Assessing the Feasibility of Landsat-Driven NO_2 Prediction: A Spatial Cross-Validation Framework

Amir Tahooni[1], Ata Kakroodi[1], Majid Kiavarz[1], Hossein Mansourian[2]

[1] Department of Remote Sensing and GIS, Faculty of Geography, University of Tehran, Tehran, Iran
[2] Department of Human Geography and Planning, Faculty of Geography, University of Tehran, Tehran, Iran

Keywords: Nitrogen Dioxide (NO_2), Landsat, Machine Learning, Spatial Cross-Validation, Air Pollution Mapping, Tehran.

Abstract: Accurate high-resolution mapping of nitrogen dioxide (NO_2) is critical for environmental health studies. While many models rely on direct satellite NO_2 column data, this study investigates an alternative approach: predicting ground-level NO_2 using Landsat imagery combined with topographic and proximity variables. We developed and evaluated Random Forest and XGBoost models on a dataset from Tehran, Iran, using 21 predictors derived from Landsat 8/9, ASTER DEM, and OpenStreetMap. To rigorously assess spatial generalization, we employed three cross-validation strategies. The results highlight a critical dependence of performance metrics on the validation method. Traditional 10-fold CV yielded optimistic results (R^2 = 0.21-0.26), while rigorous spatial methods like leave-one-station-out (LOSO) and cluster-based CV exposed significant generalization challenges. LOSO CV revealed that while models achieved a pooled RMSE of ≈45.5 μg/m³, their mean R^2 was negative, indicating predictions at unseen locations were often worse than using the simple global mean. Both algorithms showed comparable accuracy, though XGBoost exhibited greater robustness to overfitting. We conclude that Landsat-derived proxies offer a viable but limited pathway for NO_2 estimation, as the models captured broad patterns but failed to resolve fine-scale, local variations. This work underscores that rigorous spatial cross-validation is non-negotiable for obtaining a realistic assessment of model performance in air pollution mapping, especially in complex urban environments.

1. Introduction

Nitrogen dioxide (NO_2) is a major gaseous air pollutant and a key component of nitrogen oxides (NO_x), particularly in areas beyond immediate traffic influence (Huang et al., 2022). Epidemiological studies have consistently demonstrated that exposure to NO_2 is associated with both acute and chronic health outcomes, including cardiovascular and respiratory diseases (de Hoogh et al., 2019; Faustini et al., 2014). Notably, research indicates that the health impacts of NO_2 exposure on cardiovascular and respiratory systems can be as significant as those of fine particulate matter ($PM_{2.5}$) (Faustini et al., 2014). As a criteria air pollutant regulated under national ambient air quality standards worldwide, NO_2 not only directly harms human health but also serves as a precursor for secondary pollutants including tropospheric ozone and particulate nitrate, amplifying its public health significance (Ashmore, 2005; Kim et al., 2024).

The conventional approach to NO_2 monitoring relies on ground-based air quality monitoring stations (AQMS), which provide high temporal resolution measurements but suffer from sparse spatial coverage (Guay et al., 2011; He et al., 2023). This limitation is especially pronounced in developing regions where monitoring networks are often concentrated in metropolitan centers, leaving significant gaps in spatial assessment (Siddique et al., 2024). Satellite remote sensing has emerged as a powerful complementary technology, offering comprehensive spatial coverage and regular revisit capabilities. Instruments such as the Ozone Monitoring Instrument (OMI) and more recently the TROPOspheric Monitoring Instrument (TROPOMI) have enabled global monitoring of tropospheric NO_2 column densities (Lamsal et al., 2021; van Geffen et al., 2020). However, these satellite-derived column densities represent integrated atmospheric concentrations rather than surface-level exposures, requiring sophisticated modeling approaches to bridge the scale difference (Dang et al., 2023; Holloway et al., 2021).

Current methodologies for estimating surface NO_2 concentrations encompass diverse approaches including land use regression models, spatiotemporal interpolation techniques, and increasingly, machine learning algorithms (Beelen et al., 2013; Di et al., 2020). Among these, tree-based machine learning methods such as Random Forest and XGBoost have demonstrated superior performance in capturing the complex nonlinear relationships between NO_2 concentrations and predictive variables (Ghahremanloo et al., 2021; Lu et al., 2020). Recent advances have further explored ensemble modeling strategies that combine multiple algorithms to enhance prediction accuracy and robustness (Di et al., 2020; He et al., 2023). Nevertheless, a critical challenge remains in the effective integration of multi-source geospatial data while maintaining high spatial resolution capabilities.

