Estimation of Surface-Level Carbon Monoxide (CO) Concentrations Based on Satellite CO mass with Meteorological and Ancillary Variables over the Netherlands

ARSEN ORDOKOV, University of Twente, The Netherlands

It is crucial to monitor CO concentrations due to harmful effects on human lives and their indirect role in climate change, especially in the case of deactivating air quality monitoring stations across the Netherlands. Moreover, there is a lack of frameworks designed to estimate surface-level CO concentrations in the Netherlands. To address these issues, this study aims to develop and evaluate linear and machine learning models for estimating surface CO concentrations in the Netherlands, utilizing Sentinel-5P satellite data in conjunction with meteorological and ancillary variables. The observation included daily satellite-based observations from 19 locations and was derived on a fixed timescale from January 1, 2022, to December 31, 2023. The validation results demonstrate an R^2 of 0.53 for random forest with location-based cross-validation. Other approaches yielded R^2 values of 0.44, 0.30, 0.20, and 0.12 for extreme gradient boosting, support vector machine, multilinear regression, and linear regression models, respectively. Furthermore, the feature analysis highlighted the considerable importance of meteorological variables for all models that included all variables. Ancillary variables had minimal influence, suggesting a need for further methodological improvement.

Additional Key Words and Phrases: Sentinel-5P, carbon monoxide, meteorological variable, ancillary variable, machine learning, variable importance

1 INTRODUCTION

Carbon monoxide (CO) was chosen for this study due to a research gap in estimating CO at the surface level using machine learning approaches. According to Raub et al. [37], even low concentrations of CO in the air can cause cardiovascular and neurological diseases in human bodies by altering oxygen. In addition, CO produces ozone, which harms the respiratory systems of animals in several ways [50].

CO is a colorless and toxic emission that originates from the combustion of vehicles, industrial activities, and biomass burning. Vehicles that burn fossil fuels and industrial processes, controlled by humans, contribute 27% to atmospheric CO [51]. An additional 45% of CO originates from chemical reactions that involve methane (CH_4) [51]. Burning biomass adds another 19% CO to the air [51].

In terms of climate, CO indirectly affects global warming by combining with hydroxyl radicals (·OH), extending the life of the greenhouse gas - methane (CH_4) [10]. At the tropospheric level, CO forms ozone, which has a negative influence on global warming [12].

Therefore, the demand for air quality monitoring was raised to monitor CO concentrations. The Dutch National Air Quality Monitoring Network originally had 21 stations distributed throughout the Netherlands in urban, non-urban, and road areas [44]. However, after staining within the established threshold of 10 mg/m³ for 8 hours, as set by the European Union and the World Health Organization, the decision was made to limit the number of stations [8, 44]. Therefore, CO is only monitored at six stations in the North Holland region of the Netherlands, leaving other regions underrepresented in CO surveillance [48].

Consequently, the satellite-driven solution came up. In 2017, the Sentinel-5P satellite was launched, equipped with a TROPOMI sensor (Tropospheric Monitoring Instrument) that traces atmospheric gases, including CO [18]. It uses the SWIR band to measure CO, which is captured from solar radiation reflected off the Earth's surface [18]. Moreover, a hyperspectral imaging spectrometer inside the band measures the vertical column densities of atmospheric gases from the surface to the top of the atmosphere [24, 36]. The CO extracted from the satellite has been carefully evaluated by researchers in terms of accuracy and applicability. Martinez-Alonso et al. [30] evaluated satellite CO values with AirCore vertical column samples, confirming the precision of satellite CO with a relative bias of 2% under different conditions. Borsdorff et al. [1] demonstrated a high correlation between satellite CO and ground-based Fourier Transform Infrared (FTIR) CO observations, ensuring minimal bias across numerous validation sites. Schneising et al. [40] conducted a comparison in the California wildfire extreme case, which demonstrated high consistency between the satellite CO and ground measurements even in the extreme case.

However, despite positive evaluations, the usability of the satellite is limited by certain drawbacks. The first drawback is the extracted CO format, which is provided in vertical column format rather than direct surface measurements, making it difficult to assess what people actually breathe. The second issue is the unavailability of CO observations due to meteorological factors, especially cloud cover, which is usual in the Netherlands. Petetin et al. [32] raised concerns about the availability of data, stating that only 30-45% of the data were available during cloudy conditions and 70-80% during clear summer days over the Iberian Peninsula. Grzybowski [14] also confirmed the missing values of the Sentinel-5P product due to cloud concerns, stating that only 170-180 observations were extracted in 2019 in the southeastern part of Poland.

1.1 Related work

As these challenges occur not only for the satellite CO product, the trend of applying machine learning approaches has increased to estimate certain atmospheric emissions using satellite data, meteorological, and ancillary data sources. Thus, there are four studies on the estimation of ground CO mass concentrations using machine learning, which involve satellite CO data and additional variables.

TScIT 43, July 4, 2025, Enschede, The Netherlands
© 2022 University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Firstly, Chen et al. [3] used the extra tree model to predict the surface CO level in China and found R^2 = 0.712, which involved the use of other satellites (FY-4A and FY-4B). Secondly, Wang et al. [49] applied the light gradient boosting method to predict CO using the Sentinel-5P CO product in China, achieving an R^2 value of 0.55. Thirdly, Liu et al. [26] obtained an R^2 of 0.51 in predicting ground CO mass in China using random forest and spatio-temporal kriging models, but with another satellite, MOPPIT. Finally, according to Rahnama and Abkooh [35], the CO estimate at the surface level was carried out using the multiple linear regression approach in Mashhad, Iran, involving only five environmental-social variables, which accounted for 0.42 of the variance. All studies used fixed daily CO observations from the satellite.

1.2 Research questions

The above-mentioned issues and the absence of similar studies that operated in the Netherlands were the main motivation to conduct this research and to answer the following questions:

RQ1: How can machine learning models be used to accurately estimate surface-level CO concentrations based on satellite CO values that have gaps, together with meteorological and ancillary data?

RQ2: Which approach has the highest accuracy in predicting ground CO mass?

RQ3: What are the influences of meteorological and ancillary variable sets in estimating surface-level CO concentrations?

This study aims to develop and evaluate linear and machine learning models for predicting surface CO concentrations in the Netherlands, incorporating the Sentinel-5P CO product with meteorological and ancillary factors. The second objective is to understand the contribution of each input variable that influences the estimation of ground CO levels.

2 METHODOLOGY

2.1 Study area

The study area comprised multiple urban, semi-urban, and non-urban locations within the Netherlands. The selection criteria for the locations were based on diversity, population, urban levels, and location itself. A total of 19 points were observed, which can be found in Fig. 1. According to the territorial typology of Eurostat [9], 86.6% of the Dutch population resides in urban and semi-urban areas, while 13.4% live in non-urban areas. However, only 5.3% of the total area can be classified as urban centers and 16.4% as urban clusters, while 83.3% of the country remains non-urban [9].

2.2 Materials

Estimation of ground CO mass concentration included datasets from various sources and was derived on a fixed-range timescale from 1 January 2022 to 31 December 2023. All data management processes, including handling, extraction, and filtering, were executed using Python, utilizing open-source libraries and APIs.

