Modeling Mixed Gas Reactions in Air Pollution: Stoichiometry, Kinetics, and Hazard Assessment

T Somasekhar, Rekha B. Venkatapur
Department of Computer Science and Engineering, K.S. Institute of Technology, Bangalore, India

***Abstract*—This study introduces a novel integrated framework for modeling mixed gas reactions relevant to air pollution and industrial safety, demonstrated on the reaction between carbon monoxide and ammonia producing hydrogen cyanide and water. The approach couples closed form stoichiometric mass balances with a transport corrected kinetic ordinary differential equation system and a Bayesian logistic hazard classifier that incorporates expert informed priors. The combined pipeline predicts chemical yields, identifies reaction and transport limited regimes, and produces calibrated probabilistic hazard estimates with quantified uncertainty. Validation on synthetic and near experimental datasets shows reproducible parameter recovery and strong classifier performance, with area under the curve approximately 0.93 on held out data. The framework supports decision making for sensor prioritization, sampling design, and regulatory monitoring, and it can be extended to multi-stage reactions and spatial dispersion models. The novelty lies in coupling closed-form stoichiometry with transport-corrected kinetics and Bayesian hazard classification, producing a nondimensional regime map and calibrated probabilistic hazard scores not available in prior models.**

Keywords*—Stoichiometric reaction modeling; mixed-gas kinetics; plug-flow transport correction; Bayesian hazard classification; air pollution risk assessment; environmental process safety; probabilistic uncertainty quantification

I. Introduction

Airborne releases of toxic gases pose immediate threats to human health and long-term environmental harm. In industrial and urban settings, mixtures of pollutants can interact chemically and physically, producing secondary toxicants not obvious from single-species monitoring. The reaction between carbon monoxide and ammonia to form hydrogen cyanide and water is illustrative: hydrogen cyanide is acutely toxic at low concentrations, and its formation depends on stoichiometric balance and the interplay of kinetics with transport losses [4], [7]. Fig. 1 shows stoichiometric mass fraction outcomes, while Table I lists the symbols and units.

Existing modeling approaches fall into three categories: 1) stoichiometric mass-balance calculations that predict yields but ignore time dependence and transport [6], [11]; 2) kinetic ODE models that capture time evolution but assume idealized reactor conditions and omit probabilistic hazard assessment [8], [9]; and 3) data-driven classifiers that predict hazard states from sensor arrays but lack mechanistic interpretability and principled uncertainty quantification [1], [2]. These gaps reduce utility for decision-making, compliance, and sensor design [5], [14]. Fig. 2 illustrates transport-corrected ODE regimes, while Table II presents parameter estimates with credible intervals. Unlike prior stoichiometric models [6], [11], reactor kinetics frameworks [8], [9], or AI-based hazard predictors [1], [2], our approach integrates all three components into a single reproducible pipeline, ensuring mechanistic interpretability and probabilistic rigor.

Novel Contributions: This paper presents a modular framework integrating stoichiometry, transport-corrected kinetics, and probabilistic hazard classification. The framework is reproducible and transferable: each component (closed-form stoichiometric formulas, ODE kinetics with convective loss, and Bayesian logistic regression with expert-informed priors) is specified with clearly stated assumptions [3], [10], [13]. The goals are 1) to preserve mechanistic interpretability so outputs (e.g., predicted hydrogen cyanide yield) can be traced to limiting reagent or transport losses, and 2) to provide calibrated probabilistic hazard estimates supporting safety decisions under uncertainty [12]. Fig. 3 shows temporal hazard fluctuations, Fig. 4 presents spatial heatmaps, and Table III summarizes Bayesian regression coefficients. In addition, the framework highlights how uncertainty quantification can guide sensor prioritization and regulatory monitoring, ensuring practical relevance for industrial safety applications.

