High Resolution Nitrogen Dioxide Estimation and Applications for Louisiana

Amanda Murray, Alberto Lopez Rueda, Kory Kirshenbaum, and Alexandra Hurst
University of California, Berkeley
School of Information

Increased monitoring of criteria air pollutants in Louisiana is vital to address residents' concerns of heightened levels of industrial air pollution in the state. Recent monitoring efforts in Louisiana have largely focused on ground-level sensors (US EPA, 2022). While these sensors provide literal ground-truth information, they face several limitations: they are costly, fixed in place, and cannot provide comprehensive, statewide data. To fill this gap, we combine high-resolution satellite data collected by the TROPOspheric Monitoring Instrument onboard the Sentinel-5 Precursor satellite (Veefkind et al. 2012) with an ensemble of machine learning models (Di et al. 2020; Ghahremanloo et al. 2021; Chi et al. 2022) to generate satellite-derived, 1-kilometer-squared estimates of ground-level nitrogen dioxide in Louisiana for every week from 2019 to the present. Using this data, we show the disparate impact of air pollution among demographic groups in the state and develop a simple algorithm that can be used to identify areas that should be prioritized for the placement of ground-level sensors.

1. Introduction and Background Work

Residents of Louisiana have raised concerns for decades about high levels of air pollution that are believed to be linked to heightened cancer levels and other health risks (Baurick 2019). Of particular concern is "Cancer Alley"–an 85-mile stretch of land in Louisiana along the Mississippi river that contains over 150 industrial plants and is named for having cancer rates nearly 47 times the national average (Naugra 2021; Parker 2023).

Some research has been conducted in the area to begin to understand patterns of air pollution and investigate the link between industrialization, pollution, and health risks. The United Network for Human Rights used statistical methods to show the improbability of the current cancer rates around Denka's neoprene manufacturing facility in St. John's Parish, and that these cancer rates were correlated with proximity to the plant (University Network for Human Rights 2019). Terrell and Julien (2022) identified a correlation between cancer, race, and air pollution in Louisiana, providing the first evidence that air pollution may be a key factor contributing to the inequitable health outcomes in the state.

However, a lot still remains unknown about air pollution and its impact in Louisiana, in part due to a lack of adequate data (Kofman 2022). For monitoring air pollutants, the Louisiana Department of Environmental Quality (LDEQ) releases data on 29 air monitoring sites for ozone, sulfur dioxide, and PM2.5 (*https://airquality.deq.louisiana.gov/*), and only 9 sites for monitoring nitrogen dioxide (NO_2) (*https://www.epa.gov/aqs*). Residents have also raised concerns about the location of these sensors, with some advocating–thus far unsuccessfully–for fence line air quality monitoring around industrial plants (Sullivan 2023). To address the lack of data, the EPA recently funded over 2 million dollars worth of air quality monitoring projects in the state (US EPA, 2022). Yet increased ground monitoring is only part of the solution, since these sensors are costly, fixed in place, and cannot provide the high resolution maps of air pollution necessary to understand the impact of air pollution across the state–particularly for pollutants like nitrogen dioxide that can vary significantly across space and time (Niepsch et al. 2021).

To fill the gaps left by ground monitoring stations, prior research has used models to generate high resolution estimates of NO_2 pollution. These models generally utilize remote sensing estimates of tropospheric concentrations of NO_2, along with ground sensing data and additional variables, to predict ground-level concentrations. The current existing global and national NO_2 datasets predominantly use data from the Ozone Measurement Instrument (OMI) on the Aura satellite, and apply various methodologies for modeling. For example, Anenberg et al. (2022) rescaled a land use regression model to develop a dataset of annual 1 km^2 resolution NO_2 predictions from 1990-2019 to test for a relationship between pollution and pediatric asthma. Alternatively, Kim et al. (2020) used integrated empirical geographic regression models to develop annual predictions of all critical air pollutants by census block from 1979 to 2015. Di et al. (2020) innovated on these

methods by using an ensemble of machine learning models on top of chemical transport models, generating 1km^2 daily NO_2 estimates. While these datasets have contributed significantly to research on NO_2 pollution, they have limitations, including only being available up to 2019, and, with the exception of Di et al (2020)., only publishing annual predictions.

Additionally, the current global and national NO_2 datasets all rely on data from OMI, a satellite that has been in orbit for nearly 20 years, and provides data at a resolution of 13 km x 25 km (Levelt et al. 2006). The 2018 launch of TROPOspheric Monitoring Instrument (TROPOMI) onboard Sentinel-5 Precursor satellite (Copernicus Sentinel-5p, 2021), which collects NO_2 tropospheric column density data at a resolution of 3.5 km x 5.5 km, provides an exciting opportunity for improvement on air pollution estimation. In addition to the higher resolution, tropospheric column densities from TROPOMI have a much higher correlation with ground sensor measurements than OMI measurements (R=0.67 for TROPOMI compared to R=0.35 for OMI) (Griffin et al. 2019). In practice, the Atmospheric Composition Analysis Group updated their NO_2 predictions for one year using TROPOMI satellite data from 2018-2019, and found that this change significantly increased the correlation and lowered the bias in their model (Cooper et al. 2020), supporting the hypothesis the incorporation of TROPOMI satellite data in models can provide more accurate ground-level estimates.

Recent regional studies across various locations that utilize TROPOMI data, and machine learning methods, have been successful in generating high accuracy and high spatial and temporal resolution ground-level predictions (Chen et al. 2019; Chi et al. 2022; Chan et al. 2021; Scheibenreif et al. 2022). Machine Learning models can combine the advantages of satellite data, in situ observations, and statistical models (Kerr et al. 2023). Given that global and national models for air pollutant estimates often have a lot of variation in accuracy between regions (Anenberg et al. 2022; Hu et al. 2014; Di et al. 2022), these regional models can also be helpful in generating higher accuracy predictions for a specific location of interest. Ghahremanloo et al. (2021) applied this machine learning methodology in Texas with a relatively limited number of ground sensors compared to other papers and got accurate results. Considering the geographic and climatic similarities between Texas and Louisiana, these results suggest the viability of applying similar methods to Louisiana.

The goal of this study is to apply innovative machine learning techniques, using data from the TROPOMI satellite, to develop a dataset of up-to-date, high resolution estimates of NO_2 in Louisiana that can be updated and maintained for future research. We chose NO_2 as the focus pollutant for this research given the limited existing monitoring capacity in Louisiana, the high cost of increased monitoring, the connection to industrialization, and the high spatial variability of the pollutant which makes it easier to estimate disparities and determine the impact of point source emissions (Wang et al. 2023). This methodology could easily be adapted to develop estimates of other pollutants measured by TROPOMI, including ozone and sulfur dioxide.

Upon deriving this dataset, we aim to provide a preliminary analysis of the impact of nitrogen dioxide pollution in the state. We analyze the data for temporal and seasonal patterns in an effort to determine the time of year and locations where people may be most at risk. We then analyze the data for regional disparities in NO_2 exposure, using census block racial demographic data from 2020. Finally, we propose how this dataset could be used to identify locations for future ground-level monitors.

Our research contributes to current literature by providing the most up-to-date estimates for regional NO_2 in the state with accuracy comparable to state-of-the-art models. Additionally, we demonstrate regional disparities in pollution exposure at the block-level for the state, and provide an algorithm for sensor placement prioritization.

2. Study Area and Data Sources

2.1 Study Area

Due to limited NO_2 sensors in Louisiana, we selected a bounding box that included all sensors in east and central Texas. Two sensors in west Texas were excluded from the bounding box due to distance as well as differences in elevation and

climate from Louisiana. The approximate final bounding box stretched from -88.2° to -98.8° longitude and 28.7° to 33.3° latitude. Figure 1 shows the region incorporated in this study, with the points representing locations of ground-level monitors.

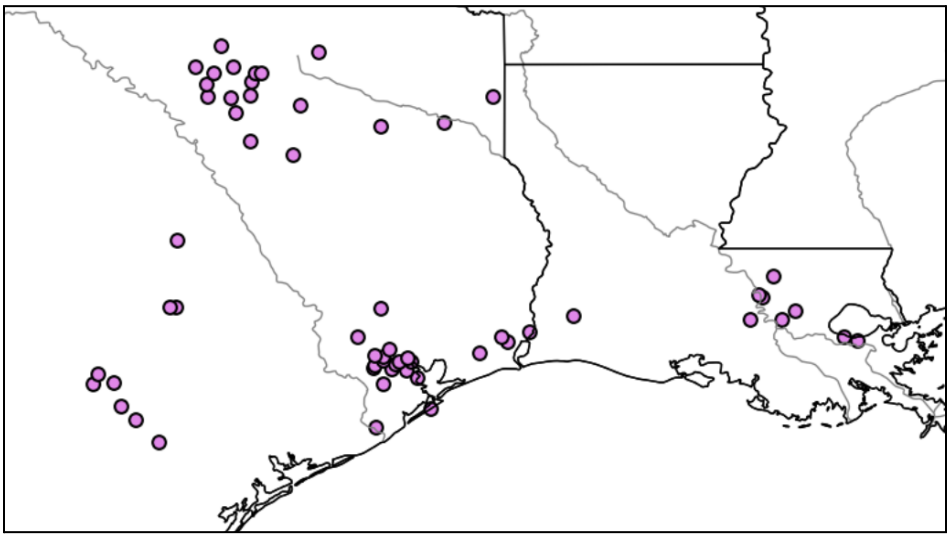


Figure 1: Bounding box of area incorporated in model training. Points represent the locations of all EPA ground-level NO_2 monitors.

2.2 Data Sources

Dta was collected for the timeframe of January 6, 2019 to October 1, 2023. While satellite data was available starting in April 2018, the mobility dataset selected as a feature variable was only available beginning in 2019, and therefore, earlier weeks were removed from the dataset.

As feature variables for the model, we collected weekly averaged TROPOMI NO_2 Tropospheric Vertical Column Density (TVCD), vegetation and land use, meteorological variables, and various anthropogenic variables. The target variable for the model was NO_2 ground sensor measurements.

Weekly averages were collected instead of daily images due to the amount of missing data in daily TROPOMI satellite images from cloud cover and low quality assurance. Additionally, previous research using machine learning with TROPOMI NO_2 TVCD to predict ground-level measurements has shown higher correlation when using weekly averages (Grzybowski et al. 2023). While selecting weeks instead of days removes the ability to observe daily trends within a week, we were still able to capture seasonal variation observed in other studies (Beirle et al. 2003; Roberts–Semple et al. 2012; Meena and Jadhav 2007). Due to the limited number of sensors in the area of interest, monthly averages would not have generated sufficient data for training a machine learning model.

2.2.1 Tropomi Satellite Data

The TROPOspheric Monitoring Instrument (TROPOMI) on board the Sentinel 5-precursor satellite provides the highest resolution and most accurate estimates of tropospheric NO_2 column density currently available (Copernicus Sentinel-5p 2021). The Sentinel-5p satellite has a revisit period of less than one day, providing daily images with a spatial resolution of 5.5 km x 3.5 km. The 2022 Reprocessed version of the L2 NO_2 products were selected for this study, as recommended for time series analyses in the Sentinel-5 precursor/TROPOMI Level 2 Product User Manual (Eskes et al. 2022).

The swath images were collected from Amazon Open Datasets (*https://registry.opendata.aws/sentinel5p*) and processed using the HARP package from Copernicus's Atmospheric Toolbox (Niemeijer 2017). With HARP, the images were averaged by week, filtered to include only pixels with quality assurance values over 0.75 (as recommended in the product

user manual to remove cloud-covered scenes and problematic retrievals; Eskes et al. 2022), and regridded to a spatial resolution of 0.01°.

Due to removing pixels with low quality assurance, some weekly averaged images contained missing pixels. Appendix 1 shows the breakdown of missing pixels by week. Images with greater than 40% of pixels missing were removed from the dataset, which resulted in the removal of two weeks: the week of 2/3/19 and 2/24/19. Prior studies have used various inpainting methods to fill in missing pixels (Wang et al. 2021; Sarafanov et al. 2020). In this study, biharmonic inpainting was used on the remainder of images to fill in missing pixels (Hudagi et al. 2022). To ensure the accuracy of this inpainting method, we conducted simulated experiments, similar to those used in related studies (Wang et al. 2021). First, we created masks of null values for the images that contained missing pixels. Since there were 95 images that had no missing pixels, we randomly selected 95 of these masks, matched each mask to a full coverage image, and removed pixels where those masks were located. We then implemented biharmonic inpainted to fill in the missing values, and compared the in-painted values to the values in the ground truth image. The Pearson correlation between ground truth and in-painted values for all images was 0.91. Appendix 2 shows examples of images from the masking experiments.

2.2.2 Vegetation and Land Use

NO_2 concentrations have been shown to vary with land use and vegetation (Grundstrom and Pleijel 2014). In order to account for these effects, Enhanced Vegetation Index and Land Use data were collected.

Enhanced Vegetation Index was collected from the MODIS instrument on the TERRA satellite. Enhanced Vegetation Index is similar to the commonly used metric Normalized Difference Vegetation Index (NDVI), but has additional adaptations to correct for atmospheric conditions. These differences can make it more sensitive in areas with dense vegetation (Huete et al. 2016). The MOD13A1 product was accessed from Planetary Computer (Microsoft Open Source 2022), and contains 16-day swaths at a resolution of 500m x 500m. The swaths were mosaiced by month using the stackstac python package (Joseph 2023). The images were then regridded from 500m x 500m to 0.01°x0.01° using bilinear interpolation to match the satellite coordinates.

National Land Cover Database (NLCD) Land Cover data for 2019 was accessed using the Pygeohydro python package (Chegini et al. 2022). This data was regridded to match the satellite coordinates (0.01° x 0.01°) using nearest neighbor interpolation. The land cover data contained 18 micro categories, 13 of which were represented in the training data. Due to the large number of categories compared to the small number of sensors, this data was left out of the model training but used to stratify the sensor split, as described in the modeling section.

2.2.3 Meteorological Data and Elevation

Prior research has also incorporated numerous meteorological variables that could impact the relationship between tropospheric NO_2 and ground-level NO_2. Due to the similarity of regions and methodology in the study, we collected the same meteorological variables used by Ghahremanloo et al. (2021). These variables included temperature (K), surface humidity (kg/kg), surface pressure (Pa), U and V wind (10 meters above ground, m/s), downward shortwave radiation flux (W/m^2), downward longwave radiation flux (W/m^2), fraction of total precipitation that is convective (unitless), convective available potential energy (J/kg), surface potential evaporation (kg/m^2), and surface precipitation hourly total (kg/m^2). This data was collected from the North American Land Data Assimilation System (NLDAS) -2 Forcing Dataset (*https://ldas.gsfc.nasa.gov/nldas/v2/forcing*). All NLDAS data is presented hourly in a 0.125° x 0.125° grid. We aggregated the hourly data to weekly averages from Sunday midnight to Sunday midnight, local time.

