A Two-Stream CNN Framework for Spatiotemporally Resolved NO_2 Estimates Using TEMPO and Sentinel-2 Satellite Data

Adam Bloom[1][0009−0006−0847−4720], Zhen Qu[2][0000−0002−3766−9838], and Ranga Raju Vatsavai[1,3][0000−0002−7083−0267]

[1] Center for Geospatial Analytics
[2] Department of Marine, Earth and Atmospheric Sciences
[3] Department of Computer Science
North Carolina State University, Raleigh, NC, USA
{abloom,zqu5,rrvatsav}@ncsu.edu

Abstract. A comprehensive understanding of ground-level air pollutants is essential for developing policies and strategies that effectively reduce associated human health, environmental, and economic risks. Nitrogen dioxide is a key air pollutant that poses its own risks to human health and is also essential to the formation of secondary air pollutants, including ozone and particulate matter. Air quality monitoring technologies enable concentration measurements at high temporal frequencies but are spatially constrained by site locations. In contrast, remote sensing satellites expand spatial coverage, yet accurately capturing short-term exposure variability remains challenging. In 2023, NASA successfully launched the TEMPO (Tropospheric Emissions: Monitoring of Pollution) geostationary satellite, which enables high spatial resolution retrievals at hourly intervals. Advancements in deep learning and computer vision effectively leverage spatial context to more accurately estimate ground-level pollution. Our research integrates these technological advancements within a deep learning framework capable of estimating hourly ground-level NO_2 concentrations at spatial scales as fine as 10 meters across the continental United States. This methodology can expand knowledge of sources driving diurnal variation of ground-level NO_2, inform emissions control policy and technologies, and deepen understanding of drivers of atmospheric chemical processes. Our thorough experimental evaluation offers several useful insights. Code[4] and data[5] are publicly available.

Keywords: Air pollution · deep learning · diurnal NO_2 · remote sensing.

1 Introduction

Air pollution is a major challenge to global public health and the environment. Nitrogen dioxide (NO_2) is designated by the EPA as a criteria air pollutant

[4] https://zenodo.org/records/19474505
[5] Scripts used to obtain the data are included in the code, and a tarball of the data can be made available upon request.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

with distinct environmental and public health impacts [1, 7]. Exposure to NO_2 increases the risk of respiratory infections, asthma, and lung cancer, and it plays a key role in acid rain formation [1, 7]. Moreover, NO_2 is an important precursor to other criteria pollutants, including ozone (O_3) and atmospheric aerosols that form secondary particulate matter [7]. Monitoring sites effectively measure air pollutants but are spatially limited [4]. Remote sensing data offer a means to expand spatial coverage by combining vertical column density retrievals with land use analysis. Existing statistical, machine learning, and physics-based methods can estimate ground-level NO_2 effectively, but they are largely constrained to coarser temporal resolutions [4, 5]. Understanding short-term (hourly) ground-level NO_2 concentrations is crucial for assessing its role as a precursor to secondary pollutants, identifying pollution sources, and characterizing cycles that influence community exposures [5]. A new generation of geostationary satellites, such as NASA's Tropospheric Emissions: Monitoring of Pollution (TEMPO), provides hourly vertical column densities (VCDs) for NO_2, ozone, and formaldehyde at a $0.02^\circ \times 0.02^\circ$ spatial resolution [5]. These retrievals present a novel opportunity to model hourly NO_2 concentrations across the United States using remote sensing.

Computer vision and convolutional neural networks (CNNs) have proven effective for quantifying ground-level air pollution using Sentinel-2 imagery and pollutant-specific retrieval products [4, 10]. These networks require minimal effort to set up, yield highly accurate predictions, and generalize well to new areas. Employing multiple convolutional backbones to extract features from different remote sensing inputs enhances accuracy and generalizability for ground-level NO_2 predictions [10]. Although CNNs provide flexible and efficient pathways to accurate spatial estimations, they have historically been constrained by the daily revisit rates of sun-synchronous satellites, limiting their ability to capture short-term variability [4, 10]. New geostationary satellites address this gap by providing estimates at higher temporal resolution [6].

