Robust Forecasting of Greenhouse Gas Emissions with Change Point and Anomaly Detection

Yiteng Zhang[1,2], Arjun Pakrashi[2,3], Soumyabrata Dev[1,2,3]

[1]The ADAPT SFI Research Centre, Dublin, Ireland
[2]School of Computer Science, University College Dublin, Ireland
[3]Beijing-Dublin International College, Beijing, China
yiteng.zhang@ucdconnect.ie, arjun.pakrashi@ucd.ie, soumyabrata.dev@ucd.ie

***Abstract*—Greenhouse gas time series are often affected by anomalies and structural breaks that arise from sensor errors, calibration periods, and episodic meteorological events. These irregularities can bias model training and reduce the accuracy of forecasts. In this study, we develop a robust forecasting framework that integrates anomaly detection, change point detection, and robustness-aware modeling strategies. Using more than six years of hourly CO_2 measurements from the IPR monitoring site, we evaluate three representative forecasting approaches of increasing complexity: Seasonal Naive, SARIMAX, and GRU. The results show that robustness mechanisms, including anomaly substitution, weighted estimation, and robust loss functions, significantly reduce forecast errors for the Seasonal Naive and SARIMAX models. In contrast, the GRU model already delivers highly accurate predictions, and additional robustness adjustments yield only marginal improvements. These findings demonstrate that explicit robustness mechanisms are crucial for simple and statistical forecasting methods. The study highlights the importance of anomaly and change point detection as practical quality control steps for reliable greenhouse gas forecasting.**

***Index Terms*—Greenhouse gas forecasting, Anomaly detection, Change point detection, Robust modeling, GRU neural network.**

I. Introduction

Accurate forecasting of greenhouse gas (GHG) emissions and atmospheric concentrations is critical for evaluating climate policies, monitoring mitigation efforts, and assessing progress toward carbon neutrality goals [1], [2]. Time series models are widely used in this context, as they can capture temporal dynamics and provide both short-term forecasts and long-term trends that are useful for policymakers and scientists. However, the reliability of these forecasts strongly depends on the quality of observational data.

In real-world monitoring networks, GHG time series are often subject to irregularities such as anomalies, abrupt structural changes, and missing values. These issues can arise from sensor downtime, environmental disturbances, or sudden external shocks, for instance extreme weather events or policy interventions. If left untreated, such irregularities can distort statistical patterns, bias model training, and lead to unreliable forecasts. Existing studies on GHG prediction have largely focused on improving forecasting models, but have paid less attention to data quality control, especially the detection and treatment of anomalies and change points.

In this study[1], we address this gap by integrating anomaly and change point detection into the GHG forecasting pipeline. We first identify abnormal observations and structural shifts in the time series using statistical decomposition and detection algorithms. The identified irregularities are then incorporated into model training through down-weighting. By comparing against conventional approaches that directly train on raw data, we demonstrate that our quality-controlled pipeline produces more accurate forecasts and more reliable predictive intervals.

The contributions of this work are threefold. First, we propose a simple yet effective framework that links anomaly and change point detection with robust forecasting. Second, we provide a systematic evaluation on high-frequency CO_2 monitoring datasets, showing consistent improvements across multiple baselines. Finally, our findings highlight the value of data quality control in environmental time series forecasting, offering practical insights for the design of climate monitoring and reporting systems.

II. Related Works

Recent progress in greenhouse gas (GHG) forecasting has combined statistical and machine learning approaches. Traditional models such as ARIMA, SARIMA and exponential smoothing remain widely used [3]. Deep learning architectures including LSTM, TCN, CNN BiLSTM and CNN BiLSTM KAN [4], [5] have also gained attention because they capture complex temporal patterns and nonlinear relationships (Zhang, 2025) [6]. Beyond these general advances, Zhang et al. investigated interconnected drivers of CO_2 emissions using principal component analysis in the Indian context [7], and further explored deep learning-based prediction of short-term CO_2 emissions in Ireland [8], demonstrating the value of both statistical and neural methods in diverse regional settings. A review of 75 urban studies confirms this trend and reports steady improvements in predictive accuracy, yet it also notes that most works assume clean and complete data with little focus on quality issues (Jin, 2025) [9]. In parallel, anomaly detection methods have been applied to high frequency GHG records. Kasatkin and Krinitskiy (2023)

[1]In the spirit of reproducible research, the code to reproduce the results of this paper can be found here: https://github.com/Yiteng369/Anomaly.

proposed a machine learning routine to detect and label anomalies in eddy covariance data, reducing the need for manual filters. Hernandez Mejia et al. (2024) combined LSTM autoencoders with Isolation Forests for anomaly detection in geological carbon sequestration datasets [10].