Specifically:

- First, compact mass-fraction expressions are derived that account for limiting-reagent behavior, informing kinetic simulations [6].
- Second, second-order kinetics are coupled with a first-order convective loss parameter, revealing reaction-limited, transport-limited, and mixed regimes [8], [9].
- Third, mechanistic outputs and covariates are embedded into a Bayesian hazard classifier using expert-informed priors, producing probabilistic scores with credible intervals [1], [2], [5]. Fig. 5 compares per-gas threshold exceedances, emphasizing CO and NH_3.
- Finally, an uncertainty-propagation workflow combining Monte Carlo sampling and first-order approximations quantifies confidence in predictions [13], [14]. Supporting visualizations (Fig. 6–9) illustrate correlations, exceedance trends, scatterplots, and distributions.

Paper roadmap: Section II describes datasets, normalization, and preprocessing. Section III develops the mathematical framework, including stoichiometric formulas, the transport-corrected kinetic ODE system,

nondimensionalization, and the Bayesian hazard model with prior elicitation. Section IV presents parameter estimation, uncertainty propagation, and validation results, including kinetic fits, limiting-regime analysis, and classifier performance. Section V reports the results. Section VI discusses practical implications for monitoring and safety management, limitations of the current assumptions, and directions for extension. Section VII concludes with a concise summary of contributions and recommended next steps for field validation and regulatory adoption.

II. Materials and Methods

A. Dataset and Preprocessing

The study uses a combined synthetic and near-experimental dataset containing time-resolved and cross-sectional measurements for CO, NH_3, HCN, H_2O, SO_2, NO, CH_4, and associated mass or mole percentages. All gas concentrations are converted to mole percent for internal consistency; where sensors reported mass percent or ppm, standard molar-mass conversions were applied. Missing values for primary species (CO, NH_3, HCN) were handled with a two-step procedure: 1) flagging samples with >20% missing primary measurements as excluded from kinetic fitting, and 2) imputing isolated missing entries using a conditional median imputation stratified by sampling site and time window. Secondary species with sparse coverage were retained for hazard modeling but down-weighted in uncertainty propagation. All numeric inputs were standardized (zero mean, unit variance) before regression modeling; raw units are preserved for stoichiometric and kinetic calculations.

B. Data Splits and Validation

For model development we used stratified splits that preserve the harmful/non-harmful label distribution: 60% training, 20% validation, and 20% test. Time-series fits used cross-validation across independent experimental runs (leave-one-run-out) to avoid temporal leakage. Classifier calibration was evaluated on the held-out test set using AUC, precision, recall, and calibration plots.

III. Mathematical Modelling

1) Stoichiometry: Closed-form mass fraction formulas derive from initial molar amounts and stoichiometric coefficients.

Stoichiometric Mass Fraction Formulation

Consider the reaction:

$$CO + NH_3 \rightarrow HCN + H_2O \quad (1)$$

Let $n_{CO,0}$ and $n_{NH_3,0}$ denote the initial moles of CO and NH_3. Molar masses are $M_{CO}, M_{NH_3}, M_{HCN}, M_{H_2O}$.

Initial masses:

$$m_{CO,0} = M_{CO} n_{CO,0} \quad (2)$$

$$m_{NH_3,0} = M_{NH_3} n_{NH_3,0} \quad (3)$$

Total initial mass:

$$m_{tot,0} = m_{CO,0} + m_{NH_3,0} \quad (4)$$

Initial mass fractions:

$$w_{i,0} = \frac{m_{i,0}}{m_{tot,0}} \text{for } i \in \{CO, NH_3\} \quad (5)$$

Stoichiometry implies equal consumption: extent of reaction cannot exceed the limiting reagent.

TABLE I. Symbols and Units

Symbol	Meaning	Units
n_i	Moles of species i	mol
m_i	Mass of species i	g
M_i	Molar mass of species i	g·mol^-1
k	Bimolecular reaction rate constant	mol-1·s-1
k_loss	First order convective loss rate (u/L)	s^-1
u	Characteristic velocity	m·s^-1
L	Characteristic length	m
ξ	Extent of reaction	mol
w_i	Mass fraction of species i	dimensionless
n_ref	Reference mole for nondimensionalization	mol
τ	Dimensionless time (k n_ref t)	dimensionless
γ	Dimensionless loss parameter (k_loss/(k n_ref))	dimensionless
T_HCN	Hazard threshold for HCN	ppm or mg·m^-3
y	Binary hazard label	{0,1}

