

AI-Enhanced Carbon Emissions Monitoring Framework for Saudi Arabia

Research Methodology

Mohammad Wajeeh Bader (ID: 251000462)
Supervisor: Dr. Amjad Qataish
Master of Science in Artificial Intelligence
College of Informatics, Midocean University, UAE
October 2024

Table of Contents

Executive Summary	3
1. Research Design Overview	4
1.1 Research Philosophy.....	4
1.2 Research Questions.....	4
1.3 Research Approach.....	4
1.4 Modified Approach Justification	4
1.5 Implementation Timeline.....	5
2. Data Collection Methodology	6
2.1 Data Sources and Justification.....	6
2.1.1 CO ₂ Emissions Data	6
2.1.2 Economic and Energy Indicators.....	6
2.1.3 Oil Price Data	6
2.1.4 Satellite NO ₂ Data	6
2.1.5 Instrument Harmonization Protocol	7
2.2 Data Collection Protocol.....	7
2.3 Data Integration Strategy	7
3. Data Preprocessing and Feature Engineering.....	8
3.1 Data Cleaning Pipeline	8
3.1.1 Outlier Detection and Treatment	8
3.1.2 Saudi-Specific Data Cleaning.....	8
3.2 Feature Engineering.....	8
3.3 Feature Selection	8
3.4 Data Normalization and Splitting	9
3.5 Stationarity Assessment.....	9
4. Model Development Framework	10
4.1 Baseline Models	10
4.2 Multi-Modal Integration	10
4.3 Ensemble Methods	10
5. Validation and Evaluation Strategy	11
5.1 Temporal Cross-Validation	11
5.2 Hold-Out Test Set.....	11
5.3 Evaluation Metrics.....	11
6. Uncertainty Quantification and Limitations.....	12
6.1 Bootstrap Confidence Intervals	12
6.2 Uncertainty Validation	12
6.3 Limitations.....	12
6.3.1 Data Limitations.....	12
6.3.2 Methodological Limitations	12
6.3.3 Generalizability Limitations.....	12
7. Ethical Considerations	13
8. Conclusion.....	14
Appendix A: Data Sources Summary	15
References	16

Executive Summary

This research methodology presents a systematic approach to developing an AI-enhanced carbon emissions monitoring framework specifically tailored for Saudi Arabia. The study employs a positivist research paradigm with quantitative methodology, integrating multiple data sources including national-level economic indicators, energy consumption data, and satellite-derived environmental measurements.

The framework adopts a three-tier progressive methodology: individual baseline models, ensemble methods for performance enhancement, and multi-modal integration incorporating satellite NO₂ data. This structured approach enables systematic comparison of model architectures, isolated assessment of satellite data contribution, and transparent identification of performance improvements.

Key methodological innovations include Saudi-specific adaptations for dust storm correction, extreme temperature adjustments, and Vision 2030 policy integration. The study employs bootstrap-based uncertainty quantification and comprehensive validation strategies including temporal cross-validation and hold-out testing on 2024 data. All methodological decisions are justified through practical constraints and scientific rationale, with transparent acknowledgment of limitations and assumptions.

1. Research Design Overview

1.1 Research Philosophy

This study adopts a positivist research paradigm with quantitative methodology to develop and validate an AI-enhanced carbon emissions monitoring framework. The research follows a systematic experimental design to investigate the benefits of multi-modal data integration for CO₂ forecasting accuracy. This philosophical stance assumes that observable phenomena can be measured objectively and that systematic analysis of empirical data can reveal underlying patterns and relationships in emissions dynamics.

1.2 Research Questions

Primary Research Question:

How can proven high-performance AI forecasting methods be systematically integrated with satellite-derived environmental indicators to develop a robust, uncertainty-quantified CO₂ emissions monitoring framework for Saudi Arabia?

Secondary Research Questions:

- What baseline forecasting accuracy can be achieved using national-level indicators with monthly temporal resolution?
- To what extent does satellite NO₂ integration improve forecasting performance beyond baseline approaches?
- How can uncertainty be effectively quantified and validated in multi-modal emissions forecasts?
- What Saudi-specific environmental adaptations are necessary for accurate predictions in arid climates with frequent dust events?