In this work, we propose a two-stream network with a late fusion module to effectively exploit multisource data. Our specific contributions are:

1. A flexible CNN framework that can effectively exploit multisource data.
2. More accurate monthly NO_2 concentration estimates compared to prior work, achieved using higher-temporal-resolution retrievals.
3. High-resolution ground-level NO_2 estimates at both high temporal (hourly) and spatial resolution.

2 Related Work

Statistical Methods and Traditional Machine Learning: Land-use regression (LUR) and kriging are commonly used to model ground-level NO_2 concentrations at monitoring sites [7, 10]. LUR estimates long-term exposures using geospatial variables but cannot capture short-term variability and requires extensive GIS data [7, 10]. Kriging interpolates spatially, treating NO_2 as a random variable with the best linear unbiased estimator, and provides variance

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

estimates and confidence intervals [10]. While remote sensing mitigates traditional data limitations, high-dimensional satellite data has shifted the field toward ML algorithms. Tree-based and ensemble models successfully leverage remote sensing products [7], but they typically require manual feature engineering for spatiotemporal context. Consequently, Deep Learning has become the state of the art for generalizable NO_2 estimation.

Deep Learning: Convolutional neural networks (CNNs) have been shown to extract spatial features from remote sensing data for ground-level pollution modeling [10, 4], while LSTMs can further improve forecasting by encoding temporal dependencies [8]. Our work is motivated by two-stream architectures [10, 3], where Sentinel-2 imagery and satellite NO_2 retrievals modeled ground-level NO_2 at high spatial but coarse temporal scales (monthly to multi-year) with robust out-of-domain generalizability [10]. However, Sentinel-5P's near-daily revisit is limited to a short-term context. We retain the two-stream structure, replace the Sentinel-5P backbone with hourly TEMPO Level-3 NO_2 retrievals, and incorporate temporal, demographic, and land-use predictors to improve spatiotemporal estimation.

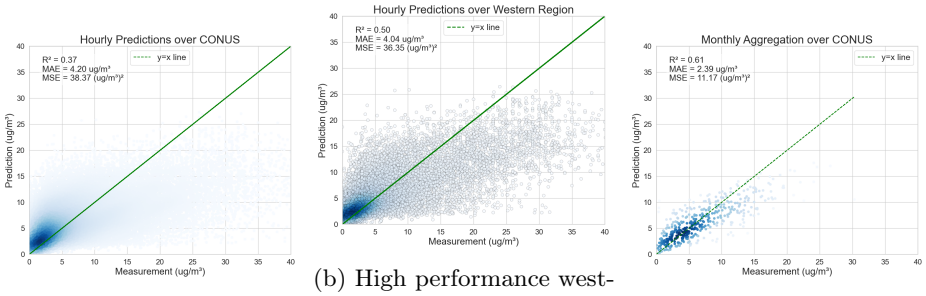
Regionally Limited Hourly Estimation Studies: High-temporal-resolution NO_2 estimation using remote sensing has historically been limited by the low revisit rates of available satellite products. Several hourly ground-level NO_2 approaches have been developed using the geostationary GEMS satellite, which covers East Asia and the Asian Pacific [6]. Launched in 2020, three years before NASA's TEMPO satellite, GEMS has enabled multiple hourly estimation efforts. The most successful models ($R^2 = 0.72$) incorporate region-specific features and are not generalizable to other locations [6]. A similar local approach for the Netherlands modeled hourly PM2.5 and NO_2 with strong performance ($R^2 = 0.35 - 0.78$), but relied on physical chemical transport models [9]. These approaches demonstrate the promise of machine learning for high-resolution estimation but remain largely limited to small geographic regions.

Chemical Transport Modeling: Chemical transport models simulate four-dimensional atmospheric chemistry processes and can model the distribution of chemicals over time and space. These models achieve high performance in ground-level pollution monitoring by leveraging emissions inventories to estimate exposures [9]. Remote sensing observations can constrain these physical models to reduce simulation errors [9]. However, physical models remain computationally intensive and may introduce simulation errors into machine learning estimates. For these reasons, this study focuses on a purely data-driven approach to ground-level NO_2 estimation.