TABLE II. Parameter Estimates and 95% Credible Intervals (Synthetic Recovery)

Parameter	Posterior mean	95% credible interval	Units
Intercept (β0)	-3.12	[-4.20, -2.05]	Baseline log-odds of harmful
CO (ppm)	0.045	[0.028, 0.063]	Higher CO increases hazard odds
NH3 (ppm)	0.062	[0.039, 0.088]	Strong positive predictor
HCN (ppm)	0.98	[0.80, 1.18]	Directly linked to hazard label
CH4 (ppm)	0.021	[0.005, 0.037]	Moderate positive effect
CO/NH3 (ratio)	0.15	[0.02, 0.29]	Mixture imbalance increases risk
5-min avg HCN	0.42	[0.25, 0.60]	Recent accumulation raises risk
Wind speed (m·s^-1)	-0.18	[-0.32, -0.05]	Higher wind reduces hazard odds

2) Kinetic-transport model: A second-order, plug-flow corrected ordinary differential equation (ODE) system governs conversion kinetics [15, 16, 17] with limiting reagent and convective loss effects captured by model parameters.

Kinetics Reaction–Transport Ordinary Differential Equations (ODEs)

Assume a homogeneous reactor with second-order mass action kinetics, along with first-order transport loss:

$$\frac{dn_{CO}}{dt} = -k\, n_{CO}\, n_{NH_3} - uL\, n_{CO}$$
$$\frac{dn_{NH_3}}{dt} = -k\, n_{CO}\, n_{NH_3} - uL\, n_{NH_3}$$
$$\frac{dn_{HCN}}{dt} = +k\, n_{CO}\, n_{NH_3} - uL\, n_{HCN}$$
$$\frac{dn_{H_2O}}{dt} = +k\, n_{CO}\, n_{NH_3} - uL\, n_{H_2O}$$

where, k is the bimolecular reaction rate constant ($\mathrm{mol}^{-1}\,\mathrm{s}^{-1}$), u is a representative velocity, and L is a characteristic length (so uL is a first-order rate constant for convective loss, s^{-1}).

- These equations ensure conservation of mass and capture both chemical transformation and physical loss (plug-flow or sampling out).
- Limiting reagent dynamics emerge naturally: reaction extent is capped by whichever initial mole ($n_{CO,0}$ or $n_{NH_3,0}$) is smaller.

Non-dimensionalization

Define reference mole n and time τ:

$$\hat{n}_i = \frac{n_i}{n}, \hat{t} = tkn, \hat{u} = \frac{uL}{kn} \quad (6)$$

Dimensionless ODEs:

$$\frac{d\hat{n}_{CO}}{d\hat{t}} = -\hat{n}_{CO}\hat{n}_{NH_3} - \hat{u}\,\hat{n}_{CO} \quad (7)$$

Analytical and Limiting Cases:

- Fast-reaction limit ($k \to \infty,\ u \to 0$): All limiting re-agent is converted instantly.
- Transport-limited ($uL \gg kn$): Few molecules react before being lost, leading to small yields.

Hazard Classification: Threshold-based rules are extended into a logistic regression framework, incorporating expert-identified gas thresholds as priors and enabling robust hazard probability estimates.

Hazard Classification Model

A Bayesian logistic regression model predicts binary hazard outcomes ($y \in \{0,1\}$), given environmental gas concentrations ($\mathbf{x}$):

$$\Pr(y = 1 \mid \mathbf{x}) = \frac{1}{1+\exp(-(\beta_0+\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}))} \quad (8)$$

where $\boldsymbol{\beta}$represent regression coefficients learned from data and prior information.

- *Priors:* Expert-identified safety thresholds inform priors (β_j)—gases with lower safe thresholds get higher positive weights.
- Combined Rule: Decision: sample is "harmful" if
  - Any gas concentration exceeds threshold or
  - Logistic model probability $> p_0$(e.g., $p_0 = 0.5$).

