lower weights to patch edges and higher weights to central regions. For each grid cell, weighted values from all overlapping patches were summed and normalized by the total weights. This reconstruction method reduced edge artifacts in overlapping areas and ensured smooth transitions across patch boundaries.



2.3 Model Architecture and Design

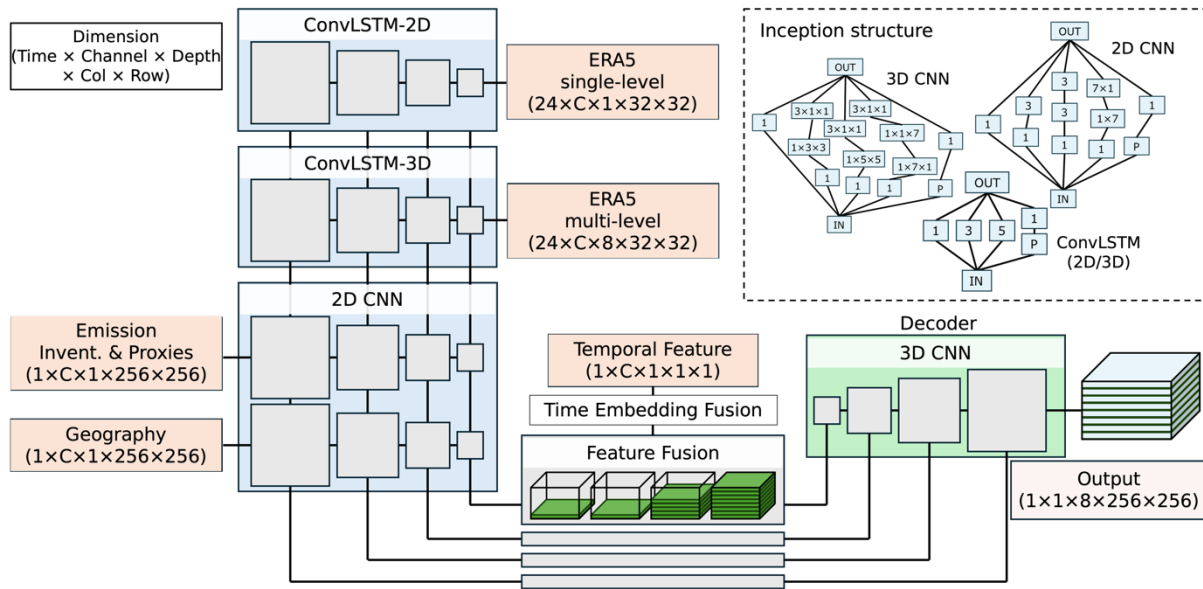


Figure 2. DACNO₂ model architecture. The model features a multi-branch encoder–decoder design for daily 3D NO₂ prediction. Five input groups are processed separately: ERA5 single-level meteorological variables (ConvLSTM-2D), ERA5 multi-level meteorological variables (ConvLSTM-3D), emission inventories and proxies (2D CNN), geography (2D CNN), and temporal features (MLP-based embedding fusion). Outputs from all encoder branches are fused and passed into a unified 3D CNN decoder to generate high-resolution NO₂ fields. The architecture enables the extraction of spatial, temporal, and multi-level atmospheric features, supporting fine-scale NO₂ modeling. Input and output dimensions are indicated for each module.

The architecture of the DACNO₂ model is illustrated in Fig. 2. The model adopts an encoder–decoder framework with residual connections (He et al., 2016) to map multi-source input features to the daily 3D NO₂ field. DACNO₂ integrates several types of neural network modules, including multilayer perceptron (MLP), convolutional neural network (CNN), and convolutional long short-term memory (ConvLSTM), to process and fuse heterogeneous input tensors. Each module is chosen for its specific strengths in handling different data structures. ConvLSTM is for spatiotemporal sequences, CNN is for spatial hierarchies, and MLP is for tabular feature vectors. Inception-style structures are applied in several neural network modules to enable the model to capture both local-scale and broader-scale spatial features.



2.3.1 Encoder and Decoder

DACNO₂ encodes ERA5 single-level (hourly 2D) and multi-level (hourly 3D) meteorological data using ConvLSTM-2D and ConvLSTM-3D modules, respectively. Both modules are based on the ConvLSTM architecture proposed by Shi et al. (2015), which combines convolutional layers for spatial feature extraction with long short-term memory (LSTM) units for temporal sequence modeling. ERA5 data are processed using a progressive upscaling strategy, where the horizontal grid size increases stepwise from 32×32 to 64×64 , 128×128 , and 256×256 , while the vertical dimension remains at 8 for multi-level inputs. This upscaling preserves spatial detail and enables residual connections to the decoder, unlike conventional encoders that downsample feature maps. To manage computational cost, the temporal sequence length is halved after each ConvLSTM block through subsampling, resulting in sequence lengths of 24, 12, 6, and 3 at successive stages. At each stage, the last time slice is extracted for feature fusion.

Emission and geographic variables are encoded by dedicated 2D CNN blocks, which extract hierarchical spatial features as the resolution decreases from 256×256 to 32×32 . At the 32×32 latent stage, features from all four branches are passed through CNN-based transition layers, each forming a 3D tensor. For each branch, feature values are assigned only to physically relevant vertical levels within the tensor, while all other levels are set to zero. Specifically, emission and geographic features are assigned to the surface level, ERA5 single-level features are placed in the lowest five levels, and ERA5 multi-level features span all vertical levels. The resulting tensors are concatenated along the channel dimension and fused using a 3D CNN block. Temporal indicators are encoded by an MLP, then expanded to match the latent spatial dimensions, and integrated at this stage, allowing the model to capture both spatial and temporal context. The same feature fusion scheme is applied to residual connections between the encoder and decoder at multiple spatial scales, although temporal embedding is used only at the 32×32 stage.

The decoder uses 3D CNN modules with hierarchical upscaling from 32×32 to 256×256 in the horizontal dimension, while maintaining a vertical size of 8. This structure learns spatial correlations across multiple altitude levels and captures both horizontal and vertical dependencies in NO₂ distributions. All 2D and 3D CNN blocks use the sigmoid linear unit (SiLU) activation function (Elfwing et al., 2017), while the output layer uses the softplus activation function to ensure non-negative estimates of the 3D NO₂ field.



2.3.2 Inception-Based Modules

To enhance multi-scale feature extraction, DACNO₂ incorporates inception modules throughout its architecture (Fig. 2), inspired by the work of inception architecture (Szegedy et al., 2014; Szegedy et al., 2015). The core concept of this architecture is to use parallel convolutional operations with varying kernel sizes, enabling the model to efficiently capture both fine-scale and broad features simultaneously. In the ConvLSTM-2D and ConvLSTM-3D branches, each inception block applies parallel convolutional operations with varying kernel sizes (1×1, 3×3, 5×5) and a max-pooling branch, enabling the model to capture both local and broader spatiotemporal patterns. The 2D CNN modules extend this approach, combining parallel 1×1, 3×3, and 5×5 convolutions, a factorized 7×7 path (decomposed into 1×7 and 7×1 convolutions), and a pooling branch. For 3D CNN modules, inception blocks use parallel convolutions with different spatial and vertical kernel shapes, such as 1×1×1, 1×3×3, and 3×1×1, along with a pooling branch. In all cases, each parallel branch includes its own batch normalization, activation, and dropout, after which the outputs are concatenated along the channel dimension. A similar design has been applied in a previous deep learning model for NO₂ estimation (Zhang et al., 2022a). It enables the model to effectively integrate information across multiple spatial and vertical scales, improving the representation of complex atmospheric NO₂ distributions.



2.4 Three-Phase Training Strategy

The DACNO₂ model development employs a three-phase training strategy, including pre-training, multi-constraint fine-tuning, and adaptive fine-tuning. Such a strategy enables the model to learn general patterns (e.g., a-priori knowledge) from a broad dataset and then transfer this internal knowledge to improve its performance on a new, more specific task. Similar approaches have been widely adopted in the development of artificial intelligence (AI) models across various domains such as earth system modeling, large language models, and biomedical image analysis (Zhuang et al., 2019; Zhou et al., 2017; Ding et al., 2023; Bodnar et al., 2024).

2.4.1 Phase-1

In the first phase, the DACNO₂ model was pre-trained on the CAMS NO₂ data. This dataset provides physically consistent 3D NO₂ distributions by assimilating real-world observations into chemical transport models (Inness et al., 2019), enabling the model to learn comprehensive 3D NO₂ patterns governed by broad-scale atmospheric processes. This approach is inspired by recent progress in AI weather modeling (Bi et al., 2023; Lam et al., 2023) and the earth system foundation model (Bodnar et al., 2025), which uses ERA5 and CAMS data for 3D forecasting of weather and air quality.

In this step, the training loss is defined as the sum of the Mean Squared Error (MSE) loss and the Structural Similarity Index Measure (SSIM) loss (Zhao et al., 2017; Zhou et al., 2004) between the DACNO₂ prediction and the CAMS NO₂ data at 8 km resolution.

$$Loss = Loss_{MSE_{DACNO_2-CAMS}} + Loss_{SSIM_{DACNO_2-CAMS}} \quad (1)$$

MSE quantifies the absolute differences in NO₂ concentrations, while SSIM evaluates the similarity of spatial patterns between model outputs and the CAMS reference. SSIM is computed independently at each vertical level by comparing normalized 2D horizontal slices of the predicted and reference NO₂ fields. Specifically, each slice is min-max normalized to the range of 0 to 1 prior to SSIM calculation, ensuring that the SSIM loss reflects only structural similarity rather than magnitude differences. The final SSIM loss is calculated as one minus the mean SSIM across all vertical levels. This dual-loss formulation encourages the model to match both the overall concentration values and the detailed spatial structures of 3D NO₂ fields.

The model was trained and validated using a random sample split from the 2019, 2021, and 2022 datasets (80% for training, 20% for validation), with 2023 reserved as an independent test set. Data from 2020 was excluded



from this process because a preliminary experiment showed that its inclusion substantially degraded the model performance on the unknown period (i.e., 2022 data, which was initially held out as an independent validation year in that experiment). This might be due to the unexpectedly higher NO₂ concentrations above 1000 m in that year (Fig. S1), which is also documented in the CAMS Evaluation and Quality Control (EQC) report (Meleux et al., 2023). While the cause remains unclear, we speculate that this anomaly is related to the substantial decrease in NO_x emissions during 2020 due to the COVID-19 pandemic (Levelt et al., 2022) and not well accounted for in the CAMS model. We evaluate and discuss DACNO₂ performance for that special year in Section 4.4.