3 Methods

We now present technical details of our two-stream CNN architecture.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34



(a) Hourly predictions (b) High performance western region (c) Monthly aggregation

Fig. 1: Model performance for hourly predictions across the continental United States and a high-performing western region, along with monthly aggregated results over the continental United States

3.1 Two-stream CNN Architecture

Two-stream (or multi-stream) architectures [10, 3] are a popular approach for modeling multi-source (multi-resolution) data. Individual networks are designed specifically for each input type to extract discriminative features, which are then fused with any additional sources before the prediction head. Our base model extends the two-stream neural network presented in [10] by introducing an hourly TEMPO retrieval backbone. This backbone comprises two convolutional layers with 10 and 15 channels, respectively, each followed by a rectified linear unit (ReLU) activation layer, along with two max pooling layers and a final linear layer. The output is a 128-dimensional feature vector. This shallow architecture is sufficient to model the low spatial resolution (2 km (north-south) by 4.75 km (east-west)) of the TEMPO satellite image.

To extract land use features from multi-spectral Sentinel-2 images (10 meter resolution), we employ a ResNet50 model fine-tuned on land use classification with the BigEarthNet dataset [11]. This feature extraction knowledge is directly transferable to identifying features that drive short- and long-term variations in NO_2. Given the number of input bands, the resulting vector (2048 dimensions) is substantially larger than the TEMPO backbone output.

The feature vectors from both backbones are fused and concatenated with encoded temporal variables, population density, and degree of urbanization, before the regression head. Although various fusion strategies (e.g., addition or multiplication) could be used, concatenation was chosen due to differing feature vector lengths. The regression head comprises two fully connected layers separated by a ReLU activation, and the full network is trained with the Adam optimizer to minimize MSE loss.

3.2 Evaluation

We evaluate our model using R^2, MSE, MAE, and mean bias (MB) to identify regional trends. Data are split by station (not by individual sample) to pre-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

vent memorization. Metrics are computed for raw hourly predictions and for daily/monthly aggregates, enabling direct benchmarking against prior studies [10]. Since high-frequency estimates at this large scale are largely unexplored, we compare against the deep learning satellite models (Sentinel-2/Sentinel-5P at monthly, quarterly, and two-year spans) and the land use regression OSM-only model from [10] to assess whether our algorithm achieves comparable accuracy at higher temporal frequencies. We also highlight strong performance across six western U.S. states (Colorado, California, Utah, Arizona, Washington, and Nevada), defining this group as the "Western Region" and including its statistics in our analysis.

4 Experiments

4.1 Data and Preprocessing

Ground Truth: We use hourly NO_2 measurements from the EPA Air Quality System (AQS) as training targets. Samples were collected over five months in 2024 at 432 stations, totaling 185,323 hourly measurements. Measurements flagged for data quality issues are excluded.

Sentinel-2: We retrieve Sentinel-2 [2] images centered on each AQS site, with dimensions of $12 \times 120 \times 120$. All bands are resampled to a consistent 10 m spatial resolution. A single image is used per site, as significant land use changes are not anticipated within a single year, and this backbone is not expected to capture short-term NO_2 variability. The image with the lowest cloud cover in 2024 is selected to ensure optimal quality.

TEMPO: We use Level 3 TEMPO [5] NO_2 vertical column densities, as their consistent gridding reduces preprocessing. These images have a pixel resolution of approximately $0.02° \times 0.02°$. For each AQS station and hour with a recorded measurement, we retrieve a 10×10 grid of NO_2 retrievals centered on the station and resample it to 120×120 to match Sentinel-2 dimensions. Observations with data-quality flags are excluded, as are retrievals with large gaps caused by cloud cover or reflective surfaces.

Temporal Variables: Diurnal and weekly cycles in NO_2 concentrations arise from anthropogenic sources. To help the model learn these patterns, we encode time of day, day of week, and month of year using cyclical encoding:

$$\begin{aligned} \text{hour}_{\sin} &= \sin\left(\tfrac{2\pi\,\text{hour}}{24}\right) & \text{hour}_{\cos} &= \cos\left(\tfrac{2\pi\,\text{hour}}{24}\right) \\ \text{month}_{\sin} &= \sin\left(\tfrac{2\pi\,\text{month}}{12}\right) & \text{month}_{\cos} &= \cos\left(\tfrac{2\pi\,\text{month}}{12}\right) \\ \text{day}_{\sin} &= \sin\left(\tfrac{2\pi\,\text{day\_of\_week}}{7}\right) & \text{day}_{\cos} &= \cos\left(\tfrac{2\pi\,\text{day\_of\_week}}{7}\right) \end{aligned}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