1.3 Research Approach

The study employs a three-tier progressive methodology designed to enable systematic comparison and transparent performance assessment. Each tier builds upon the previous level, allowing isolated evaluation of specific contributions:

Tier 1: Individual Models

Baseline establishment using XGBoost, Random Forest, and Linear Regression with polynomial features to assess fundamental forecasting capabilities using national-level indicators only.

Tier 2: Ensemble Methods

Performance enhancement through weighted averaging and stacking techniques to leverage complementary strengths of individual models while maintaining interpretability.

Tier 3: Multi-Modal Integration

Satellite data addition using TROPOMI/Sentinel-5P NO₂ observations to evaluate independent observational validation and assess incremental improvement from space-based measurements.

1.4 Modified Approach Justification

Several key modifications were made from the original research proposal to address practical constraints while maintaining scientific rigor. Each modification is justified through a combination of data availability, computational feasibility, and scientific validity:

Temporal Resolution Enhancement:

Modified from annual to monthly resolution to provide sufficient data points for deep learning models while maintaining Saudi-specific relevance. This increases training samples from 34 to 120-300 points, addressing the fundamental challenge of limited historical data for model training.

Sequential Model Development:

Changed from simultaneous development to progressive complexity approach. Starting with models proven effective for limited datasets reduces overfitting risk and enables clear attribution of performance improvements to specific methodological choices.

Focused Satellite Integration:

Concentrated on NO₂ only rather than multiple indicators within the three-month timeline constraint. NO₂ serves as proxy for fossil fuel combustion based on global observational studies, though the correlation strength will be validated using Saudi Arabia data as part of this research.

Realistic Performance Targets:

Established monthly forecasting targets balancing ambition with complexity. Literature shows annual emissions achieve $R^2=0.95-0.99$ (Nassef et al., 2023), but monthly data exhibits higher temporal variability. Target performance ranges: baseline (national data) $R^2=0.80-0.90$, multi-modal (with satellite) $R^2=0.85-0.92$, with success threshold $R^2 \geq 0.80$ demonstrating framework viability. These ranges acknowledge monthly prediction challenges while setting achievable excellence standards.

Pragmatic Uncertainty Method:

Adopted bootstrap confidence intervals instead of Bayesian Model Averaging. While Bayesian approaches offer theoretical advantages, bootstrap methods provide robust uncertainty estimates within the three-month timeline while maintaining statistical validity and requiring less computational overhead for parameter estimation.

1.5 Implementation Timeline

The three-month implementation is structured to ensure systematic progression with clear milestones and decision points:

Month 1: Data Preparation and Baseline Establishment

Data collection, cleaning, and preprocessing including interpolation validation. Feature engineering and selection. Baseline model development (XGBoost, Random Forest, Linear Regression) with hyperparameter optimization. Initial performance assessment against success criteria. Checkpoint: If baseline $R^2 < 0.75$, extend feature engineering or adjust targets.

Month 2: Model Enhancement and Integration

Ensemble method implementation (weighted averaging and stacking). Satellite data integration with harmonization protocol. Multi-modal model training and comparison. Statistical significance testing. Checkpoint: If satellite integration yields no significant improvement, document findings and focus on ensemble optimization.

Month 3: Validation and Documentation

Comprehensive uncertainty quantification via bootstrap. Hold-out test set evaluation (2024 data). Residual diagnostics and assumption validation. Sensitivity analyses for interpolation and calibration. Final documentation and thesis writing. Optional LSTM exploration if timeline permits.

2. Data Collection Methodology

Data collection employs a multi-source strategy combining authoritative national statistics, international economic databases, and satellite remote sensing observations. All sources are publicly accessible, ensuring reproducibility and transparency in the research process. Detailed source specifications are provided in Appendix A.

2.1 Data Sources and Justification

2.1.1 CO₂ Emissions Data

Primary Source: Our World in Data (OWID) CO₂ Dataset provides comprehensive historical coverage from 1949-2023 with annual resolution. Data compiled from authoritative sources including Global Carbon Project, BP Statistical Review, and Carbon Dioxide Information Analysis Center (CDIAC). Quality assurance includes cross-validation across multiple independent methodologies following IPCC guidelines.