2.4.2 Phase-2

In the second phase, we fine-tuned DACNO₂-Phase-1 by introducing an additional MSE constraint based on EEA NO₂, while maintaining the CAMS NO₂ constraints, as shown in Equation (2). The EEA NO₂ MSE was computed at the surface level and only for 2 km grids with available EEA data

$$Loss = LossMSE_{DACNO_2-CAMS} + LossSSIM_{DACNO_2-CAMS} + LossMSE_{DACNO_2-EEA} \quad (2)$$

The EEA NO₂ data were split into training and validation sets using the same spatiotemporal alignment as the CAMS NO₂ split. Most training settings remained consistent with the first phase, except that the learning rate was reduced and the EEA NO₂ MSE term was added to both the training loss and the validation metric. The model with the best validation performance was selected and is referred to as DACNO₂-Phase-2 for subsequent use. Although Phase-2 includes the same CAMS constraint as Phase-1, which may make Phase-1 appear redundant, we recommend retaining Phase-1. Skipping directly to Phase-2 can cause the model to overfit local EEA observations and limit its ability to learn generalizable NO₂ patterns from process-based data.

2.4.3 Phase-3

Recent changes in air quality policies and emission technologies (Castellanos and Boersma, 2012; Wang et al., 2021; Chang et al., 2023) may introduce systematic NO₂ variations that are not well represented in the historical training dataset. To ensure the DACNO₂ model remains adaptable to such real-world changes, we introduced a third phase. In this step, we adopted a strategy inspired by the data assimilation system in the CAMS model (Inness et al., 2019). DACNO₂-Phase-2 was fine-tuned using only EEA NO₂ data from training stations during the test period (2023 in this study), to mimic real-world application. To maintain spatial patterns learned from earlier phases, a regularization term based on SSIM was added to both the training loss and validation metric. SSIM was computed at 8 km resolution between predictions from the updated model and DACNO₂-Phase-2 (equation 3):

$$Loss = LossMSE_{DACNO_2(new)-EEA} + LossSSIM_{DACNO_2(new)-DACNO_2(Phase-2)} \quad (3)$$



325 This approach allows the updated model to adjust prediction magnitudes in response to new measurements while preserving spatial patterns established in previous phases, since the CAMS constraint is no longer available in Phase-3. The model with the best validation performance was selected and is referred to as DACNO₂-Phase-3, which incorporates recent real-world NO₂ variations while retaining consistency with patterns learned during earlier training.

330 **2.4.4 Training and Implementation**

DACNO₂ was trained and implemented in Python using PyTorch on two NVIDIA A30 GPUs. Training was performed with a batch size of 56, achieved by gradient accumulation. The first and second training phases each required approximately three weeks to complete 200 epochs on three years of data. The third training phase required about one week for 100 epochs on a single year of data. Once trained, the model generates daily NO₂ estimates for the whole area within minutes. Further efficiency improvements are possible through hardware upgrades or model optimization.

335



3. Assessing DACNO₂ Performance and Evolution

3.1 Model Performance Across Training Phases

Table 2. Performance of DACNO₂ on the 2023 test dataset.

DACNO ₂ -Phase-1					DACNO ₂ -Phase-2				DACNO ₂ -Phase-3				CAMS-2km			
EEA	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)
Total	5.88	0.75	0.52	-0.93	5.81	0.79	0.54	1.67	4.99	0.82	0.66	-0.38	5.32	0.80	0.61	-1.15
Urban	6.45	0.76	0.47	-2.57	5.64	0.80	0.59	0.49	5.34	0.82	0.64	-1.42	5.89	0.82	0.56	-2.89
Suburban	5.11	0.79	0.58	-0.04	6.01	0.81	0.42	2.87	4.55	0.83	0.67	0.58	4.39	0.84	0.69	-0.23
Rural	5.22	0.76	0.42	1.96	5.96	0.76	0.25	3.17	4.59	0.78	0.55	1.05	4.79	0.80	0.51	1.98
DACNO ₂ -10km																
Total levels	0.99	0.96	0.91	-0.05	1.17	0.96	0.87	0.25	1.02	0.95	0.90	-0.10	CAMS (10km)			
L0	1.92	0.94	0.85	-0.36	2.29	0.93	0.79	0.75	1.94	0.93	0.85	-0.63				
L50	1.58	0.95	0.88	-0.25	1.94	0.94	0.82	0.66	1.62	0.94	0.87	-0.31				
L250	1.05	0.93	0.83	0.10	1.19	0.93	0.78	0.41	1.09	0.93	0.82	0.01				
L500	0.70	0.92	0.79	0.09	0.71	0.92	0.79	0.16	0.79	0.90	0.74	0.05				
L1000	0.31	0.89	0.72	0.03	0.30	0.89	0.74	0.03	0.39	0.86	0.56	0.06				
L2000	0.08	0.81	0.61	-0.01	0.08	0.81	0.64	-0.01	0.09	0.79	0.48	0.00				



L3000	0.04	0.78	0.55	-0.01	0.03	0.77	0.58	0.00	0.04	0.74	0.40	0.00
L5000	0.01	0.67	0.31	0.00	0.01	0.66	0.34	0.00	0.02	0.61	-0.01	0.00

Note: For the comparison against EEA NO₂ (shown in the upper panel), both DACNO₂ outputs and CAMS NO₂ were evaluated at 2 km grids. The CAMS NO₂ data was interpolated to a 2 km resolution (CAMS-2km) and used as a baseline in this comparison. For evaluating DACNO₂ using CAMS NO₂ (shown in the lower panel), DACNO₂ outputs were downsampled and evaluated at the original 10 km resolution of CAMS across all vertical levels as well as for individual levels.

The performance of the DACNO₂ model was evaluated using both EEA NO₂ and CAMS NO₂ test data from 2023. For the comparison against EEA NO₂ (results in the upper panel of Table 2), both DACNO₂ outputs and CAMS NO₂ were evaluated at 2 km grids. The CAMS NO₂ data was interpolated to 2 km resolution (CAMS-2km) and served as a baseline in this comparison. For the evaluation of DACNO₂ using CAMS NO₂ (results in the lower panel of Table 2), DACNO₂ outputs were evaluated at the CAMS original 10 km resolution across all vertical levels as well as for individual levels. Evaluation metrics included root mean squared error (RMSE), Pearson correlation coefficient (r), coefficient of determination (R^2), and bias.

Comparisons with EEA NO₂ indicate progressive improvement across DACNO₂ training phases. DACNO₂-Phase-3 achieves the best overall agreement (RMSE = 4.99 ug/m³, r = 0.82, R^2 = 0.66, bias = -0.38 ug/m³), outperforming both DACNO₂-Phase-1 (RMSE = 5.88 ug/m³, r = 0.75, R^2 = 0.52, bias = -0.93 ug/m³) and DACNO₂-Phase-2 (RMSE = 5.81 ug/m³, r = 0.79, R^2 = 0.54, bias = -1.67 ug/m³). This demonstrates the added value of incorporating local EEA constraints and adaptive fine-tuning. Compared to the interpolated CAMS-2km dataset (RMSE = 5.32 ug/m³, r = 0.80, R^2 = 0.61, bias = -1.15 ug/m³), DACNO₂-Phase-3 shows improved accuracy and reduced bias. Station-type analysis further highlights the advantages of DACNO₂-Phase-3, especially at urban and rural sites. For urban locations, DACNO₂-Phase-3 achieves better agreement (RMSE = 5.34 ug/m³, r = 0.82, R^2 = 0.64, bias = -1.42 ug/m³) compared with CAMS-2km (RMSE = 5.89 ug/m³, r = 0.82, R^2 = 0.56, bias = -2.89 ug/m³). In rural areas, DACNO₂-Phase-3 reduces the bias (RMSE = 4.59 ug/m³, bias = 1.05 ug/m³) compared to CAMS-2km (RMSE = 4.79 ug/m³, bias = 1.98 ug/m³). These results suggest that DACNO₂ is more effective than CAMS at capturing localized emission patterns and small-scale spatial variability in areas with either dense emission sources or diffuse background levels.



Comparisons with CAMS NO₂ across all levels show that DACNO₂ effectively learns and preserves 3D NO₂ distributions across all training phases. Near the surface (Level 0 m), DACNO₂-Phase-3 maintains strong agreement with CAMS (RMSE = 1.94 µg/m³, $r = 0.93$, $R^2 = 0.85$, bias = -0.63 ug/m³), and performance remains robust at mid-altitudes (Level 500 m: RMSE = 0.79 ug/m³, $r = 0.90$, $R^2 = 0.74$, bias = 0.05 ug/m³), similar to earlier phases. At higher levels, particularly 5000 m, the agreement decreases, with DACNO₂-Phase-3 yielding a near-zero R^2 (-0.01) compared to 0.31 in DACNO₂-Phase-1 and 0.34 in DACNO₂-Phase-2. However, Pearson correlation remains moderate ($r \geq 0.6$), indicating that differences at these heights are mainly due to magnitude adjustment rather than loss of spatial structure. The very low NO₂ concentrations (approximately 0.1 µg/m³ at 3000 m and 0.03 µg/m³ at 5000 m; see Fig. S1) at these high levels also increase relative noise and contribute to metric variability. This reduction in agreement at upper levels mainly reflects the greater uncertainty and lack of direct constraints at high altitudes, which remains a key challenge for ML-based 3D air quality modeling.