2.2.1 Satellite CO. The satellite sensor (Sentinel-5P TROPOMI) does not take images, but it records sunlight that is absorbed by gases [27, 41]. Each gas has its own special way of absorbing sunlight [41]. Consequently, the sensor detects CO concentrations by identifying this "special" way [41]. Due to sunlight reflected off the Earth's surface, CO measurements are presented in a vertical column format, which does not accurately represent the concentrations people actually breathe [41]. Thus, the satellite captures this vertical column of CO ($CO_{satellite}$) at 13:30 local time with a spatial resolution of $5.5\times7km^2$ [41].

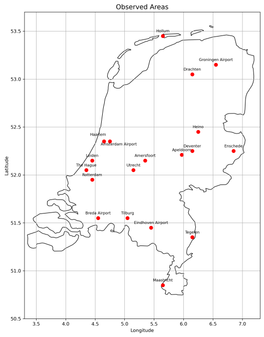

Fig. 1. Observed locations for estimating ground CO concentrations

Google Earth Engine was used to extract CO column densities ($CO_{satellite}$) every hour, providing direct access to satellite products (Sentinel-5P) without barriers or limitations [13].

2.2.2 Meteorological data. The impact of meteorological conditions can influence the prediction of all atmospheric gases [3, 14, 21, 22, 26, 33, 47, 49, 52]. According to the last related works (Section 1.1), the following variables were recommended to include in the study: temperature ($T_{temperature}$) in Kelvin, pressure ($PRS_{pressure}$) in Pascals, radiation accumulated over one hour ($RAD_{radiation}$) in W/m^2, U zonal wind component (U WIND), meridional V wind component (V WIND), planetary boundary layer height (PBLH) in meters, precipitation ($PR_{precipitation}$) in mm, cloud fraction ($CF_{cloudiness}$) from 0-1. All variables, except PBLH, were extracted from CDS at a spatial resolution of 0.1° × 0.1° [16]. The PBLHs were derived from the ERA5-Land hourly data. [31]. U WIND and V WIND were used to calculate the wind speed ($WS_{wind\ speed}$): $WS_{wind\ speed}=U^2+V^2$

2.2.3 Ancillary data. Additional variables also influence the prediction of atmospheric gases [3, 17, 23, 26, 29, 34]. Due to time and resource constraints, only these variables were used in the study: population density ($PD_{population}$), elevation ($ELVTN_{elevation}$) in meters, and intensity of the nighlights ($NLS_{nightlights}$) [7]. $PDs_{population}$ were collected from the 2015 Global Human Settlement Layers, including the number of people within a spatial grid of 250 m x 250 m [39]. The $ELVTN_{elevation}$ variable is extracted from the Shuttle Radar Topography Mission with a 30 m × 30 m spatial resolution [20]. Finally, the NL variable was acquired from VIIRS at a spatial resolution of 464 m x 464 m [13]. These variables are slowly changing, which makes them different from other data.

2.2.4 Ground-truth data. The validated surface CO concentrations of CAMS for 2022 and the interim reanalysis for 2023 were used as actual values (CO_{actual}) in this study. The truth values were obtained for the fixed time range and each location, as described in Section 2.1. The spatial resolution is 1 km x 1 km and was derived for a fixed time

- 13:00, at a vertical level of 500 m from the surface [19]. The validated reanalysis is provided as fully quality-controlled observations, while the interim is nearly validated and has a one-year delay to undergo full quality control [4]. However, interim reanalysis in this research can be used as truth observations due to the completion of quality control.

2.3 Preprocessing of data

After collecting all the necessary data for this study, it was crucial to preprocess the diverse datasets for reliable machine learning modeling. Diverse sources caused missing and inaccurate values in the merged dataset. Preprocessing was applied using Python to ensure consistency, accuracy, and robustness of the model. It aimed to align all the above-mentioned datasets across common time and location, eliminate invalid or extreme values, and reduce noise.

2.3.1 Harmonization of timescales. Datasets provided different temporal resolutions. The meteorological values and satellite CO observations were interpolated to a daily resolution (13:00). This process resulted in a unique combination of date and location rows with all required values from the mentioned data sources.

2.3.2 Keeping only clear-sky data. For other atmospheric gases, there is a possibility of retrieval of qa > 0.75 (high-quality) emission concentrations, ensuring low cloud fraction retrievals [45]. Meanwhile, satellite CO concentration collection does not provide this option, resulting in manual removal of values with a high cloud fraction. The official documentation on cloud property recommends using the 0.5 quality assurance value threshold [42]. Consequently, observations with cloud fraction > 0.5 were excluded.

2.3.3 Removing outliers. Outliers can result from sensor malfunctions or environmental spikes, leading to biased perception and reduced accuracy. The presence of cloud causes blockage of view in the SWIR band, resulting in incorrect CO values [27, 36]. Furthermore, climate extremes such as heat waves, droughts, and heavy precipitation can also indirectly affect CO levels, causing outliers in datasets [5, 53]. In addition, removing outliers is an effective practice to enhance the predictive precision of models [43]. Ouliers were removed using the interquartile range (IQR) method (Q1, Q3, IQR, ±1.5 × IQR), which is endorsed in several studies applying modeling in geological contexts [6, 28, 46].

2.3.4 Preparing data for modeling. Building a unified "final" dataset was an essential step in proceeding with modeling. Each dataset represented a different dimension. All variables were aligned by date and location. Furthermore, duplicate dates were removed to eliminate overlapping measurements, avoiding model bias [15]. During preprocessing, missing values were identified due to merging of different sources, particularly due to cloud cover issues, as mentioned in Section 1. Therefore, removing rows with empty values was a necessary step to prevent machine learning models from training on incomplete data, which can lead to misleading predictions [38].

In general, the preprocessing pipeline ensured consistency, cleanliness, and dynamism with atmospheric and satellite conditions. Consequently, the study phase can proceed to the model training phase.

2.4 Methods

2.4.1 Final dataset. After preprocessing the data, the final dataset included a total of 2,795 daily satellite-based observations from 19 locations, mentioned in Section 2.1. The dataset consisted of observations from February, April, June, August, October, and December.

As mentioned in Section 2.2 and Section 2.3.2, only data for 2022-2023 with clear-sky data were chosen for observation.

2.4.2 Modeling methods. To predict ground CO concentrations, the final dataset from Section 2.4.1 was used. Furthermore, the features of $CO_{\text{satellite}}$, $T_{\text{temperature}}$, PRS_{pressure}, $PR_{\text{precipitation}}$, $WS_{\text{wind speed}}$, $RAD_{\text{radiation}}$, Planetary Boundary Layer Height (PBLH), $CF_{\text{cloudiness}}$, $NLS_{\text{nightlights}}$, $PD_{\text{population}}$, and $ELVTN_{\text{elevation}}$ were included as input variables. In total, five types of machine learning were applied in this study.

- Linear regression with a single independent variable (LR) – $CO_{\text{satellite}}$.
- Multiple linear regression with all independent input variables (MLR) – all features.
- Random forest with all independent input variables (RF) - all features.
- Extreme gradient boosting with all independent input variables (XGB) - all features.
- Support vector machine with all independent input variables (SVM) – all features.

2.4.3 Validation of performance. Location-based cross-validation was used, which is well-suited for spatially clustered data [26, 49], particularly when different locations are involved. Each iteration consisted of a specific location for testing, ensuring that only one area out of 19 areas from Section 2.1 was tested. The results of the prediction of the ground CO concentrations were validated by the following statistical parameters: R^2, mean absolute error (MAE), root mean squared error (RMSE), bias, and mean percentage absolute error (MAPE) (Appendix A).