While previous studies have heavily relied on satellite NO_2 column products from instruments like TROPOMI as primary predictors for both large-scale mapping and downscaling applications—such as monitoring urban pollution hotspots (Rahman et al., 2024), validating air quality data (Ngcoliso et al., 2024), and analyzing correlations with land cover (Hazaymeh et al., 2024)—a significant research gap remains regarding the feasibility of high-resolution NO_2 mapping using alternative satellite data sources. This is particularly critical when direct NO_2 column measurements are unavailable or limited by factors like cloud cover or the absence of dedicated atmospheric sensors (Li and Wu, 2021). Landsat imagery, with its moderate spatial resolution (30-100m) and extensive historical archive, presents a promising alternative through its capacity to capture surface characteristics and environmental proxies related to NO_2 distribution patterns. When integrated with topographic derivatives and infrastructure proximity metrics, these multi-source geospatial datasets offer a comprehensive framework for NO_2 exposure assessment.

This study addresses three key research objectives: (1) to evaluate the feasibility of high-resolution NO_2 mapping using Landsat imagery combined with topographic and proximity variables; (2) to assess the performance of machine learning algorithms in capturing spatial NO_2 patterns without direct

satellite NO_2 column inputs; and (3) to compare different cross-validation strategies for robust spatial model evaluation. By focusing on Tehran as a case study of a rapidly urbanizing metropolis with complex air quality challenges, we aim to develop a transferable framework for NO_2 exposure assessment that can be applied in regions with limited dedicated air quality monitoring infrastructure.

2. Materials and Methods

2.1. Study area

The study area is Tehran, Iran (Figure 1), a metropolis of approximately 16 million people situated on the southern slopes of the Alborz Mountain range. This unique topography, with mountains to the north and plains to the south, creates a semi-arid basin that severely traps air pollutants, limiting dispersion. The city experiences a semi-arid climate with low precipitation and is characterized by rapid urbanization, high traffic volumes, and limited green space (~4.5 m² per person). These factors, combined with the prevailing wind patterns that transport emissions but are blocked by the northern mountain barrier, result in persistent air quality challenges, making Tehran an ideal case study for investigating NO_2 distribution patterns.

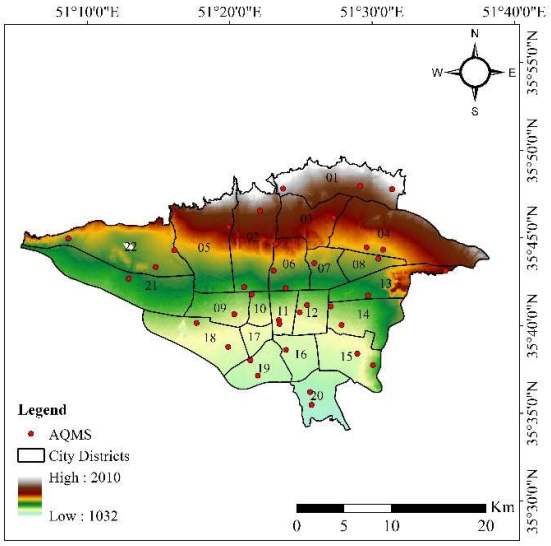

Figure 1. Study area

2.2. Data

A total of 22 Landsat 8 and 9 satellite images were acquired for the summer periods (Tir, Mordad, and Shahrivar months) of 2022 and 2023, with temporal extension to include images from late Khordad and early Mehr to maximize data coverage. All images were processed in Google Earth Engine (GEE) to derive ten spectral indices and thermal characteristics at 100-meter spatial resolution. Each processed raster contained the following bands: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Albedo, Normalized Difference Built-up Index (NDBI), Urban Index (UI), Modified Normalized Difference Water Index (MNDWI), Land Surface Temperature in Celsius (LST), and Tasseled Cap transformations (Brightness, Greenness, and Wetness) (Table 1). All derived products were exported as GeoTIFF files with consistent spatial referencing and 100-meter grid resolution to maintain compatibility across the dataset.