Research on structural breaks has developed tools such as PELT (Killick et al., 2012) and Bayesian Online Change Point Detection (Adams and MacKay, 2007), both of which are well established for identifying changes in mean or variance [11] [12]. RobustTAD extends this line by combining seasonal trend decomposition with convolutional neural networks to detect anomalies in complex time series, although its applications remain focused on industrial and commercial data (Gao et al., 2020) [13].

These advances show clear progress in both forecasting and detection. However, GHG forecasting pipelines rarely integrate anomaly or change point detection into the training of predictive models. Detection is usually treated as a separate preprocessing step without systematic evaluation of its effect on forecast accuracy and interval reliability. Our work addresses this gap by embedding anomaly and change point detection into the forecasting process and by providing empirical evidence that this integration improves point predictions as well as uncertainty quantification.

III. Methodology

The problem of forecasting greenhouse gas emissions is complicated by several challenges in observational data. Hourly CO_2 series are prone to missing values, measurement outliers, and structural changes in long-term trends. If not addressed, these issues can degrade the accuracy of both statistical and machine learning models. To overcome these limitations, we designed a robust forecasting framework consisting of four major components: data preprocessing, anomaly detection, change point detection, and forecasting models enhanced with robustness mechanisms. The overall logic is to first ensure data quality, then identify structural characteristics, and finally build predictive models capable of handling irregularities.

A. Data Preprocessing

Greenhouse gas monitoring systems often produce sentinel values such as -999.990 or -9.990 to indicate missing or invalid records. These values were replaced with missing markers. For short gaps lasting up to six hours, we applied linear time interpolation to ensure local continuity. Longer gaps were imputed using forward and backward filling, preventing discontinuities while avoiding artificial trends. The result was a continuous hourly series, which served as the foundation for subsequent analysis.

B. Anomaly Detection

Greenhouse gas time series are often contaminated by instrumental errors, calibration periods, or meteorological extremes. To mitigate their influence, we applied a hybrid anomaly detection strategy that combines decomposition-based residual analysis with a machine learning approach.

We first decomposed the CO_2 series into trend, seasonal, and residual components using Seasonal-Trend decomposition via Loess (STL):

$$y_t = T_t + S_t + R_t, \tag{1}$$

where T_t denotes the long-term trend, S_t the seasonal component, and R_t the residual. Under normal conditions, R_t should behave approximately as white noise. To identify outliers, we computed robust z-scores of residuals:

$$z_t = \frac{R_t - \text{median}(R)}{1.4826 \times \text{MAD}(R)}, \tag{2}$$

where MAD(R) denotes the median absolute deviation of residuals, scaled by 1.4826 to make it comparable to the standard deviation. Observations were flagged as anomalous if $|z_t| > 4$.

In parallel, we employed the Isolation Forest algorithm, which isolates points through recursive random partitioning of the feature space. Anomalies are expected to have shorter average path lengths in the trees. We trained the model using both raw CO_2 values and STL residuals as features, with contamination parameter set to 0.01.

Finally, a sample was labeled as anomalous if it was detected by either method. This conservative strategy ensured that both local irregularities (captured by residual analysis) and global inconsistencies (captured by Isolation Forest) were effectively identified.

C. Change Point Detection

Long-term CO_2 dynamics are often influenced by structural shifts such as instrument recalibrations, policy interventions, or seasonal regime transitions. To capture these shifts, we applied the Pruned Exact Linear Time (PELT) algorithm to the STL trend component [14]. The PELT method formulates change point detection as a segmentation problem: given a cost function $C(\cdot)$ and a penalty term β, the objective is to find the set of change points $\{\tau_1, \tau_2, \ldots, \tau_m\}$ that minimizes the following criterion:

$$\min_{\{\tau_i\}} \left\{ \sum_{j=0}^{m} C\big(y_{\tau_j+1:\tau_{j+1}}\big) + m\beta \right\}, \tag{3}$$

where $\tau_0 = 0$, $\tau_{m+1} = n$, and $y_{\tau_j+1:\tau_{j+1}}$ denotes the segment between consecutive change points. The cost function was specified using an rbf kernel, which allows the method to detect both mean and slope changes.