3.2 Model Evolution in the Multi-constraint Strategy

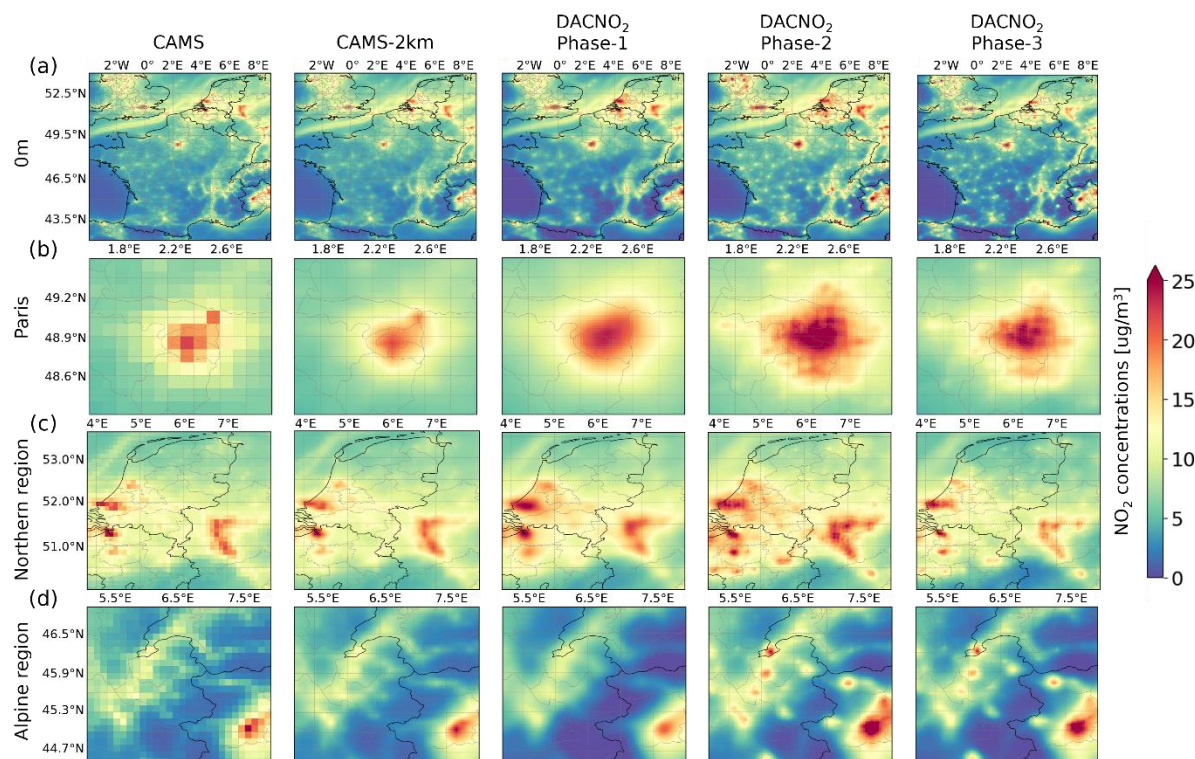


Figure 3. Spatial comparison of surface NO₂ estimates for 2023 from multiple models. (a) Annual mean surface NO₂ fields over the entire study region from DACNO₂-Phase-1, DACNO₂-Phase-2, DACNO₂-Phase-3, CAMS (10 km), and CAMS-2km (bilinearly interpolated to 2 km). (b–d) Enlarged views for three representative local areas: (b) Paris, (c) the northern region (NO₂ hotspot area encompassing the Netherlands, Belgium, and the Ruhr area), and (d) the Alpine region.

To further illustrate the evolution of estimated NO₂ spatial distributions achieved through a phased training, multi-constraint strategy, Fig. 3 compares average surface NO₂ estimates for 2023 from DACNO₂-Phase-1, DACNO₂-Phase-2, DACNO₂-Phase-3, CAMS, and CAMS-2km. Results are shown for the full study region and three representative local areas of Paris, the northern region (NO₂ hotspot area encompassing the Netherlands, Belgium, and the Ruhr area), and the Alpine region.

Across the study region (Fig. 3a), all models exhibit broad and similar NO₂ patterns over land and ocean, consistent with the high spatial agreement between DACNO₂ and CAMS NO₂ reported in Section 3.1.



395 Nonetheless, DACNO₂-Phase-2 and DACNO₂-Phase-3 yield sharper spatial contrasts and more clearly defined
local NO₂ hotspots than CAMS and DACNO₂-Phase-1. As an additional experiment, we trained the model using
only EEA NO₂ data, resulting in the DACNO₂-onlyobs version. As shown in Fig. S2, this model yields effective
NO₂ estimates primarily limited to the land surface and cannot reproduce the shipping track patterns, which are
visible in the CAMS and DACNO₂ results. Meanwhile, this model produces obvious artifacts over the ocean and
400 at higher altitudes due to the lack of training constraints. These differences highlight the significance of the CAMS
NO₂ constraint in facilitating broad spatial generalization in ML-based models.

Differences between models become more pronounced when focusing on local regions (Figs. 3b–d). CAMS NO₂
exhibits visible pixelation effects in these areas due to its coarse native resolution. While bilinear interpolation (as
405 in CAMS-2km) can smooth these effects, it does not introduce additional spatial detail, resulting in oversmoothed
patterns. DACNO₂-Phase-1 shows a spatial NO₂ distribution similar to CAMS-2km, despite using high-resolution
input features from emission proxies and geography. This indicates that constraints from CAMS NO₂ alone are
insufficient for the model to capture fine-scale local NO₂ variability. Incorporating the EEA NO₂ constraint in
DACNO₂-Phase-2 addresses this limitation, inspired by approaches in recent ML-based high-resolution surface
410 NO₂ modeling studies using ground measurements as targets (Sun et al., 2024; Wei et al., 2022; Kim et al., 2021;
Ghahremanloo et al., 2023). DACNO₂-Phase-2 reconstructs spatial patterns of NO₂ that better match urban layout
in Paris (Fig. 3b), identifies more small-scale emission hotspots in the northern region (Fig. 3c), and enhances
hotspot signals in the Alpine region (Fig. 3d). DACNO₂-Phase-3 retains these spatial characteristics and primarily
adjusts concentration magnitudes by assimilating new measurements to better represent actual NO₂ levels during
415 the application period. For example, the average surface NO₂ concentration estimate in Paris decreases from 11.89
µg/m³ in DACNO₂-Phase-2 to 10.08 µg/m³ in DACNO₂-Phase-3. This evolution demonstrates the value of
integrating multiple constraints and adaptive fine-tuning for high-resolution NO₂ estimation.



3.3 Global and Local Differences Between DACNO₂ and CAMS

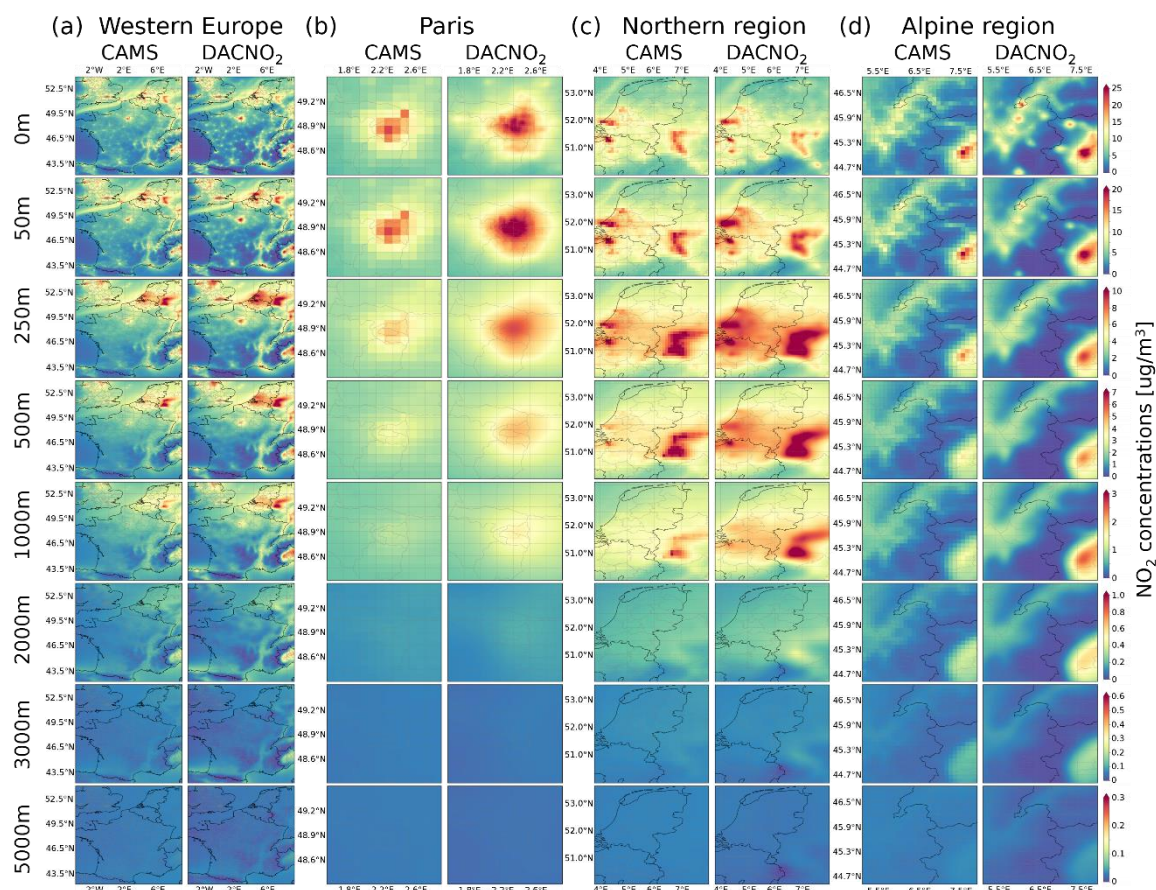


Figure 4. Annual mean NO₂ distributions for 2023 estimated from DACNO₂-Phase-3 (2 km) and CAMS (10 km) at multiple vertical levels. Level-wise average NO₂ distributions over (a) Western Europe (entire study region), (b) Paris, (c) the northern region, and (d) the Alpine region.

To further analyze differences in 3D NO₂ estimates between DACNO₂ and CAMS, Fig. 4 compares their annual average NO₂ distributions for 2023 across all vertical levels over the entire study region and three selected local areas. At the regional scale (Fig. 4a), DACNO₂ and CAMS show strong overall agreement at all altitudes, demonstrating that DACNO₂ effectively learns and reproduces large-scale 3D NO₂ structures from CAMS. However, DACNO₂ provides enhanced spatial detail, presenting sharper gradients and better-defined urban and industrial hotspots, particularly from the surface up to 250 m. At higher altitudes, the differences between the two models gradually diminish, accompanied by a decrease in NO₂ concentrations. Nevertheless, subtle magnitude discrepancies persist, with DACNO₂ estimates reaching lower values, sometimes approaching zero.