Topographic characteristics were derived from ASTER Global Digital Elevation Model (GDEM) data, including elevation (DEM) and slope calculations, resampled to 100-meter resolution to match the Landsat-derived products. Proximity variables were generated using OpenStreetMap (OSM) data, calculating Euclidean distances to major roads and industrial areas. These distance metrics serve as proxies for anthropogenic emission sources, with road distance capturing traffic-related NO_2 contributions and industrial distance representing point source emissions. All topographic and proximity layers were normalized to a common 100-meter grid and scaled to a range between 0 and 1 to ensure comparability across datasets.

Ground-level NO_2 concentration measurements were obtained from Tehran's air quality monitoring network, comprising 38 stations distributed throughout the metropolitan area. Data were collected for the summer months of 2022 and 2023, corresponding with the satellite acquisition period. After quality control and removal of missing records (NO_2 = 0), the final dataset contained 491 valid measurements across both years. Not all stations maintained continuous operation throughout the study period, resulting in variable temporal coverage across the network.

A spatiotemporal matching procedure was implemented to associate satellite-derived predictors with ground-based NO_2 measurements. For each Landsat acquisition date, the corresponding NO_2 concentrations recorded at monitoring stations on the same day were compiled from station records stored in Excel files. Predictor variable values were then extracted at station locations using point sampling techniques, ensuring precise spatial alignment between satellite-derived features and ground measurements.

The final integrated dataset contained 21 variables, including station identifiers, NO_2 measurements, 10 Landsat-derived indices, 4 topographic/proximity variables, and spatiotemporal metadata. The predictor set encompassed spectral vegetation indices (NDVI, EVI), urban characteristics (NDBI, UI), hydrological features (MNDWI), thermal properties (LST), landscape transformations (Tasseled Cap components), and anthropogenic influence proxies (road and industrial distances). Land Surface Temperature and DEM were normalized to a 0-1 scale to enhance model stability, while spatial coordinates (X, Y) were included to capture geographic trends.

While the primary analysis utilized synchronous Landsat-NO_2 pairs, temporal features were engineered to capture seasonal patterns within the summer period. Day of Year (DOY) was transformed into cyclic features using sine and cosine transformations to model seasonal progression while maintaining continuity between year-end and year-beginning. This approach allowed the models to capture intra-seasonal variations in NO_2 concentrations related to meteorological patterns and human activity cycles during the summer months.

Comprehensive quality checks were performed to ensure data integrity, including validation of spatial alignment, assessment of temporal consistency, and screening for missing values. The final dataset of 491 observations with complete predictor information provided a robust foundation for subsequent

machine learning modeling, representing a comprehensive integration of remote sensing, topographic, and urban infrastructure data for NO_2 concentration estimation.

Variable Category	Specific Variables	Description	Source
Vegetation Indices	NDVI, EVI	Vegetation health and density	Landsat 8/9
Urban Characteristics	NDBI, UI, Albedo	Built-up areas and urban surface properties	Landsat 8/9
Thermal Properties	LST_Celsius	Land surface temperature	Landsat 8/9
Landscape Components	TC_Brightness, TC_Greenness, TC_Wetness	Spectral landscape transformations	Landsat 8/9
Topographic Factors	DEM, Slope	Elevation and terrain characteristics	ASTER GDEM
Proximity Metrics	Road Distance, Industrial Distance	Distance to emission sources	OpenStreet Map
Temporal Features	DOY_sin, DOY_cos	Seasonal progression within summer	Date transformation
Spatial Coordinates	X, Y	Geographic location	

Table 1. Summary of predictor variables used in the study

2.3. Methods

2.3.1. Machine Learning Framework

Two advanced tree-based machine learning algorithms were employed for NO_2 concentration estimation: Random Forest (RF) (Bartkowiak et al., 2019) and eXtreme Gradient Boosting (XGBoost) (Xiao et al., 2023). These ensemble methods were selected for their demonstrated superiority in capturing complex nonlinear relationships between environmental predictors and air pollutant concentrations (Dong et al., 2020; (Aksoy et al., 2025). Random Forest operates on the bagging principle, constructing multiple decision trees during training and outputting the mean prediction of individual trees, while XGBoost employs a gradient boosting framework that sequentially builds trees to correct errors from previous iterations.