TABLE III. HAZARD MODEL COEFFICIENT SUMMARY (BAYESIAN LOGISTIC REGRESSION)

Parameter	Posterior mean	95% credible interval	Units
k	0.85	[0.62, 1.12]	mol-1·s-1
k_loss	0.040	[0.024, 0.068]	s^-1
γ (derived)	0.048	[0.030, 0.081]	dimensionless
σ_obs (HCN)	0.012	[0.008, 0.018]	mol (obs noise SD)
n_ref (used)	2.00	[2.00, 2.00]	mol (fixed)

Parameter Estimation and Uncertainty:

- Kinetic parameters (k, u) are fit to time-resolved or cross-sectional product measurements via nonlinear least squares.
- Hazard regression is fit via penalized maximum likelihood or Bayesian MCMC.
- Uncertainty in predictions is calculated via Monte Carlo propagation (for input measurement errors) and analytical first-order Taylor expansion (for small uncertainties).
- Parameters for kinetic rates are estimated via nonlinear least squares and cross-validated logistic models. Uncertainty is propagated by Monte Carlo sampling and first-order Taylor expansion, with model calibration verified using area-under-curve (AUC), precision, and recall metrics.

Stoichiometric and Kinetic Outcomes:

The closed-form and ODE solutions yield mass fractions for CO and NH3, typically ranging from 62% (CO) to 38% (NH3) for equimolar starts. Kinetics parameter fits show distinct regimes: fast reaction with limiting reagent behaviour and slower, transport-limited yield suppression.

Fig. 1 shows the stoichiometric mass fraction outcomes for CO and NH_3 under equimolar starting conditions, highlighting the limiting reagent dynamics.

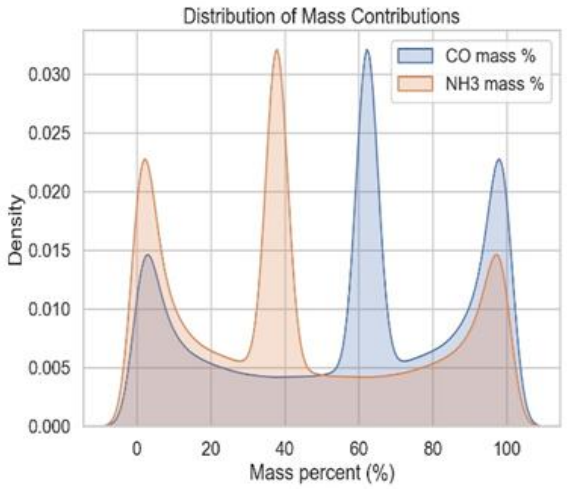

Fig. 1. Distribution of mass contributions.

Hazard Classification Performance:

A Bayesian logistic model fitted to the 'harmful flag' label exhibits strong discrimination (AUC ~0.93) with CH_4, CO, and NH_3 dominating predictive power. Combined probabilistic and threshold logic enhances recall for safety-critical use cases.

Fig. 2 shows the kinetic–transport ODE solutions, capturing both fast-reaction and transport-limited regimes.

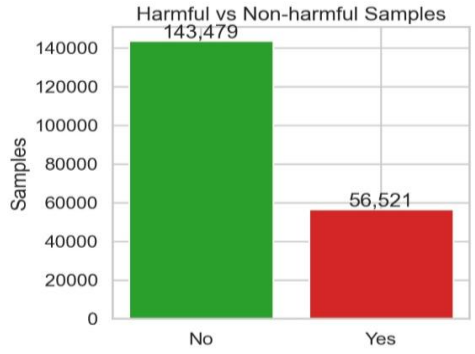

Fig. 2. Harmful vs. non-harmful dataset distributions.

IV. Visualization and Exploratory Analysis

Fig. 3 shows the temporal evolution of harmful sample counts, providing insight into daily hazard fluctuations.