In practice, the minimum segment length was set to 72 hours to avoid over-segmentation, and the penalty value was fixed at $\beta = 8.0$ to balance sensitivity and stability. The resulting change points were encoded as categorical regime labels, which served as exogenous variables in the forecasting models, enabling them to adapt to different structural regimes.

D. Robust Weighting Strategy

Instead of discarding anomalous CO_2 observations, we reduced their influence during model training by assigning lower weights. This strategy preserves the full temporal structure of the data while limiting the impact of irregular samples.

For the SARIMAX model, anomalies were replaced with missing values (NaN). In the state-space formulation, missing observations are naturally handled by the Kalman filter, which effectively assigns them zero weight in the likelihood update step. This allows the model to propagate information from past states without being distorted by anomalous values.

For the GRU model, we introduced a weighted Huber loss. The Huber loss is quadratic for small errors and linear for large errors, thereby reducing the influence of extreme deviations. It is defined as

$$L_\delta(e_t) = \begin{cases} \frac{1}{2}e_t^2, & \text{if } |e_t| \leq \delta, \\ \delta\left(|e_t| - \frac{1}{2}\delta\right), & \text{if } |e_t| > \delta, \end{cases} \tag{4}$$

where $e_t = y_t - \hat{y}_t$ is the prediction error and δ is a robustness threshold. To further down-weight anomalous samples, each observation was assigned a soft weight $w_t \in [0.2, 1]$ derived from its anomaly score. The final weighted loss function is

$$L = \frac{1}{N}\sum_{t=1}^{N} w_t \, L_\delta(e_t). \tag{5}$$

This formulation penalizes anomalous points less heavily while still learning from the overall sequence, thereby enhancing robustness without discarding data.

E. Forecasting Models

We evaluated three forecasting approaches of increasing complexity: a Seasonal Naive baseline, a SARIMAX model, and a GRU neural network. The Seasonal Naive method simply repeats daily CO_2 values, with a robust variant that replaces anomalous references with the median of recent valid days. SARIMAX integrates autoregressive dynamics, seasonality, and exogenous variables such as diurnal cycles, weekends, and change point regimes. The GRU model captures nonlinear and long-term dependencies using 48-hour input sequences and is trained with a weighted Huber loss to down-weight anomalous samples. Together, these models form a spectrum from simple statistical baselines to deep learning architectures, enabling a systematic evaluation of robustness mechanisms across different forecasting paradigms.

F. Evaluation Protocol

Model performance was assessed through a walk-forward validation procedure. The final seven days of the hourly CO_2 series were reserved as an independent test set, while all preceding data were used for model training. This setup reflects a realistic forecasting scenario, where models are trained on historical observations and then applied to predict future emissions under evolving conditions.

Forecast accuracy was measured using three complementary error metrics. The mean absolute error (MAE) captures the average magnitude of deviations without emphasizing direction, providing a direct measure of overall accuracy. The root mean squared error (RMSE) penalizes large deviations more heavily, making it sensitive to extreme errors and thus informative about robustness against spikes or sudden shifts. The mean absolute percentage error (MAPE) normalizes errors relative to the observed values, offering a measure of relative accuracy that facilitates comparison across different time scales or magnitudes.

By combining these three metrics, we obtain a comprehensive evaluation of each model's forecasting performance, balancing absolute precision, sensitivity to extreme events, and proportional accuracy. This multi-faceted assessment enables us to rigorously analyze the benefits of robustness mechanisms across simple, statistical, and neural forecasting approaches.

IV. Experiments

The experimental design aimed to evaluate the effectiveness of the proposed robust forecasting framework using hourly CO_2 concentration data from the IPR monitoring site, spanning December 2017 to March 2024. After preprocessing and anomaly detection, the dataset contained approximately 55,000 valid hourly observations. Figure 1 shows the complete time series. The data clearly exhibit strong diurnal and seasonal cycles, along with long-term upward trends. Occasional sharp spikes are also present, which can be attributed to either measurement irregularities or episodic meteorological events. The final seven days were reserved as an independent test set, while the preceding data were used for model training. This setup reflects a realistic forecasting scenario, in which models are required to generalize from historical data to short-term future emissions under non-stationary conditions.