Elevation data was collected from the Climatology Lab at University of California Merced (*https://www.climatologylab.org/gridmet.html*). Elevation data was a static 4km x 4 km grid formulated to be on the same meteorological grid as the weather model produced by the Climatology Lab (GRIDMET). Since the data was static, it was expanded over time to line up with the other features of our dataset.

All meteorological and elevation data was regridded to match the satellite coordinates (0.01° x 0.01°) using bilinear interpolation.

2.2.4 Anthropogenic Variables

We selected various anthropogenic variables to incorporate into the model. As used in similar research, we included impervious surface measurements, population density, and road density (Grzybowski et al. 2023; Chi et al. 2022; Ghahremanloo et al. 2021). Due to the impact of COVID-19 on air pollution (Berman and Ebisu 2020), we also chose to incorporate mobility data by county so the model could learn patterns that may have arisen from localized stay-at-home orders and decreased mobility due to the pandemic.

Impervious Surface data for 2019 was also accessed from NLCD. Impervious surface is a measurement of how developed an area is, giving the percentage of developed impervious surface for each 30m x 30m land use pixel (Multi-Resolution Land Characteristics Consortium). This data was gathered at a 1km x 1km resolution. Population density data was collected from Global Human Settlement Layers (GHSL) which provides gridded population density at 30 arcsecond resolution (approximately 0.008° x 0.008°) (Schiavina et al. 2023). We collected road density from the Global Roads Inventory Project (Meijer et al., 2018) at a spatial resolution of 8 km^2. Impervious surface, population density, and road density were regridded using bilinear interpolation to match the satellite coordinates (0.01° x 0.01°).

Mobility data was collected from the Bureau of Transportation statistics, which records the number of trips taken by county, aggregated by week (Bureau of Transportation Statistics 2021). In order to reduce correlation between number of trips and population density, the percentage of people who took trips was calculated by dividing the recorded number of people who did not stay at home by the sum of the recorded number or people who did and did not stay home. We plotted this data over time to see if this calculated value was showing trends from COVID-19, and the results show an expected dip during peak pandemic months (see appendix 3). Mobility data was joined to gridded data by joining satellite coordinates that fell within the given county boundary based on the U.S. Census Tiger Shapefiles (U.S. Census Bureau 2021).

2.2.5 Ground Sensor Measurements

Ground-level NO_2 measurements were collected from the United States Environmental Protection Agency's Air Quality System (AQS) API, which provides access to validated data from thousands of sensors around the country maintained by state and federal agencies (*https://www.epa.gov/outdoor-air-quality-data/*). We chose to utilize this data over the more frequently updated AirNow API, also maintained by the EPA, because the AirNow data is not fully validated and thus "cannot be used to formulate or support regulation, guidance or any other Agency decision or position" (US EPA). Additionally, the AQS data was only a few months behind the AirNow data at the time of data extraction, so there was minimal impact on the recency of our results. Using the AQS API, we pulled daily mean NO_2 values in parts per billion (ppb) for the 59 sensors within the study area bounding box between the start of 2019 and the 3rd quarter of 2023. Next, we dropped any values marked as invalid, as indicated by a validity indicator of "N". Finally, we aggregated the values for each sensor on a weekly basis using the arithmetic mean of the daily measurements in order to match our other data.

3. Methodology

3.1 Data Pre-Processing and Train/Test Split

As described above, all feature data sources were re-gridded using bilinear interpolation to match the satellite data coordinates with a resolution of 0.01° x 0.01°. The data was then cropped to exclude data points outside of the boundaries of Louisiana and Texas. Null values for weather data and mobility data were dropped, as these points fell along the coast or over water.

The feature variables were then joined to the ground sensor measurements using inverse distance weighting (IDW), as has been used in similar studies (Yu and Liu 2021). Rather than linking sensors to the nearest satellite coordinate, IDW weighs the nearest coordinates by distance, to better take into account sensors that may be between two centroids.

Once the data was joined, the correlation between sensor and satellite data was plotted (see appendix 4). The overall correlation (pearson correlation coefficient=0.60) was the same as has been found in similar studies (Yu and Liu 2021) and close to the overall reported correlation for TROPOMI (Griffin et al. 2019). When plotting the correlation by sensor, we noticed two outlier sensors from Louisiana that had a lower correlation with satellite values. We chose to drop these two sensors from the dataset until we had further information, in case these sensors were picking up hyper-local variation, or other noise that would not be beneficial for training a model to generate ~1km x 1km predictions (see Appendix 5). The final dataset for modeling contained 12,387 rows with ground sensor values. Appendix 6 contains a diagram of the data processing pipeline.

To ensure that we could measure how well our model performed on unseen sensors, we held out a test set containing 6 sensors. To select these sensors, we stratified the data into “developed” and “undeveloped” areas based on the NLCD 2019 land cover data. To formulate the stratified grouping, we split the total number of sensors into 90% for training, and 10% for test. We maintain the overall ratio of developed to undeveloped sensors over each set. This stratification ensures that both the test and training set had a mix of urban and rural sensors since prior research has suggested these models may perform better in urban areas given the higher density of sensors (Ghahremanloo et al. 2021). While similar studies have exclusively relied on cross-validation using data from all sensors, we choose to hold out a test set to ensure we have a subset of the data that is not being used during hyperparameter tuning to give a metric of how the model will perform on unseen locations. Holding out 10% of the data for a test set resulted in a training set of 11,055 rows and 51 sensors, and a test set of 1,332 rows and 6 sensors.

3.2 Feature Engineering

Additional features were added or transformed prior to modeling in order to improve the performance of the model. In order to incorporate spatial relationships, a 19 x 19 pixel (~21 km^2) average of the satellite data was calculated around each center location. To reduce autocorrelation with the point location satellite value, the point satellite value was subtracted from this average, and the difference between the point and surrounding pixel was included. Similarly, temporally lagged satellite values were included to account for temporal relationships in the data. The prior week satellite value was subtracted from the current week satellite value to reduce autocorrelation. The time variable was also transformed into one-hot encodings for the month and year in order to account for seasonal and annual variation without learning patterns for a specific date.

UWind and VWind were transformed into wind direction and magnitude using the following equation:

$$wind\_direction = \text{atan2}(\text{v}, \text{u}) + 180^\circ$$

$$wind\_magnitude = \sqrt{\text{u}^2 + \text{v}^2}$$

Finally, All features were scaled using scikit-learn’s Robust Scaler to account for outlier values (Pedregosa et al. 2011). Robust Scaler works by centering the data around the median value rather than the mean, making it less sensitive to outliers. After centering the data around the median, the algorithm will divide it by the interquartile range (between the 75th and 25th percentiles). A Robust Scaler is also less sensitive to feature distribution than a Standard Scaler, which assumes a Gaussian distribution. Scaling the features is necessary to account for differences in units between features and consistent learning in the gradient-based algorithms. The formula describing robust scaling for a given feature X is detailed below.

$$x_{i,\text{scaled}} = \frac{x_i - \text{median}(X_i)}{\text{IQR}(X_i)}$$

3.3 Machine-Learning Modeling

For estimating ground-level pollution, we test four machine-learning models: a Feed-Forward Neural Network, XGBoost, a 1-Dimensional Convolutional Neural Network (1DCNN), and TabNet. Ordinary Least Squares (OLS) regression was used as a baseline for comparison. The sections below describe each of these individual models in further detail.

Each model was trained by minimizing Root Mean Square Error (RMSE) using 10-fold spatial cross-validation on the training set. To implement this approach, all sensors were split into ten non-overlapping groups using stratified random sampling by land use, as was done with the test set. For each fold, we trained the models on 90% of the sensors and evaluated the performance on the validation set consisting of 10% of sensors not seen during training. This spatial cross-validation methodology aligns well with our purpose of estimating pollution for locations without sensors. In contrast, other methodologies, such as standard or temporal cross-validation, could artificially inflate results due to data leakage.

For evaluation, we compare the estimated surface NO2 concentrations from all 10 folds to the observed concentrations, calculating the RMSE as well as MSE-based out of sample R2 for comparisons with existing literature. Out of sample R2 provides a measure of fit to the 1-1 line as opposed to squaring the correlation coefficient which tends to be typically higher (Keller at al. 2014).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$R^2_{\text{out}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

For each individual model, we use hyperparameter search using Optuna (Akiba et al. 2019) to select the optimal hyper-parameters for each model. The models were tuned to minimize the cross-validated RMSE. The sections below indicate the specific hyperparameters for each model selected with this methodology.

3.3.1 Baseline Models

Each model utilized the same set of features described in the section above. We utilized OLS for a baseline model. OLS regression models capture linear relationships between an independent variable, ground sensor NO_2, and a series of dependent variables. OLS regression is fit by solving for the set of coefficients that minimize the sum of squared residuals. The formula, with beta representing the resulting vector of coefficients, is described below.

$$\beta = (X^T X)^{-1} X^T Y$$

We chose OLS regression as a baseline model because of the simplicity of implementing a closed-form solution. However, unlike our final deep learning techniques, OLS regression cannot model non-linear relationships between independent and dependent variables without variable transformation. We describe the model below, with each beta representing a vector of coefficients by category. A full detailed table of results can be found in appendix 7.

$$NO_2 = \beta_0 + \beta_1 \cdot \text{Satellite} + \beta_2 \cdot \text{Weather} + \beta_3 \cdot \text{Land Description} + \beta_4 \cdot \text{Population} + \beta_5 \cdot \text{Temporal Fixed Effects} + \epsilon$$

3.3.2 Feed-Forward Neural Network

A feed-forward neural network model (FFNN) is a multilayered model consisting of interconnected nodes (units) with associated weights and biases. The FFNN uses backpropagation and gradient loss to minimize a specified loss function (gradient descent). Our model is optimized to minimize the RMSE of the predicted NO_2 values versus the ground truth sensors. The models consist of an input layer (m rows in data by n features), hidden layers, and an output layer of various sizes that all pass through the weights and biases to activation functions. Through activation functions and hidden layers, the FFNN can pick up nonlinear and complex relationships that OLS regression will miss (van Hemmen 1988). Additional hyperparameters that are associated with an FFNN are epochs (passes through the training set for optimization), learning rate (step size of the gradient descent process), dropout rate (a regularization to randomly drop units at a rate p), and optimizer (algorithm describing the process to adjust weights and biases). Through Optuna experiments, we used OLS regression to observe the hyperparameters and model structure associated with the lowest spatial cross-validation RMSE. We experimented with different numbers of units, hidden layers, learning rates, activation functions, optimizers, and dropout rates. After observing the experiments' results, we fit a complex model (many layers) and a simple model (a single hidden layer) with a single-unit regression output layer. When testing for generalizability, the simpler model performed better on spatial cross-validation (see appendix 8 for a table of all hyperparameters for the final model).

3.3.31-Dimensional Convolutional Neural Network

The 1-Dimensional Convolutional Neural Network (1DCNN) relies on the same properties of FFNN to minimize a specified loss function and set associated weights and biases. The primary difference between the 1DCNN and the FFNN is the addition of convolutional layers that process sequences of features (Kiranyaz et al. 2021). We hypothesized that this network architecture could pick up local spatial relationships that the FFNN could not capture. Due to the convolution layer, we adjust the input to the network from an m by n matrix to m 1-dimensional vectors containing data representing each of our features. After the convolutional layer, we utilize a max pooling and flattening layer to reduce the dimensions of our model. Once we reduce the model's dimensions, we pass the output to hidden and output layers similar to the FFNN. When experimenting with the optimal hyperparameters for the 1DCNN, we used the same OLS covariates as the FFNN but with additional variables of the number of convolutions (filters) and the size of the convolutions (kernel size). Like with the FFNN, we experimented with simpler and more complex architectures. In this case, we found no difference in results between the two structures and utilized the smaller model for simplicity. See appendix 8 for a table of all hyperparameters for the final model.

3.3.4 XGBoost

XGBoost is an ensemble learning method which combines the predictive power of multiple learners. In this algorithm, decision trees are built sequentially such that each subsequent tree aims to reduce the pseudo-residual errors of the previous tree. This is in contrast to bagging techniques such as Random Forest where each tree predicts a subsample of bootstrapped data. To find the optimal system parameters, we use Optuna, focusing on the number of estimators (or trees), max depth of each tree, minimum child weight to split a node and the learning rate which controls how quickly the predictions are updated after each weak learner. In order to improve generalization, we also sample data by columns and rows for each tree, include regularization parameters to prune the trees (lambda, gamma and alpha), and apply early stopping after 20 trees without improvement in RMSE (see appendix 8 for a table of all hyperparameters for the final model).

3.3.5 TabNet

TabNet is a deep learning encoder-decoder algorithm developed by Google for tabular data (Arik and Pfister 2020). The algorithm combines the features of neural networks to fit complex functions and the feature selection benefits of decision trees (see appendix 9 for a diagram of the model architecture). One unique difference is that TabNet uses a sequential attention mechanism in order to select the most appropriate model features at each step in the model while decision-trees only do feature-selection globally. We also use Optuna to find the best hyperparameters in cross-validation focusing on

key parameters such as the dimensions of the prediction and attention layer (n_a and n_d), the type of mask applied when projecting results, the number of successive steps in the network, and the number of independent GLU layer in each GLU block (n_shared). In our model we use Adam optimizer with a learning rate of 0.02 applying weight_decay and a learning rate scheduler as well as patience of 20 (see appendix 8 for a table of all hyperparameters for the final model).

3.3.6 Ensemble Averaging

Prior research has demonstrated that ensemble averaging can improve the performance of the individual models (Di et al. 2020). After the tuned models were generated, we ran experiments with unweighted ensemble model averages to determine the best combination of models. The best model was chosen using cross-validated RMSE, and then this model was evaluated on the test set.

4. Modeling Results and Discussion

Table 1 details the accuracy of each model architecture and the top-performing ensemble model. Of the individual model types, XGBoost performed the best in RMSE and R^2 values. However, when experimenting across equally weighted combinations of XGBoost and other deep learning techniques, ensembles outperformed the individual models. Ensembles of XGBoost and deep learning techniques exceeding other deep learning techniques on tabular data is consistent with previous research (Shwartz-ziv and Armon 2022).