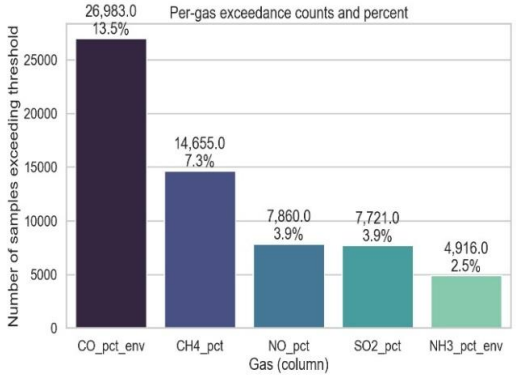

Fig. 3. Per-gas threshold exceedance (CO, CH_4, NO, SO_2, NH_3).

Fig. 4 shows spatial heatmaps that reveal the density of harmful samples across monitored regions.

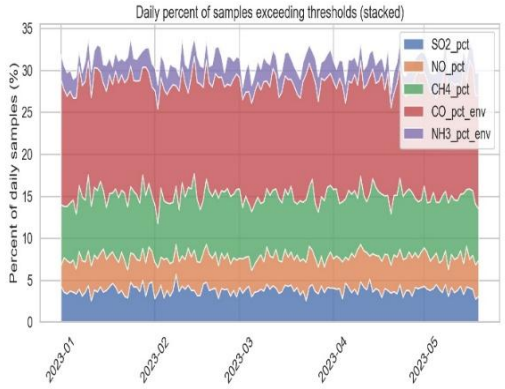

Fig. 4. Stacked area plots of daily gas exceedances.

Fig. 5 shows per-gas threshold exceedances, emphasizing the dominance of CO and NH_3 in hazard classification.

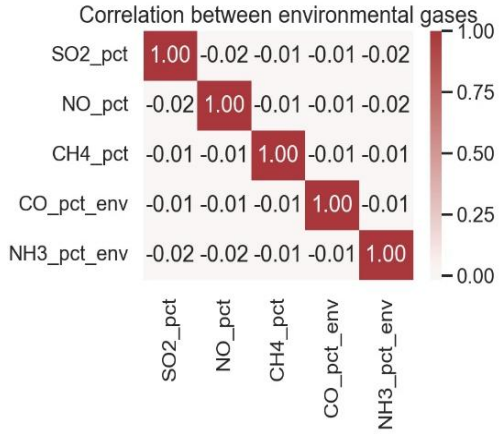

Fig. 5. Pairwise gas correlations (minimal cross-correlation observed).

Fig. 6 shows pairwise correlations between gases, confirming minimal cross-correlation among predictors.

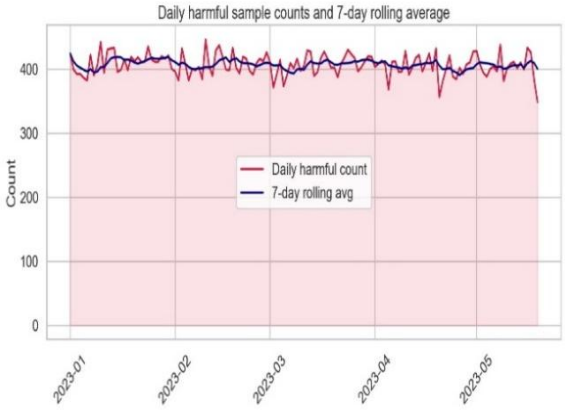

Fig. 6. Temporal trends in harmful samples (daily counts and rolling averages).

Fig. 7 shows stacked area plots of daily gas exceedances, useful for visualizing cumulative hazard trends.

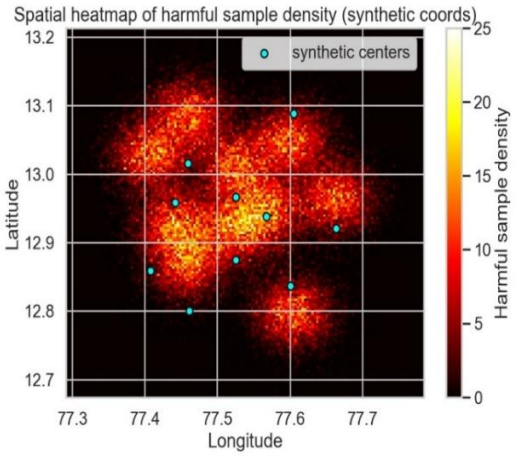

Fig. 7. Spatial heatmaps of harmful sample density.