Table 1: Comparison of Prior Studies and Our Model

Input	Span	R^2	MAE	MSE
Prior Studies				
Sen.-2/5P (Monthly)	18–20	0.51	6.54	78.96
Sen.-2/5P (Quarterly)	18–20	0.53	6.05	72.53
Sen.-2/5P	18–20	0.57	5.50	58.47
OSM Only	18–20	0.34	7.22	88.29
Our Model				
Hourly	2024	0.37	4.20	38.37
Western Region	2024	0.50	4.04	36.35
Daily Aggregation	2024	0.40	3.84	29.27
Monthly Aggregation	2024	0.61	2.39	11.17

Note: Prior studies did not include mean bias, so we only compare R^2, MAE, and MSE in this table.

Population Density and Degree of Urbanization: From the 2022 American Community Survey, we extracted population density within a three-mile radius of each AQS site. Additional contextual information came from AQS data in the form of a categorical degree of urbanization, specifically rural, suburban, or urban, which is based on U.S. Census block groups and therefore supports inference at unmonitored locations. We applied one-hot encoding to this categorical variable.

4.2 Experimental Set Up

We trained the model on 185,323 hourly observations from 432 monitoring stations, using a station-level split of 60% for training, 20% for validation, and 20% for testing to prevent spatial leakage. Training was conducted on a single NVIDIA A10 GPU (24 GB memory) with two CPU cores and 40 GB of system memory. The model was implemented in PyTorch and trained with a batch size of 50 using the Adam optimizer, a learning rate of 5×10^{-5}, and a mean squared error (MSE) loss function. Dropout with a probability of 0.05 was applied to the final two fully connected layers. Each model was trained for up to 30 epochs with early stopping based on validation loss.

5 Results and Discussion

Monthly, Daily, and Hourly Performance: We iteratively optimized our architecture by adding features to the two-backbone baseline from [10]. Initial results showed over-prediction of low NO_2 and under-prediction of high NO_2; adding cyclical temporal variables, population density, and urbanization progressively improved performance. The final model achieves hourly $R^2 = 0.37$, MAE=4.20 $\mu g/m^3$, and MSE=38.37 $(\mu g/m^3)^2$ across the continental US (Figure 1a). While R^2 is lower than Sentinel-2/Sentinel-5P models [10] (expected due to higher temporal resolution and variance), it exceeds the OSM-only baseline

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

for long-term exposures, and MSE/MAE are consistently lower than all Sentinel-2/Sentinel-5P models (Table 1). Aggregating hourly estimates to daily and monthly averages yields further gains: monthly aggregation achieves $R^2 = 0.61$, MAE=2.39 $\mu g/m^3$, MSE=11.17 $(\mu g/m^3)^2$ (Figure 1c), significantly outperforming the monthly results in [10]. This demonstrates that higher-resolution estimates better represent long-term exposures than directly predicting averages from lower-frequency products.

Regional Trends: The spatial analysis of station-level mean prediction bias indicates systematic under-prediction in the Mid-Atlantic and slight over-prediction in the Northeast. The largest over-predictions occur in Oklahoma, Wyoming, and western Kentucky—stations with below-average NO_2—reflecting "regression to the mean," common in large spatial models. At the Kentucky site, bias may arise from land features outside the 1.2×1.2 km Sentinel-2 patch: although the patch appears urban, the broader region is rural with extensive forest cover, highlighting spatial context limitations. Stations in Michigan, Illinois, and Ohio show significant underestimation, likely due to few training examples near major highways and pretraining on European imagery (BigEarthNet) [11], which may not encode North American highway interchanges and low-density sprawl as distinctly. Conversely, high performance is observed across six western states: Colorado, California, Utah, Arizona, Washington, and Nevada (Figure 1b).