Feature scaling was used for prediction purposes because of the varying units of input variables. Thus, z-score normalization was applied to the support vector machine due to the sensitivity to the input scale (Appendix B).

2.4.4 Feature importance in the modeling. To assess the contribution of each predictor to the predicted ground-level CO mass, an analysis of feature importance was performed, helping to understand the drivers behind the variation in ground-level CO. Two methodologies were used for different models:

- Linear regression: not used due to a single independent variable, $CO_{\text{satellite}}$.
- Multilinear regression: absolute values of the standardized regression coefficients were used [2].
- Random forest and XGB: permutation importance was used, which is more accurate compared to default impurity-based scores [11, 25].
- Support vector machine: Permutation importance was used, which is the only option for feature importance in this model.

3 RESULTS

3.1 Available observations

As mentioned in Section 2.4.1, a total of 2,795 observations were extracted for modeling, representing at least one useful satellite estimation based on a low cloud fraction.

The average number of clear-sky observation days per month for 2022-2023 was 18.04 (18 days). Reflects the availability of 60.1% useful satellite measurements in one month based on the 30-day format. The mean number of valid observation days per month for each location during the 2022-2023 period is presented in the Appendix C. The highest monthly averages were recorded in Leiden, Deventer, and Maastricht with 19.36, 18.64, and 18.55 days, respectively. The lowest monthly averages were captured in Breda Airport, Tegelen, and Enschede with 17.36, 17.36, and 16.05 days, respectively.

3.2 Estimation of the surface-level CO mass concentration

Predicting ground CO concentrations was one of the objectives. After collecting data, preprocessing it, and preparing models, the validation and evaluation of the results can proceed. All input and actual variables were taken from Section 2.2.

3.2.1 *Linear regression with single variable.*

A linear regression model was used to estimate the ground CO concentrations, involving a single independent variable – $CO_{satellite}$. The results showed R^2 = 0.122 with MAE, MAPE, and RMSE values of 20.05 $\mu g/m^3$, 14.01%, and 26.4 $\mu g/m^3$ for daily estimates, respectively (Table 1). Linear regression showed the highest values for MAE, MAPE, and RMSE, and the lowest R^2 between models (Table 1). Thus, this model is the most unsuccessful. However, a low R^2 represents the low correlation between $CO_{satellite}$ and CO_{actual}, ensuring the validity of the actual values. Furthermore, according to Fig. 2, the linear regression model focuses on predicted values in the range of 140 to 160 $\mu g/m^3$, failing to capture the high and low extremes of CO_{actual}. A large proportion of data points in Fig. 2 that exceeded 200 $\mu g/m^3$ were predicted to have much lower values (130-160 $\mu g/m^3$), confirming a solid underestimation. Meanwhile, the actual low values around 130 $\mu g/m^3$ were mainly overpredicted (Fig. 2).

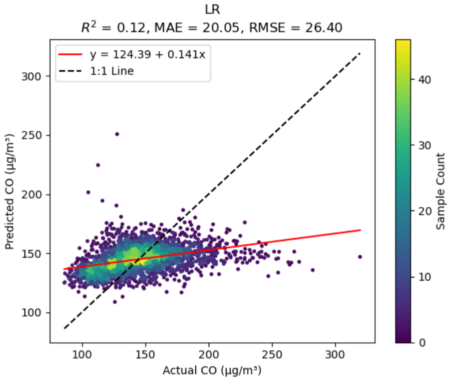

Fig. 2. Scatter plot for Linear regression ($CO_{satellite}$)
Note: Red line - fitted line, black dashed line - 1:1.

3.2.2 *Multilinear regression.*

The second linear model is a multilinear regression with eleven predictors. The results appear slightly more accurate than the linear regression model, considering R^2 = 0.2, MAE = 19.46 $\mu g/m^3$, MAPE = 13.6%, and RMSE = 25.15 $\mu g/m^3$ (Table 1). The low R^2 in this model represents limitations in linear interactions, while some reduced errors (MAE = 19.46 $\mu g/m^3$, RMSE = 25.15 $\mu g/m^3$) suggest that additional variables (meteorological and ancillary) minimally increase the fit of the model. According to Fig. 3, the multilinear regression model improves the spread of predicted values in the midrange (130-200 $\mu g/m^3$), but it does not yet capture actual values in the high range, exceeding 200 $\mu g/m^3$ as in the linear regression model. The parallel trend to the linear regression model can be highlighted, indicating the underprediction of high values and the overprediction of low values (Fig. 3).

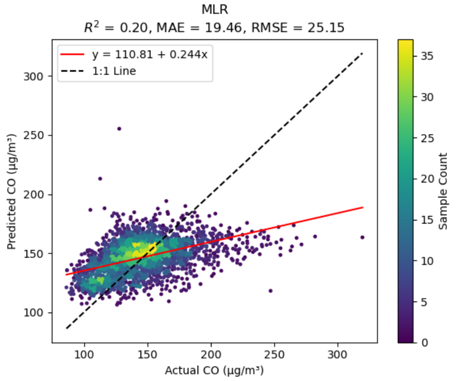

Fig. 3. Scatter plot for Multilinear regression
Note: Red line - fitted line, black dashed line - 1:1.

3.2.3 *Random forest.*

The third approach is a random forest with all predictors, which significantly outperformed linear models with R^2 = 0.53, MAE = 14.07 $\mu g/m^3$, MAPE = 9.86%, and RMSE = 18.75 $\mu g/m^3$ (Table 1). Moreover, it resulted in the highest R^2 and the lowest MAE, MAPE, and RMSE values across all models. This model shows the ability to represent complex data and nonlinear patterns. Furthermore, the random forest showed a small positive bias of 1.27 $\mu g/m^3$, which means that there is a slight trend of overestimation (Table 1). The random forest shows that the predictions are most closely clustered above the dashed line (Fig. 4). A large concentration of close predicted values can be found in the range of 100-200 $\mu g/m^3$. However, Fig. 4 indicates that the random forest spread after exceeding the actual value of 210 $\mu g/m^3$. In general, the random forest model provided better performance for extreme values than other models.

3.2.4 *Extreme gradient boosting.*

The XGB model also used eleven features. According to Table 1, it performed slightly worse than the random forest with R^2 = 0.44, MAE = 15.95 $\mu g/m^3$, MAPE = 11.04%, and RMSE = 20.88 $\mu g/m^3$, but performed better than the support vector machine and significantly better than the linear models. However, the bias in this model prediction was the lowest in all models (0.60 $\mu g/m^3$), illustrating that it neither overestimates nor underestimates ground CO concentrations with significant changes. The

Dataset	**Mean**	**STD**	**Min**	**Max**	**Q1 / Q3**
Testing Dataset (n=2,795)	144.80	16.05	98.16	236.38	134.12 / 155.20

Model	R^2	**MAE [μg/m^3]**	**RMSE [μg/m^3]**	**MAPE [%]**	**Bias [μg/m^3]**
Linear Regression (CO$_{\text{satellite}}$)	0.12	20.05	26.40	14.01	-0.03
Multilinear regression	0.20	19.46	25.15	13.60	1.22
Random Forest	0.53	14.07	18.75	9.86	1.27
XGB	0.44	15.95	20.88	11.04	0.60
Support vector machine	0.30	17.00	23.34	11.40	-3.31

Table 1. Surface CO Mass Concentration Predictions statistics.