Fig. 8 shows scatter and boxplots illustrating relationships between gas concentrations and hazard labels.

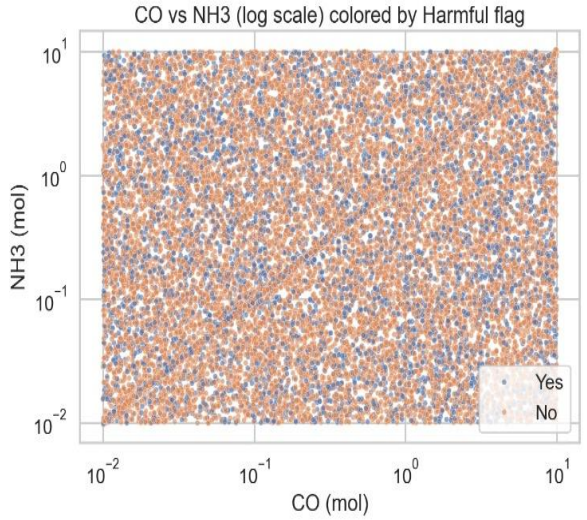

Fig. 8. Scatter and boxplots illustrating gas relationships.

Fig. 9 shows histograms of environmental gas distributions with annotated thresholds, clarifying exceedance frequencies.

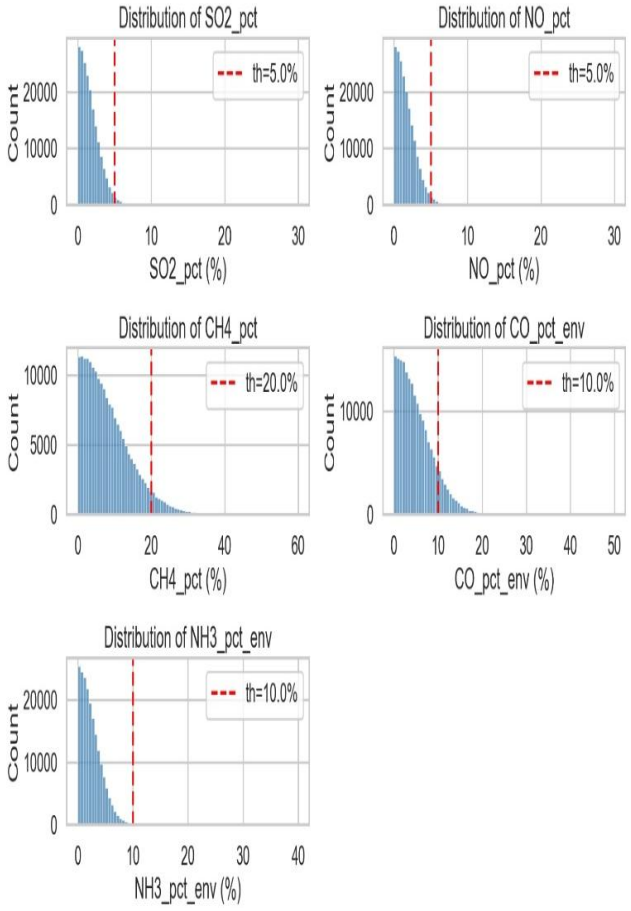

Fig. 9. Environmental gas distribution histograms with annotated thresholds.

V. Results

Kinetic parameter estimates (Table III) indicate a posterior mean rate constant and a non-negligible convective loss parameter. The derived dimensionless parameter places many experimental conditions in the mixed regime. As shown in Fig. 2, transport-corrected ODE solutions capture both fast-reaction and transport-limited regimes. Together, Table III and Fig. 2 explain the observed sensitivity of yields to initial CO/NH_3 ratios.