Local-scale comparisons further highlight the advantages of DACNO₂ (Figs. 4b–d). In the Paris region (Fig. 4b), DACNO₂ provides finer spatial detail and stronger NO₂ signals at lower altitudes (e.g., 0 m: 10.08 µg/m³; 50 m:



8.94 $\mu\text{g}/\text{m}^3$; 250 m: 4.65 $\mu\text{g}/\text{m}^3$), whereas CAMS results remain coarser with generally lower estimates (0 m: 8.43 $\mu\text{g}/\text{m}^3$; 50 m: 7.15 $\mu\text{g}/\text{m}^3$; 250 m: 3.63 $\mu\text{g}/\text{m}^3$). In the northern region (Fig. 4c), DACNO₂ more distinctly resolves localized emission sources at low levels, capturing a greater number of hotspots than CAMS. As a result, the average NO₂ concentration from DACNO₂ is elevated throughout the boundary layer (up to 1000 m), with mean values 8.8% higher than those from CAMS. In the Alpine region (Fig. 4d), DACNO₂ more effectively represents terrain-driven gradients and captures NO₂ signals within mountainous areas, demonstrating greater sensitivity to complex topographic influences. At higher altitudes, fine-scale variability diminishes in both models and their predicted NO₂ fields become more similar. This is because the influence of local emissions and surface features weakens, while regional-scale processes and long-range transport dominate (see Section 4.1). This reduced difference is also accompanied by much lower NO₂ concentrations at higher altitudes.



4. Insights and Implications of DACNO₂

4.1 Feature Importance and Data-driven Insights



Figure 5. Relative importance of each input feature group for DACNO₂ model predictions, evaluated using the integrated gradients (IG) method. (a) Feature group contributions to RMSE between DACNO₂ surface NO₂ estimates and EEA ground-based measurements for 2023. (b) Feature group contributions to RMSE between DACNO₂ and CAMS NO₂ estimates at different vertical levels for 2023. The five feature groups are: temporal indicators, emission inventories and proxies, geography, ERA5 single-level meteorology, and ERA5 multi-level meteorology. Results are shown for each model training phase (Phase-1, Phase-2, and Phase-3), illustrating how the relative influence of input feature groups varies with training constraints and altitude. See Figure S3 for the contributions of individual features within each group.

We assessed the relative importance of input feature groups in DACNO₂ using the integrated gradients (IG) method (Sundararajan et al., 2017) implemented via the Captum interpretability library (Kokhlikyan et al., 2020). IG quantifies the effect of varying each input feature from a zero baseline to its actual value on a selected target function. In this analysis, we computed IG at two targets: (1) the RMSE between DACNO₂ predictions and 2023 EEA NO₂ training measurements at the surface, and (2) the RMSE between DACNO₂ predictions and 2023 CAMS



460 NO₂ at multiple vertical levels. Feature group results are shown in Fig. 5, and results for individual features are provided in Fig. S3.

For surface NO₂ predictions evaluated against EEA measurements, DACNO₂ relies primarily on emission proxies, geographic features, and multi-level meteorological variables, while temporal indicators and single-level
 465 meteorological features play a lesser role. The addition of the EEA NO₂ constraint in Phase-2 and Phase-3 increases the importance of geographic data, highlighting its value for high-resolution surface NO₂ estimation. As shown in Fig. S3, land cover emerges as the most influential single feature in DACNO₂-Phase-3, accounting for 36.6% of the total importance. Multi-level meteorological variables dominate the meteorological contribution, although the radiation variable from the single-level meteorology group also remains relevant (1.0% in Phase-3,
 470 Fig. S3). This likely reflects some redundancy between single-level and multi-level meteorological features, with multi-level data providing more comprehensive atmospheric information.

For NO₂ estimates by level evaluated against CAMS, the distribution of input feature importance at lower levels (up to 1000 m) is similar to that for surface NO₂ evaluated against EEA, suggesting that DACNO₂ remains
 475 relatively stable across training phases with different constraints. Differences between the three-phase models are most apparent near the surface but gradually diminish with height. The importance of geographic features steadily decreases as height increases, while emission features reach their highest influence around 500 m before declining. Above 3000 m, both become negligible, reflecting the transition from the Planetary Boundary Layer (PBL), which is influenced by local surface features, into the free troposphere, which is dominated by broad-scale processes.

480 In contrast, temporal indicators, single-level meteorology, and especially multi-level meteorological features become increasingly important with height. For example, in DACNO₂-Phase-3, the contribution from multi-level meteorology rises from 16.5% at the surface to 71.8% at 5000 m, while temporal indicators and single-level meteorology also show steady increases. This shift highlights the greater reliance on temporal and large-scale
 485 atmospheric information for NO₂ estimates at higher levels. Among these features, radiation flux (3.3% at 3000 m, 2.4% at 5000 m) primarily drives the increase in single-level meteorology, while wind becomes the dominant variable among all meteorological features (approximately 21% at both 3000 m and 5000 m, Fig. S3). Given the consistently low overall contribution of single-level meteorological variables, future model development may consider reducing or refining the use of this feature group to streamline the input space.



4.2 Enhanced Vertical NO₂ Profile Representation

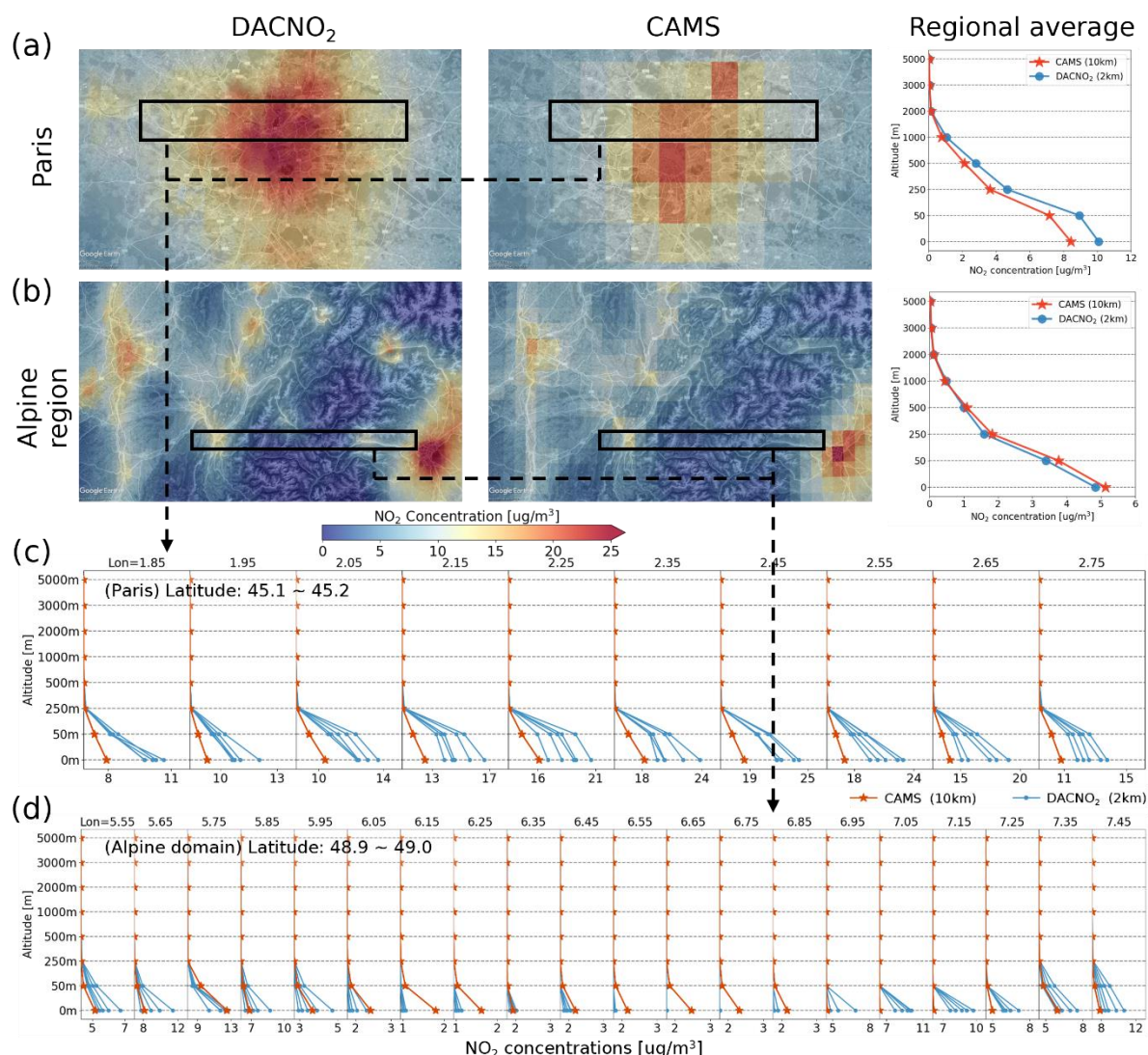


Figure 6. Comparison of NO₂ profile estimates from DACNO₂ (2 km resolution) and CAMS (10 km resolution) for the Paris and Alpine regions in 2023. Regional average vertical profiles and surface NO₂ distributions for Paris (a) and the Alpine region (b), with results shown over Google Earth imagery for geographic context. Interpixel variability of NO₂ profiles from DACNO₂ and CAMS along a 100 km transect (black boxes) in the Paris area (c) and a 200 km transect in the Alpine region (d), illustrating local-scale differences in vertical structure.