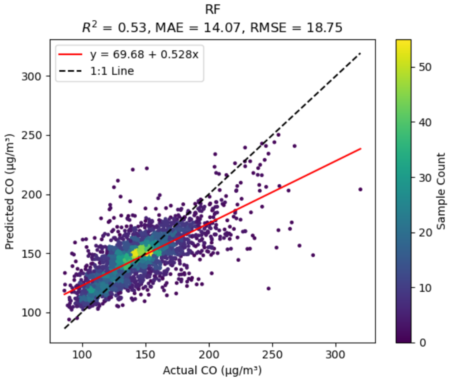

Fig. 4. Scatter plot for Random Forest
Note: Red line - fitted line, black dashed line - 1:1.

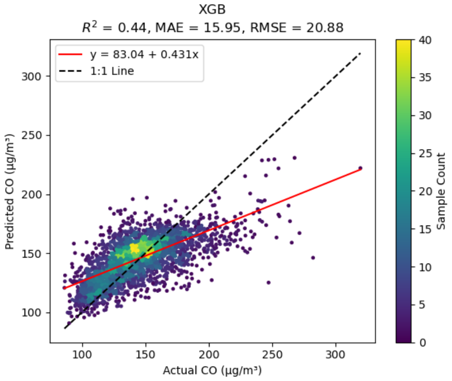

Fig. 5. Scatter plot for Extreme Gradient Boosting
Note: Red line - fitted line, black dashed line - 1:1.

XGB exhibits moderate MAE and RMSE values compared to other models, confirming the reliability of the model. The performance of the XGB shows a trend similar to the random forest, but with a greater dispersion on the right side of the diagonal, especially at values greater than 180 μg / m³, which means that it underpredicts the highest values (Fig. 5). Generally, predictions of the XGB are well within the central distribution but have a flattening effect on extremes.

3.2.5 *Support vector machine.*

The support vector machine model outperformed linear models but underperformed compared to tree-based models (Random forest and Extreme gradient boosting). The result was R² = 0.30, MAE = 17 μg/m³, MAPE = 11.4%, and RMSE = 23.34 μg/m³, requiring additional work on optimization (Table 1). Furthermore, it indicated a negative bias of −3.31 μg/m³, illustrating a consistent underestimation of CO values. The support vector machine has problems with underpredicting high values, as other models, especially in the range of 190-250 μg/m³ actual values by predicting values below 180 μg/m³ (Fig. 6). A small detail can be observed in Fig. 6, where the support vector machine performs well in the range of low actual values (90–120 μg/m³). A deviation to the right side of the 1:1 line proposes the limitation of capturing all nonlinearities (Fig. 6).

As the most accurate model for the prediction of the surface CO mass, Random Forest was chosen for further exploration.

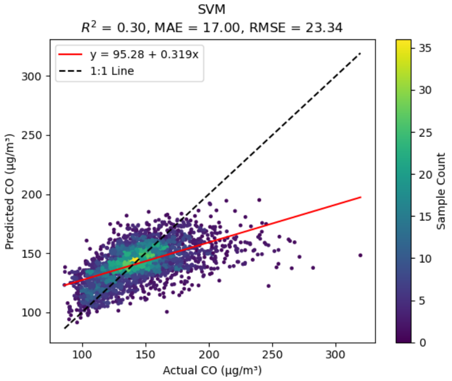

Fig. 6. Scatter plot for Support vector machine
Note: Red line - fitted line, black dashed line - 1:1.

3.2.6 *Residual random forest distribution.*

Fig. 7 suggests a quantile-quantile graph that illustrates the distribution of residuals of the random forest model. In the case of a perfect model fit, the blue line would lie perfectly on the red diagonal. The points below the red line represent underestimations, whereas the points above the line represent overestimations. Significant deviations can be distinguished from both tails, indicating that the random forest model predicts some asymmetrically distributed residuals (Fig. 7). Tails are heavier in underestimations than in overestimations, which can be confirmed with extreme points of −80 $\mu g/m^3$ and +70 $\mu g/m^3$. However, alignment of most observations with the normal line in the center portion of the quantiles between -1.5 and +2.4 ensures that the model performs well.

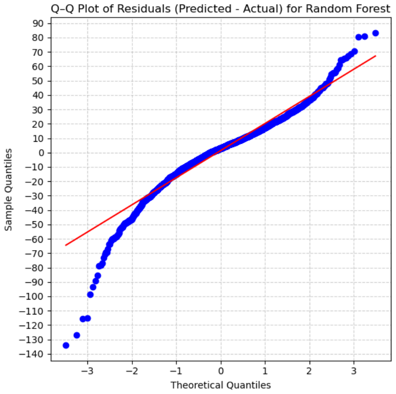

Fig. 7. Quantile-quantile Plot for Random Forest

3.2.7 *Answer to RQ1.*

Standard tree-based machine learning methods, such as random forest (0.53) and extreme gradient boosting (0.44), demonstrated solid predictive performance in estimating the ground CO mass using satellite observations, with the help of meteorological and ancillary variables. Detailed preprocessing, including harmonization of timescales (13:00), interpolation of all variables, and removal of outliers (IQR) and cloud-covered values (cloud fraction < 0.5), effectively addressed the drawbacks caused by the unclear vertical column format from the satellite and cloud-related gaps. The models were trained and validated using location-based cross-validation.

3.2.8 *Answer to RQ2.*

After completing all prediction approaches, the highest predictive accuracy was achieved by random forest with R^2 of 0.53 (MAE = 14.07 $\mu g/m^3$, MAPE = 9.86%, and RMSE = 18.75 $\mu g/m^3$). The substantial difference between random forest and linear models demonstrates the strong nonlinearity of the data. According to Figs. 2-3, linear models overestimated the ground CO concentrations during clean or calm conditions and underestimated during spikes or pollution events. The random forest model performed well in estimating surface CO mass during clean conditions, but slightly underperformed in predictions during pollution events due to the underestimation of high values.

3.3 Importance of features

Importance analysis ensures an understanding of the contribution of each input variable that predicts ground-level CO concentrations. Each calculation methodology is described in Section 2.4.4. The scores of each variable are expressed as percentages.

3.3.1 *Importance of the multilinear regression variables.*

The importance profile of the multilinear regression model reveals that the most influential variable is $CO_{satellite}$ with 34% of the total difference (Fig. 8). The second-ranked variable is $RAD_{radiation}$ with 16%, while the third is PBLH with 14.4% importance. $NLS_{nightlights}$, $PRS_{pressure}$, $CF_{cloudiness}$, $T_{temperature}$, and $WS_{wind\ speed}$ explained approximately 8%, 7%, 7%, 4.7%, and 4.5%, respectively. The lowest influences captured in the variables $ELVTN_{elevation}$, $PD_{population}$, and $PR_{precipitation}$ with 2.5%, 0.8%, and 0.4%, respectively. Therefore, the summarized meteorological variables ($RAD_{radiation}$, PBLH, $PRS_{pressure}$, $CF_{cloudiness}$, $T_{temperature}$, $WS_{wind\ speed}$, $PR_{precipitation}$) outweigh the importance of $CO_{satellite}$ with 54.3%, while the ancillary variables ($NLS_{nightlights}$, $PD_{population}$, $ELVTN_{elevation}$) have the least influence with 11.7% in total (Fig. 8).