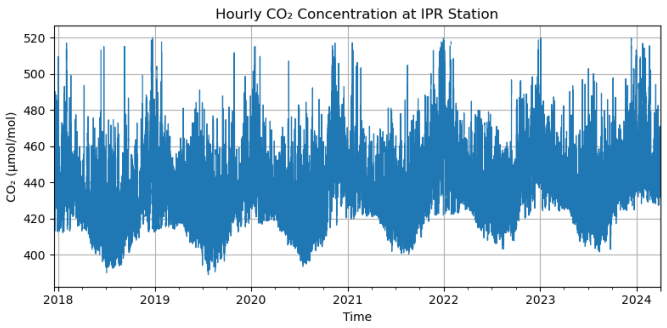


Fig. 1: Hourly CO_2 concentration at the IPR monitoring station from December 2017 to March 2024. The series exhibits clear diurnal and seasonal cycles as well as occasional spikes, reflecting both natural variability and measurement irregularities.

For time series cross-validation, we adopted a rolling-window design. The initial training window consisted of two weeks of hourly data, and forecasts were generated up to 24 hours ahead. The window was then shifted forward by one day and the process repeated. This rolling scheme

allowed us to evaluate forecasting performance across multiple overlapping periods while ensuring that each prediction relied only on information available at the time. In total, fifty rolling evaluations were performed, covering a wide range of temporal contexts and data conditions.

The forecasting horizon was fixed at 24 hours in all experiments, consistent with the diurnal periodicity of CO_2 concentrations. Performance was assessed on one-step-ahead through 24-step-ahead forecasts, with errors aggregated over the horizon to compute mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). These metrics capture complementary aspects of predictive accuracy.

Each model was implemented with specific configurations chosen to balance predictive power and computational efficiency. The Seasonal Naive model required no parameters beyond the lag of 24 hours. The SARIMAX model was specified as $(1,0,1)(1,1,1)_{24}$, with seasonal differencing set to one cycle of 24 hours. The GRU model consisted of a single recurrent layer with 64 hidden units and a dense linear output layer. Input sequences were fixed at 48 hours, which provided sufficient historical context for learning diurnal and short-term temporal patterns. The GRU was trained using the Adam optimizer with a batch size of 64 and a maximum of 50 epochs, while early stopping based on validation loss was employed to prevent overfitting.

This experimental design ensured consistency across models, controlled the forecasting horizon, and allowed systematic comparison of robustness mechanisms under uniform evaluation conditions.

V. Results

We first analyzed the temporal distribution of anomalies to better understand their systematic characteristics. Table I summarizes the monthly anomaly rates from 2018 to 2024. Anomaly rates tend to peak during the winter months, with particularly high values in January and December. These seasonal peaks are likely related to a combination of harsh meteorological conditions (e.g., stable boundary layers, strong heating-related emissions) and sensor stability issues during cold weather.

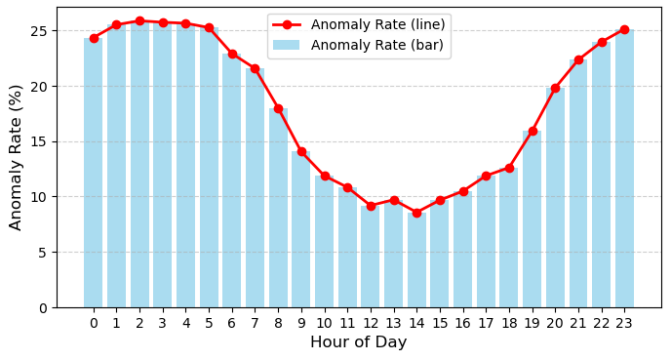


Fig. 2: Hour-of-day anomaly rate (%) grouped by time of day

Figure 2 further illustrates the anomaly rate as a function of the hour of day. The pattern shows a distinct diurnal cycle: anomaly rates are highest during late night and early morning hours (00:00–05:00, around 25%), gradually decrease during the morning hours, and reach a minimum around midday and early afternoon (12:00–15:00, below 10%). After sunset, anomaly rates rise again, exceeding 20% in the evening hours (20:00–23:00). This cycle is consistent with the influence of atmospheric boundary layer dynamics: shallow nocturnal layers amplify concentration variability and sensor sensitivity, whereas daytime mixing suppresses extremes and reduces the likelihood of anomalies [15], [16].