To ensure a robust evaluation of model generalizability, we implemented and compared two distinct hyperparameter tuning strategies within our spatial validation framework:

Global Tuning: A single set of optimal hyperparameters was identified for each algorithm using Randomized Search with 5-fold GroupKFold cross-validation (grouped by station) on the entire dataset. The parameter search space for Random Forest included the number of trees (n_estimators: 100-400), maximum tree depth (max_depth: 6-20), and minimum samples required to split nodes (min_samples_split: 2-10). For XGBoost, the search encompassed learning rate (learning_rate: 0.01-0.1), maximum tree depth (max_depth: 3-10), and regularization parameters (reg_alpha: 0-0.5, reg_lambda: 1-2). These globally optimal parameters were then frozen and applied across all folds of the primary Leave-One-Station-Out Cross-Validation (LOSO CV).

Nested Tuning: A more computationally intensive but rigorous approach was employed, where hyperparameter optimization was nested within each fold of the LOSO CV. For each held-out station, a separate Randomized Search with 5-fold CV was performed exclusively on the training stations to find the best parameters for that specific training set. This method prevents any potential information leakage from the test station during tuning and assesses the model's ability to self-optimize for new spatial contexts.

2.3.2. Feature Engineering and Selection

A comprehensive feature engineering pipeline was implemented to enhance model performance and interpretability. Initial predictors included 10 Landsat-derived spectral indices, 4 topographic/proximity variables, spatial coordinates, and temporal features. The pipeline incorporated several preprocessing steps:

First, near-zero variance filtering was applied to remove predictors with minimal information content (variance threshold: 1e-6). Second, highly correlated features (Pearson correlation > 0.95) were identified and reduced, prioritizing retention of features with stronger correlations to NO_2 concentrations. Third, adaptive feature selection was performed within each cross-validation fold using Random Forest importance rankings, retaining features contributing to 95% cumulative importance or the top 15 features as a fallback.

Feature engineering included derivation of inverse distance transformations for road and industrial proximity metrics, logarithmic transformations of distance variables, and interaction terms between key predictors such as elevation-slope and elevation-land surface temperature combinations. Temporal features were encoded using cyclic transformations of day of year to capture seasonal patterns while maintaining continuity. All continuous variables were scaled using RobustScaler to mitigate the influence of outliers.

2.3.3. Spatial Cross-Validation Framework

To comprehensively evaluate model performance and address the critical challenge of spatial autocorrelation in environmental data, three distinct cross-validation approaches were implemented with increasing spatial rigor:

2.3.3.1 Leave-One-Station-Out Cross-Validation (LOSO CV)

As our primary evaluation method, LOSO CV iteratively held out all data from one monitoring station as the test set while training the model on data from the remaining stations. This process was repeated until each of the 29 stations had served as the test set once. For each fold, a rigorous feature engineering pipeline was applied independently to the training data to prevent data leakage, including near-zero variance filtering, removal of highly correlated features (Pearson's r > 0.95), and adaptive

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-751-2026 | © Author(s) 2026. CC BY 4.0 License.

feature selection based on Random Forest importance (retaining features contributing to 95% cumulative importance).

LOSO CV provides the most realistic assessment of model performance when predicting at completely unseen locations, effectively preventing spatial autocorrelation from inflating performance metrics (Pahlevan et al., 2021). By implementing both global and nested tuning within this framework, we were able to evaluate not only spatial generalizability but also the impact of hyperparameter stability on model performance.

2.3.3.2 Cluster-Based Spatial Cross-Validation

Monitoring stations were grouped into five spatial clusters using K-means clustering based on geographic coordinates. The model was trained on data from four clusters and tested on the remaining cluster, with this process repeated for all cluster combinations. This method evaluates regional generalization capabilities while maintaining spatial independence between training and test sets (Ansari et al., 2023).

2.3.4. Traditional 10-Fold Cross-Validation

As a baseline comparison, standard 10-fold CV with random data splitting was employed. While computationally efficient, this conventional approach potentially overestimates model performance due to spatial autocorrelation when samples from the same geographic area may appear in both training and test sets (Ansari et al., 2023).

2.3.5. Performance Metrics and Evaluation

Model performance was evaluated using three standard metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). These metrics provide complementary insights into model accuracy, with RMSE emphasizing larger errors, MAE providing a robust measure of average error, and R^2 indicating the proportion of variance explained by the model.