6 Conclusion

In this study, we develop a deep learning architecture that estimates hourly NO_2 concentrations across the continental United States. Expanding spatial coverage of hourly air pollution is critical for evaluating public exposures and advancing understanding of atmospheric chemistry. We leverage NASA's geostationary TEMPO satellite, which provides hourly NO_2 retrievals covering the entire continental U.S. By incorporating these retrievals into a two-backbone CNN architecture, we generate efficient and accurate NO_2 estimates nationwide. Compared to prior approaches using similar architectures at coarser temporal resolutions, we achieve lower MSE and MAE, despite a slightly lower R^2. This lower R^2 reflects regional variability in hourly NO_2 and the diversity of environments across the U.S. We identify regions where the model performs particularly well, including six western states where hourly estimates achieve an R^2 of 0.50—nearly identical to monthly aggregated R^2 in prior studies [10]. These findings highlight strong regional performance and opportunities for future enhancement.

While our model demonstrates significant promise, particularly in specific regions, several limitations and future directions remain. Progressive performance gains from added features suggest that incorporating 3D spatiotemporal retrievals, meteorological variables, or traffic data could further enhance accuracy. Regional fine-tuning may also better capture localized NO_2 dynamics.

Acknowledgments. Vatsavai is supported in part by the "AI Institute for Land, Economy, Agriculture and Forestry (AI-LEAF)," which is supported by USDA National

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34

Institute of Food and Agriculture (NIFA) and the National Science Foundation (NSF) National AI Research Institutes Competitive Award no. 2023-67021-39829. Project website: https://cse.umn.edu/aileaf.
Generative AI Usage Disclosure: The authors used Generative AI tools exclusively for language editing to improve readability. No AI tools were used to generate any new content in this paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, T.M., Kuschner, W.G., Gokhale, J., Shofer, S.: Outdoor air pollution: nitrogen dioxide, sulfur dioxide, and carbon monoxide health effects. The American journal of the medical sciences **333**(4), 249–256 (2007)
2. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P.: Sentinel-2: Esa's optical high-resolution mission for gmes operational services. Remote Sensing of Environment **120**, 25–36 (2012). https://doi.org/https://doi.org/10.1016/j.rse.2011.11.026
3. Gadiraju, K.K., Ramachandra, B., Chen, Z., Vatsavai, R.R.: Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery. In: Rajesh Gupta, e.a. (ed.) KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020. pp. 3234–3242. ACM (2020). https://doi.org/10.1145/3394486.3403375
4. Hong, K.Y., Pinheiro, P.O., Weichenthal, S.: Predicting global variations in outdoor pm2.5 concentrations using satellite images and deep convolutional neural networks (2019), https://arxiv.org/abs/1906.03975
5. Jin, X., Yang, Y., Gonzalez Abad, G., Nowlan, C., Liu, X.: Observing the diurnal variations of ozone-nox-voc chemistry over the u.s. from the geostationary tempo instrument. Geophysical Research Letters **52**(14) (2025)
6. Lee, H.J., Kim, N.R., Shin, M.Y.: Capabilities of satellite geostationary environment monitoring spectrometer (gems) no2 data for hourly ambient no2 exposure modeling. Environmental Research **261**, 119633 (2024)
7. Liu, J., Chen, W.: First satellite-based regional hourly no2 estimations using a space-time ensemble learning model: A case study for beijing-tianjin-hebei region, china. Science of The Total Environment **820**, 153289 (2022). https://doi.org/10.1016/j.scitotenv.2022.153289
8. Mahmudimanesh, M., Mirzaee, M., Dehghan, A., Bahrampour, A.: Forecasts of cardiac and respiratory mortality in tehran, iran, using arimax and cnn-lstm models. Environmental Science and Pollution Research **29**(19), 28469–28479 (2022)
9. Ndiaye, A., Shen, Y., Kyriakou, K., Karssenberg, D., Schmitz, O., Flückiger, B., de Hoogh, K., Hoek, G.: Hourly land-use regression modeling for no2 and pm2. 5 in the netherlands. Environmental Research **256**, 119233 (2024)
10. Scheibenreif, L., Mommert, M., Borth, D.: Toward global estimation of ground-level no2 pollution with deep learning and remote sensing. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–14 (2022)
11. Sumbul, G., Charfuelan, M., Demir, B., Markl, V.: Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In: IEEE international geoscience and remote sensing symposium. pp. 5901–5904. IEEE (2019)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_34