Secondary Source: EDGAR v7.0 (Emissions Database for Global Atmospheric Research) offers higher temporal resolution with monthly data from 2000-2023. This source serves dual purposes: cross-validation of OWID annual estimates and primary source for monthly resolution requirements enabling time series analysis.

2.1.2 Economic and Energy Indicators

World Bank Open Data provides annual indicators from 1960-2023 accessed programmatically via API. Selection criteria prioritized direct relevance to emissions drivers, high data quality with international comparability, and consistent methodology over time. Key indicators include GDP per capita, fossil fuel energy consumption percentage, electric power consumption per capita, renewable energy consumption share, energy intensity, and total population for per-capita normalization.

2.1.3 Oil Price Data

Federal Reserve Economic Data (FRED) provides Europe Brent Spot Price from 1987-2024 at daily resolution, aggregated to monthly for analysis. For Saudi Arabia, oil prices serve as multifaceted proxy: economic activity intensity indicator, energy sector investment pattern tracker, and government revenue metric affecting renewable energy initiative funding.

2.1.4 Satellite NO₂ Data

NASA Giovanni Platform provides NO₂ tropospheric column density from two complementary instruments covering the period 2004-2024:

- TROPOMI/Sentinel-5P (2018-2024) offers high spatial resolution and improved signal-to-noise ratio for recent observations
- OMI/Aura (2004-2017) provides extended historical coverage enabling longer-term trend analysis

Scientific justification: NO₂ serves as proxy for fossil fuel combustion based on mechanistic understanding of nitrogen oxide formation during high-temperature combustion processes. The correlation strength will be empirically validated using Saudi Arabia data. NO₂ provides independent observational data source not derived from emission inventories, and operational satellite continuity ensures consistent long-term monitoring capability.

2.1.5 Instrument Harmonization Protocol

The transition from OMI to TROPOMI instruments in 2018 requires harmonization to ensure consistent time series. During the overlap period (2018-2020), both instruments observed Saudi Arabia simultaneously, enabling calibration:

Harmonization Procedure:

- Extract monthly NO₂ averages for identical spatial domain (15°N-32°N, 35°E-55°E) from both instruments during 2018-2020
- Establish linear calibration relationship: $TROPOMI_equivalent = \alpha + \beta \times OMI_observed$ using ordinary least squares regression
- Validate calibration quality: target correlation $R^2 > 0.85$ and systematic bias $< 5\%$ during overlap period
- Apply calibration to pre-2018 OMI data with propagated uncertainty estimates included in final prediction intervals
- If calibration quality insufficient ($R^2 < 0.85$), instruments will be analyzed separately with temporal gap acknowledged in limitations

2.2 Data Collection Protocol

National-level data collection employs automated programmatic access through established APIs ensuring consistency and reproducibility. Quality control measures include:

- Automated validation of data ranges preventing negative values for non-growth variables
- Consistency checks between sources comparing OWID and EDGAR CO₂ estimates
- Temporal completeness verification identifying missing periods requiring imputation
- Outlier detection using interquartile range method with manual verification

2.3 Data Integration Strategy

Temporal alignment protocol establishes common monthly time index from 2004-2024 for multi-modal analysis. This integration presents a key methodological challenge: reconciling annual economic indicators with monthly emissions and satellite data.

Annual economic indicators undergo cubic spline interpolation to monthly resolution. This method preserves annual totals while providing smooth monthly estimates. However, interpolation introduces assumptions of gradual intra-annual simplifying assumption of gradual variation that may underestimate short-term economic fluctuations. Sensitivity analysis will assess model robustness by comparing cubic spline against linear interpolation and seasonal decomposition approaches. If model performance variance across interpolation methods exceeds 0.02 R^2 , interpolation uncertainty will be explicitly reported.