Fig. 6 compares the average 2023 NO₂ profile estimates from DACNO₂ (2 km) and CAMS (10 km) for the Paris and Alpine regions, with results overlaid on Google Earth imagery for geographic context. In Paris, the regional



average profile (Fig. 6a) indicates that DACNO₂ yields higher near-surface NO₂ concentrations and steeper vertical gradients compared to CAMS. This enhancement likely results from DACNO₂'s use of high-resolution emission proxies and land cover information, allowing the model to resolve smaller and more localized emission sources (Kuik et al., 2018; Shahrokhishahraki et al., 2022). At the local scale, we take transect over the grids of 100 km over Paris to compare the interpixel profile variability from CAMS (10 km) and DACNO₂ (2 km) (Fig. 6c). It is observed that DACNO₂ more clearly distinguishes spatial variability in the vertical structure, showing sharper contrasts and more pronounced local peaks than CAMS, particularly below 250 m. The regional average profiles for the Alpine area are similar between DACNO₂ and CAMS (Fig. 6b), which is due to the overall lower concentrations over this region with limited emission sources. However, local differences remain visible across a 200 km transect (Fig. 6d). DACNO₂ captures higher NO₂ concentrations around urban and small-scale hotspots, especially in valleys and canyons where pollutants tend to accumulate. Conversely, DACNO₂ provides lower NO₂ estimates in areas between the mountains with few sources. Overall, DACNO₂ provides more spatially detailed 3D NO₂ fields, revealing greater variability in the vertical profiles across different grids in this complex terrain. This refinement is important, as small point and line sources can contribute significant NO₂ in mountainous regions (Kim et al., 2021).



4.3 Implications for Satellite NO₂ Retrievals

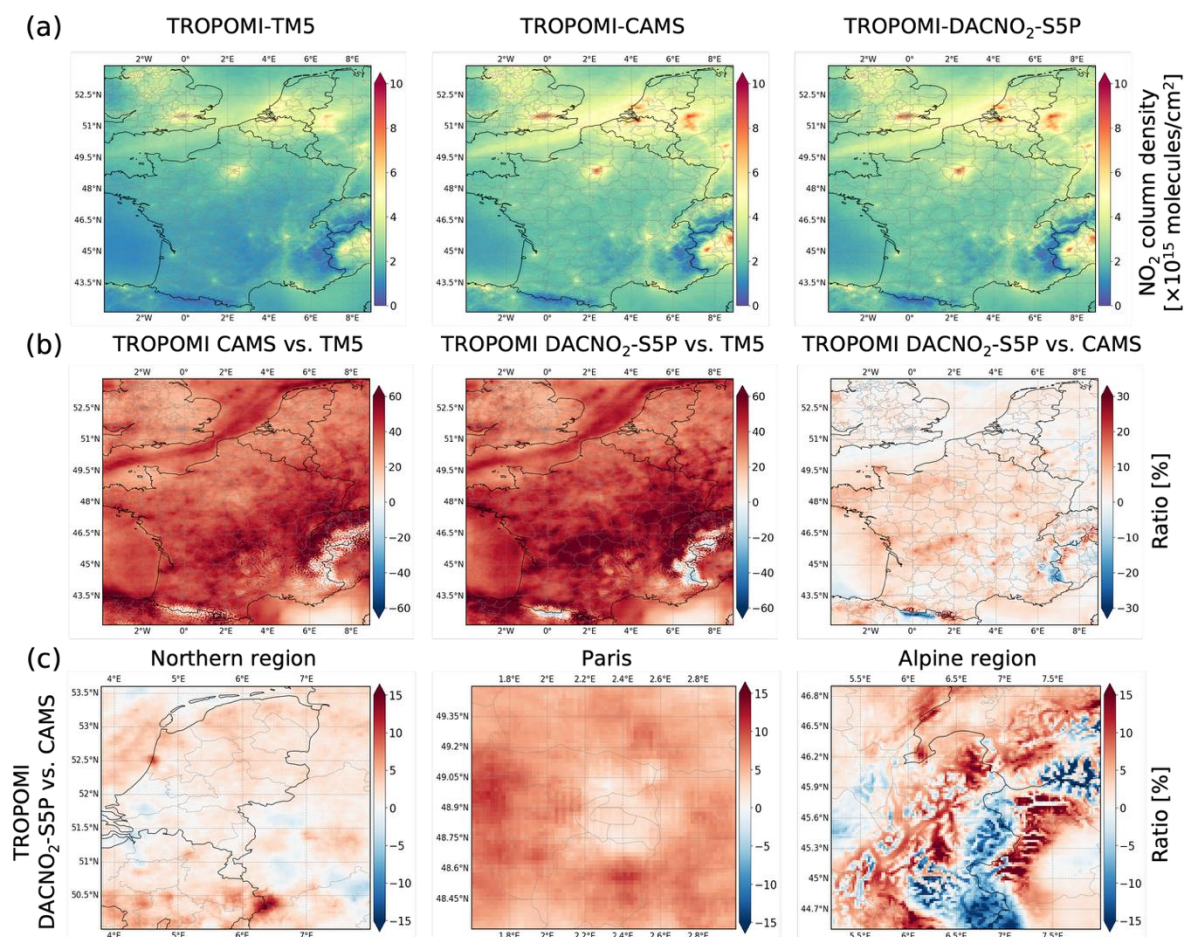


Figure 7. Impact of a-priori profile selection on TROPOMI tropospheric NO₂ column retrievals for 2023. (a) Annual mean TROPOMI NO₂ columns retrieved using the original TM5 (about 100 km), CAMS-S5P (10 km), and DACNO₂-S5P (2 km) a-priori profiles. (b) Spatial distribution of the relative difference (%) in TROPOMI NO₂ columns retrieved with three profiles. (c) The relative change in retrieved NO₂ columns across three subregions (the northern region, Paris, and the Alpine region) when using DACNO₂-S5P versus CAMS-S5P profiles.

To assess the potential of DACNO₂ for satellite NO₂ product improvement and development, we tested its use as a source of a-priori NO₂ profiles in TROPOMI retrievals. For this, a dedicated version of the model (DACNO₂-S5P) was developed for the TROPOMI overpass time, predicting a 3-hour average NO₂ (11:00–13:00 UTC) using the same three-phase strategy. The model targets, named CAMS-S5P and EEA-S5P, represent process-based and measured NO₂ data during this period.



Model evaluation (Table S1) shows DACNO₂-S5P agrees well with CAMS-S5P (RMSE = 0.98 µg/m³, $r = 0.94$, $R^2 = 0.88$, bias = 0.03 µg/m³) at 10 km grids. Compared to EEA-S5P measurements, DACNO₂-S5P achieves better agreement (RMSE = 5.07 µg/m³, $r = 0.77$, $R^2 = 0.59$, bias = 0.05 µg/m³) than CAMS-S5P (RMSE = 5.27 µg/m³, $r = 0.76$, $R^2 = 0.55$, bias = -0.94 µg/m³).

We replaced the original TM5 a-priori profiles (about 100 km resolution) in the TROPOMI retrievals with CAMS-S5P and DACNO₂-S5P profiles, following the approach described in Douros et al. (2023) and focusing on the troposphere. Fig. 7a presents the annual mean TROPOMI NO₂ columns retrieved using these different a priori profiles, with inter-comparisons shown in Figs. 7b and 7c. Both CAMS-S5P and DACNO₂-S5P profiles lead to substantial increases in the retrieved NO₂ columns, by 36.2% and 39.8% on average, respectively. The increase associated with CAMS-S5P is consistent with previous findings (Douros et al., 2023) and is primarily attributable to the improved spatial resolution of the a-priori profile, which enhances the sensitivity of TROPOMI retrievals to NO₂ hotspots (Tack et al., 2021; Ialongo et al., 2020).

Compared to the CAMS-S5P profile, using DACNO₂-S5P as the a priori increases retrieved NO₂ columns by about 3.0% on average (Fig. 7b), associated with the reduced negative bias against EEA-S5P measurements reported above. In central-western France (0°E–2.6°E, 45.6°N–46.3°N), a distinct southwest-to-northeast line of reinforced NO₂ columns appears because DACNO₂-S5P enhances the emission signals from the cities of Angoulême, Guéret, and Montluçon. Regional comparisons (Fig. 7c) show that the DACNO₂-S5P profile leads to a 1.8% increase in the northern region and 5.9% in Paris, with the most significant increases surrounding major emission hotspots. This is likely due to DACNO₂'s enhanced ability, based on finer resolution, to capture small-scale emission sources and resolve strong spatial gradients around NO₂ hotspots. In the Alpine region, the average increase reaches 1.7%, ranging from -18.1% (5th percentile) to +24.1% (95th percentile) between the surrounding areas and the central mountains, with a similar pattern observed in the Pyrenees. The absolute difference remains small, ranging from -3.94×10^{14} molecules/cm² (5th percentile) to 5.39×10^{14} molecules/cm² (95th percentile). This large fluctuation reflects the complex NO₂ distribution in the mountainous region and benefits from high-resolution modeling, as DACNO₂ estimates can reach lower background values or enhance the hotspots signal in this region.



These results illustrate the potential of using DACNO₂ profiles to improve satellite NO₂ retrievals, particularly for evolving high-resolution instruments. However, the DACNO₂ product remains a prototype, and a roadmap for operational deployment is outlined in the conclusion.



4.4 Generalization Capability and Data Quality: Insights from the COVID-19 Period

560 **Table 3. Performance of DACNO₂ on the 2020 dataset.**

Year 2020	DACNO ₂ -Phase-2				DACNO ₂ -Phase-3-2020				CAMS-2020-2km			
EEA	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)	RMSE (ug/m ³)	r	R ²	bias (ug/m ³)
Total	5.78	0.80	0.63	-0.27	5.47	0.82	0.67	0.01	4.99	0.88	0.73	-2.05
Urban	6.06	0.81	0.63	-1.53	5.62	0.83	0.68	-1.28	5.83	0.88	0.65	-3.44
Suburban	5.81	0.80	0.58	1.12	5.63	0.82	0.60	1.55	3.97	0.91	0.80	-1.08
Rural	4.97	0.81	0.63	1.08	4.84	0.82	0.65	1.24	3.80	0.89	0.78	0.14
DACNO ₂ -10km												
Total levels	1.66	0.91	0.80	-0.43	1.62	0.91	0.81	-0.34	CAMS-2020 (10 km)			
L0	2.18	0.93	0.84	0.43	2.09	0.94	0.86	0.48				
L50	1.95	0.93	0.84	0.57	2.06	0.94	0.82	0.77				
L250	1.97	0.89	0.74	-0.70	1.89	0.88	0.76	-0.50				
L500	2.06	0.83	0.50	-1.01	1.94	0.82	0.55	-0.81				
L1000	1.64	0.72	0.00	-1.01	1.57	0.70	0.09	-0.93				
L2000	1.42	0.49	-0.81	-1.01	1.40	0.47	-0.77	-0.99				
L3000	0.81	0.46	-0.99	-0.59	0.80	0.45	-0.95	-0.59				



L5000	0.17	0.55	-0.70	-0.12	0.17	0.56	-0.64	-0.12
-------	------	------	-------	-------	------	------	-------	-------

Note: Similar to Table 2, but for the year 2020.