The overall performance for each CV method was calculated as both the average across all folds and through pooled predictions aggregating all test set predictions. Training metrics were additionally computed to assess model overfitting, with performance gaps between training and test sets indicating generalization capability.

All analyses were implemented in Python using scikit-learn for machine learning algorithms, XGBoost for gradient boosting. The nested cross-validation design ensured complete separation between hyperparameter tuning and final model evaluation, preventing optimistic bias in performance estimates.

3. Results

This study evaluated the feasibility of predicting ground-level NO_2 concentrations using Landsat-derived indices, topographic, and proximity variables, without relying on direct satellite NO_2 column products. The performance of Random Forest (RF) and XGBoost (XGB) models was rigorously assessed using multiple spatial cross-validation strategies to ensure robust and generalizable results.

3.1. Model Performance and Generalization Ability

The model performance varied significantly based on both the cross-validation strategy and the hyperparameter tuning approach, underscoring the critical importance of rigorous spatial evaluation. The results from the Leave-One-Station-Out Cross-Validation (LOSO CV) provide the most realistic estimate of model performance when predicting at entirely unseen locations.

A key finding was the comparable performance between the two hyperparameter tuning strategies. The Global Tuning approach (using a single, fixed parameter set) and the more computationally intensive Nested Tuning (re-optimizing parameters for each fold) yielded similar results. For the pooled predictions across all stations, both methods showed modest performance, with RMSE values of approximately 45.5 µg/m³ and R^2 values around 0.04 for both Random Forest and XGBoost (Table 2). The nested tuning offered only a marginal improvement for XGBoost (RMSE: 45.14 vs. 45.48 µg/m³), suggesting that the globally tuned parameters were already near-optimal for spatial generalization.

Validation Method	Tuning Strategy	Model	RMSE (µg/m³)	MAE (µg/m³)	R^2
LOSO CV (Pooled)	Global	RF	45.41	24.84	0.041
	Global	XGB	45.48	25.33	0.038
LOSO CV (Pooled)	Nested	RF	45.55	25.02	0.036
	Nested	XGB	45.14	24.91	0.053
10-Fold CV	Global	RF	35.91	21.25	0.212
	Global	XGB	35.57	22.29	0.257
Cluster-Based CV	Global	RF	45.15	25.47	0.052
	Global	XGB	46.38	27.39	0.000

Table 1. Overall model performance across different cross-validation and tuning strategies.

When performance was averaged across the 29 held-out stations, the mean RMSE was lower (Global Tuning - RF: 34.12 µg/m³, XGB: 34.80 µg/m³; Nested Tuning - RF: 34.24 µg/m³, XGB: 33.53 µg/m³). However, the mean R^2 values were consistently negative across both LOSO approaches. This indicates that while the models' average error magnitude was acceptable, their predictions for specific, unseen stations were often worse than simply using the global mean concentration, highlighting the fundamental challenge of spatial extrapolation in a complex urban environment.

The 10-fold CV, which does not account for spatial autocorrelation, yielded deceptively optimistic results with R^2 values above 0.21 for both models. This performance inflation is a classic symptom of data leakage, where training and test sets contain samples from the same spatial regions, allowing the model to "memorize" local patterns rather than learning generalizable relationships. The cluster-based spatial CV, which tests generalization to entire regions, confirmed the findings of the LOSO approach, with R^2 values close to zero. This consistent pattern across rigorous spatial CV methods confirms that our models capture broad spatial trends but struggle with precise predictions at specific, unseen locations.

3.2. Analysis of Spatial Extrapolation and Overfitting

The LOSO CV results revealed significant station-to-station variation in prediction error, highlighting the challenges of spatial extrapolation. For instance, both models and tuning strategies performed exceptionally poorly at Station 0 (RMSE > 85 µg/m³) and Station 29 (RMSE > 190 µg/m³), suggesting these stations are influenced by highly localized emission sources or unique micro-environments not captured by our 100m-resolution predictors. Conversely, models achieved good accuracy (RMSE < 20 µg/m³) at several stations (e.g., Stations 1, 4, 16, 17), indicating the feature set was sufficiently informative in these areas.

An analysis of the performance gap between training and test sets provides clear evidence of overfitting. As summarized in Table 3, under the Global Tuning strategy, the Random Forest model showed a substantial Train-Test RMSE gap of -3.37, compared to -1.46 for XGBoost. This indicates that the RF model was more prone to learning station-specific noise in the training data, which degraded its performance on unseen stations. For over a third of the stations, this overfitting exceeded 10 µg/m³ in RMSE.