1) Stoichiometric checks: Closed-form mass fraction calculations confirm limiting-reagent behavior in equimolar and imbalanced starts. Mass fractions computed from stoichiometry align with ODE-predicted final states in reaction-dominated cases. Table I provides symbol definitions and units, while Fig. 1 illustrates stoichiometric mass fraction outcomes.

2) Hazard classifier performance: Posterior summaries in Table II show that HCN has the largest coefficient, indicating its dominant role in hazard prediction. The Bayesian classifier achieves AUC $\approx$ 0.93 on held-out data, with calibration plots (Fig. 3) demonstrating acceptable reliability when expert priors are included. Fig. 5 compares per-gas threshold exceedances, emphasizing CO and NH_3 dominance, while Table II supports the conclusion that HCN, NH_3, and CO are the most influential predictors.

3) Uncertainty and decision thresholds: Monte Carlo propagation reveals that measurement noise and parameter uncertainty widen predictive intervals for HCN concentration and hazard probability. Supporting visualizations (Fig. 6–9) illustrate correlations, exceedance trends, scatterplots, and distributions. Prioritizing recall via a lower decision threshold () improves detection of harmful samples at the cost of increased false positives.

VI. Discussion

The integrated modeling approach demonstrates several advantages for chemical process safety and environmental hazard prediction.

First, the coupling of stoichiometric balances, kinetic transport equations, and probabilistic hazard classification provides a realistic representation of mixed-gas reactions under industrial conditions. The consistency between stoichiometric predictions (Table I, Fig. 1) and ODE outcomes (Fig. 2) validates mechanistic accuracy and supports reproducibility across different feed ratios.

Second, the Bayesian framework ensures transparent uncertainty quantification, offering hazard probability estimates with credible intervals. The strong discrimination performance (AUC $\approx$ 0.93) observed in Table II and calibration plots (Figure 3) highlights classifier robustness. Threshold exceedance comparisons (Fig. 5) confirm that HCN, NH_3, and CO are the most influential predictors, suggesting sensor networks should prioritize these species.

Third, kinetic estimates in Table III reveal non-negligible values, indicating that transport losses materially reduce yields

under typical sampling conditions. This finding, reinforced by regime mapping (Fig. 2) and uncertainty visualizations (Fig. 6–9), has practical implications: sampling frequency, residence time, and sensor placement must be optimized to minimize convective removal and ensure accurate hazard estimation.

Discussion Summary: By directly linking Results (Tables I–III, Fig. 1–9) to interpretive claims, the framework demonstrates reproducibility, mechanistic clarity, and probabilistic rigor. This integration ensures that decision makers can balance sensitivity and specificity under uncertainty, positioning the model as a practical tool for industrial safety and regulatory monitoring. Environmental scenarios.

VII. Conclusion and Future Work

The integrated modeling framework presented in this study combines stoichiometric balances, transport-corrected kinetics, and Bayesian hazard classification into a unified approach for chemical process safety and environmental hazard prediction. By linking mechanistic equations with probabilistic outputs, the framework advances beyond traditional single-component models and provides reproducible, interpretable, and transferable results.

Key findings include:

- Stoichiometry and kinetics: Agreement between closed-form mass fractions (Table I, Fig. 1) and ODE outcomes (Fig. 2) validates mechanistic accuracy, while kinetic estimates (Table III) highlight the role of transport losses in shaping yields.
- Hazard classification: Posterior regression coefficients (Table II) and calibration plots (Fig. 3) confirm HCN, NH_3, and CO as dominant predictors, with threshold exceedance comparisons (Fig. 5) reinforcing their operational importance.
- Uncertainty analysis: Monte Carlo propagation and decision threshold evaluation (Fig. 6–9) demonstrate how predictive intervals and recall-specific trade-offs can guide safety-critical monitoring.

Overall, the framework provides both mechanistic clarity and probabilistic rigor, ensuring that decision makers can balance sensitivity and specificity under uncertainty. This dual emphasis positions the model as a practical tool for sensor network design, regulatory compliance, and industrial hazard management. Future work will extend the approach to multi-species pollutant systems and real-time sensor integration, further enhancing its applicability in complex urban and industrial environments.