Missing data handling applies hierarchical strategy based on gap magnitude:

- Less than 5% missing data: Linear interpolation with maximum gap constraint of three consecutive months
- 5-10% missing data: Multiple Imputation by Chained Equations (MICE) algorithm with 10 iterations and pooled results across 5 imputations
- Greater than 10% of missing data: Variable excluded or replaced with proxy indicator

3. Data Preprocessing and Feature Engineering

Data preprocessing transforms raw collected data into analysis-ready format through systematic cleaning, transformation, and feature engineering. This process is particularly critical for Saudi Arabian data given unique environmental challenges including frequent dust storms and extreme temperature variations.

3.1 Data Cleaning Pipeline

3.1.1 Outlier Detection and Treatment

Hybrid outlier detection combines Interquartile Range (IQR) and Z-score methods (threshold $|z| > 3$). Values flagged by either method undergo manual verification to distinguish genuine events from measurement errors. Treatment protocol prioritizes preserving real phenomena:

- Real events such as COVID-19 economic impacts are retained with annotation
- Confirmed errors are replaced with interpolated values or removed if occurring late in series
- All treatments are documented in metadata file ensuring transparency and reproducibility

3.1.2 Saudi-Specific Data Cleaning

Saudi Arabia presents unique environmental challenges requiring specialized data cleaning approaches not typically found in standard preprocessing pipelines.

Dust Storm Correction for NO₂:

Saudi Arabia experiences over 50 dust storm days annually, concentrated in peak season (March-August). These events cause artificially low NO₂ retrievals through two mechanisms: increased light scattering reducing signal strength, and reduced satellite sensitivity due to aerosol interference. Correction methodology employs three-stage approach:

- Identification of high Aerosol Optical Depth periods using auxiliary satellite data
- More stringent outlier removal during dust season using tighter IQR bounds (1.0x instead of standard 1.5x multiplier)
- Seasonal adjustment normalization by monthly baseline to account for systematic dust-related biases

Extreme Temperature Adjustment:

Temperature variations exceeding 40°C seasonally drive significant air conditioning demand fluctuations. Feature engineering incorporates temperature-dependent energy consumption proxy applying 30% increase factors for summer months (June-September) when cooling loads dominate electricity demand patterns.

3.2 Feature Engineering

Feature engineering transforms raw variables into informative predictors capturing temporal dynamics, cyclical patterns, and domain-specific phenomena relevant to emissions forecasting. Features include temporal lags (1, 3, and 12-month), rolling statistics (moving averages and standard deviations), cyclical encodings (sine-cosine month transformation), Saudi-specific indicators (Vision 2030 timeline, Hajj season flags, dust season markers), and interaction terms combining economic and environmental variables.

3.3 Feature Selection

Multi-stage selection process balances statistical criteria with domain knowledge:

- Correlation-based filtering removes weak predictors ($|r| < 0.1$) and highly collinear features ($|r| > 0.95$)
- Variance Inflation Factor check identifies multicollinearity (VIF > 10 threshold)
- Recursive Feature Elimination with cross-validation selects top 10-15 features
- Domain knowledge validation ensures scientific justification and interpretability

3.4 Data Normalization and Splitting

Neural network models employ StandardScaler normalization fitted exclusively on training data. Tree-based models use original scales leveraging scale-invariance. Temporal splitting preserves chronological order: training (1990-2022), validation (2023), hold-out test (2024).

3.5 Stationarity Assessment

Time series modeling assumptions require stationary data where statistical properties remain constant over time. Augmented Dickey-Fuller (ADF) test will assess stationarity of CO₂ emissions and predictor variables. The null hypothesis assumes non-stationarity (unit root presence); rejection at 5% significance level indicates stationarity. Complementary Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test provides confirmatory evidence with reversed hypothesis structure. Non-stationary variables will undergo differencing transformations or detrending as appropriate. These tests ensure modeling assumptions validity and guide feature transformation decisions.

4. Model Development Framework

Model development follows the three-tier progressive methodology: Tier 1 establishes baseline performance using individual models (XGBoost, Random Forest, Linear Regression), Tier 2 enhances performance through ensemble methods (weighted averaging and stacking), and Tier 3 integrates multi-modal satellite data evaluating incremental improvement from NO₂ observations. Each tier builds upon previous results with clear success criteria and statistical validation.