Table 1 Performance Metrics for Different Models

	Spatial Cross Validation R^2		Spatial Cross Validation RMSE	
Model	**Cross Validation R^2**	**Difference from Baseline**	**Cross Validation RMSE**	**Difference from Baseline**
Baseline	0.42	0	3	0
XGBoost	0.59	+0.17	2.72	-0.28
TabNet	0.52	+0.1	2.94	-0.06
1 Layer 1D CNN	0.52	+0.1	2.94	-0.06
Multilayer FFNN	0.54	+0.12	2.89	-0.11
1 Layer FFNN	0.54	+0.12	2.86	-0.14
XGBoost + TabNet	0.60	+0.18	2.70	-0.30
XGBoost + 1 Layer 1D CNN	0.58	+0.16	2.74	-0.26
XGBoost + 2 Layer 1D CNN	0.59	+0.17	2.73	-0.27
XGBoost + 1 Layer FFNN	0.60	+0.18	2.69	-0.31
XGBoost + Multi FFNN	0.59	+0.17	2.70	-0.30
TabNet + 1 Layer 1D CNN	0.55	+0.13	2.84	-0.16
TabNet + 2 Layer 1D CNN	0.55	+0.13	2.83	-0.17
TabNet + 1 Layer FFNN	0.57	+0.15	2.80	-0.20
TabNet + Multi FFNN	0.56	+0.14	2.80	-0.20
XGBoost + TabNet + 1 Layer CNN	0.59	+0.17	2.72	-0.28
XGBoost + TabNet + 2 Layer CNN	0.59	+0.17	2.72	-0.28
XGBoost + TabNet + 1 Layer FFNN	**0.60**	**+0.18**	**2.69**	**-0.31**
XGBoost + TabNet + Multi FFNN	0.60	+0.18	2.70	-0.30

Table 1: Results of all models and ensembles on cross-validation R^2 and RMSE. "Difference from Baseline" is the difference between the metrics for the model and the OLS regression baseline model. The model in bold (XGBoost + TabNet +1 Layer FFNN) is the model that was selected for inference.

Based on performance across the spatial cross-validation, we selected the equally weighted combination of XGBoost, TabNet, and the Simple FFNN as the final model. We evaluated the final model on the held-out stratified test set to understand the generalizability of our model. When applied to the test set, the ensemble model R^2 value of .70 and an RMSE of 2.04.

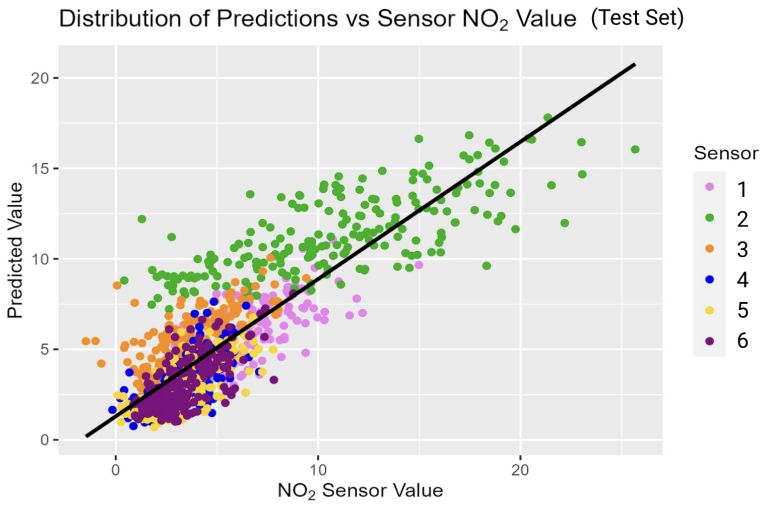


Figure 2: Plot of predicted and actual values for the test set, colored by sensor. The black line represents the 1:1 line, which would indicate points whose estimated values match the predicted values.

To further understand the generalizability of our model, we examined the distribution of the residuals across each fold of the spatial cross-validation and the test set. While the distribution of residuals was consistent across most sensors, there were some sensors with a wider distribution. The sensors with the wider distribution of residuals tended to coincide with extreme data point sensors (very high or low values). The performance of these sensors highlights the need for increased training data in the models and additional ground sensors to calibrate accuracy. We also examined the distribution of residuals across time. Our model performed consistently across time, even accounting for the pandemic year with different ground truth dynamics. A chart of residual distribution in the test set by sensor and time is available in appendix 10.

4.2 Measures of Uncertainty

To measure the uncertainty of our predictions, we utilized a bootstrapping technique with the predictions from the test set by randomly sampling with replacement rows of features from the test set and predicting ground truth value. The residual values were recorded and stored for each pass through the test set. After many iterations, we calculated a 95% confidence interval around the results. Table 2 contains the results of our bootstrap analysis, bucketed in 20th percentile increments of ground truth values to understand better how our model performs with various sensor values (see appendix 11 for graph of bounds). As you can see, the model tends to overpredict for lower sensor values and underpredict for higher values with much wider confidence intervals. These results again highlight the need for additional sensor readings for model improvement.

Table 2 Residual Uncertainty 95% Confidence Intervals

Ground Truth Bucket	Lower Bound	Average Value	Upper Bound
−1.48 to −2.29	−6.859	−0.975	0.740
2.29 to 3.13	−5.927	−0.312	1.525
3.13 to 4.25	−5.013	−0.314	2.073
4.25 to 6.41	−4.824	0.008	2.842
6.41 to 25.7	−4.327	0.889	6.787

Table 2: To calculate the above table, we randomly sample with replacement from the test set and an equivalent number of rows to the test set, predict each datapoint, compare to the ground truth value, store the residual values, and repeat the process 10,000 times. We then separate the results based on 20th percentile buckets of the ground truth values and calculate a 95% confidence interval for each bucket. The column "Ground Truth Bucket" represents the range of each percentile bucket, the "Lower Bound" and "Upper Bound" represent the end of the 95% confidence interval, and the "Average Value" is the mean residual value.

4.3 Experiments

We conducted a number of experiments to determine the potential effect of our data pre-processing decisions. Overall, all experiments had a negligible effect on evaluation metrics. For example, in order to test the impact of inpainting satellite images we trained our models using non-inpainted data achieving very similar metrics. In order to account for the presence of outliers across many features, particularly satellite and meteorological variables, we tested clipping values based on different standard deviations from the mean of each variable (both with inpainted and non-inpainted data). Lastly, model results were also not affected by our decision to stratify sensor cross validation as well as our validation and tests sets by land use.

5. Analysis and Applications

After selecting the best ensemble model, we used this model to predict weekly NO_2 values for the state of Louisiana from January 2019 through October 2023. Using this inference dataset, we conduct a preliminary analysis on spatial and temporal patterns of nitrogen dioxide pollution, as well as an analysis on disparate impacts for various demographic groups. Lastly, we develop a simple, linear weighting algorithm that could be used to determine the placement of future sensors.

For some sections of this analysis, we rely on monthly or annual averages. Monthly averages are calculated by assigning weekly values to each day in the week, then averaging by month. We note that there are three temporal gaps in the weeks that could impact some monthly values: the first 5 days of January 2019 as well as the weeks that were removed due to missing pixels: 2/3/19 and 2/24/19. Annual averages are calculated by averaging the monthly values. 2023 annual averages are not reported since the dataset does not span the full year.

5.1 Spatial and Temporal Patterns in Regional NO_2 Pollution

Figure 3 shows the annual average concentrations of nitrogen dioxide for 2019- 2022 in Louisiana. NO_2 values appeared to be heightened around population dense areas such as New Orleans, Baton Rouge, Lake Charles, and Shrevport.

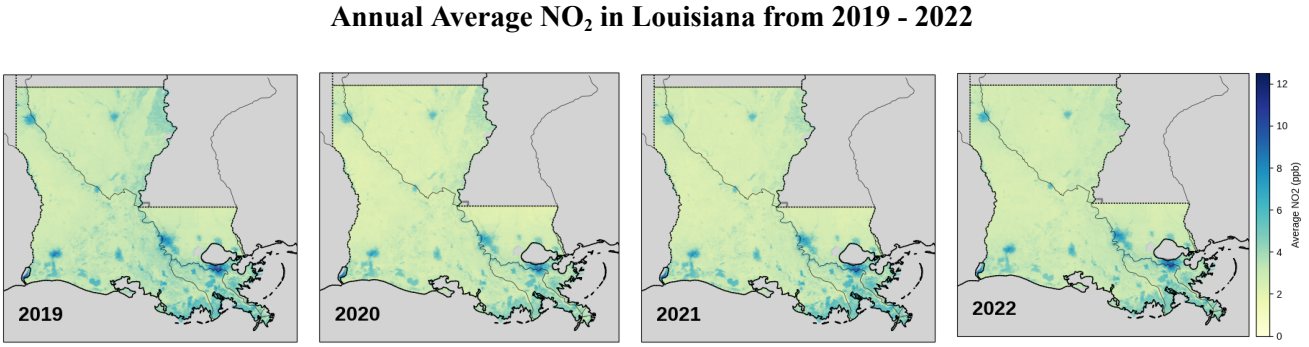


Figure 3: Mean annual NO_2 for Louisiana from 2019-2022. 2023 is not included since the data does not span the entire year. Estimates for annual NO_2 are averaged by month and do not contain data before 1/6/2019 or the weeks of 2/3/19 and 2/24/19.

We find that nitrogen dioxide is generally higher during winter months, as has been widely documented (US EPA 2023; Jarvis et al. 2010). Figure 4 shows the monthly average NO_2 compared to annual average NO_2. The average NO_2 concentration in winter months (November - February) from 2019-2022 was 3.97 ppb, which was 32% than the average concentration in the other months.

Comparing Monthly and Annual NO_2 Averages in Louisiana from 2019-2022

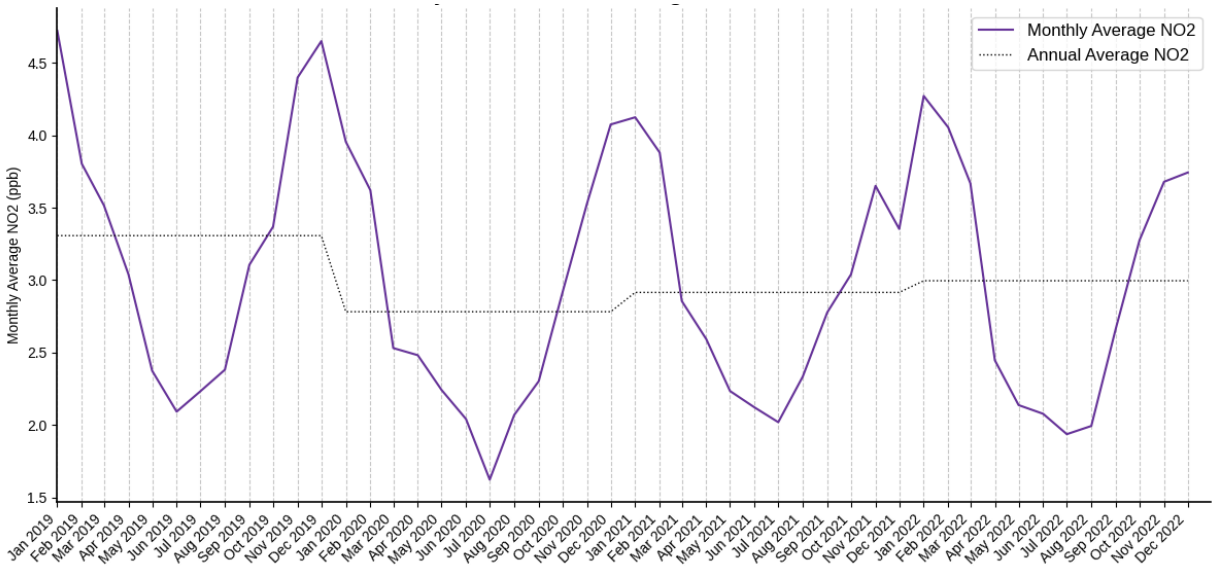


Figure 4: Monthly and annual average NO_2 for Louisiana from 2019-2022. 2023 is not included since the data does not span the entire year. Estimates for annual NO_2 are averaged by month and the graph does not contain data before 1/6/2019 or the weeks of 2/3/19 and 2/24/19.

We also see heightened NO_2 pollution along the southern stretch of the Mississippi river, in the area referred to as "Cancer Alley." We define cancer alley to be 10 miles on either side of the Mississippi river, stretching form a latitude of 29.57 to 31.08. Figure 5 shows monthly NO_2 for this region compared to the state average.

Monthly NO_2 Averages in Cancer Alley compared to Overall Louisiana Average

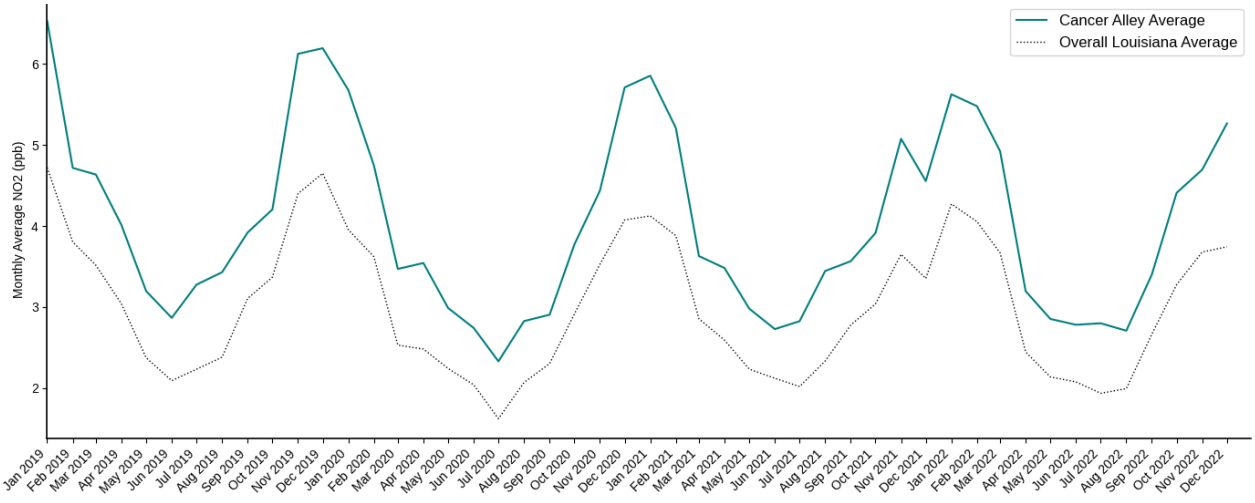


Figure 5: Monthly and annual average NO_2 for Louisiana from 2019-2022. "Cancer Alley" is defined as 10 miles to either side of the Mississippi River from a latitude of 29.57 to 31.08. Overall average is inclusive of this region. 2023 is not included. Estimates for annual NO_2 are averaged by month and the graph does not contain data before 1/6/2019 or the weeks of 2/3/19 and 2/24/19.