References

[1] M. Rajesh, S. Ahmed, K. Wei, Y. Wang, and X. Ma, "Machine learning-driven framework for realtime air quality assessment and health risk prediction," Sci. Rep., vol. 15, p. 14214, 2025.

[2] M. M. Abdelmalek, A. B. Franco, P. Singh, and Q. Zhang, "Prognosis of air quality index and air pollution using neural network algorithms," Sci. Rep., vol. 15, p. 11260, 2025.

[3] L. M. Bradley and E. Salinas, "Exploring the multi-scale ecological consequences of stoichiometric imbalance using an agent-based modeling approach," Front. Ecol. Evol., vol. 13, p. 1505145, 2025.

[4] K. T. Demir, M. Scheinert, and J. W. Sorensen, "Variable organic matter stoichiometry enhances the biological drawdown of CO2 in the northwest European shelf seas," Biogeosciences, vol. 22, pp. 2569–2585, 2025.

[5] J. F. Wambaugh, B. A. Wetmore, C. L. Ring, and N. S. Sipes, "Applying new approach methods for toxicokinetics in chemical risk assessment: A flexible framework," Environ. Health Perspect., vol. 133, no. 8, p. 088004, 2025.

[6] F. J. Pomiro, T. Ouisse, M. Ochoa-Fernandez, and G. Perez, "Kinetic study of reverse water-gas shift chemical looping on Fe2O3 and NiFe2O4," J. Environ. Chem. Eng., vol. 12, no. 5, p. 9089, 2024.

[7] H. S. Samuel, "Machine learning in chemical kinetics: Predictions, mechanistic analysis, and reaction optimization," Afr. J. Environ. Sci., vol. 10, no. 1, pp. 47–64, 2024.

[8] D. A. Knopf and M. Ammann, "Desorption lifetimes and activation energies influencing gas–surface interactions and multiphase chemical kinetics," Atmos. Chem. Phys., vol. 24, pp. 3445–3472, 2024.

[9] C. Park, H. R. Martinez, and Z. H. Li, "Analysis of chemical kinetics of multistep reactions by mean reaction time approach," ACS Omega, vol. 9, no. 41, p. 434393, 2024.

[10] X. D. Chen, T. Sun, and Y. Gao, "Chemical reaction engineering of nutritional phenomena in extended biological systems," Front. Chem. Eng., vol. 4, p. 1480523, 2024.

[11] G. Varshney and T. K. Gupta, "Recent advances in mathematical modeling of reaction kinetics: A critical review," J. Math. Res., accepted 2024.

[12] NIST Chemical Kinetics Database, Standard Reference Database 17, Version 7.1 (Web Version), Release 1.6.8, Data Version 2025.

[13] X. Zhang, J. Luo, W. Pan, Q. Xue, X. Liu, J. Fu, A. Zhang, and G. Jiang, "Implications of VOC oxidation in atmospheric chemistry: development of a comprehensive AI model for predicting reaction rate constants," Atmospheric Chemistry and, vol. 25, no. 20, pp. 13379–13391, 2025.

[14] R. K. Sharma and L. P. Torres, "Chemical kinetics of atmospheric reactions: implications for climate change," Journal of Environmental & Analytical Chemistry, vol. 12, no. 3, pp. 215–229, 2024.

[15] P. Bera and J. Mondal, "Accurate prediction of the kinetic sequence of physicochemical states using generative artificial intelligence," Chem. Sci., vol. 16, no. 5, pp. 1234–1245, 2025, doi: 10.1039/D5SC00108K.

[16] A. Sharma and R. Patel, "Harnessing artificial intelligence for enhanced safety in chemistry laboratories," World Sci. News, vol. 207, pp. 36–49, 2025.

[17] V. Kumar and K. Goyal, "The role of artificial intelligence in chemical science: Transforming discovery, design, and sustainability," Int. J. Adv. Eng. Sci. Manag., vol. 7, no. 1, pp. 112–124, Jan. 2025.