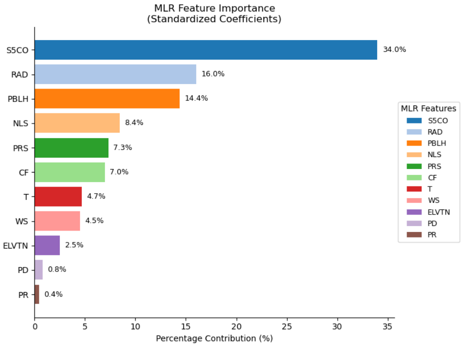

Fig. 8. Feature Importance for Multilinear regression

3.3.2 *Importance of the random forest variables.*

The random forest model also has a well-distributed feature importance portfolio, like other models. As was the case for multilinear regression, the $CO_{satellite}$ variable outperforms other variables with an importance of 27. 7% (Fig. 9). The variables PBLH, $T_{temperature}$, $RAD_{radiation}$, and $PRS_{pressure}$ follow after $CO_{satellite}$ with 17.5%, 15.8%, 14.6%, and 8.1%, respectively (Fig. 9). Other variables resulted in low contributions. Thus, well-distributed variables demonstrate the ability to learn patterns, ignoring nonlinearity. Compared to multilinear regression, it has less influence of ancillary variables (4. 2%) and more influence of meteorological variables (68.1%), outperforming $CO_{satellite}$ (27. 7%).

3.3.3 *Importance of the XGB variables.*

This model is also well-distributed with the influence of the feature sets. The most influential variable remains $CO_{satellite}$ with 34.3%, which is higher than in the multilinear regression and random forest models (Fig. 10). For the first time, $T_{temperature}$ dominates over other variables in the analysis of other models with 19.9% importance. Other key contributors are $RAD_{radiation}$, PBLH, and $PRS_{pressure}$, each contributing to a range of 5-16%. Other features contribute a small fraction, ranging from 0.3% to 3.2% (Fig. 10). Therefore, this approach effectively utilizes nonlinear features such as the random forest model. The

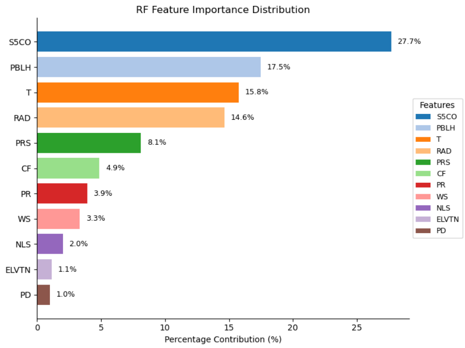

Fig. 9. Feature Importance for Random Forest

total meteorological influence accounts for 61.6%, dominating the satellite-based observation influence (34.3%), while the ancillary influence is only 4.1%.

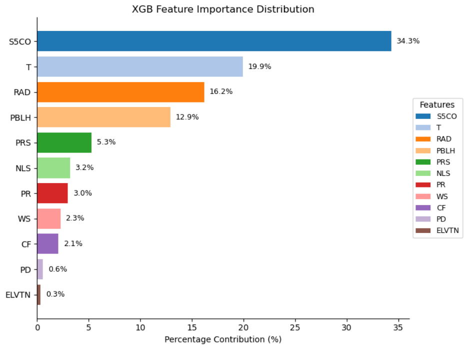

Fig. 10. Feature Importance for Extreme Gradient Boosting

3.3.4 Importance of the support vector machine variables. The support vector machine approach uses the $CO_{\text{satellite}}$ variable for prediction more than other approaches with 35.8% influence (Fig. 11). However, the distribution still exists for additional variables. PBLH, PRS_{pressure}, $T_{\text{temperature}}$, and $RAD_{\text{radiation}}$ are important with impact fractions of 14%, 10.8%, 8.3%, and 8.2%, respectively. Other variables affect prediction performance with less than 5.25% (Fig. 11). Furthermore, the meteorological influence is the lowest in the models, at 54.2%. At the same time, the ancillary features sum up to 9.9%, making them the least important set in the support vector machine feature importance analysis.

3.3.5 Answer to RQ3. After analyzing the predictors, insights can be derived about the variables and sets. According to Sections 3.3.1-3.3.4, all models demonstrated a similar pattern with variable importance, showing that satellite-based observation has the highest influence, followed by individual meteorological features, and ancillary variables hold the least impact. However, according to the means of multilinear regression, XGB, and SVM approaches, Table 2 indicates that the whole meteorological set of variables outperforms the single $CO_{\text{satellite}}$ with a difference of 22.01%. Meanwhile, ancillary variables collectively have the lowest stake in the average, at 11.17%, indicating a need for further investigation.

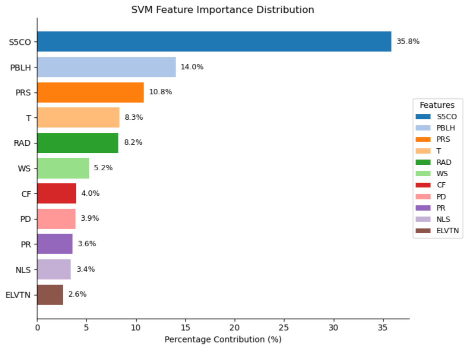

Fig. 11. Feature Importance for Support Vector Machine

However, the most predictive model, random forest ($R^2 = 0.53$, MAE = 14.07 $\mu g/m^3$), illustrated a similar behavior to others, accounting for the highest importance of the meteorological set with 68.15% (Table 2). The random forest stated the lowest importance of the ancillary set (4.18%), confirming the requirement to focus on this set (Table 2).

In general, $CO_{\text{satellite}}$ is universally crucial for predictive performance, while meteorological variables must be included to achieve accurate ground CO concentrations. Ancillary variables contributed modestly under the present methodology.

Feature Set	**Mean Importance Across Other Models [%]**	**Importance in Random Forest [%]**
$CO_{\text{satellite}}$	34.70	27.68
Meteorological set	56.71	68.15
Ancillary set	11.17	4.18

Table 2. Comparison Between Mean of Multilinear regression, Extreme gradient boosting, Support vector machine and Random forest Models Feature Importance sets

4 DISCUSSION

According to Table 3, Liu et al. [26] achieved slightly higher precision than the current study and exhibited higher RMSE using the light gradient boosting model. Both studies used Sentinel-5P as the primary data source for CO concentrations. However, the results are incomparable due to the difference in the number of observed locations (1640 and 19 sites) and the dissimilarity of the study areas (China and the Netherlands), suggesting that Liu et al. [26] trained the model on a richer dataset and indicated a high diversity of geography, affecting the results.