We find that NO_2 pollution is, on average, 34% higher in this region compared to the state average. This difference is at least in part due to the inclusion of two major cities, Baton Rouge and New Orleans. Future research is needed to investigate the link between heightened pollution in this area and industrial emissions in the area.

5.2 Analysis of Demographic Disparities in NO_2 Exposure

One of the recommended use cases for statistical models of high-resolution nitrogen dioxide estimates is disparity analysis, which is particularly impactful in Louisiana given the recent EPA investigation into disparate environmental impacts (Kerr et al. 2023; EPA 2023). Prior research comparing emissions between census tracts in LA has found that black communities with industrial activity are exposed to 7 times higher nitrogen dioxide emissions than average (Terrell et al. 2023). We expand on this research by using our high resolution NO_2 pollution dataset to estimate demographic disparities in overall NO_2 pollution exposure estimated at the individual level using census block data.

We utilize demographic data and shapefiles from the 2020 U.S. Census. To analyze disparity, we first calculate pollution at the census block centroid, as has been done in similar research (Kim et al. 2020). Since the majority of census blocks are smaller than our 1 km^2 grid, we use inverse distance weighting of the four nearest gridded estimates to estimate the pollution at our centroid. We note that, because census blocks do not have consistent land area, some blocks can be much larger than our gridded estimates, with the largest block having an area of 340 km^2. We ran our analyses with and without larger blocks and did not find noticeable differences since only 15% of the population live in blocks over 2 km^2, and only 2.5% live in blocks over 10 km^2.

After we calculate pollution by census block, we assign this value to all individuals by demographic group in that block. We then aggregate to the desired geographic level (tract, county, or state) and calculate the overall average pollution exposure as well as the average by demographic group. To calculate disparity, we calculate the percent difference between the average concentration of NO_2 pollution in a demographic and the overall average concentration in the region, or (C_race - C_overall) / C_overall, as has been done in prior research (Chambliss et al. 2021). For analyzing within tract and county disparities, we calculate the percent difference between the exposure of black and white residents in the area, since these groups are the two largest demographic groups and the primary focus of prior research (Terrell et al. 2023). This metric is calculated as (C_black - C_white) / C_overall.

5.2.1 Disparities by Demographic Group

Calculating disparities across the whole state, we find that, on average, white residents are exposed to 8% *less* pollution than average, and black, asian, and hispanic or latinx residents are exposed to 10-20% *more* pollution than average. The table on the below shows the percent disparity broken down by race/ethnicity.

Table 4 Racial Disparities in NO_2 Pollution Exposure

Race/Ethnicity	Percentage Change
White (NH)*	−8%
Other (NH)**	−3%
Black (NH)*	+10%
Hispanic or Latinx	+17%
Asian (NH)*	+20%
Average Exposure: 4.60 ppb	

*Table 3: Disparities by demographic group across the state. The table shows percent disparity, or (C_race - C_average)/C_average using the modeled NO_2 estimates from 1/2019 - 10/2023. *NH means non-hispanic or Latinx. **Other includes american indian or alaskan native, native hawaiian or pacific islander, other, or two or more races*

These disparities have been relatively consistent for the past 5 years. Figure 6 shows the percent disparity by demographic group since 2019. Asian and Hispanic/Latinx populations show some seasonal variation in percent disparity, but overall, there does not appear to be an increase or decrease in disparities over the past 5 years.

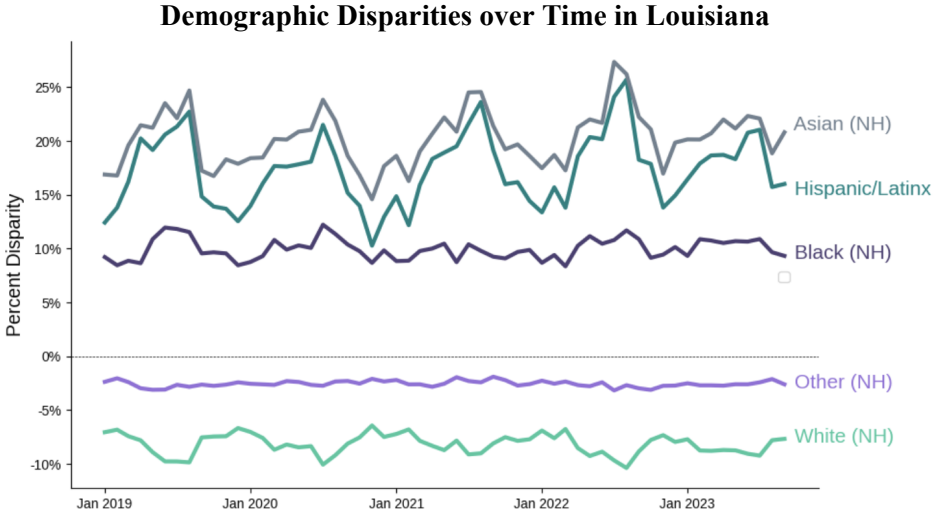


Figure 6: Percent disparity for all demographic groups over time. NH signifies non-hispanic and "Other" includes american indian or alaskan native, native hawaiian or pacific islander, other, or two or more races.

5.2.2 Within-Tract and County Disparities

Prior research has shown that level of spatial aggregation can be critical to accurately estimating disparities (Clark et al. 2022). To determine whether aggregating by tract or county may be impacting existing disparity estimates, we calculate *within-tract* and *within-county* disparities. In order to have one metric that can be compared across all tracts/counties, we calculate black/white disparity, which is the percentage difference in exposure of NO_2 between black and white residents.

Table 4 shows the population-weighted average of within-tract, and within-county disparities. The average black-white disparity within a tract is calculated by averaging the disparity values for all 1,382 tracts, weighing the values by the total population of the tract. The same methodology is applied for the 64 counties. The overall disparity in the state is calculated using the average black and white exposure estimate for the entire state.

Table 3 Black-White Disparities at Different Geographic Levels

Average Black-White Disparity **within a tract**	Average Black-White Disparity **within a county**	Overall Black-White Disparity **within the state**
+1.1%	+7.0%	+19.4%

Table 4: Black-White disparities in NO_2 exposure at different geographic levels. This disparity is calculated as (C_black - C_white)/C_white. Tract/county average is calculated by taking the population-weighted disparity for all tracts or counties. The within state average represents the overall black-white disparity for all residents.

As expected, the larger the geographic area, the higher the internal disparity since this value encompasses disparity both within, and between, the lower level of aggregation. All values are positive, suggesting that even within tracts or counties, black residents have a disparate level of exposure. This table suggests that aggregating to the tract or county level when

analyzing disparities, as has been done in prior research, could lead to underestimating disparate impacts on black residents since it does not include the disparities that exist within those geographic locations. With that said, aggregations by tract would underestimate disparities much less than aggregations by county.

Figure 7 shows the percent black-white disparity within tracts and counties, with purple representing counties where black residents were exposed to higher levels of pollution, and orange representing counties where white residents are exposed to higher levels of pollution. The darker colors represent larger disparities. As shown in the graphs, disparities within counties are much more apparent than disparities within tracts.

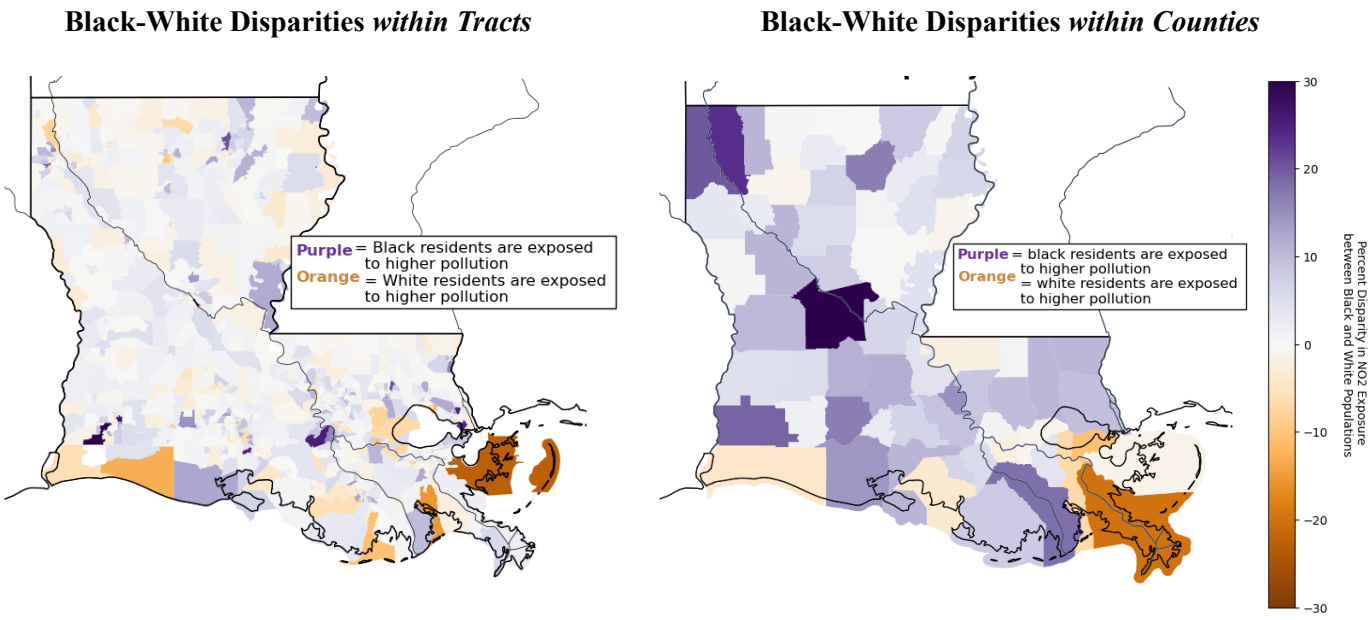


Figure 7: Map of disparities within tracts and census blocks. Counties are shaded by the percent disparity between black and white residents, or (C_black - C_white)/C_white, with purple representing that the black population in an area is exposed to higher levels of pollution, and orange representing that the white population in an area is exposed to higher levels of pollution. White means not disparity exists. Tracts without people have been removed from the map.

Overall, the majority of both tracts and counties have a positive black-white disparity. For tracts, 60% of tracts had positive black-white disparities, compared to 39% with negative black-white disparities, with ~1% of tracts not containing individuals in one or both demographics. For counties, 78% of counties had a positive black-white disparity, and only 22% had negative disparities. Appendix 12 contains a table of the counties with the most positive, and most negative disparities. Appendix 13 contains maps of race and pollution in Rapides and Plaquemines parishes: the parishes with the most positive and most negative black-white disparities. These within tract and county disparities highlight the need to increase the resolution of disparity analyses in Louisiana to the census block level, to ensure measurements are accurately representing the magnitude of disparate impacts.

5.3 Sensor Placement Algorithm

Lastly, we design an interpretable and customizable algorithm using our NO_2 estimates that can be implemented to help inform the intentional placement of future ground-level sensors.

First, we determine key variables that could be considered when determining the placement of NO_2 sensors. While these could be adjusted based on the priorities of a group or organization, we select mean estimated NO_2 to identify overall concentrations, 90th percentile estimated NO_2 to capture peak values, population density to capture people impacted , and Climate Vulnerability Index (CVI) (Tee Lewis et al. 2023) by tract as a metric for “Susceptible and Vulnerable

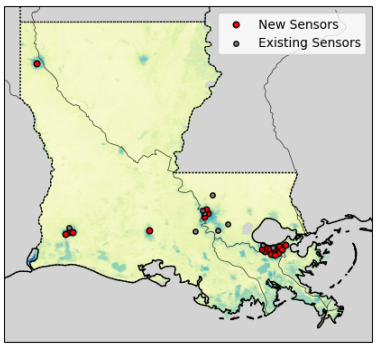
populations," as highlighted in the EPA recommendations (EPA 2023). To ensure equal weighting of variables, we normalize all values to be between 0 and 1. We then calculate a score for each 1km^2 gridded location using the following linear equation, where $\boldsymbol{\beta}_i$ are the different, adjustable, weights, and the covariates are the normalized mean NO_2, 90th percentile NO_2, population density, and CVI:

$$\text{sensor\_placement\_score} = \beta_1 \cdot \text{mean\_NO2} + \beta_2 \cdot \text{90th\_perc\_NO2} + \beta_3 \cdot \text{pop\_dens} + \beta_4 \cdot \text{CVI}$$

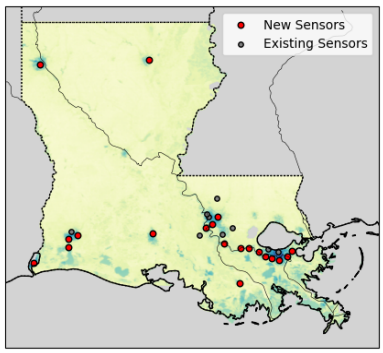
Next, we rank these scores from least to greatest and identify the location with the highest score. The distance between this location and all existing sensors is checked, and if the location is further than the minimum difference from all existing sensors, the sensor is added and the next highest score is checked. If the location is within the minimum distance, the location is removed and the next highest score is checked. The graphs below demonstrate various examples of sensor location recommendations using this algorithm, with existing sensors in gray and new sensors in red. Additional experiments, with coordinates of new sensors, can be found in appendix 14.