4.1 Baseline Models

XGBoost serves as primary baseline with conservative hyperparameters (shallow trees, strong regularization) preventing overfitting. Random Forest provides alternative baseline enabling ensemble uncertainty estimation. Linear regression with polynomial features establishes interpretable benchmark. Hyperparameter tuning employs timeseries Split cross-validation ensuring temporal validity.

4.2 Multi-Modal Integration

Early fusion strategy augments baseline features with satellite-derived NO₂ measurements, lags, rolling averages, seasonal adjustments, and interaction terms. Two-stage training isolates satellite contribution: Stage 1 trains on baseline features, Stage 2 incorporates satellite data. Performance difference undergoes permutation testing (1000 iterations, $p < 0.05$ threshold) validating statistical significance of improvements.

4.3 Ensemble Methods

Weighted averaging assigns inverse RMSE weights to individual models. Stacking ensemble employs two-level architecture: base models (Level 0) provide diverse predictions, Ridge meta-learner (Level 1) learns optimal combination through cross-validation.

5. Validation and Evaluation Strategy

5.1 Temporal Cross-Validation

Timeseries Split with 8 folds employs expanding window strategy. Each fold adds one year to training set while testing on subsequent 12 months, preventing look-ahead bias and providing multiple validation points. Mean and standard deviation of fold performances quantify model stability.

5.2 Hold-Out Test Set

Entire 2024 year reserved as untouched hold-out test set, never used during training or hyperparameter tuning. Final model trained on complete data through 2023 undergoes single evaluation on 2024, providing unbiased generalization performance estimate simulating one-year-ahead forecasting.

5.3 Evaluation Metrics

Primary metrics include:

- Coefficient of Determination (R^2) measuring explained variance proportion
- Root Mean Squared Error (RMSE) in original units (million tons CO₂)
- Mean Absolute Percentage Error (MAPE) for scale-independent comparison

6. Uncertainty Quantification and Limitations

6.1 Bootstrap Confidence Intervals

Bootstrap methodology generates 95% confidence intervals through 1000 iterations of resampling with replacement, training models on bootstrap samples, and calculating percentile-based intervals. This approach quantifies epistemic uncertainty (model uncertainty reducible with more data) without distributional assumptions.

6.2 Uncertainty Validation

Coverage probability validates calibration: well-calibrated 95% intervals should contain 90-97% of true values. Sharpness measures precision through average interval width, targeting relative width below 15% while maintaining proper coverage.

6.3 Limitations

6.3.1 Data Limitations

- Limited historical satellite data: NO₂ observations from 2004 only with instrument transition affecting consistency
- Annual-to-monthly interpolation: Cubic spline interpolation introduces smoothing potentially oversimplifying intra-annual economic fluctuations. Sensitivity analysis will quantify this uncertainty.
- Incomplete emission sources: Focus on fossil fuel combustion excludes flaring, land use change, and agricultural emissions

6.3.2 Methodological Limitations

- Correlation versus causation: NO₂-CO₂ relationship reflects shared combustion driver rather than direct causality, limiting mechanistic understanding
- Stationarity assumption: Stable feature-target relationships assumed, potentially violated by Vision 2030 structural changes
- Spatial aggregation: National-level analysis loses sub-national heterogeneity

6.3.3 Generalizability Limitations

- Saudi-specific optimization may not generalize beyond Gulf Cooperation Council region
- Temporal generalizability: Historical relationships may not capture future regime shifts from renewable energy transition

7. Ethical Considerations

All data sources are publicly available under appropriate open licenses. No personally identifiable information is involved. Potential dual-use concerns (carbon market manipulation, competitive intelligence) are mitigated through transparent methodology publication and emphasis on policy support applications. Algorithmic fairness addressed through stratified performance evaluation and explicit uncertainty reporting.

8. Conclusion

This methodology presents a systematic approach to AI-enhanced carbon emissions monitoring specifically adapted for Saudi Arabian conditions. The three-tier progressive framework enables transparent assessment of baseline capabilities, ensemble enhancements, and multi-modal integration benefits. Key strengths include pragmatic modifications addressing constraints while maintaining rigor, extensive Saudi-specific adaptations, robust uncertainty quantification, and transparent limitation acknowledgment.