Wang et al. [49] follow a pattern similar to Liu et al. [26], but using MOPPIT instead of Sentinel-5P. Wang et al. [49] applied a methodology similar to the present study, using the random forest

Reference	Model	Metric	Location-based CV	Sample-based CV	Satellite	Temp	Area	Sites	Feature Set	Importance [%]
Proposed	Random Forest	R^2 RMSE	0.53 0.019 mg/m^3	– –	Sentinel-5P	Daily	Netherlands	19	Satellite CO Meteorological set Ancillary set	27.68 68.15 4.18
Liu et al. [26]	Light gradient boosting	R^2 RMSE	0.55 0.33 mg/m^3	– –	Sentinel-5P	Daily	China	1640	Satellite CO Meteorological set Ancillary set	9.40 35.6 25.0
Wang et al. [49]	Random Forest-STK	R^2 RMSE	0.51 0.54 mg/m^3	– –	MOPITT	Daily	China	1656	Satellite CO Meteorological set Ancillary set	5.30 55.00 –
Chen et al. [3]	Extra tree	R^2 RMSE	– –	0.71 0.17 mg/m^3	FY-4A&4B	Daily	China	1600	Satellite CO Meteorological set Ancillary set	19.60 42.00 20.00
Rahnama & Abkooh [35]	Multilinear regression	R^2 RMSE	0.42 (No CV) –	– –	Sentinel-2A	Daily	Iran	25	– –	– –

Table 3. Comparison with related work on surface-level CO prediction using various models, satellites, and feature types.

model, which led to slightly worse results that could be affected by the use of another satellite (Table 3). However, the results are still incomparable due to differences in satellite selection and the aforementioned reasons, such as the number of sites (1656 and 19 sites) and diverse geography (China and the Netherlands).

Chen et al. [3] used the extra tree model, utilizing the FY-4A and FY-4B satellites, to extract CO estimates. This approach led the authors to the highest accuracy under cross-validation based on the sample (0.712), indicating the effectiveness of the satellite and the methodology (Table 3). However, the current study applied location-based cross-validation. Thus, the current study remains incomparable to the research of Chen et al. [3] due to the factors mentioned above, such as different satellites (FY-4A&4B), the number of sites (1600 and 19), and geographical differences (China and the Netherlands).

Rahnama and Abkooh [35] demonstrated higher results (0.42) in terms of applying the multilinear regression model than in the current study (0.20), despite having almost the same number of sites observed (Table 3). Their results differ due to the absence of cross-validation, the number of variables applied (5), the use of another satellite to extract CO estimates (Sentinel-2A), and the difference in area (Iran and the Netherlands). Thus, the results cannot be compared.

Feature analysis in all studies except Rahnama and Abkooh [35] demonstrated that the meteorological set dominated the satellite CO and ancillary sets in influencing the prediction of ground CO, illustrating the high importance of including meteorological variables in the modeling (Table 3). Liu et al. [26] and Chen et al. [3] illustrated the substantial importance of the ancillary set to predict the ground CO mass, reflecting the great use and availability of the variables (Table 3). Furthermore, Liu et al. [26] and Wang et al. [49] demonstrated a low dependence on satellite CO, despite using different satellites, which means that these studies utilized additional variables (meteorological and ancillary) more effectively than the current study (Table 3).

5 CONCLUSION

The study demonstrated the effectiveness of using machine learning approaches and high spatial resolution CO observations based on the Sentinel-5P satellite to improve CO monitoring at the surface level in the Netherlands. Among the methods, the random forest emerged as a high-quality predictive model, demonstrating its ability to handle complex and nonlinear data. Furthermore, the results highlighted the substantial importance of meteorological variables, increasing the demand for atmospheric data in emission modeling. At the same time, the ancillary variables performed poorly, reinforcing the need to implement a more effective methodology for ancillary variables. Ultimately, this research can serve as a foundation for other environmental scientists to implement satellite-based CO observations, thereby promoting more accurate CO assessments in regions similar to the Netherlands.

The proposed framework is an end-to-end solution that excludes the chemical aspect of CO formation, thereby limiting the comprehensive understanding of the role of CO in this research. Secondly, the ancillary variables could have been expanded to include more variables, potentially limiting the full potential of social and environmental features. In future studies, the framework will be enhanced by incorporating chemical mechanisms of CO and other ancillary variables to improve the predictive performance of the models.

ACKNOWLEDGMENTS

The author would like to thank N. Bouali at the University of Twente for his assistance and valuable review.

REFERENCES

[1] Tobias Borsdorff, Joannes aan de Brugh, Hong Hu, Otto P. Hasekamp, and Ilse Aben. 2019. Sentinel-5 Precursor/TROPOMI Level 2 Product for Carbon Monoxide: Algorithm Theoretical Basis. *Atmospheric Measurement Techniques* 12, 10 (2019), 5443–5465. https://doi.org/10.5194/amt-12-5443-2019

[2] J. Bring. 1994. How to standardize regression coefficients. *The American Statistician* 48, 3 (1994), 209–213.

[3] B. Chen, J. Hu, and Y. Wang. 2024. Synergistic observation of FY-4A&4B to estimate CO concentration in China: combining interpretable machine learning to reveal the influencing mechanisms of CO variations. *npj Climate and Atmospheric Science* 7 (2024), 9. https://doi.org/10.1038/s41612-023-00559-0

[4] Copernicus Atmosphere Monitoring Service. 2021. CAMS European air quality reanalyses. https://doi.org/10.24381/7cc0465a. Accessed on 25-05-2025.

[5] P. Cristofanelli, C. Fratticioli, L. Hazan, M. Chariot, C. Couret, O. Gazetas, D. Kubistin, A. Laitinen, A. Leskinen, T. Laurila, M. Lindauer, G. Manca, M. Ramonet, P. Trisolino, and M. Steinbacher. 2023. Identification of spikes in continuous ground-based in situ time series of CO_2, CH_4 and CO: an extended experiment within the European ICOS Atmosphere network. *Atmospheric Measurement Techniques* 16, 24 (2023), 5977–5994. https://doi.org/10.5194/amt-16-5977-2023

[6] B. Dastjerdy, A. Saeidi, and S. Heidarzadeh. 2023. Review of Applicable Outlier Detection Methods to Treat Geomechanical Data. *Geotechnics* 3, 2 (2023), 375–396. https://doi.org/10.3390/geotechnics3020022

[7] C.D. Elvidge, K. Baugh, M. Zhizhin, F.C. Hsu, and T. Ghosh. 2017. VIIRS night-time lights. *International Journal of Remote Sensing* 38, 21 (2017), 5860–5879.

https://doi.org/10.1080/01431161.2016.1254297

[8] European Union. 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0050. Official Journal of the European Union, L 152, 11.6.2008, pp. 1–44.

[9] Eurostat. 2024. Urban-rural Europe – introduction. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Urban-rural_Europe_-_introduction. Accessed: 2025-05-21.

[10] B. Gaubert, A. F. Arellano, J. Barré, H. M. Worden, L. K. Emmons, S. Tilmes, R. R. Buchholz, F. Vitt, K. Raeder, N. Collins, J. L. Anderson, C. Wiedinmyer, S. M. Alonso, D. P. Edwards, M. O. Andreae, J. W. Hannigan, C. Petri, K. Strong, and N. Jones. 2016. Toward a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition. *Journal of Geophysical Research: Atmospheres* 121, 12 (2016), 7310–7343. https://doi.org/10.1002/2016JD024863

[11] R. Genuer, J. M. Poggi, and C. Tuleau-Malot. 2010. Variable selection using random forests. *Pattern Recognition Letters* 31, 14 (2010), 2225–2236.