As noted in Section 2.4.1, CAMS NO₂ data for 2020 were excluded from the training set based on preliminary experiments showing that their inclusion substantially reduced model generalization at higher levels. Since 2020 was marked by the COVID-19 pandemic and large reductions in anthropogenic emissions (Levelt et al., 2022), we specifically evaluated DACNO₂'s predictive performance for this atypical year. To this end, the DACNO₂-Phase-2 model was fine-tuned on 2020 EEA NO₂ data, following the same phased development strategy, to produce DACNO₂-Phase-3-2020.

Table 3 summarizes the 2020 evaluation results, following the format of Table 2. Both DACNO₂-Phase-2 and DACNO₂-Phase-3-2020 reproduced observed surface NO₂ concentrations well (e.g., DACNO₂-Phase-3-2020: RMSE = 5.47 µg/m³, $r = 0.82$, $R^2 = 0.67$, bias = 0.01 µg/m³), with performance comparable to CAMS (RMSE = 4.99 µg/m³, $r = 0.88$, $R^2 = 0.73$, bias = -2.05 µg/m³) but notably lower bias. This demonstrates robust generalization by DACNO₂ under the emission anomalies of the pandemic year. Agreement between DACNO₂ and CAMS remains strong at low altitudes (e.g., surface: RMSE = 2.09 µg/m³, $r = 0.94$, $R^2 = 0.86$, bias = 0.48 µg/m³), but declines rapidly above 1000 m, where R^2 values approach zero or become negative, indicating a failure to reproduce high altitude CAMS NO₂ distributions for 2020.

A comparison of CAMS NO₂ vertical distributions from 2019 to 2023 (Fig. S1) shows generally consistent annual patterns, except for 2020, which is characterized by anomalously high values above 1000 m. This anomaly is also noted in the CAMS 2020 annual evaluation report (Meleux et al., 2023), which attributes it to some sub-models producing unexpectedly high NO₂ at high levels, resulting in inflated tropospheric column estimates. The underlying causes remain unresolved and require further investigation. These findings highlight the importance of data screening, such as checking distributions and identifying outliers, before model training. Including biased or anomalous target data can introduce noise, increase the risk of overfitting, and reduce generalization performance.



5. Conclusions and outlook

This study presents the Deep Atmospheric Chemistry NO₂ model (DACNO₂), a deep learning model for daily, high-resolution (2 km) 3D NO₂ estimation. DACNO₂ integrates multi-source and multi-modal input features, including emissions, geography, meteorology, and temporal indicators. It uses a multi-constraint and phased training approach to learn from both process-based CAMS NO₂ and measured EEA NO₂ data. This approach allows DACNO₂ to reproduce broad-scale, process-based NO₂ patterns and capture local NO₂ gradients. Results show that DACNO₂ significantly improves the ability to resolve fine-scale spatial patterns, near-surface NO₂ variability, and vertical distribution. It also generalizes well across different spatial areas (urban, rural, mountainous, and emission hotspot regions) and periods of anomalous emissions. Furthermore, the framework demonstrates transferability and flexibility, allowing the model to be fine-tuned to adapt to future emission scenarios and to be adjusted to produce outputs for specific satellite overpass times in addition to daily averages.

A systematic evaluation shows that DACNO₂ outperforms the state-of-the-art regional CAMS product in reproducing measured surface NO₂ concentrations. Overall, DACNO₂ achieves a lower RMSE (4.99 vs. 5.32 µg/m³), higher correlation ($r = 0.82$ vs. 0.80, $R^2 = 0.66$ vs. 0.61), and a substantially reduced bias (-0.38 vs. -1.15 µg/m³). The improvement is particularly significant in urban areas with high spatial variability, where DACNO₂ yields a higher R^2 (0.64 vs. 0.56) and halves the negative bias (-1.42 vs. -2.89 µg/m³). In rural areas with very low background levels, DACNO₂ also shows a better R^2 (0.55 vs. 0.51) and maintains a small positive bias, in contrast to the large one from CAMS (1.05 vs. 1.98 µg/m³). Vertical profile analysis indicates that DACNO₂ provides greater spatial detail and variation than CAMS, capturing small-scale emission sources and topographic influences more effectively. Feature importance analysis highlights the contribution of high-resolution emission proxies, land cover, and multi-level meteorological information to resolving spatial and vertical NO₂ patterns. In contrast, single-level meteorological variables contribute minimally, likely because part of their information is already captured by the more comprehensive multi-level data. This redundancy suggests opportunities for future model optimization.

Application to satellite NO₂ retrievals demonstrates that using DACNO₂-generated a-priori profiles enhances the sensitivity of TROPOMI NO₂ products to near-surface concentrations and emission hotspots, particularly in small-scale emission sources and complex geographic regions. These findings underscore the potential of high-resolution ML-based profiles for future high-resolution satellite retrievals. However, DACNO₂ remains a prototype, and further work is needed for operational deployment. First, this would involve extending the model's



output to continuous hourly profiles over a broader geographic domain. Second, the model would need to be operated on a robust GPU computational platform with automated data pipelines. Third, a routine validation framework would need to be established to continuously monitor performance against various data, such as CAMS NO₂, EEA NO₂, and vertical measurements (e.g., MAX-DOAS). Finally, this operational system would require a strategy for periodic model fine-tuning to adapt to evolving emission patterns and maintain long-term accuracy.

Analysis of model performance during COVID-19 indicates that DACNO₂ consistently generalizes well despite emission anomalies. The inconsistencies observed in CAMS reanalysis for 2020 at high levels highlight the need for screening and quality assurance in model training data to avoid learning biased patterns and degrading model reliability.

Future development of DACNO₂ may include integrating high-resolution 3D process-based NO₂ fields from models such as WRF-Chem and column observations from satellites to strengthen constraints at higher altitudes (e.g., above 2000 m), exploring transformer architectures for improved scalability and multi-modal data processing, and embedding additional physical constraints into the loss function. Extension to continental or global applications (including data-poor regions such as the African continent) will further support large-scale air quality management and atmospheric chemistry research.

Data and code availability

The daily number of flights is accessible at <https://www.eurocontrol.int/Economics/DailyTrafficVariation-States.html>. The CAMS global emission inventories are accessible at <https://ads.atmosphere.copernicus.eu/>. The GRIP global roads database can be downloaded from <https://www.globio.info/download-grip-dataset>. The VIIRS nighttime light data can be accessed from <https://eogdata.mines.edu/products/vnl/>. The population dataset is provided by <https://ec.europa.eu/eurostat/web/gisco/geodata/population-distribution/population-grids>. The MERIT DEM data is accessible via https://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT_DEM/. The CORINE land cover dataset can be downloaded from <https://land.copernicus.eu/en/products/corine-land-cover/clc2018>. The single-level and multi-level meteorological data are provided by the fifth-generation ECMWF atmospheric reanalysis of the global climate product (ERA5), which can be accessed via <https://cds.climate.copernicus.eu/>. The CAMS European air quality reanalyses dataset is accessible via <https://ads.atmosphere.copernicus.eu/>. The



EEA AirBase dataset can be downloaded from <https://eeadmz1-downloads-webapp.azurewebsites.net/>. The official TROPOMI NO₂ product is accessible via the Copernicus Data Space Ecosystem (<https://dataspace.copernicus.eu/>). The data generated for this study can be accessed from the Zenodo data archive (Sun et al. (2025), <https://doi.org/10.5281/zenodo.16986854>).

The DACNO₂ model and its framework are built using the Pytorch library (<https://pytorch.org/>) in the Python environment. All code related to model design and data processing is available upon request from the corresponding author.

Author contribution

WS, FT, and MVR conceived the study. WS built the model, performed all analyses, and wrote the initial draft of the manuscript. FT, LC, and MVR reviewed and revised the draft. All authors substantially contributed to the final manuscript.

Competing interests

The co-author, MVR, is a member of the editorial board of Atmospheric Chemistry and Physics. The other authors declare no competing interests.

Acknowledgements

The Belgian Federal Science Policy Office is gratefully appreciated for funding part of this work in the framework of the Terrascope-S5P PRODEX project (PEA 4000136290); L.C. is a research associate supported by the Belgian F.R.S.-FNRS. We used AI-assisted tools to polish the manuscript. The authors are solely responsible for the scientific content and interpretations.



References

- Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K. T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M. J., Dèdelé, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., and de Hoogh, K.: Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project, *Atmospheric Environment*, 72, 10-23, 10.1016/j.atmosenv.2013.02.037, 2013.
- Bernd, S., Valerie, F., Bézy, J. L., Meijer, Y., Durand, Y., Courrèges-Lacoste, G. B., Pachot, C., Löscher, A., Nett, H., Minoglou, K., Boucher, L., Windpassinger, R., Pasquet, A., Serre, D., and Hennepe, F. t.: The Copernicus CO₂M mission for monitoring anthropogenic carbon dioxide emissions from space, *Proc.SPIE*, 118523M, 10.1117/12.2599613,
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q. B., Liu, H. G. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *Journal of Geophysical Research-Atmospheres*, 106, 23073–23095, 10.1029/2001jd000807, 2001.
- Bézy, J.-L., Sierk, B., Caron, J., Veihelmann, B., Martin, D., and Langen, J.: The Copernicus Sentinel-5 mission for operational atmospheric monitoring: status and developments, *SPIE Remote Sensing*, SPIE, 10.1117/12.2068177, 2014.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, *Nature*, 619, 533-538, 10.1038/s41586-023-06185-3, 2023.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A foundation model for the Earth system, *Nature*, 641, 1180-1187, 10.1038/s41586-025-09005-y, 2025.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Vaughan, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A Foundation Model for the Earth System, 10.48550/arXiv.2405.13063, 2024.
- Castellanos, P. and Boersma, K. F.: Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession, *Scientific Reports*, 2, 265, 10.1038/srep00265, 2012.
- Chang, S. Y., Huang, J., Chaveste, M. R., Lurmann, F. W., Eisinger, D. S., Mukherjee, A. D., Erdakos, G. B., Alexander, M., and Knipping, E.: Electric vehicle fleet penetration helps address inequalities in air quality and improves environmental justice, *Communications Earth & Environment*, 4, 135, 10.1038/s43247-023-00799-1, 2023.
- CAMS Regional: European air quality reanalyses data documentation: <https://confluence.ecmwf.int/display/CKB/CAMS+Regional%3A+European+air+quality+reanalyses+data+documentation>, last access: June 2024.