The framework acknowledges limitations transparently: limited historical satellite data, interpolation assumptions, correlation-based rather than causal relationships, and Saudi-specific optimization potentially constraining generalizability. Three-month implementation timeline includes strategic checkpoints enabling adaptive decision-making with contingency plans for common challenges.

This methodology establishes foundation for reproducible, scientifically-sound emissions monitoring supporting Vision 2030 objectives while contributing to broader climate mitigation through enhanced measurement capabilities.

Appendix A: Data Sources Summary

Comprehensive listing of all data sources used in this research, including access methods, licensing terms, and temporal coverage specifications.

Data Source	Provider	Access URL	License	Coverage	Resolution	Access Method
CO ₂ Emissions	OWID	github.com/owid/co2-data	CC BY 4.0	1949-2023	Annual	Direct download (CSV)
CO ₂ (monthly)	EDGAR v7.0	edgar.jrc.ec.europa.eu	Public	2000-2023	Monthly	Web portal download
Economic Indicators	World Bank	databank.worldbank.org	CC BY 4.0	1960-2023	Annual	API (Python wbdata)
Oil Prices	FRED	fred.stlouisfed.org	Public Domain	1987-2024	Daily	API (fredapi)
NO ₂ (OMI)	NASA Giovanni	giovanni.gsfc.nasa.gov	Public Domain	2004-2017	Monthly	Web interface extract
NO ₂ (TROPOMI)	NASA Giovanni	giovanni.gsfc.nasa.gov	Public Domain	2018-2024	Monthly	Web interface extract

References

- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192-213. <https://doi.org/10.1016/j.ins.2011.12.028>
- Cao, E. L. (2023). National ground-level NO₂ predictions via satellite imagery driven convolutional neural networks. *Frontiers in Environmental Science*, 11, Article 1285471. <https://doi.org/10.3389/fenvs.2023.1285471>
- Dou, X., Wang, Y., Ciais, P., Chevallier, F., Davis, S. J., Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Solazzo, E., Yan, F., Zheng, B., & Liu, Z. (2023). Near-real-time global gridded daily CO₂ emissions 2021. *Scientific Data*, 10(1), Article 69. <https://doi.org/10.1038/s41597-023-01963-0>
- Duncan, B. N., Lamsal, L. N., Thompson, A. M., Yoshida, Y., Lu, Z., Streets, D. G., Hurwitz, M. M., & Pickering, K. E. (2016). A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005–2014). *Journal of Geophysical Research: Atmospheres*, 121(2), 976-996. <https://doi.org/10.1002/2015JD024121>
- European Space Agency. (n.d.). TROPOMI/Sentinel-5P nitrogen dioxide data product. Retrieved October 14, 2024, from <https://sentinel.esa.int/web/sentinel/missions/sentinel-5p>
- Federal Reserve Bank of St. Louis. (n.d.). Crude Oil Prices: Brent - Europe [DCOILBRENTU]. FRED, Federal Reserve Economic Data. Retrieved October 14, 2024, from <https://fred.stlouisfed.org/series/DCOILBRENTU>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- NASA Goddard Earth Sciences Data and Information Services Center. (n.d.). Giovanni: The bridge between data and science. Retrieved October 14, 2024, from <https://giovanni.gsfc.nasa.gov/giovanni/>
- Nassef, A. M., Hussein, T. M., & Mokhiamar, O. (2023). Application of artificial intelligence to predict CO₂ emissions: Critical step towards sustainable environment. *Sustainability*, 15(9), Article 7648. <https://doi.org/10.3390/su15097648>
- Ritchie, H., Roser, M., & Rosado, P. (2020). CO₂ and greenhouse gas emissions [Data set]. Our World in Data. Retrieved October 14, 2024, from <https://github.com/owid/co2-data>
- World Bank. (n.d.). World Development Indicators. Retrieved October 14, 2024, from <https://databank.worldbank.org/source/world-development-indicators>