These findings indicate that anomalies in greenhouse gas time series are not randomly distributed but instead exhibit both seasonal and diurnal structures. Recognizing these regularities provides an opportunity to design more context-aware anomaly detection strategies, for instance by applying stricter thresholds in winter and nighttime periods while relaxing them during midday. Such refinements can further enhance the robustness of forecasting pipelines.

The proposed framework was evaluated through anomaly detection, change point detection, and forecasting experiments. The anomaly detection module identified a total of 6,465 anomalous observations, corresponding to 11.7% of the dataset. These anomalies included both short-term spikes and longer-lasting irregularities, consistent with the expected influence of sensor noise, calibration periods, and episodic meteorological events. Change point detection applied to the STL trend component revealed 380 structural breaks across the study period, highlighting frequent shifts in the long-term dynamics of atmospheric CO_2 at the IPR site.

Table II and Figure 3 summarize the forecasting performance across all models. The basic Seasonal Naive model achieved an MAE of 12.25, RMSE of 16.04, and MAPE of 2.79%. Incorporating robustness through median substitution of anomalous values substantially improved performance, reducing errors to MAE 8.05, RMSE 11.09, and MAPE 1.80%. This demonstrates that even a simple robustness mechanism can yield significant gains.

SARIMAX models further improved accuracy, with the baseline SARIMAX configuration producing an MAE of 7.64, RMSE of 9.41, and MAPE of 1.73%. The weighted version, in which anomalous values were treated as missing within the state-space framework, achieved MAE 7.00, RMSE 9.10, and MAPE 1.58%. This confirms the benefit of down-weighting anomalies in a statistical model.

The GRU neural network delivered the strongest performance overall, with MAE 2.95, RMSE 4.50, and MAPE 0.66%. When trained with the weighted Huber loss, the GRU achieved MAE 3.09, RMSE 4.75, and MAPE 0.69%, performing comparably to the unweighted model. These results suggest that the GRU architecture is inherently robust to outliers, such that additional weighting does not substantially improve predictive accuracy.

The results demonstrate that robustness mechanisms are particularly effective for simple and statistical forecasting

TABLE I: Monthly anomaly rate (%) by Year-Month

Year/Month	1	2	3	4	5	6	7	8	9	10	11	12	Year Avg
2018	25.81	17.86	22.31	10.14	10.35	17.22	19.22	17.34	16.94	12.50	23.89	26.75	18.36
2019	35.48	20.09	13.44	11.81	12.77	19.58	17.88	19.49	17.22	22.18	11.67	23.92	18.79
2020	26.08	19.83	12.37	8.61	15.99	14.17	17.88	14.78	15.00	15.99	28.19	25.00	17.82
2021	23.66	20.98	10.22	8.75	13.71	14.03	13.84	13.44	16.11	15.73	25.56	36.42	17.70
2022	35.89	19.49	12.63	9.17	13.17	12.78	11.96	12.37	13.47	14.25	20.00	26.61	16.82
2023	23.52	22.92	12.63	10.42	15.32	17.36	18.41	13.71	12.92	17.47	19.72	28.90	17.78
2024	23.12	18.82	13.44	–	–	–	–	–	–	–	–	–	18.46
Month Avg	27.65	20.00	13.86	9.82	13.55	15.86	16.53	15.19	15.28	16.35	21.50	27.93	17.96

TABLE II: Forecasting performance before and after applying robustness mechanisms.

Model	Robustness	MAE	RMSE	MAPE (%)
Seasonal Naive	Before	12.248	16.044	2.79
	After	8.050	11.087	1.80
SARIMAX	Before	7.644	9.408	1.73
	After	7.002	9.099	1.58
GRU	Before	2.952	4.504	0.66
	After	3.087	4.748	0.69

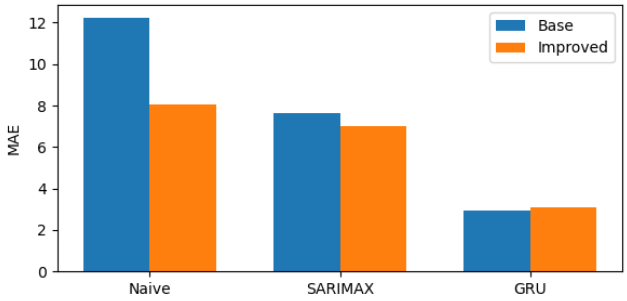
models, where they substantially reduce error. For deep learning models such as GRU, the improvements are less pronounced, suggesting that such architectures may already capture irregular patterns and mitigate their impact during training.