- 710 Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U.,
 Olivier, J. G. J., Pagliari, V., and Janssens-Maenhout, G.: Gridded emissions of air pollutants for the
 period 1970–2012 within EDGAR v4.3.2, *Earth System Science Data*, 10, 1987–2013, 10.5194/essd-10-
 1987-2018, 2018.
- Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.-M., Chen, W., Yi, J., Zhao, W.,
 Wang, X., Liu, Z., Zheng, H.-T., Chen, J., Liu, Y., Tang, J., Li, J., and Sun, M.: Parameter-efficient fine-
 tuning of large-scale pre-trained language models, *Nature Machine Intelligence*, 5, 220–235,
 10.1038/s42256-023-00626-4, 2023.
- 715 Douros, J., Eskes, H., van Geffen, J., Boersma, K. F., Compernelle, S., Pinardi, G., Blechschmidt, A. M., Peuch,
 V. H., Colette, A., and Veeffkind, P.: Comparing Sentinel-5P TROPOMI NO₂ column observations with
 the CAMS regional air quality ensemble, *Geoscientific Model Development*, 16, 509–534, 10.5194/gmd-
 16-509-2023, 2023.
- 720 Elfwing, S., Uchibe, E., and Doya, K.: Sigmoid-Weighted Linear Units for Neural Network Function
 Approximation in Reinforcement Learning, 10.48550/arXiv.1702.03118, 2017.
- Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., and Taneja, J.: Annual Time Series of Global VIIRS Nighttime
 Lights Derived from Monthly Averages: 2012 to 2019, *Remote Sensing*, 13, 10.3390/rs13050922, 2021.
- Eurocontrol: Daily traffic variation for states [dataset], [https://www.eurocontrol.int/Economics/2020-
 DailyTrafficVariation-States.html](https://www.eurocontrol.int/Economics/2020-DailyTrafficVariation-States.html), 2025.
- 725 AirBase - The European air quality database: <https://eeadmz1-downloads-webapp.azurewebsites.net/>, last
 Feranec, J., Soukup, T., Hazeu, G., and Jaffrain, G.: European Landscape Dynamics CORINE Land Cover Data,
 in, CRC Press, Boca Raton, 367, 10.1201/9781315372860, 2016.
- 730 Ghahremanloo, M., Lops, Y., Choi, Y., Mousavinezhad, S., and Jung, J.: A Coupled Deep Learning Model for
 Estimating Surface NO₂ Levels From Remote Sensing Data: 15-Year Study Over the Contiguous United
 States, *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037010, 10.1029/2022JD037010,
 2023.
- Giovannini, L., Ferrero, E., Karl, T., Rotach, M. W., Staquet, C., Trini Castelli, S., and Zardi, D.: Atmospheric
 Pollutant Dispersion over Complex Terrain: Challenges and Needs for Improving Air Quality
 Measurements and Modeling, *Atmosphere*, 11, 646, 10.3390/atmos11060646, 2020.
- 735 Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled
 “online” chemistry within the WRF model, *Atmospheric Environment*, 39, 6957–6975,
 10.1016/j.atmosenv.2005.04.027, 2005.
- 740 Gulde, S., Kolm, M., Smith, D., Maurer, R., Courrèges-Lacoste, G. B., Sallusti, M., and Bagnasco, G.: Sentinel
 4: a geostationary imaging UVN spectrometer for air quality monitoring: status of design, performance
 and development, *International Conference on Space Optics—ICSO 2014, Tenerife, Canary Islands,
 Spain*, 1158–1166, 10.1117/12.2304099,
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, 2016 IEEE Conference
 on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016, 1,
 10.1109/cvpr.2016.90,



- 745 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999-2049, 10.1002/qj.3803, 2020.
- 750 Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H. W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, *Geosci. Model Dev.*, 3, 445-473, 10.5194/gmd-3-445-2010, 2010.
- 755 Ialongo, I., Virta, H., Eskes, H., Hovila, J., and Douros, J.: Comparison of TROPOMI/Sentinel-5 Precursor NO₂ observations with ground-based measurements in Helsinki, *Atmos. Meas. Tech.*, 13, 205-218, 10.5194/amt-13-205-2020, 2020.
- 760 Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A. M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V. H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19, 3515-3556, 10.5194/acp-19-3515-2019, 2019.
- 765 Jiang, Z., Wang, S., Yan, Y., Zhang, S., Xue, R., Gu, C., Zhu, J., Liu, J., and Zhou, B.: Constructing the 3D Spatial Distribution of the HCHO/NO₂ Ratio via Satellite Observation and Machine Learning Model, *Environmental Science & Technology*, 59, 4047-4058, 10.1021/acs.est.4c12362, 2025.
- 770 Kim, J., Jeong, U., Ahn, M.-H., Kim, J. H., Park, R. J., Lee, H., Song, C. H., Choi, Y.-S., Lee, K.-H., Yoo, J.-M., Jeong, M.-J., Park, S. K., Lee, K.-M., Song, C.-K., Kim, S.-W., Kim, Y. J., Kim, S.-W., Kim, M., Go, S., Liu, X., Chance, K., Chan Miller, C., Al-Saadi, J., Veihelmann, B., Bhartia, P. K., Torres, O., Abad, G. G., Haffner, D. P., Ko, D. H., Lee, S. H., Woo, J.-H., Chong, H., Park, S. S., Nicks, D., Choi, W. J., Moon, K.-J., Cho, A., Yoon, J., Kim, S.-k., Hong, H., Lee, K., Lee, H., Lee, S., Choi, M., Veefkind, P., Levelt, P. F., Edwards, D. P., Kang, M., Eo, M., Bak, J., Baek, K., Kwon, H.-A., Yang, J., Park, J., Han, K. M., Kim, B.-R., Shin, H.-W., Choi, H., Lee, E., Chong, J., Cha, Y., Koo, J.-H., Irie, H., Hayashida, S., Kasai, Y., Kanaya, Y., Liu, C., Lin, J., Crawford, J. H., Carmichael, G. R., Newchurch, M. J., Lefer, B. L., Herman, J. R., Swap, R. J., Lau, A. K. H., Kurosu, T. P., Jaross, G., Ahlers, B., Dobber, M., McElroy, C. T., and Choi, Y.: New era of air quality monitoring from space: Geostationary Environment Monitoring Spectrometer (GEMS), *Bulletin of the American Meteorological Society*, 101, E1-E22, 10.1175/bams-d-18-0013.1, 2020.
- 775 Kim, M., Brunner, D., and Kuhlmann, G.: Importance of satellite observations for high-resolution mapping of near-surface NO₂ by machine learning, *Remote Sensing of Environment*, 264, 112573, 10.1016/j.rse.2021.112573, 2021.
- 780 Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., and Reblitz-Richardson, O.: Captum: A unified and generic model interpretability library for PyTorch, 10.48550/arXiv.2009.07896, 2020.



- 785 Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F.,
 and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and
 applications, *Atmos. Chem. Phys.*, 5, 417-432, 10.5194/acp-5-417-2005, 2005.
- Kuhn, L., Beirle, S., Osipov, S., Pozzer, A., and Wagner, T.: NitroNet – a machine learning model for the prediction
 of tropospheric NO₂ profiles from TROPOMI observations, *Atmos. Meas. Tech.*, 17, 6485-6516,
 790 10.5194/amt-17-6485-2024, 2024a.
- Kuhn, L., Beirle, S., Kumar, V., Osipov, S., Pozzer, A., Bösch, T., Kumar, R., and Wagner, T.: On the influence of
 vertical mixing, boundary layer schemes, and temporal emission profiles on tropospheric NO₂ in WRF-
 Chem – comparisons to in situ, satellite, and MAX-DOAS observations, *Atmos. Chem. Phys.*, 24, 185-
 217, 10.5194/acp-24-185-2024, 2024b.
- 795 Kuik, F., Kerschbaumer, A., Lauer, A., Lupascu, A., von Schneidmesser, E., and Butler, T. M.: Top-down
 quantification of NO_x emissions from traffic in an urban area using a high-resolution regional
 atmospheric chemistry model, *Atmos. Chem. Phys.*, 18, 8203-8225, 10.5194/acp-18-8203-2018, 2018.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T.,
 Eaton-Rosen, Z., Hu, W., Meroze, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed,
 800 S., and Battaglia, P.: Learning skillful medium-range global weather forecasting, *Science*, 382, 1416-
 1421, doi:10.1126/science.adi2336, 2023.
- Landgraf, J., Rusli, S., Cooney, R., Veeffkind, P., Vemmix, T., de Groot, Z., Bell, A., Day, J., Leemhuis, A., and
 Sierk, B.: The TANGO mission: A satellite tandem to measure major sources of anthropogenic
 greenhouse gas emissions, *EGU General Assembly Conference Abstracts*, 19643,
- 805 Levelt, P. F., Stein Zweers, D. C., Aben, I., Bauwens, M., Borsdorff, T., De Smedt, I., Eskes, H. J., Lerot, C.,
 Loyola, D. G., Romahn, F., Stavrou, T., Theys, N., Van Roozendaal, M., Veeffkind, J. P., and Verhoelst,
 T.: Air quality impacts of COVID-19 lockdown measures detected from space using high spatial
 resolution observations of multiple trace gases from Sentinel-5P/TROPOMI, *Atmospheric Chemistry
 and Physics*, 22, 10319-10351, 10.5194/acp-22-10319-2022, 2022.
- 810 Li, S. and Xing, J.: DeepSAT4D: Deep learning empowers four-dimensional atmospheric chemical concentration
 and emission retrieval from satellite, *The Innovation Geoscience*, 2, 100061, 10.59717/j.xinn-
 geo.2024.100061, 2024.
- Lu, L., Cohen, J. B., Qin, K., Li, X., and He, Q.: Identifying missing sources and reducing NO_x emissions
 uncertainty over China using daily satellite data and a mass-conserving method, *Atmos. Chem. Phys.*,
 815 25, 2291-2309, 10.5194/acp-25-2291-2025, 2025.
- Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., and Schipper, A. M.: Global patterns of current and future
 road infrastructure, *Environmental Research Letters*, 13, 064006, 10.1088/1748-9326/aabd42, 2018.
- Meleux, F., Raux, B., Ung, A., Colette, A., Gauss, M., Douros, J., Tsikerdekis, T., and Bennouna, Y.: Annual report
 on the evaluation of validated reanalyses VRA2020,
 820 [https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-
 regional/VRA/CAMS283_2021SC1_D83.2.2.1-2020_202303_VRA2020_evaluation_v2.pdf](https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-regional/VRA/CAMS283_2021SC1_D83.2.2.1-2020_202303_VRA2020_evaluation_v2.pdf), 2023.
- Meleux, F., Raux, B., Ung, A., Colette, A., Gauss, M., Douros, J., Tsikerdekis, T., Eskes, H., Bennouna, Y., and
 Thouret, V.: Annual EQC report for the interim reanalysis for 2023, for each of the operational systems