[12] A. Gonzalez, D. B. Millet, X. Yu, K. C. Wells, T. J. Griffis, B. C. Baier, P. C. Campbell, Y. Choi, J. P. DiGangi, A. Gvakharia, H. S. Halliday, E. A. Kort, K. McKain, J. B. Nowak, and G. Plant. 2021. Fossil Versus Nonfossil CO Sources in the US: New Airborne Constraints From ACT-America and GEM. *Geophysical Research Letters* 48, 11 (2021). https://doi.org/10.1029/2021GL093361

[13] Noel Gorelick, Matt Hancher, Mike Dixon, Sergey Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202 (2017), 18–27. https://doi.org/10.1016/j.rse.2017.06.031

[14] P. T. Grzybowski, K. M. Markowicz, and J. P. Musiał. 2023. Estimations of the ground-level NO2 concentrations based on the Sentinel-5P NO2 tropospheric column number density product. *Remote Sensing* 15, 2 (2023), 378. https://doi.org/10.3390/rs15020378

[15] M. S. Hamdard and H. Lodin. 2023. Effect of Feature Selection on the Accuracy of Machine Learning Model. *International Journal of Multidisciplinary Research and Analysis* 6, 9 (2023). https://doi.org/10.47191/ijmra/v6-i9-66

[16] H. Hersbach, A. Cobb, M. Kaandorp, P. Poli, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut. 2025. ERA5 hourly time-series data on single levels from 1940 to present. Accessed on 24-05-2025.

[17] I. Ialongo, H. Virta, H. Eskes, J. Hovila, and J. Douros. 2020. Comparison of TROPOMI/Sentinel-5 Precursor NO2 observations with ground-based measurements in Helsinki. *Atmospheric Measurement Techniques* 13 (2020), 205–218. https://doi.org/10.5194/amt-13-205-2020

[18] A. Inness et al. 2022. Assimilation of S5P/TROPOMI carbon monoxide data with the global CAMS near-real-time system. *Atmospheric Chemistry and Physics* 22, 21 (2022), 14355–14376. https://doi.org/10.5194/acp-22-14355-2022

[19] Institut national de l'environnement industriel et des risques (Ineris), Aarhus University, Norwegian Meteorological Institute (MET Norway), Jülich Institut für Energie- und Klimaforschung (IEK), Institute of Environmental Protection – National Research Institute (IEP-NRI), Koninklijk Nederlands Meteorologisch Instituut (KNMI), METEO FRANCE, Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek (TNO), Swedish Meteorological and Hydrological Institute (SMHI), Finnish Meteorological Institute (FMI), Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), and Barcelona Supercomputing Center (BSC). 2022. CAMS European air quality forecasts, ENSEMBLE data. https://ads.atmosphere.copernicus.eu/datasets/cams-europe-air-quality-reanalyses. Accessed on 25-05-2025.

[20] A.H.I. Jarvis, A. Reuter, E. Nelson, and Guevara. 2008. Hole-Filled SRTM for the Globe Version 4. https://srtm.csi.cgiar.org. Accessed on 20 May 2025.

[21] J.A. Kamińska. 2019. A random forest partition model for predicting NO2 concentrations from traffic flow and meteorological conditions. *Science of The Total Environment* 651 (2019), 475–483. https://doi.org/10.1016/j.scitotenv.2018.09.155

[22] H. Kang, B. Zhu, C. Zhu, G. de Leeuw, X. Hou, and J. Gao. 2019. Natural and anthropogenic contributions to long-term variations of SO2, NO2, CO, and AOD over East China. *Atmospheric Research* 215 (2019), 284–293. https://doi.org/10.1016/j.atmosres.2018.09.016

[23] Y. Kang, H. Choi, J. Im, S. Park, M. Shin, C.K. Song, and S. Kim. 2021. Estimation of surface-level NO2 and O3 concentrations using TROPOMI data and machine learning over East Asia. *Environmental Pollution* 288 (2021), 117711. https://doi.org/10.1016/j.envpol.2021.117711

[24] M. Kawka, J. Struzewska, and J. W. Kaminski. 2021. Spatial and Temporal Variation of NO_2 Vertical Column Densities (VCDs) over Poland: Comparison of the Sentinel-5P TROPOMI Observations and the GEM-AQ Model Simulations. *Atmosphere* 12, 7 (2021), 896. https://doi.org/10.3390/atmos12070896

[25] A. Liaw and M. Wiener. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.

[26] D. Liu, B. Di, Y. Luo, X. Deng, H. Zhang, F. Yang, M. L. Grieneisen, and Y. Zhan. 2019. Estimating ground-level CO concentrations across China based on the national monitoring network and MOPITT: potentially overlooked CO hotspots in the Tibetan Plateau. *Atmospheric Chemistry and Physics* 19, 19 (2019), 12413–12430. https://doi.org/10.5194/acp-19-12413-2019

[27] D. G. Loyola, S. Gimeno García, R. Lutz, A. Argyrouli, F. Romahn, R. J. D. Spurr, M. Pedergnana, A. Doicu, V. Molina García, and O. Schüssler. 2018. The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor. *Atmospheric Measurement Techniques* 11, 1 (2018), 409–427. https://doi.org/10.5194/amt-11-409-2018

[28] M. Mahajan, S. Kumar, B. Pant, and U. K. Tiwari. 2020. Incremental Outlier Detection in Air Quality Data Using Statistical Methods. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*. IEEE, Sakheer, Bahrain, 1–5. https://doi.org/10.1109/ICDABI51230.2020.9325683

[29] S. Marinello, M.A. Butturi, and R. Gamberini. 2021. How changes in human activities during the lockdown impacted air quality parameters: A review. *Environmental Progress Sustainable Energy* 40, 5 (2021), e13672. https://doi.org/10.1002/ep.13672

[30] S. Martínez-Alonso, M. N. Deeter, B. C. Baier, K. McKain, H. Worden, T. Borsdorff, C. Sweeney, and I. Aben. 2022. Evaluation of MOPITT and TROPOMI carbon monoxide retrievals using AirCore in situ vertical profiles. *Atmospheric Measurement Techniques* 15 (2022), 4751–4765. https://doi.org/10.5194/amt-15-4751-2022

[31] Joaquín Muñoz Sabater. 2019. ERA5-Land hourly data from 1950 to present. https://doi.org/10.24381/cds.e2161bac Accessed on 24-05-2025.

[32] H. Petetin et al. 2023. Potential of TROPOMI for understanding spatio-temporal variations in surface NO2 and their dependencies upon land use over the Iberian Peninsula. *Atmospheric Chemistry and Physics* 23, 6 (2023), 3905–3935. https://doi.org/10.5194/acp-23-3905-2023

[33] H. Plaisance, A. Piechocki-Minguy, S. Garcia-Fouque, and J.C. Galloo. 2004. Influence of meteorological factors on the NO2 measurements by passive diffusion tube. *Atmospheric Environment* 38, 4 (2004), 573–580. https://doi.org/10.1016/j.atmosenv.2003.10.001

[34] K. Qin, L. Rao, J. Xu, Y. Bai, J. Zou, N. Hao, S. Li, and C. Yu. 2017. Estimating ground level NO2 concentrations over Central-Eastern China using a satellite-based geographically and temporally weighted regression model. *Remote Sensing* 9, 9 (2017), 950. https://doi.org/10.3390/rs9090950

[35] M. R. Rahnama and S. S. Abkooh. 2023. Prediction of CO pollutant in Mashhad metropolis, Iran: Using multiple linear regression. *The Geographical Journal* (2023). https://doi.org/10.1111/geoj.12534