Figure 4 presents a visual comparison of baseline and robustness-enhanced forecasts for the three model families. The Seasonal Naive model (Figure 4a) shows clear deviations from the true CO_2 series, particularly during periods of irregular fluctuations. The robustness-enhanced version produces a noticeably smoother trajectory and aligns more closely with the observed series, indicating that substituting anomalous reference values effectively reduces forecast distortions.

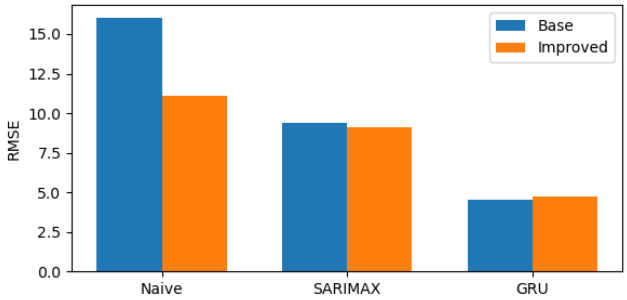
In the SARIMAX model (Figure 4b), both baseline and improved forecasts capture the main seasonal dynamics. However, the baseline model exhibits localized mismatches during volatile periods. The weighted version alleviates these discrepancies by down-weighting anomalous samples, yielding trajectories that remain closer to the true series when variability is high.

For the GRU model (Figure 4c), the forecasts from both the baseline and the robustness-enhanced variants are visually almost indistinguishable from each other, and both closely match the true CO_2 dynamics. This suggests that the GRU architecture is inherently robust to anomalies, such that additional weighting offers little visual improvement.

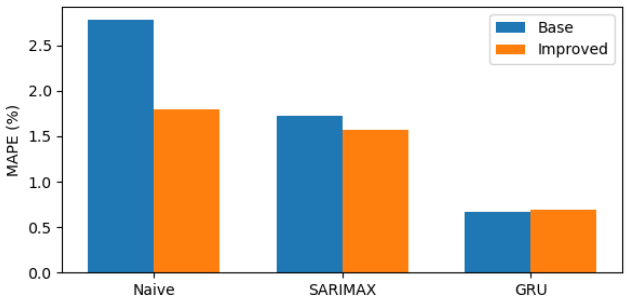
Overall, the line plots confirm that robustness mechanisms primarily benefit simpler models such as Naive and SARIMAX, while deep learning models like GRU already maintain stable predictions in the presence of anomalous observations.



(a) MAE comparison



(b) RMSE comparison

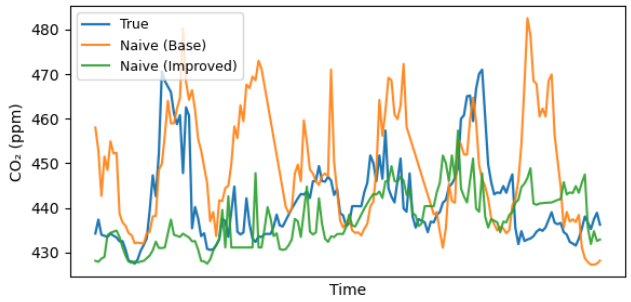


(c) MAPE comparison

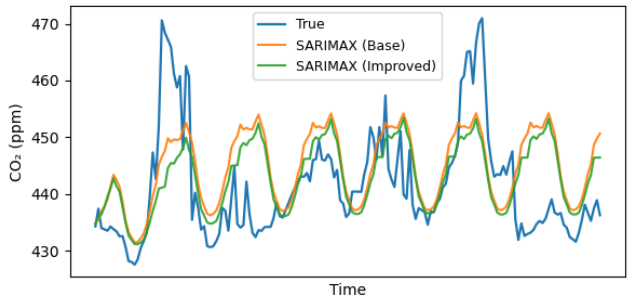
Fig. 3: Forecasting performance before (Base) and after (Improved) applying robustness mechanisms, evaluated across Seasonal Naive, SARIMAX, and GRU models.