Example 1: Changing Minimum Distance
20 new sensors; All weights set to 1; *Minimum distance variable*

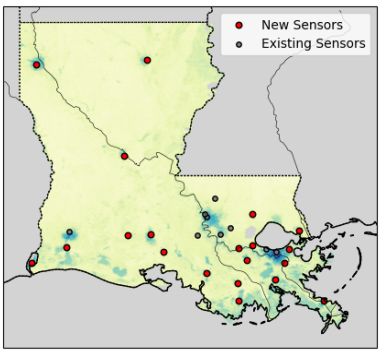
Minimum distance = 5 km



Minimum distance = 10 km

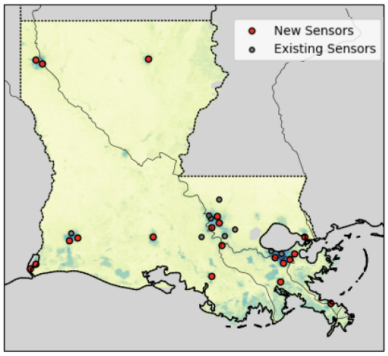


Minimum distance = 20 km

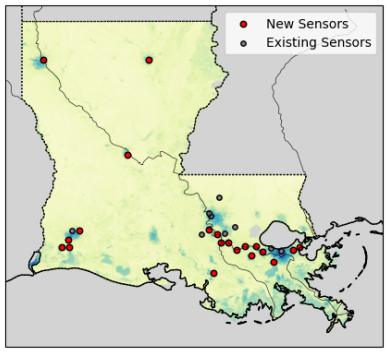


Example 2: Changing Weights
20 new sensors; Minimum Distance of 10 km; *Weights variable*

Mean NO_2=1 | *90th % NO_2 = 0*
Pop Density =1 | *CVI = 0*



Mean NO_2=1 | *90th % NO_2 = 0*
Pop Density = 0 | ***CVI = 1***



Mean NO_2=1 | 90th % NO_2 = 1
Pop Density = 0 | CVI = 0

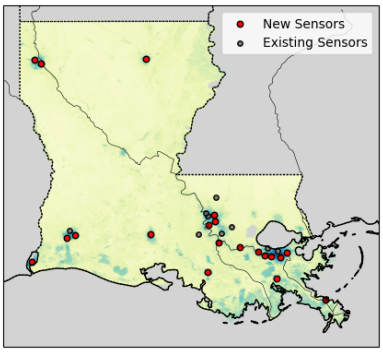


Figure 8: Experiments with the sensor placement algorithm to demonstrate how different parameters or weights can determine the locations of the sensors. Existing sensors are in gray, and new sensor locations determined by the algorithm are in red. The first 3 graphs show the impact of changing the minimum distance between sensors, and the second 3 graphs show the impact of changing the weights. Latitude and longitude coordinates for these values can be found in appendix 13.

The algorithm has an intentional simplicity, to ensure it can be understood, trusted, and adjusted by organizations needing to place the sensors. Additional variables could be added, or removed, to highlight an organization's priorities for sensor placement.

6. Limitations

Machine learning models of air pollution can provide a unique opportunity to take advantage of the ground truth values from ground-level monitoring, the complete coverage of satellite observations, and the impact of meteorological and anthropogenic patterns (Kerr et al. 2023). However, they also have limitations that should be understood when considering applications of the modeled estimates. First, these models are dependent on the quality and amount of sensor data available. Our model struggled to learn values of specific sensors, which was likely due to localized pollution patterns around the sensors. Additionally, given that most sensors are located in urban areas, the estimates may be less accurate in rural areas. Incorporating additional ground-level measurements can help to reduce these issues.

Additionally, "black box" machine learning models can raise concerns due to a lack of interpretability. Methods such as SHAP (Lundberg and Lee 2017) can help users understand how the model is weighing individual variables, yet are still unable to explain exactly how the model is learning.

The limitations of these models highlights the need to be judicious with the application of model-derived estimates. Kerr et al. (2023) suggest potential use cases for different types of pollution estimates, including satellite imagery, statistical, photochemical, and in-situ measurements. Given that our model has a similar accuracy to models using statistical methods (Anenberg et al. 2022), and the incorporation of satellite data, we propose use cases similar to those proposed by Kerr et al. for satellite and statistical models. We suggest that these models can be beneficial for identifying long-term spatial and temporal trends, analyzing disparities, locating hotspots, and identifying locations for future sensors. We recommend caution when using this data to estimate source sector contributions to ambient NO_2, particularly for estimating impacts from industrial emissions in rural areas. Furthermore, the estimates from these models should not be interpreted as ground-truth values for locations, as only ground-level sensors can provide that information.

7. Conclusion

Machine learning provides a unique opportunity for leveraging both high-resolution satellite imagery from Sentinel-5p TROPOMI, as well as in-situ measurements, to provide full coverage ground-level estimates. Given the heightened concern around health risks, and lack of adequate air monitoring, in Louisiana, this methodology can help provide valuable estimates of ground-level pollution that can be used to inform policy and regulatory decisions.

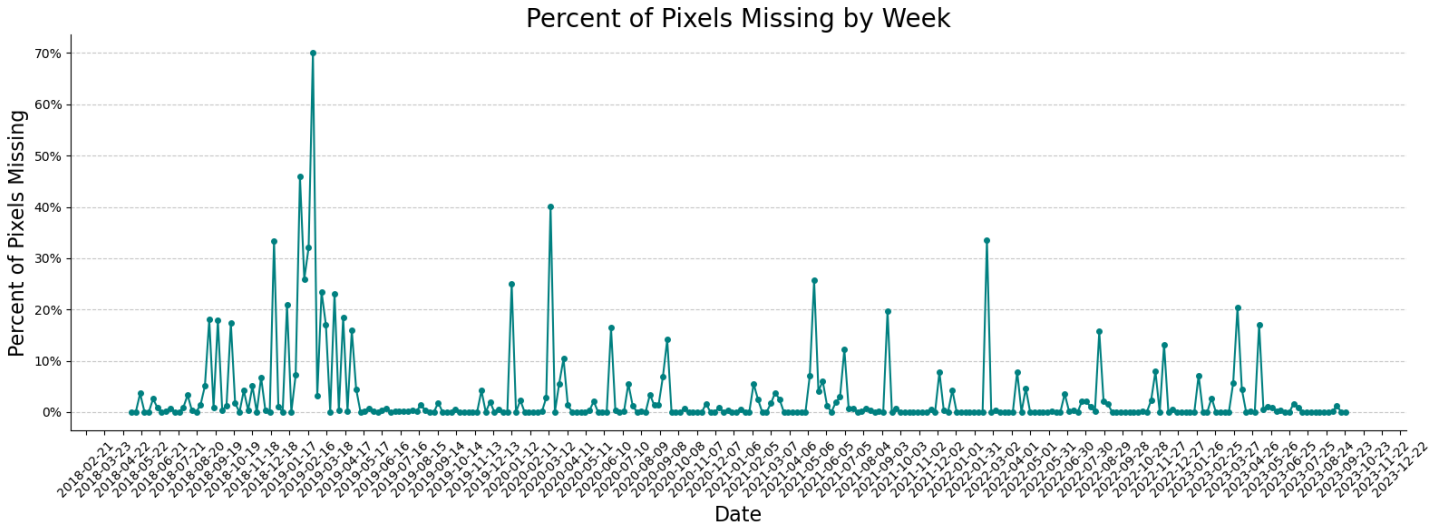
We develop an up-to-date, accessible dataset that can be used by researchers, policymakers, and relevant stakeholders to better understand spatial and temporal patterns of pollution. We then identify disparities, conducting the first regional analysis of demographic disparities in NO_2 exposure at the census block level, showing heightened exposure for black and brown residents. This analysis lays the groundwork for future high resolution regional analyses of pollution disparities, and can be used by policymakers to ensure that policy and regulatory decisions consider communities most impacted by pollution.

We propose a simple, weighted algorithm that could be used for identifying high priority locations for future ground-level monitoring. Given the recent funding for the expansion of monitoring in the state, this algorithm can be used to help ensure that these sensors are intentionally placed.

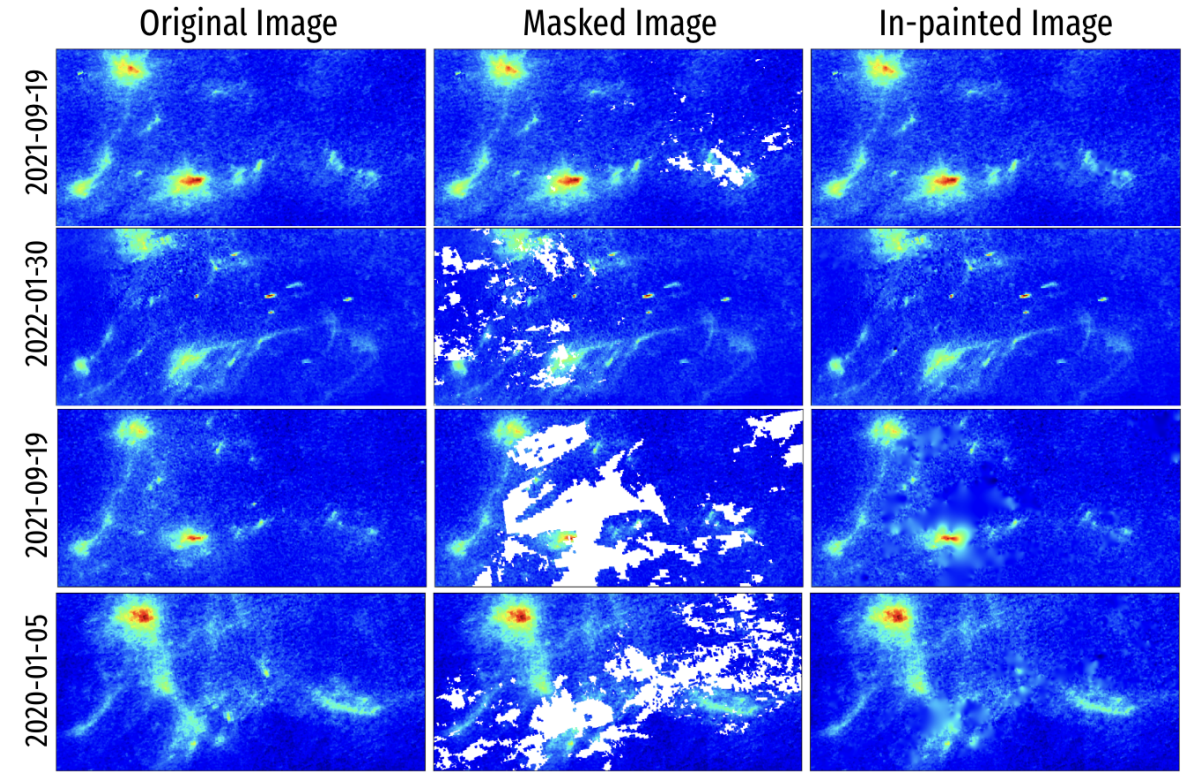
Future research is needed to identify the drivers of disparities in pollution exposure. Additionally, the incorporation of modeled emissions data, as well as additional ground-level measurements as new sensors are placed, could help improve the accuracy of the model.

Appendix

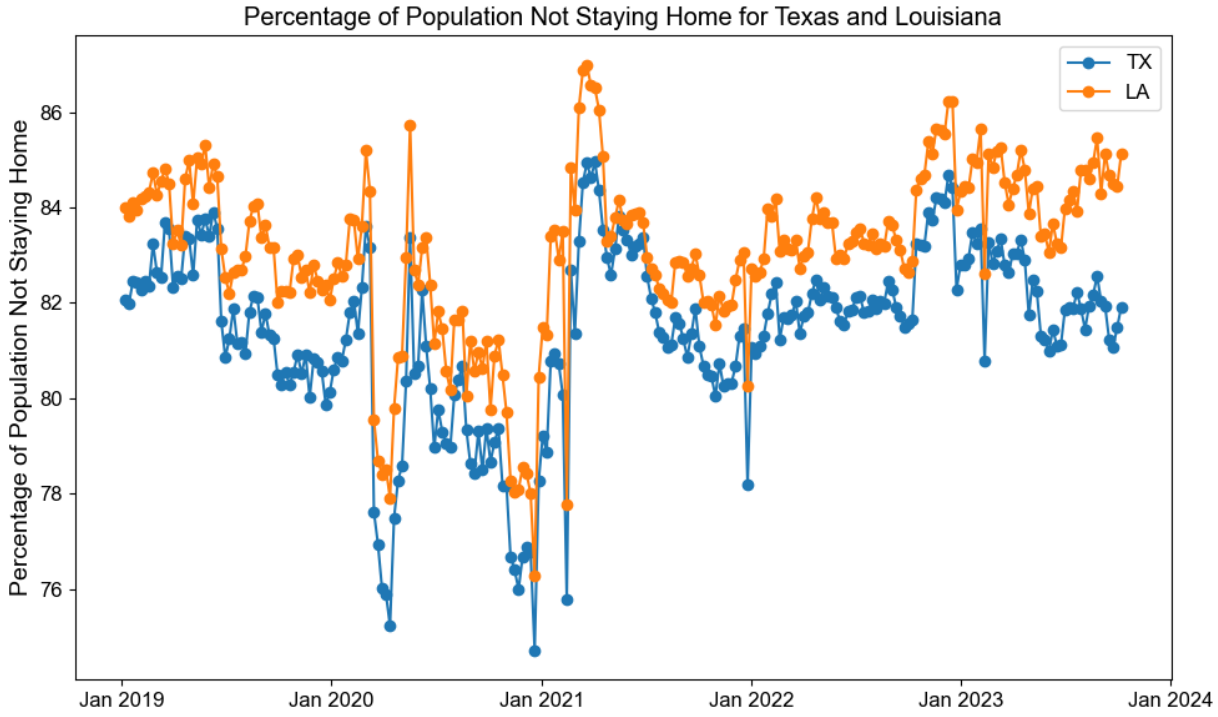
Appendix 1: Missing pixels in weekly satellite images by week



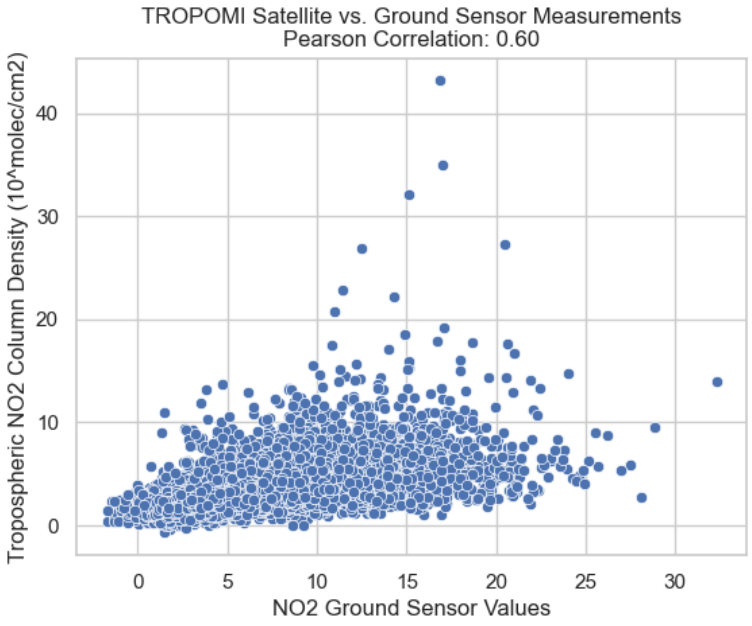
Appendix 2: Examples of masking experiments to determine accuracy of biharmonic in-painting. The first column displays the original image without missing pixels; the second column displays the masked image, masking pixels using a mask from an image with missing pixel values; the third column displays the results of the in-painting.