- 825 and the ENSEMBLE, https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-regional/IRA/CAMS283_2021SC2_D83.2.1.1-2023_202406_EQC_IRA2023_v1.pdf, 2024.
- Oppenheim, A. V.: Discrete-time signal processing, Pearson Education India 1999.
- Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., and Li, Q.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, *Journal of Geophysical Research: Atmospheres*, 106, 14539-14550, 10.1029/2000JD900772, 2001.
- 830 Peuch, V.-H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., Ades, M., Agustí-Panareda, A., Ananasso, C., Andersson, E., Armstrong, D., Barré, J., Bousserez, N., Dominguez, J. J., Garrigues, S., Inness, A., Jones, L., Kipling, Z., Letertre-Danczak, J., Parrington, M., Razinger, M., Ribas, R., Vermoote, S., Yang, X., Simmons, A., Garcés de Marcilla, J., and Thépaut, J.-N.: The Copernicus Atmosphere Monitoring Service: From Research to Operations, *Bulletin of the American Meteorological Society*, 103, E2650-E2668, 10.1175/BAMS-D-21-0314.1, 2022.
- 835 Shahrokhishahraki, N., Rayner, P. J., Silver, J. D., Thomas, S., and Schofield, R.: High-resolution modeling of gaseous air pollutants over Tehran and validation with surface and satellite data, *Atmospheric Environment*, 270, 118881, 10.1016/j.atmosenv.2021.118881, 2022.
- 840 Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, 10.48550/arXiv.1506.04214, 2015.
- Silva, F. B. e., Poelman, H., and Dijkstra, L.: JRC-GEOSTAT 2018 [dataset], 2021.
- Soulie, A., Granier, C., Darras, S., Zilbermann, N., Doumbia, T., Guevara, M., Jalkanen, J. P., Keita, S., Liousse, C., Crippa, M., Guizzardi, D., Hoesly, R., and Smith, S. J.: Global anthropogenic emissions (CAMSGLOB-ANT) for the Copernicus Atmosphere Monitoring Service simulations of air quality forecasts and reanalyses, *Earth Syst. Sci. Data*, 16, 2261-2279, 10.5194/essd-16-2261-2024, 2024.
- 845 Sun, W., Tack, F., Clarisse, L., and Roozendael, M. V.: Technical Note: DACNO2 – A Multi-Constraint Deep Learning Framework for High-Resolution 3D NO2 Field Estimation (Version 1) [dataset], 10.5281/zenodo.16986854, 2025.
- 850 Sun, W., Tack, F., Clarisse, L., Schneider, R., Stavrakou, T., and Van Roozendael, M.: Inferring Surface NO2 Over Western Europe: A Machine Learning Approach With Uncertainty Quantification, *Journal of Geophysical Research: Atmospheres*, 129, e2023JD040676, 10.1029/2023JD040676, 2024.
- Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, 10.48550/arXiv.1703.01365, 2017.
- 855 Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *arXiv e-prints*, arXiv:1512.00567, 10.48550/arXiv.1512.00567, 2015.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going Deeper with Convolutions, 10.48550/arXiv.1409.4842, 2014.
- 860 Tack, F., Merlaud, A., Iordache, M. D., Pinardi, G., Dimitropoulou, E., Eskes, H., Bomans, B., Veeffkind, P., and Van Roozendael, M.: Assessment of the TROPOMI tropospheric NO2 product based on airborne APEX observations, *Atmospheric Measurement Techniques*, 14, 615-646, 10.5194/amt-14-615-2021, 2021.



- 865 Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sensing of Environment*, 120, 70-83, 10.1016/j.rse.2011.09.027, 2012.
- 870 Wang, L., Chen, X., Zhang, Y., Li, M., Li, P., Jiang, L., Xia, Y., Li, Z., Li, J., Wang, L., Hou, T., Liu, W., Rosenfeld, D., Zhu, T., Zhang, Y., Chen, J., Wang, S., Huang, Y., Seinfeld, J. H., and Yu, S.: Switching to electric vehicles can lead to significant reductions of PM_{2.5} and NO₂ across China, *One Earth*, 4, 1037-1048, 10.1016/j.oneear.2021.06.008, 2021.
- 875 Wei, J., Liu, S., Li, Z., Liu, C., Qin, K., Liu, X., Pinker, R. T., Dickerson, R. R., Lin, J., Boersma, K. F., Sun, L., Li, R., Xue, W., Cui, Y., Zhang, C., and Wang, J.: Ground-Level NO₂ Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence, *Environmental Science & Technology*, 56, 9988-9998, 10.1021/acs.est.2c03834, 2022.
- Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, *Geosci. Model Dev.*, 10, 721-750, 10.5194/gmd-10-721-2017, 2017.
- 880 Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, *Geophysical Research Letters*, 44, 5844-5853, 10.1002/2017GL072874, 2017.
- 885 Yang, L. H., Jacob, D. J., Colombi, N. K., Zhai, S., Bates, K. H., Shah, V., Beaudry, E., Yantosca, R. M., Lin, H., Brewer, J. F., Chong, H., Travis, K. R., Crawford, J. H., Lamsal, L. N., Koo, J. H., and Kim, J.: Tropospheric NO₂ vertical profiles over South Korea and their relation to oxidant chemistry: implications for geostationary satellite retrievals and the observation of NO₂ diurnal variation from space, *Atmos. Chem. Phys.*, 23, 2465-2481, 10.5194/acp-23-2465-2023, 2023.
- 890 Yang, Y., Zhao, Y., Zhang, L., Zhang, J., Huang, X., Zhao, X., Zhang, Y., Xi, M., and Lu, Y.: Improvement of the satellite-derived NO_x emissions on air quality modeling and its effect on ozone and secondary inorganic aerosol formation in the Yangtze River Delta, China, *Atmos. Chem. Phys.*, 21, 1191-1209, 10.5194/acp-21-1191-2021, 2021.
- Zhang, C., Liu, C., Li, B., Zhao, F., and Zhao, C.: Spatiotemporal neural network for estimating surface NO₂ concentrations over north China and their human health impact, *Environmental Pollution*, 307, 119510, 10.1016/j.envpol.2022.119510, 2022a.
- 895 Zhang, S., Li, B., Liu, L., Hu, Q., Liu, H., Zheng, R., Zhu, Y., Liu, T., Sun, M., and Liu, C.: Prediction of Vertical Profile of NO₂ Using Deep Multimodal Fusion Network Based on the Ground-Based 3-D Remote Sensing, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-13, 10.1109/TGRS.2021.3061476, 2022b.
- 900 Zhang, Z., Li, Q., Hu, Q., Xue, J., Liu, T., Tang, Z., Wang, F., Zhang, C., Lu, C., Wang, Z., Gao, M., and Liu, C.: Deep learning approach for reconstructing three-dimensional distribution of NO₂ on an urban scale, *Remote Sensing of Environment*, 321, 114678, 10.1016/j.rse.2025.114678, 2025.
- Zhao, H., Gallo, O., Frosio, I., and Kautz, J.: Loss Functions for Image Restoration With Neural Networks, *IEEE Transactions on Computational Imaging*, 3, 47-57, 10.1109/TCI.2016.2644865, 2017.



- 905 Zhou, W., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, 13, 600-612, 10.1109/TIP.2003.819861, 2004.
- Zhou, Z., Shin, J., Zhang, L., Gurudu, S., Gotway, M., and Liang, J.: Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, 4761-4772, 10.1109/CVPR.2017.506,
- 910 Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q.: A Comprehensive Survey on Transfer Learning, 10.48550/arXiv.1911.02685, 2019.
- 915 Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., Hilton, B. B., Nicks, D. K., Newchurch, M. J., Carr, J. L., Janz, S. J., Andraschko, M. R., Arola, A., Baker, B. D., Canova, B. P., Chan Miller, C., Cohen, R. C., Davis, J. E., Dussault, M. E., Edwards, D. P., Fishman, J., Ghulam, A., González Abad, G., Grutter, M., Herman, J. R., Houck, J., Jacob, D. J., Joiner, J., Kerridge, B. J., Kim, J., Krotkov, N. A., Lamsal, L., Li, C., Lindfors, A., Martin, R. V., McElroy, C. T., McLinden, C., Natraj, V., Neil, D. O., Nowlan, C. R., O'Sullivan, E. J., Palmer, P. I., Pierce, R. B., Pippin, M. R., Saiz-Lopez, A., Spurr, R. J. D., Szykman, J. J., Torres, O., Veefkind, J. P., Veihelmann, B., Wang, H., Wang, J., and Chance, K.: Tropospheric emissions: Monitoring of pollution (TEMPO), *Journal of Quantitative Spectroscopy and Radiative Transfer*, 186, 17-39, 10.1016/j.jqsrt.2016.05.008, 2017.
- 920