VI. Conclusion & Future Work

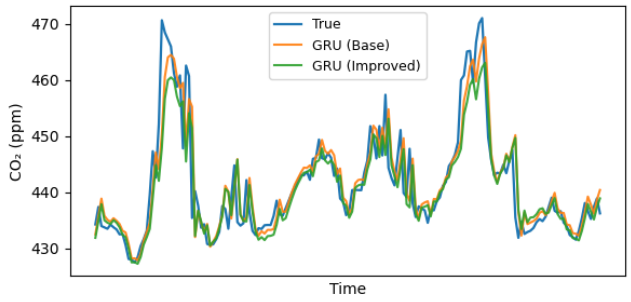
This study presented a robust framework for forecasting greenhouse gas concentrations that combines anomaly detection, change point detection, and robustness-aware forecasting models. Based on more than six years of hourly CO_2 data from the IPR station, we found that about 11.7% of the series consisted of anomalous observations and that the long-term trend contained frequent structural breaks. These characteristics motivated the development of robustness mechanisms to reduce bias and to prevent error propagation in forecasting.



(a) Naive model



(b) SARIMAX model



(c) GRU model

Fig. 4: Forecasting results of baseline (Base) and robustness-enhanced (Improved) models compared against true CO_2 values.

The experimental results showed that the introduction of robustness substantially improved the performance of simple and statistical models. The Seasonal Naive model exhibited the largest benefit, with errors reduced by more than one third after replacing anomalous reference values. The weighted SARIMAX also improved upon its baseline by lowering the influence of anomalous samples. In contrast, the GRU neural network already produced highly accurate forecasts, and the additional robustness adjustment led to only minor changes. These findings suggest that robustness mechanisms are especially important for lightweight or statistical approaches, while neural models such as GRU may already account for irregularities during training.

Future work will concentrate on two directions. The first is the inclusion of meteorological covariates such as temperature, wind speed, and boundary layer height in the forecasting framework. These variables strongly influence short-term greenhouse gas dynamics, and their integration may improve both the accuracy and the interpretability of predictions by linking observed fluctuations to underlying physical processes. The second direction is the development of probabilistic forecasting methods that provide predictive intervals in addition to point forecasts [17], [18]. By applying Bayesian state space models or quantile regression neural networks, it will be possible to quantify forecast uncertainty. This extension is essential for risk sensitive applications such as greenhouse gas monitoring, policy evaluation, and climate risk assessment.

References

[1] IPCC, *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, 2021.

[2] United Nations Framework Convention on Climate Change (UNFCCC), "National greenhouse gas inventory data," https://unfccc.int/ghg-inventories-annex-i-parties/2022, 2022, accessed: 2025-09-08.

[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018. [Online]. Available: https://arxiv.org/abs/1803.01271

[6] J. Zhang, "Efficient greenhouse gas prediction using iot data streams via cnn-bilstm-kan model," *Energy Informatics*, vol. 8, no. 1, p. 5, 2025.

[7] Y. Zhang, A. Pakrashi, and S. Dev, "Assessing interconnected factors in CO_2 emissions: A case study of india using principal component analysis," in *Proceedings of the IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 2023.

[8] ——, "Deep learning-based prediction of short-term CO_2 emissions in ireland," in *17th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2024.

[9] Y. Jin, "Machine learning for predicting urban greenhouse gas emissions: A review," *Renewable and Sustainable Energy Reviews*, 2025, in press.

[10] J. L. Hernández-Mejía, "Anomaly detection for geological carbon sequestration using lstm autoencoders and isolation forest," *Journal of Petroleum Science and Engineering*, 2024, in press.

[11] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[12] R. P. Adams and D. J. C. MacKay, "Bayesian online change-point detection," in *Proceedings of the 23rd Annual Conference on Uncertainty in Artificial Intelligence*, 2007, pp. 1–8.

[13] J. Gao, X. Song, Q. Wen, P. Wang, L. Sun, and H. Xu, "Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks," *arXiv preprint*, 2020, preprint arXiv:2002.09545.

[14] R. P. Adams and D. J. C. MacKay, "Bayesian online changepoint detection," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2007.

[15] R. B. Stull, *An Introduction to Boundary Layer Meteorology*. Springer, 1988.

[16] A. Baklanov, B. Grisogono, R. Bornstein, and et al., "Urban meteorology and atmospheric boundary layer processes," *Atmospheric Research*, vol. 132–133, pp. 363–392, 2013.

[17] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151, 2014.

[18] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "Weatherbench: A benchmark dataset for data-driven weather forecasting," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 11, p. e2020MS002203, 2020.