[36] Lanlan Rao, Jian Xu, Dmitry S. Efremenko, Diego G. Loyola, and Adrian Doicu. 2022. Hyperspectral Satellite Remote Sensing of Aerosol Parameters: Sensitivity Analysis and Application to TROPOMI/S5P. *Frontiers in Environmental Science* Volume 9 - 2021 (2022). https://doi.org/10.3389/fenvs.2021.770662

[37] J. A. Raub, M. Mathieu-Nolf, N. B. Hampson, and S. R. Thom. 2000. Carbon monoxide poisoning–a public health perspective. *Toxicology* 145, 1 (2000), 1–14. https://doi.org/10.1016/S0300-483X(99)00217-6

[38] R. Rios, R. J. H. Miller, N. Manral, T. Sharir, A. J. Einstein, M. B. Fish, T. D. Ruddy, P. A. Kaufmann, A. J. Sinusas, E. J. Miller, T. M. Bateman, S. Dorbala, M. F. Di Carli, S. D. Van Kriekinge, P. B. Kavanagh, T. Parekh, J. X. Liang, D. Dey, D. S. Berman, and P. J. Slomka. 2022. Handling missing values in machine learning to predict patient-specific risk of adverse cardiac events: Insights from REFINE SPECT registry. *Computers in Biology and Medicine* 145 (2022), 105449. https://doi.org/10.1016/j.compbiomed.2022.105449

[39] Marcello Schiavina, Sergio Freire, and Kytt MacManus. 2019. GHS-POP R2019A - GHS population grid multitemporal (1975-1990-2000-2015) - OBSOLETE RELEASE. https://doi.org/10.2905/0C6B9751-A71F-4062-830B-43C9F432370F PID: http://data.europa.eu/89h/0c6b9751-a71f-4062-830b-43c9f432370f.

[40] O. Schneising, M. Buchwitz, M. Reuter, H. Bovensmann, and J. P. Burrows. 2020. Severe Californian wildfires in November 2018 observed from space: the carbon monoxide perspective. *Atmospheric Chemistry and Physics* 20, 6 (2020), 3317–3332. https://doi.org/10.5194/acp-20-3317-2020

[41] SRON Netherlands Institute for Space Research. 2024. *Sentinel-5P Level 2 Product User Manual: Carbon Monoxide (Version 2.8.0)*. Copernicus SentiWiki. https://sentiwiki.copernicus.eu/__attachments/1673595/....

[42] TROPOMI Science Team. 2025. Cloud Data Product — TROPOMI. https://www.tropomi.eu/data-products/c/cloud. Accessed: 3 June 2025.

[43] Md Gazi Uddin, Arafatur Rahman, Rosa Taghikhah, and Agnieszka I. Olbert. 2024. Data-driven evolution of water quality models: An in-depth investigation of innovative outlier detection approaches—A case study of Irish Water Quality Index (IEWQI) model. *Water Research* 255 (2024), 121499. https://doi.org/10.1016/j.watres.2024.121499

[44] Eric Van der Swaluw and Ronald Hoogerbrugge. 2012. *CO Monitoring Strategy*. Technical Report 680704015/2012. National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands. https://www.rivm.nl Performed by order of the Directorate-General for Environmental Protection.

[45] T. Verhoelst, S. Compernolle, G. Pinardi, J.-C. Lambert, H.J. Eskes, K.-U. Eichmann, A.M. Fjæraa, J. Granville, S. Niemeijer, A. Cede, et al. 2021. Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO_2 measurements with the

NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks. *Atmospheric Measurement Techniques* 14 (2021), 481–510. https://doi.org/10.5194/amt-14-481-2021

[46] H.P. Vinutha, B. Poornima, and B.M. Sagar. 2018. Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In *Information and Decision Sciences*, S. Satapathy, J. Tavares, V. Bhateja, and J. Mohanty (Eds.). Advances in Intelligent Systems and Computing, Vol. 701. Springer, Singapore, 499–506. https://doi.org/10.1007/978-981-10-7563-6_53

[47] M. Voiculescu, D.E. Constantin, S. Condurache-Bota, V. Călmuc, A. Roșu, and C.M. Dragomir Bălănică. 2020. Role of meteorological parameters in the diurnal and seasonal variation of NO2 in a Romanian urban environment. *International Journal of Environmental Research and Public Health* 17, 17 (2020), 6228. https://doi.org/10.3390/ijerph17176228

[48] RIVM Rijksinstituut voor Volksgezondheid en Milieu. 2024. Luchtmeetnet Monitoring Network. https://www.luchtmeetnet.nl/meetpunten Accessed on 6 June 2025.

[49] Yuan Wang, Qiangqiang Yuan, Tongwen Li, Liye Zhu, and Liangpei Zhang. 2021. Estimating daily full-coverage near surface O3, CO, and NO2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP. *ISPRS Journal of Photogrammetry and Remote Sensing* 175 (2021), 311–325. https://doi.org/10.1016/j.isprsjprs.2021.03.018

[50] Junfeng Zhang, Yongjie Wei, and Zhangfu Fang. 2019. Ozone pollution: a major health hazard worldwide. *Frontiers in immunology* 10 (2019), 2518.

[51] B. Zheng, F. Chevallier, Y. Yin, P. Ciais, A. Fortems-Cheiney, M. N. Deeter, R. J. Parker, Y. Wang, H. M. Worden, and Y. Zhao. 2019. Global atmospheric carbon monoxide budget 2000–2017 inferred from multi-species atmospheric inversions. *Earth System Science Data* 11, 3 (2019), 1411–1436. https://doi.org/10.5194/essd-11-1411-2019

[52] Y. Zhou, D. Brunner, C. Hueglin, S. Henne, and J. Staehelin. 2012. Changes in OMI tropospheric NO2 columns over Europe from 2004 to 2009 and the influence of meteorological variability. *Atmospheric Environment* 46 (2012), 482–495. https://doi.org/10.1016/j.atmosenv.2011.09.054

[53] Jakob Zscheischler, Anna M Michalak, Christopher R Schwalm, Miguel D Mahecha, Deborah N Huntzinger, Markus Reichstein, Etienne Berthier, Philippe Ciais, Benjamin Poulter, Stephen Sitch, et al. 2014. Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data. *Global Biogeochemical Cycles* 28, 6 (2014), 585–600. https://doi.org/10.1002/2014GB004826

A METRICS

- R-squared (R^2):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{1}$$

- Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

- Root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

- Bias:

$$\text{BIAS} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{4}$$

- Mean percentage absolute error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

- y_i – Actual value
- $\hat{y}_i$ – Predicted value
- $\bar{y}$ – Mean of the actual values
- n – Number of observations

B Z-SCORE

$$z = \frac{X - \mu}{\sigma} \tag{6}$$

- X – Original value (for standardization)
- μ – Mean of the population (used in z-score)
- σ – Standard deviation of the population
- z – Standardized value (z-score)

C AVAILABLE OBSERVATIONS TABLE

Area	**Days**	**Area**	**Days**
Leiden	19.36	Deventer	18.64
Maastricht	18.55	Haarlem	18.45
Rotterdam	18.45	Amsterdam Airport	18.37
The Hague	18.36	Drachten	18.27
Apeldoorn	18.20	Amersfoort	18.09
Eindhoven Airport	17.97	Heino	17.91
Hollum	17.83	Utrecht	17.82
Groningen Airport	17.79	Tilburg	17.45
Breda Airport	17.36	Tegelen	17.36
Enschede	16.05		

Table 4. Average clear-sky observation days per month for each location, 2022–2023.