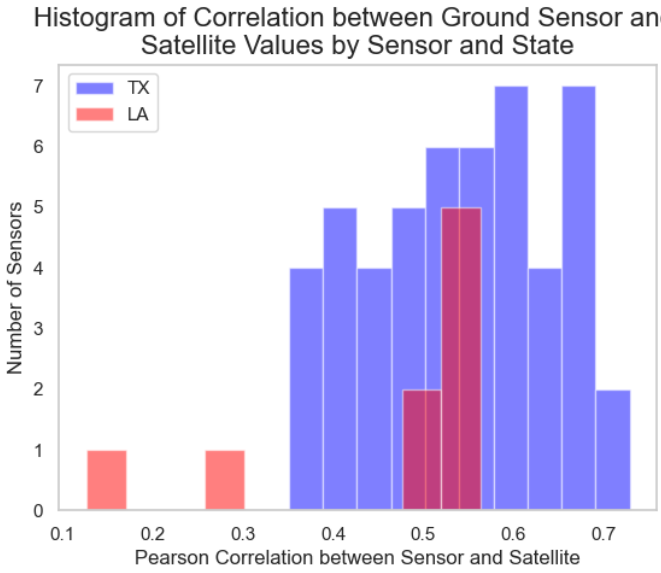
Appendix 3: Mobility data (percentage of population not staying at home) graphed by week for Louisiana and Texas. The graph shows a dip during COVID shutdowns in 2020.



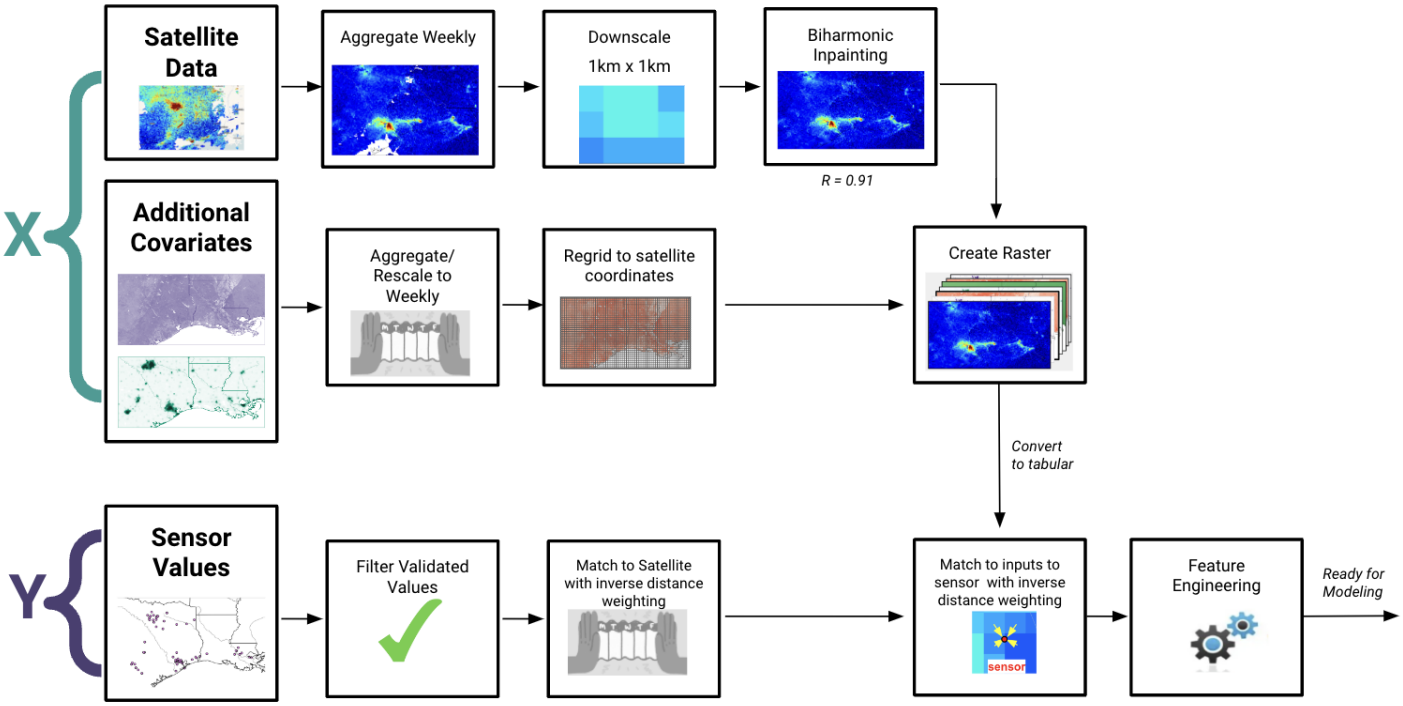
Appendix 4: Correlation between ground sensor values and satellite values for all sensor locations in the modeling set.



Appendix 5: Correlation between ground sensor and satellite values for each individual sensor. The two Louisiana sensors with low correlation were removed from the dataset.



Appendix 6: Overview of data processing pipeline.



Appendix 7: Full results from baseline model (OLS)

	Dependent variable:
	NO_2
Tropospheric N02 Density	0.737*** (0.041)
Temperature	−0.066** (0.026)
Surface Humidity	−77.136* (43.445)
Surface Pressure	0.0003** (0.0001)
Downward Longwave Radiation Flux	0.010*** (0.004)
Fraction of Convective Precipitation	1.658*** (0.573)
Convective Available Potential Energy	0.0001 (0.0001)
Surface Potential Evaporation	1.330 (0.859)
Surface Precipitation Hourly	0.288 (0.221)
Downward Shortwave Radiation Flux	0.001 (0.002)
Impervious Road	−0.009*** (0.002)
Road Density	0.0002*** (0.00001)
Population Density	0.001*** (0.00005)
Elevation	0.001 (0.002)
Enhanced Vegetation Index	−10.707*** (0.526)
Tropospheric N02 Density Local Difference	0.112 (0.115)
Tropospheric Lag 1	−0.571*** (0.085)
Tropospheric N02 Density First Difference	−0.051* (0.030)
Wind Direction	0.001* (0.001)
Wind Magnitude	−0.282*** (0.028)
Mobility	−9.373*** (1.275)
Constant	3.890 (15.294)
Temporal Fixed Effects	Yes
Observations	9,769
R^2	0.538
Adjusted R^2	0.536
Residual Std. Error	2.817 (df = 9732)
F Statistic	314.252*** (df = 36; 9732)
Note:	*p<0.1; **p<0.05; ***p<0.01

Appendix 8: Hyperparameters for Machine Learning Models, selected using Optuna Hyperparameter Search

Feed-Forward Neural Network

Parameters	Description
Input Layer	Input Features: 38
	Activation: None
Hidden Layer 1	Number of Neurons: 96
	Activation: ReLU
	Dropout Rate: 0.42
Output Layer	Number of Neurons: 1
Training Parameters	Learning Rate: 0.000016
	Batch Size: 32
	Number of Epochs: Early Stopping
	Optimizer: Adam
	Loss Function: RMSE

1D CNN

Parameters	Description
Input Layer	Input Features: 38
	Activation: None
1D Convolution	Filters: 72
	Kernel Size: 7
	Dropout Rate: 0.027192
MaxPooling1D	Pool Size: 2
Hidden Layer 1	Number of Neurons: 32
	Activation: ReLU
Output Layer	Number of Neurons: 1
Training Parameters	Learning Rate: 0.000035
	Batch Size: 32
	Number of Epochs: Early Stopping
	Optimizer: RMSprop
	Loss Function: RMSE

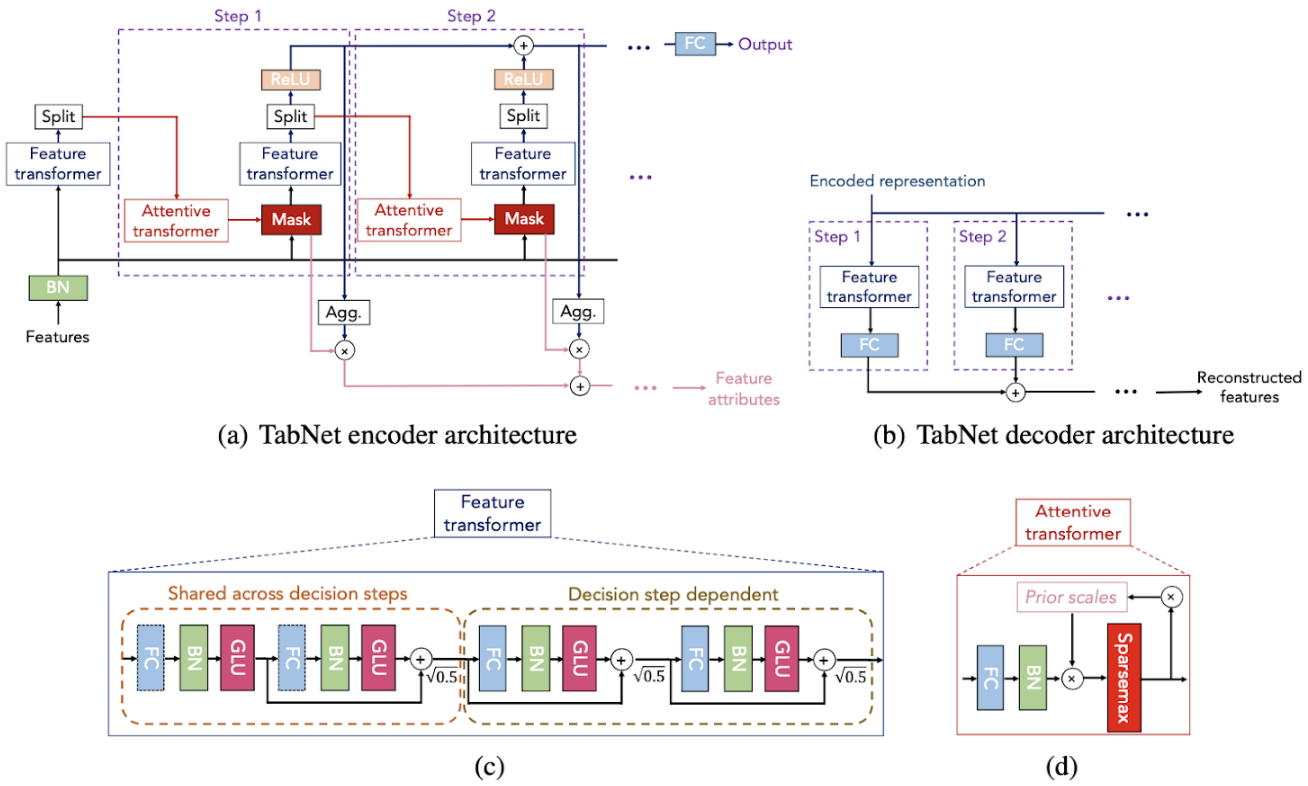
XGBoost

Parameters	Description
Main parameters	n_estimators: 992, max_depth: 8, learning_rate: 0.14
Sampling	subsample: 0.87, colsample_bytree: 0.71
Regularization	gamma: 0.56, reg_alpha: 0.69, reg_lambda: 4
Prevent overfitting	min_child_weight: 3, early_stopping: 20

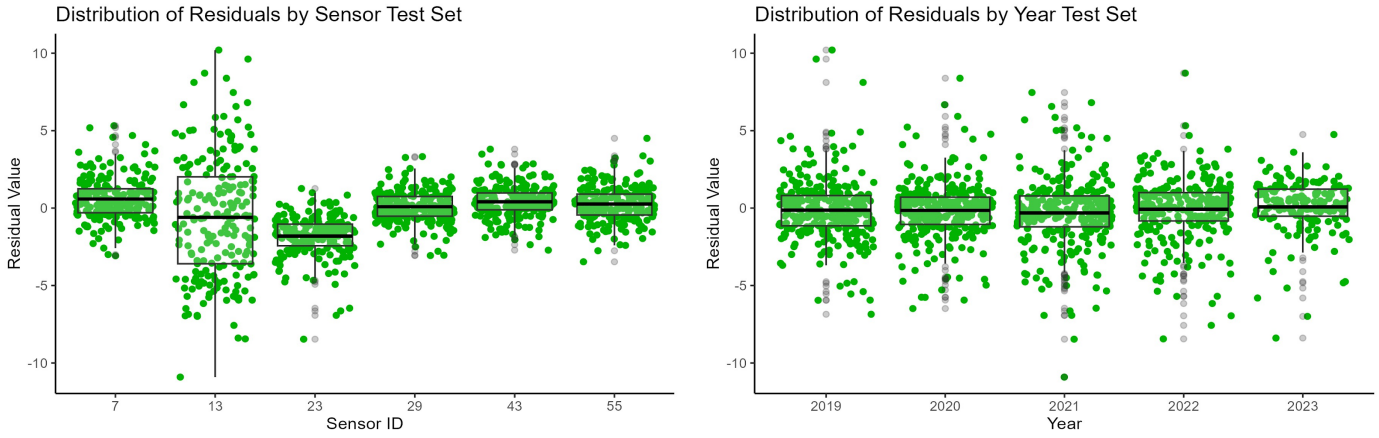
TabNet

Parameter	Description
mask_type	entmax
epochs	641
early_stopping	20
patienceScheduler	7
n_d & n_a	56
n_steps	3
n_shared	2
gamma	1
lambda_sparse	5.19E-05

Appendix 9: Diagram of TabNet model architecture from the original paper
(Source: https://ojs.aaai.org/index.php/AAAI/article/view/16826)



Appendix 10: Distribution of residuals in the test set separated by sensor (left) and year (right).



Appendix 11: Graph of predicted vs. actual values, highlighting the upper and lower bounds of the confidence interval.