Model	Tuning Strategy	Train RMSE	Train R^2	Test RMSE	Test R^2	RMSE Gap	R^2 Gap
RF	Global	30.75 ± 2.35	0.56 ± 0.03	34.12	-2.32	-3.37	2.88
XGB	Global	33.33 ± 2.28	0.48 ± 0.02	34.80	-3.73	-1.46	4.21
RF	Nested	33.50 ± 3.30	0.48 ± 0.06	34.24	-2.59	-0.74	3.06
XGB	Nested	33.95 ± 4.74	0.46 ± 0.10	33.53	-2.70	+0.41	3.15

Table 2. Training performance and overfitting analysis from LOSO CV with Global and Nested Tuning

Note: RMSE Gap = Train RMSE - Test RMSE. A negative value indicates test error > train error, i.e., overfitting.

The Nested Tuning strategy had a pronounced effect on mitigating overfitting, particularly for XGBoost. As shown in Table 3, the nested approach successfully reduced the overfitting gap for XGBoost, resulting in a positive RMSE gap (+0.41), which indicates slightly better performance on the unseen test stations than on the training set—a sign of strong generalization. This suggests that the nested tuning protocol was more effective at finding robust, generalizable parameters for XGBoost across different spatial contexts. While nested tuning also slightly reduced the overfitting for Random Forest (RMSE Gap: -0.74), it remained prone to learning spatial idiosyncrasies in the training data.

3.3. Comparative Performance of Machine Learning Models

Overall, the RF and XGB models demonstrated remarkably similar predictive accuracy in the spatial validation frameworks, with neither model establishing a clear superiority in final test performance. The choice of hyperparameter tuning strategy had a more significant impact on model behavior than the choice of the algorithm itself.

The key difference between the models lay in their training behavior and propensity for overfitting. The Random Forest model consistently achieved lower training errors and higher training R^2 values across both tuning strategies, indicating a greater inherent capacity to fit the training data. However, this strong fitting capability came at the cost of reduced generalization, as evidenced by its larger performance drop on the test sets and more significant overfitting gap, particularly under the Global Tuning strategy.

In contrast, XGBoost, with its built-in regularization framework (reg_alpha=0.5, reg_lambda=1.5), exhibited higher training errors but demonstrated greater robustness. This was especially true under the Nested Tuning strategy, where XGBoost successfully achieved a positive RMSE gap, indicating superior generalization. While its final test metrics were similar to RF, XGBoost's more consistent performance between training and testing, and its enhanced responsiveness to rigorous tuning, position it as the more reliable and robust choice for spatial extrapolation in this application.

4. Conclusion

This study demonstrates the feasibility and limitations of predicting surface NO_2 concentrations using Landsat imagery combined with topographic and proximity variables, without relying on direct satellite NO_2 column products. Implemented through an integrated Google Earth Engine and Google Colab Python workflow, this approach enabled efficient processing of satellite imagery and development of machine learning models. While the machine learning models successfully learned broad spatial patterns—as indicated by reasonable performance in traditional 10-fold cross-validation—the rigorous spatial validation frameworks revealed significant challenges in extrapolating to entirely unseen locations.

The primary methodological finding is that spatial cross-validation is non-negotiable for the realistic evaluation of air pollution mapping models. Performance metrics from traditional CV were severely inflated due to spatial autocorrelation, while LOSO and cluster-based CV provided a stark, but more accurate, assessment of model generalizability. The consistently low or negative R^2 values from these spatial tests indicate that our set of 100m-resolution predictors lacks the granularity to capture the fine-scale, local dynamics of NO_2 in a complex urban environment like Tehran. Key unmeasured factors, such as hyper-local traffic volume, building morphology, and real-time meteorological conditions, likely govern the concentrations at individual stations and are not captured by our static, moderate-resolution features.

Our comparison of modeling approaches yielded critical insights for future applications. The similar performance between Random Forest and XGBoost suggests that the model choice is less critical than the validation framework. However, XGBoost demonstrated greater robustness and a better response to nested hyperparameter tuning, making it the preferable algorithm for spatial extrapolation tasks. Furthermore, the minimal

performance gain from the computationally expensive nested tuning over the global tuning strategy indicates that a well-conducted global parameter optimization may be sufficient for practical applications, streamlining the modeling process.