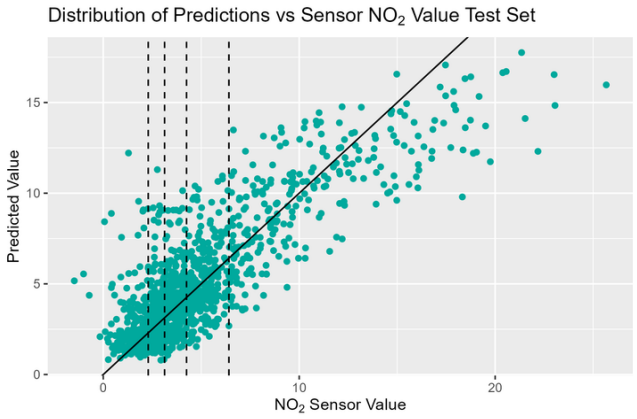


Table 2 Residual Uncertainty 95% Confidence Intervals

Ground Truth Bucket	Lower Bound	Average Value	Upper Bound
−1.48 to −2.29	−6.859	−0.975	0.740
2.29 to 3.13	−5.927	−0.312	1.525
3.13 to 4.25	−5.013	−0.314	2.073
4.25 to 6.41	−4.824	0.008	2.842
6.41 to 25.7	−4.327	0.889	6.787

Appendix 12: Exposure and disparity values for black and white residents in the counties with the 5 most positive, and 5 most negative, black-white disparity percentages.

Counties with *Most Positive* Black-White Disparity

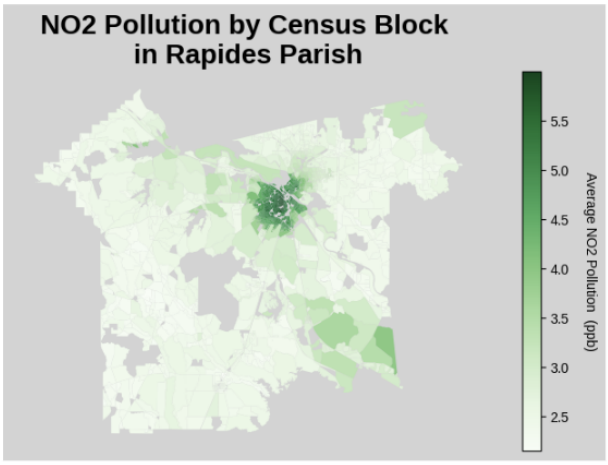
County	Average Exposure for Black Residents (ppb)	Average Exposure for White Residents (ppb)	% Black-White Disparity
Rapides	4.39	3.24	35.7%
Bossier	4.48	3.63	23.3%
Caddo	5.34	4.46	20.0%
Calcasieu	5.38	4.52	19.0%
Lafourche	3.66	3.10	18.1%

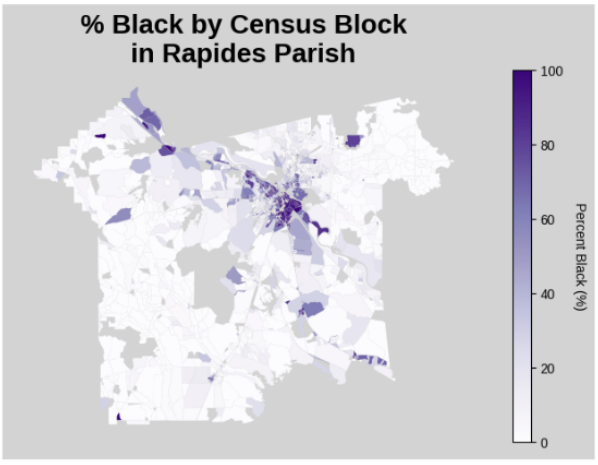
Counties with *Most Negative* Black-White Disparity

County	Average Exposure for Black Residents (ppb)	Average Exposure for White Residents (ppb)	% Black-White Disparity
Plaquemines	3.69	4.62	-20.6%
Orleans	7.69	8.60	-10.6%
Jefferson	6.89	7.40	-6.9%
Cameron	2.66	2.77	-4.2%
St. Mary	3.83	3.98	-3.7%

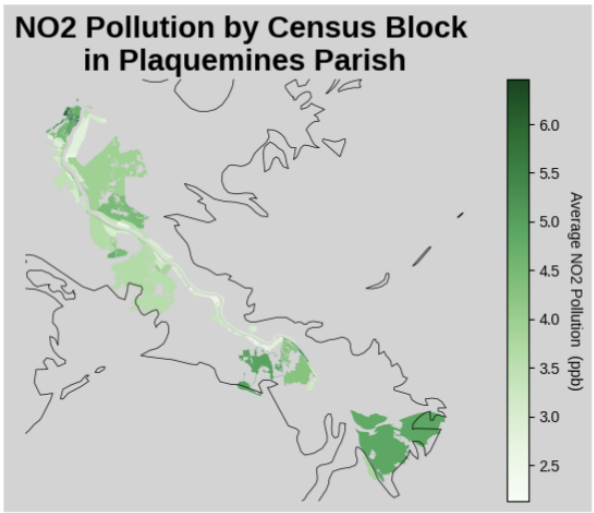
Appendix 13: Race and Pollution in Rapides and Plaquemines Parishes, the parishes with the most positive (Rapides) and most negative (Plaquemines) Black-White disparities. The maps on the left shows the estimated average NO2 pollution at the census block centroid from 1/2019 - 10/2023. The maps on the right shows the percentage of black residents by census block. In Rapides parish, pollution appears to be centered in areas with a high percentage of black residents, where as in Plaquemines, pollution is higher in areas with lower percentages of black residents. Areas without population have been removed from the map. Note that pollution scales are not consistent between maps to highlight differences within the county, rather than between counties.

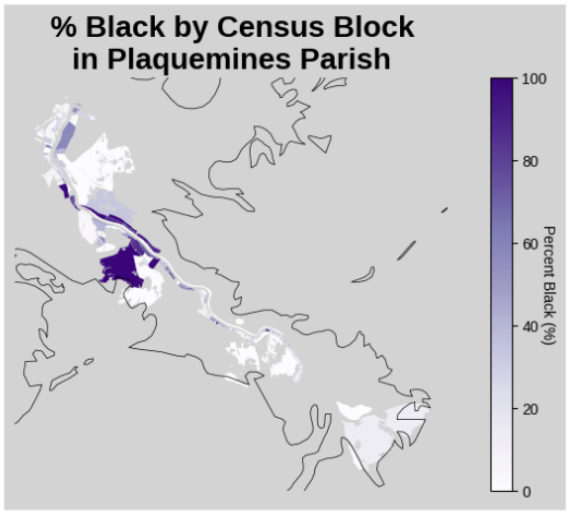
Race and Pollution in *Rapides Parish*





Race and Pollution in *Plaquemines Parish*



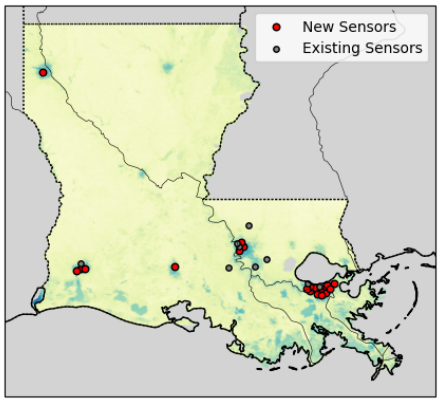


Appendix 14: Sensor placement experiments with associate latitude and longitude values for new sensors.

Example 1: Changing Minimum Distance

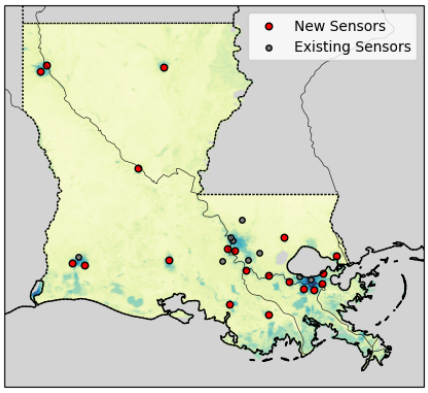
20 new sensors; All weights set to 1; *Minimum distance variable*

Minimum distance = 5 km



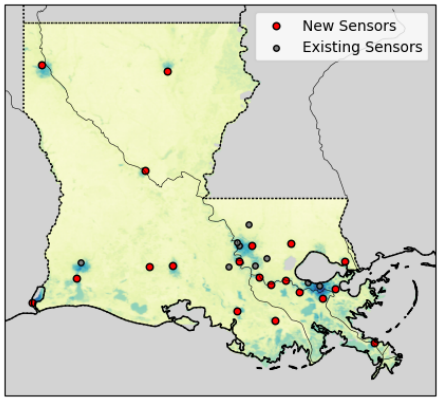
[(30.195, -92.065),
(29.865, -90.085),
(30.405, -91.035),
(32.485, -93.735),
(30.345, -91.115),
(32.395, -93.805),
(29.965, -90.375),
(29.945, -89.955),
(30.295, -89.755),
(29.625, -90.745),
(31.305, -92.465),
(30.045, -89.975),
(30.495, -91.045),
(30.195, -93.195),
(30.585, -91.145),
(30.085, -90.475),
(30.215, -92.385),
(30.145, -91.965),
(32.515, -92.155),
(29.695, -91.205)]

Minimum distance = 10 km



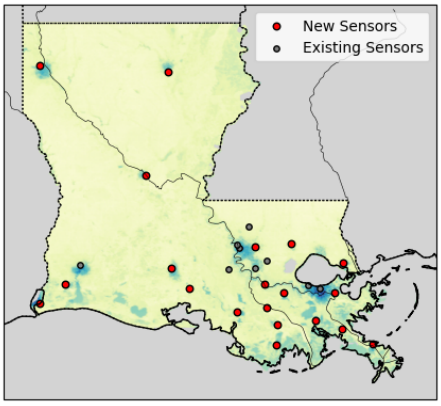
(32.445, -93.805),
(29.955, -89.955),
(29.875, -90.065),
(29.595, -90.695),
(30.045, -90.695),
(32.525, -93.725),
(32.495, -92.125),
(29.895, -90.215),
(30.335, -91.155),
(30.175, -93.205),
(30.105, -90.995),
(29.975, -90.405),
(30.235, -92.045),
(30.275, -89.765),
(31.315, -92.465),
(30.075, -89.945),
(30.495, -90.475),
(30.195, -93.375),
(30.365, -91.255),
(29.705, -91.215)

Minimum distance = 15 km



(30.015, -90.765),
(30.065, -90.555),
(32.535, -93.805),
(29.965, -89.905),
(29.855, -90.065),
(32.455, -92.135),
(31.325, -92.425),
(29.925, -90.375),
(30.285, -91.175),
(29.705, -91.205),
(30.275, -89.775),
(30.485, -90.485),
(30.085, -93.345),
(29.605, -90.705),
(30.465, -91.015),
(30.235, -92.065),
(30.215, -92.375),
(29.805, -93.925),
(29.345, -89.385),
(30.095, -90.915)

Minimum distance = 20 km

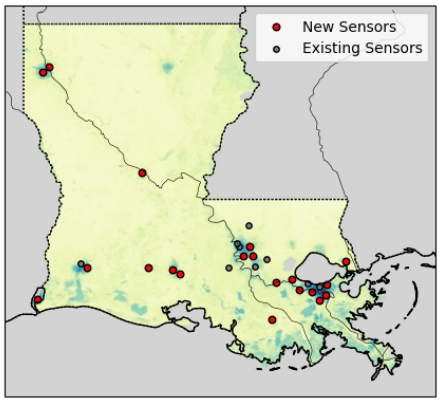


(32.535, -93.815),
(30.295, -89.805),
(29.735, -91.205),
(31.285, -92.415),
(29.955, -90.585),
(30.505, -90.485),
(32.455, -92.125),
(29.835, -93.815),
(29.595, -90.675),
(30.055, -90.845),
(29.375, -89.415),
(30.055, -93.475),
(30.235, -92.075),
(29.955, -89.915),
(30.005, -91.835),
(29.355, -90.685),
(30.475, -90.965),
(29.545, -89.815),
(29.635, -90.165),
(29.785, -90.815)

Example 2: Changing Weights

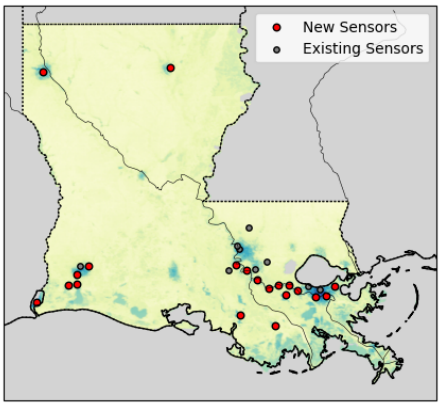
20 new sensors; Minimum Distance of 10 km; *Weights variable*

Mean NO_2=1 | 90th % NO_2 = 0
Pop Density =1 | CVI = 0



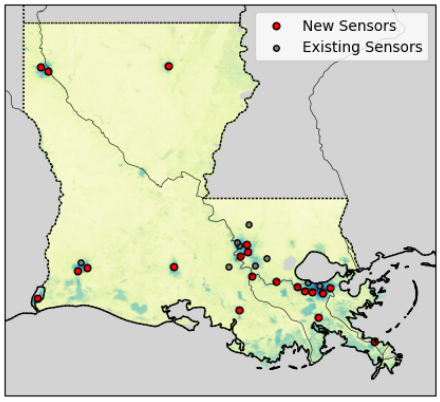
(29.905, -90.035),
(29.945, -90.205),
(30.195, -92.065),
(30.465, -91.045),
(30.025, -90.015),
(30.355, -91.125),
(29.845, -90.115),
(32.455, -93.785),
(32.515, -93.705),
(29.965, -90.375),
(30.215, -93.195),
(29.855, -93.865),
(30.355, -90.995),
(30.295, -89.755),
(29.625, -90.745),
(31.305, -92.465),
(30.045, -90.685),
(30.215, -92.385),
(30.085, -90.475),
(30.145, -91.965)

Mean NO_2=1 | 90th % NO_2 = 0
Pop Density = 0 | **CVI = 1**



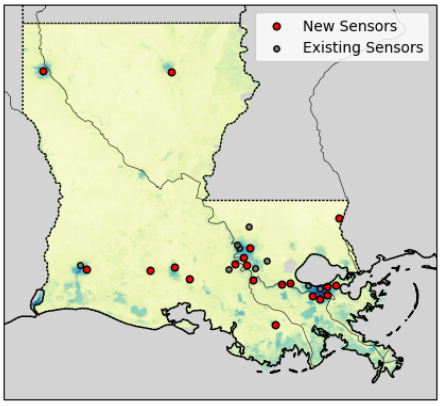
(30.035, -89.915),
(29.925, -90.025),
(29.915, -90.165),
(30.055, -90.655),
(29.995, -90.405),
(30.055, -90.515),
(30.175, -93.325),
(29.935, -90.565),
(30.285, -91.215),
(32.475, -93.775),
(30.115, -90.945),
(30.265, -93.175),
(29.855, -93.865),
(29.715, -91.165),
(30.065, -93.325),
(30.215, -91.075),
(32.525, -92.085),
(30.055, -93.445),
(30.015, -90.785),
(29.595, -90.705)

Mean NO_2=1 | 90th % NO_2 = 1
Pop Density = 0 | CVI = 0



(29.925, -90.205),
(29.915, -90.065),
(29.855, -93.865),
(29.975, -89.975),
(32.505, -93.815),
(29.635, -90.125),
(30.475, -91.075),
(30.345, -91.165),
(30.215, -92.045),
(30.175, -93.325),
(30.045, -90.685),
(30.205, -93.205),
(29.995, -90.405),
(29.355, -89.385),
(29.945, -90.315),
(30.115, -91.005),
(30.385, -91.065),
(32.455, -93.715),
(32.525, -92.115),
(29.725, -91.175)

Mean NO_2=0 | 90th % NO_2 = 1
Pop Density = 1 | **CVI = 1**



(29.925, -90.015),
(29.885, -90.115),
(30.025, -90.015),
(30.245, -92.035),
(30.065, -90.505),
(30.035, -89.905),
(30.095, -90.995),
(30.215, -93.195),
(30.205, -92.365),
(29.595, -90.705),
(30.465, -91.045),
(32.475, -93.775),
(29.915, -90.215),
(30.355, -91.125),
(30.285, -91.235),
(30.055, -90.615),
(30.115, -91.835),
(30.795, -89.855),
(32.455, -92.075),
(30.265, -91.085)

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv:1907.10902 [cs, stat]*.