For future work, several pathways could bridge the gap between model capability and real-world accuracy. Integrating very high-resolution imagery, data from street-level sensors, or incorporating physical dispersion models could provide the necessary local context. Additionally, exploring spatial feature engineering or graph neural networks that explicitly model the relationships between monitoring stations might better capture the spatial processes driving NO_2 distribution.

Despite these challenges, the framework presented provides a viable and transferable pathway for NO_2 exposure assessment, particularly in data-scarce regions where high-resolution satellite NO_2 products from instruments like TROPOMI are unavailable or limited. This work establishes a critical benchmark for the performance that can be expected when using Landsat-derived proxies for air quality mapping and underscores the paramount importance of rigorous spatial validation in environmental machine learning.

References

Abbasi, M.T., Alesheikh, A.A., Rezaie, F., 2025. A Lightweight Spatiotemporal Graph Framework Leveraging Clustered Monitoring Networks and Copula-Based Pollutant Dependency for PM2.5 Forecasting. Land 14(8), 1589. https://doi.org/10.3390/land14081589

Aksoy, S., Sertel, E., Roscher, R., Tanik, A., Hamzehpour, N., 2024. Assessment of soil salinity using explainable machine learning methods and Landsat 8 images. Int. J. Appl. Earth Obs. Geoinf. 130, 103879. https://doi.org/10.1016/j.jag.2024.103879

Ansari, M., Knudby, A., Amani, M., Sawada, M., 2023. Retrieving Inland Water Quality Parameters via Satellite Remote Sensing: Sensor Evaluation, Atmospheric Correction, and Machine Learning Approaches. Remote Sens. 17(10), 1734. https://doi.org/10.3390/rs15102505

Ashmore, M.R., 2005. Assessing the future global impacts of ozone on vegetation. Plant Cell Environ. 28(8), 949–964. https://doi.org/10.1111/j.1365-3040.2005.01341.x

Bartkowiak, P., Castelli, M., Notarnicola, C., 2019. Downscaling Land Surface Temperature from MODIS Dataset with Random Forest Approach over Alpine Vegetated Areas. Remote Sens. 11(11), 1319. https://doi.org/10.3390/rs11111319

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., Dėdelė, 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., de Hoogh, K., 2013. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmos. Environ. 72, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037

Chi, Y., Fan, M., Zhao, C., Zhang, Z., Zhu, K., Feng, R., 2021. Ground-level NO_2 concentration estimation based on OMI tropospheric NO_2 column. Remote Sens. 13(3), 436. https://doi.org/10.3390/rs13030436

Dang, R., Liao, H., Fu, Y., 2023. Estimating daily full-coverage surface NO_2 concentrations over China based on a hybrid machine learning model. Environ. Sci. Technol. 57(8), 3246–3256. https://doi.org/10.1021/acs.est.2c07398

de Hoogh, K., Gulliver, J., Donkelaar, A.V., Martin, R.V., Marshall, J.D., Bechle, M.J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., Forsberg, B., Galassi, C., Heinrich, J., Hoffmann, B., Jacquemin, B., Katsouyanni, K., Korek, M., Künzli, N., Lindley, S.J., Leander, K., Meleux, F., de Nazelle, A., Nieuwenhuijsen, M.J., Nystad, W., Raaschou-Nielsen, O., Peters, A., Peuch, V.H., Rouil, L., Udvardy, O., Slama, R., Stempfelet, M., Stephanou, E.G., Tsai, M.Y., Yli-Tuomi, T., Weinmayr, G., Brunekreef, B., Vienneau, D., Hoek, G., 2019. Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 177, 108601. https://doi.org/10.1016/j.envres.2019.108601

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2020. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 142, 105827. https://doi.org/10.1016/j.envint.2020.105827

Faustini, A., Rapp, R., Forastiere, F., 2014. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. Eur. Respir. J. 44(3), 744–753. https://doi.org/10.1183/09031936.00114713

Ghahremanloo, M., Choi, Y., Sayeed, A., Salman, A.K., 2021. Deep learning-based estimation of surface NO_2 concentrations over Texas. Remote Sens. Environ. 264, 112580. https://doi.org/10.1016/j.rse.2021.112580

Grzybowski, P., Ziółkowski, D., 2023. Estimation of surface NO_2 concentrations over Poland based on satellite data. Atmos. Environ. 294, 119508. https://doi.org/10.1016/j.atmosenv.2022.119508

Guay, P., et al., 2011. J. Expo. Sci. Environ. Epidemiol. 21(2), 115–128.