Susan C. Anenberg, Arash Mohegh, Daniel L. Goldberg, Gaige H Kerr, and Michael Brauer. 2022. Long-term trends in urban NO2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *The Lancet Planetary Health*, 6(1).

Sercan O. Arik and Tomas Pfister. 2020. TabNet: Attentive Interpretable Tabular Learning. *arXiv:1908.07442 [cs, stat]*.

Tristan Baurick, Lylla Younes, and Joan Meiners. 2019. Welcome to "Cancer Alley," Where Toxic Air Is About to Get Worse.

Jesse D. Berman and Keita Ebisu. 2020. Changes in U.S. air pollution during the COVID-19 pandemic. *Science of The Total Environment*, 739:139864.

Bureau of Transportation Statistics. 2021. Trips by Distance - Daily Average by Week | Tyler Data & Insights.

Sarah E. Chambliss, Carlos P. R. Pinon, Kyle P. Messier, Brian LaFranchi, Crystal Romeo Upperman, Melissa M. Lunden, Allen L. Robinson, Julian D. Marshall, and Joshua S. Apte. 2021. Local- and regional-scale racial and ethnic disparities in air pollution determined by long-term mobile monitoring. *Proceedings of the National Academy of Sciences*, 118(37).

Ka Lok Chan, Ehsan Khorsandi, Song Liu, Frank Baier, and Pieter Valks. 2021. Estimation of Surface NO2 Concentrations over Germany from TROPOMI Satellite Observations Using a Machine Learning Method. *Remote Sensing*, 13(5):969.

Taher Chegini, Hong-Yi Li, and L. Ruby Leung. 2021. HyRiver: Hydroclimate Data Retriever.

Zhao-Yue Chen, Rong Zhang, Tian-Hao Zhang, Chun-Quan Ou, and Yuming Guo. 2019. A kriging-calibrated machine learning method for estimating daily ground-level NO2 in mainland China. *Science of The Total Environment*, 690:556–564.

Yulei Chi, Meng Fan, Chuanfeng Zhao, Yikun Yang, Hao Fan, Xingchuan Yang, Jie Yang, and Jinhua Tao. 2022. Machine learning-based estimation of ground-level NO2 concentrations over China. *Science of The Total Environment*, 807:150721.

Lara P. Clark, Maria H. Harris, Joshua S. Apte, and Julian D. Marshall. 2022. National and Intraurban Air Pollution Exposure Disparity Estimates in the United States: Impact of Data-Aggregation Spatial Scale. *Environmental Science & Technology Letters*, 9(9):786–791.

Mark E Cooper, R M Martin, Chris A McLinden, and Jeffrey R Brook. 2020. Inferring ground-level nitrogen dioxide concentrations at fine spatial resolution applied to the TROPOMI satellite instrument. *Environmental Research Letters*, 15(10):104013–104013.

Copernicus Sentinel-5p (processed by ESA). 2021. TROPOMI Level 2 Nitrogen Dioxide total column products. *European Space Agency*, Version 2.

Qian Di, Heresh Amini, Liuhua Shi, Itai Kloog, Rachel Silvern, James Kelly, M. Benjamin Sabath, Christine Choirat, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, Loretta J. Mickley, and Joel Schwartz. 2019. An

ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment International*, 130:104909.

Henk Eskes, Jos Van Geffen, Folkert Boersma, Kai-Uwe Eichmann, Arnoud Apituley, Mattia Pedergnana, Maarten Sneep, J Veefkind, and Diego Loyola. 2022. Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Nitrogendioxide document number : S5P-KNMI-L2-0021-MA.

Gaige Hunter Kerr, Daniel L Goldberg, Maria H Harris, Barron H Henderson, Perry Hystad, Ananya Roy, and Susan C Anenberg. 2023. Ethnoracial Disparities in Nitrogen Dioxide Pollution in the United States: Comparing Data Sets from Satellites, Models, and Monitors. *Environmental Science & Technology*, 57(48):19532–19544.

Masoud Ghahremanloo, Yannic Lops, Yunsoo Choi, and Bijan Yeganeh. 2021. Deep Learning Estimation of Daily Ground-Level NO 2 Concentrations From Remote Sensing Data. *Journal of Geophysical Research: Atmospheres*, 126(21).

Debora Griffin, Xiaoyi Zhao, Chris A. McLinden, Folkert Boersma, Adam Bourassa, Enrico Dammers, Doug Degenstein, Henk Eskes, Lukas Fehr, Vitali Fioletov, Katherine Hayden, Shailesh K. Kharol, Shao-Meng Li, Paul Makar, Randall V. Martin, Cristian Mihele, Richard L. Mittermeier, Nickolay Krotkov, Maarten Sneep, et al. 2019. High-Resolution Mapping of Nitrogen Dioxide With TROPOMI: First Results and Validation Over the Canadian Oil Sands. *Geophysical Research Letters*, 46(2):1049–1060.

Maria Grundström and Håkan Pleijel. 2014. Limited effect of urban tree vegetation on NO2 and O3 concentrations near a traffic route. *Environmental Pollution*, 189:73–76.

Patryk Tadeusz Grzybowski, Krzysztof Mirosław Markowicz, and Jan Paweł Musiał. 2023. Estimations of the Ground-Level NO2 Concentrations Based on the Sentinel-5P NO2 Tropospheric Column Number Density Product. *Remote Sensing*, 15(2):378.

X Hu, Lance A Waller, Alexei Lyapustin, Yuesi Wang, and Yang Liu. 2014. 10-year Spatial and Temporal Trends of PM2.5 Concentrations in the Southeastern US Estimated Using high-resolution Satellite Data. *Atmos. Chem. Phys.*, 14(12):6301–6314.

Manjunath R Hudagi, Shridevi Soma, and Rajkumar L Biradar. 2022. Performance Evaluation of Biharmonic Function-Based Image Inpainting Approach. *Smart innovation, systems and technologies*:475–483.

A Huete, C. Justice, and W. van Leeuwen. 2016. Huete, A., Justice, C. and van Leeuwen, W. (1999) MODIS Vegetation Index (MOD13) Algorithm Theoretical Basis Document, Version 3. - References - Scientific Research Publishing.

Debbie J Jarvis, Gary Adamkiewicz, Marie-Eve Heroux, Regula Rapp, and Frank J Kelly. 2010. Nitrogen Dioxide: WHO Guidelines for Indoor Air Quality: Selected Pollutants.

Gabe Joseph. 2023. StackSTAC.

Joshua L Keller, Casey Olives, Lianne Sheppard, Paul D Sampson, Adam A Szpiro, Assaf P Oron, Johan Lindström, Sverre Vedal, Joel D Kaufman, and Sun Yeou Kim. 2014. A Unified Spatiotemporal Modeling Approach for Predicting Concentrations of Multiple Air Pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *Environmental Health Perspectives*, 123(4):301–309.

Sun-Young Kim, Matthew Bechle, Steve Hankey, Lianne Sheppard, Adam A. Szpiro, and Julian D. Marshall. 2020. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLOS ONE*, 15(2):e0228535.

Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 2021. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398.

Ava Kofman. 2022. EPA takes action to combat industrial air pollution.

P.F. Levelt, G.H.J. van den Oord, M.R. Dobber, A. Malkki, Huib Visser, Johan de Vries, P. Stammes, J.O.V. Lundell, and H. Saari. 2006. The ozone monitoring instrument | IEEE Journals & Magazine | IEEE Xplore.

Louisiana Department of Environmental Quality. Current Monitoring Data & AQI.

Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions.

G. S. Meena and D. B. Jadhav. 2007. Study of diurnal and seasonal variation of atmospheric NO2, O3, H2O and O4 at Pune, India. *Atmósfera*, 20(3):271–287.

Johan R Meijer, Mark A J Huijbregts, Kees C G J Schotten, and Aafke M Schipper. 2018b. Global patterns of current and future road infrastructure. *Environmental Research Letters*, 13(6):064006.

Multi-Resolution Land Characteristics Consortium. NLCD 2019 Percent Developed Imperviousness (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium.

Microsoft Open Source, Matt McFarland, Rob Emanuele, Dan Morris, and Tom Augspurger. 2022. microsoft/PlanetaryComputer: October 2022. *Zenodo*, 2022.10.28.

Multi-Resolution Land Characteristics Consortium. NLCD 2019 Percent Developed Imperviousness (CONUS) | Multi-Resolution Land Characteristics (MRLC) Consortium.

Ruhan Nagra, Robert Taylor, Mary Hampton, and Lance Hilderbrand. 2021. “Waiting to Die”: Toxic Emissions and Disease Near the Denka Performance Elastomer Neoprene Facility in Louisiana’s Cancer Alley. *Environmental Justice*, 14(1):14–32.

Sander Niemeijer. 2017. ESA Atmospheric Toolbox. *NASA ADS*:8286.

Daniel Niepsch, Leon J. Clarke, Konstantinos Tzoulas, and Gina Cavan. 2021. Spatiotemporal variability of nitrogen dioxide (NO2) pollution in Manchester (UK) city centre (2017–2018) using a fine spatial scale single-NOx diffusion tube network. *Environmental Geochemistry and Health*.

Halle Parker. 2023. Shuttered EPA investigation could’ve brought “meaningful reform” in Cancer Alley, documents reveal.

Fabian Pedregosa, Fabian Pedregosa, Gael Varoquaux, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Dawn Roberts–Semple, Fei Song, and Yuan Gao. 2012. Seasonal characteristics of ambient nitrogen oxides and ground–level ozone in metropolitan northeastern New Jersey. *Atmospheric Pollution Research*, 3(2):247–257.

Mikhail Sarafanov, Eduard Kazakov, Nikolay O. Nikitin, and Anna V. Kalyuzhnaya. 2020. A Machine Learning Approach for Remote Sensing Data Gap-Filling with Open-Source Implementation: An Example Regarding Land Surface Temperature, Surface Albedo and NDVI. *Remote Sensing*, 12(23):3865.

Linus Scheibenreif, Michael Mommert, and Damian Borth. 2022. Toward Global Estimation of Ground-Level NO2 Pollution With Deep Learning and Remote Sensing | IEEE Journals & Magazine | IEEE Xplore.

Marcello Schiavina, Sergio Freire, Alessandra Carioli, and Kytt MacManus. 2023. GHS-POP R2023A - GHS population grid multitemporal (1975-2030). *data.jrc.ec.europa.eu*.

Sentinel Hub. Sentinel-5P L2.

Ravid Shwartz-Ziv and Amitai Armon. 2022. Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90.

Steffen Beirle, U Platt, Mark Wenig, and Thomas Wagner. 2003. Weekly Cycle of NO2 by GOME measurements: a Signature of Anthropogenic Sources. *Atmospheric Chemistry and Physics*, 3(6):2225–2232.

Claire Sullivan. 2023. Lawmakers reject plea from Cancer Alley residents for fence line air monitoring.

P. Grace Tee Lewis, Weihsueh A. Chiu, Ellu Nasser, Jeremy Proville, Aurora Barone, Cloelle Danforth, Bumsik Kim, Jolanda Prozzi, and Elena Craft. 2023. Characterizing vulnerabilities to climate change across the United States. *Environment International*, 172:107772.

Kimberly A Terrell and Gianna St Julien. 2022. Air pollution is linked to higher cancer rates among black or impoverished communities in Louisiana. *Environmental Research Letters*, 17(1):014033.

University Network for Human Rights. 2019. “WAITING TO DIE” Toxic Emissions and Disease Near the Louisiana Denka / DuPont Plant.

US Census Bureau. 2021. TIGER/Line Shapefiles.

US EPA. 2013. Air Quality System (AQS).

US EPA. 2022. EPA Announces Louisiana to Receive Nearly $2.4 Million for Air Monitoring Projects.

US EPA. 2023. Overview of Nitrogen Dioxide (NO2) Air Quality in the United States.

J. L. van Hemmen, D. Grensing, A. Huber, and R. Kühn. 1988. Nonlinear neural networks. I. General theory. *Journal of Statistical Physics*, 50(1-2):231–257.

Yifan Wang, Pengfei Liu, Joel Schwartz, Edgar Castro, Wenhao Wang, Howard H Chang, Noah Scovronick, and Liuhua Shi. 2023. Disparities in ambient nitrogen dioxide pollution in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 120(16).

Yuan Wang, Qiangqiang Yuan, Tongwen Li, Siyu Tan, and Liangpei Zhang. 2021. Full-coverage spatiotemporal mapping of ambient PM2.5 and PM10 over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions. *Science of The Total Environment*, 793:148535.

Manzhu Yu and Qian Liu. 2021. Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations. *Science of The Total Environment*, 773:145145.