Haashemi, S., Weng, Q., Darvishi, A., Alavipanah, S.K., 2016. Seasonal Variations of the Surface Urban Heat Island in a Semi-Arid City. Remote Sens. 8(4), 352. https://doi.org/10.3390/rs8040352

He, S., Dong, H., Zhang, Z., Yuan, Y., 2023. An Ensemble Model-Based Estimation of Nitrogen Dioxide in a Southeastern Coastal Region of China. Remote Sens. 15(6), 1684. https://doi.org/10.3390/rs15061684

Holloway, T., Miller, D., Anenberg, S., 2021. Satellite Monitoring for Air Quality and Health. Annu. Rev. Biomed. Data Sci. 4, 417–447. https://doi.org/10.1146/annurev-biodatasci-031021-100511

Huang, C., Sun, K., Hu, J., Xue, T., Xu, H., Wang, M., 2022. Estimating 2013–2019 NO_2 exposure with high spatiotemporal resolution in China using an ensemble model. Environ. Pollut. 314, 120256. https://doi.org/10.1016/j.envpol.2022.120256

IEEE Trans. Geosci. Remote Sens., 2024. Estimation of Surface-Level NO_2 Using Satellite Remote Sensing and Machine Learning. 62, 1–14.

Kim, E.J., Holloway, T., Kokandakar, A., Harkey, M., Elkins, S., Goldberg, D.L., Heck, C., 2024. A Comparison of Regression Methods for Inferring Near-Surface NO_2 With Satellite Data. J. Geophys. Res. Atmos. 129, e2023JD039400. https://doi.org/10.1029/2023JD039400

Lamsal, L.N., Krotkov, N.A., Celarier, E.A., Swartz, W.H., Pickering, K.E., Bucsela, E.J., Gleason, J.F., Martin, R.V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Cohen, R.C., 2021. OMI/Aura NO_2 Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25 degree × 0.25 degree V3. NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/Aura/OMI/DATA3007

Li, T., Wu, J., 2021. A machine learning approach for estimating high-resolution NO_2 in China. Environ. Res. 201, 111456. https://doi.org/10.1016/j.envres.2021.111456

Li, X., Jia, H., Wang, L., 2023. Remote Sensing Monitoring of Drought in Southwest China Using Random Forest and eXtreme Gradient Boosting Methods. Remote Sens. 15(19), 4840. https://doi.org/10.3390/rs15194840

Lu, X., Zhang, L., Chen, Y., 2020. Estimating hourly surface $PM_{2.5}$ concentrations across China from satellite data. Atmos. Chem. Phys. 20(13), 7753–7769. https://doi.org/10.5194/acp-20-7753-2020

Pahlevan, N., et al., 2021. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sens. Environ. 270, 112860. https://doi.org/10.1016/j.rse.2021.112860

Siddique, M. A., Naseer, E., Usama, M., & Basit, A. (2024). Estimation of Surface-Level NO2 Using Satellite Remote Sensing and Machine Learning: A review. IEEE Geoscience and Remote Sensing Magazine, 12(3), 8–34. https://doi.org/10.1109/mgrs.2024.3398434

van Geffen, J., Boersma, K.F., Eskes, H., Sneep, M., ter Linden, M., Zara, M., Veefkind, J.P., 2020. Sentinel-5P TROPOMI NO_2 retrieval: impact of version v1.2 improvements and comparisons with OMI and ground-based data. Atmos. Meas. Tech. 13(3), 1315–1335. https://doi.org/10.5194/amt-13-1315-2020

Venter, Z.S., 2023. Comparing Global Sentinel-2 Land Cover Maps for Regional Species Distribution Modeling. Remote Sens. 15(7), 1749. https://doi.org/10.3390/rs15071749

Xiao, Y., Li, S., Huang, J., Huang, R., Zhou, C., 2023. A New Framework for the Reconstruction of Daily 1 km Land Surface Temperatures from 2000 to 2022. Remote Sens. 15(20), 4982. https://doi.org/10.3390/rs15204982