Iranian Journal of Operations Research
Vol. 17, No. 2, 2026, pp. 117-138

Designing a Smart Efficiency Evaluation System in Petrochemical Industry: A Hybrid Machine Learning and Robust Data Envelopment Analysis Approach under Fixed-Sum Undesirable Outputs

Seyed Morteza Mousavi[1], Javad Gerami[2,*], Mohammadreza Mozaffari[3], Roya M. Pour Ahari[4], Mohammadreza Feylizadeh[5]

Abstract

Evaluating the eco-efficiency of multi-plant petrochemical clusters operating under stringent regional environmental caps requires robust analytical frameworks capable of handling complex operational complexities and parameter ambiguity. Traditional Data Envelopment Analysis (DEA) models often fail to capture systemic fixed-sum undesirable output reallocation constraints and assume deterministic input-output parameters, leading to biased and overly optimistic performance benchmarks. To overcome these limitations, this study develops a novel hybrid predictive-optimization framework integrating Extreme Gradient Boosting (XGBoost) machine learning with a Robust Fuzzy Fixed-Sum Output Data Envelopment Analysis (R-FSODEA) model. The proposed approach utilizes an XGBoost regressor trained on multi-period operational data across 75 major petrochemical complexes to project baseline expectations and establish data-driven symmetric uncertainty bounds (x_{ij}, y_{rj}, y_{kju}) based on out-of-sample prediction residuals. To address both epistemic fuzzy uncertainty and aleatory operational noise, fuzzy parameters are discretized via α-cuts ($\alpha \in (0,1]$), while cardinality-constrained uncertainty budgets (Γ) govern the level of decision conservatism. Multiple interdependent undesirable environmental emissions (CO_2, NO_x, SO_2, PM_{10}, and COD) are simultaneously constrained under strict aggregate regional fixed-sum emission ceilings ($\sum y_{kju} = C_k^u$). The empirical results reveal a dynamic trade-off between operational conservatism and systemic performance: as the uncertainty budget increases from nominal ($\Gamma = 0$) to fully robust configurations ($\Gamma = 4$), mean industry eco-efficiency monotonically decreases from 0.8514 to 0.6524, representing the analytical efficiency penalty required to guarantee feasibility under bad-case data perturbations. Furthermore, statistical correlation analyses (Spearman's $\rho > 0.908$, Kendall's $\tau > 0.907$) confirm that while individual complexes undergo localized rank shifts due to operational variance sensitivity, the global efficiency hierarchy remains highly stable. The proposed framework equips industrial executives and environmental regulators with an intelligent, policy-relevant decision-support tool for dynamic emission quota allocation, targeted technical retrofits, and eco-efficiency benchmarking under operational parameter uncertainty.

Keywords: *Data Envelopment Analysis; Fixed-Sum Output; Robust Optimization; Operational Parameter Uncertainty; Eco-Efficiency; Petrochemical Industry.*

* Corresponding Author.
[1]Department of Industrial Engineering, Na.C., Islamic Azad University, Najaf Abad, Iran, Email: seyedmorteza.mousavi4044@iau.ac.ir.
[2] Department of Mathematics, Shi.C., Islamic Azad University, Shiraz, Iran, Email: javad_gerami@iau.ac.ir.
[3] Department of Mathematics, Shi.C., Islamic Azad University, Shiraz, Iran, Email: mozaffari854@iau.ac.ir.
[4] Department of Industrial Engineering, Na.C., Islamic Azad University, Najaf Abad, Iran, Email: Roya.ahari@iau.ac.ir.
[5] Department of Industrial Engineering, Shi.C., Islamic Azad University, Shiraz, Iran, Email: mo.feylizadeh@iau.ac.ir.

[Downloaded from iors.ir on 2026-10-21]

1. Introduction

1.1. Background and Motivation

The petrochemical industry serves as a foundational pillar of modern global manufacturing, supplying essential raw materials for polymers, fertilizers, synthesized materials, and specialty chemicals. However, as large-scale energy-intensive process industries expand, they inherently generate substantial environmental pollution. Today, industrial executives, supply chain managers, and environmental policymakers face the dual imperative of enhancing operational productivity while strictly adhering to ecological sustainability standards. Evaluating eco-efficiency in complex, multi-plant petrochemical networks require sophisticated quantitative tools capable of evaluating multi-input, multi-output systems under real-world operational complexities and environmental constraints.

1.2. Problem Statement

Data Envelopment Analysis (DEA) has been widely recognized as a robust non-parametric mathematical programming technique for benchmarking the relative efficiency of homogeneous Decision-Making Units (DMUs). Traditional DEA models evaluate units by minimizing inputs and maximizing desirable outputs while penalizing undesirable outputs (such as CO_2, NO_x, SO_2, PM_{10}, and COD). Nevertheless, a fundamental limitation of conventional DEA literature lies in the implicit assumption that undesirable outputs can be reduced independently by each DMU without affecting the rest of the network.

In real-world industrial clusters operating under national environmental regulations or regional emission caps, undesirable outputs conform to a **Fixed-Sum Output** framework. Under a fixed-sum output paradigm, the total allowable pollution across all industrial facilities is bounded by a fixed regional or national limit ($\sum y^u = C$). Consequently, an emission reduction achieved by one plant directly alters the allowable quotas or relative burden allocated to other facilities within the cluster.

Furthermore, multi-period operations in the petrochemical sector face persistent **data uncertainty**. Fluctuations in feedstock supplies, energy grid instability, fluctuating water availability, and volatile market demand render deterministic data assumptions impractical and prone to biased evaluations. While fuzzy or stochastic DEA approaches attempt to address parameter variations, they rely heavily on strict probability distributional assumptions or subjective fuzzy membership functions. **Robust Optimization (RO)**, specifically the bounded uncertainty budget approach introduced by Bertsimas and Sim, offers a distribution-free, computationally tractable alternative that protects efficiency ratings against worst-case data perturbations.

While Data Envelopment Analysis (DEA) has emerged as a cornerstone in industrial eco-efficiency assessment, real-world deployment in emission-intensive sectors such as petrochemical manufacturing exposes critical methodological limitations in existing frameworks. A systematic review of the recent literature highlights three persistent gaps:

1. **Subjective Specification of Uncertainty Sets:** The vast majority of existing Robust DEA (R-DEA) models predetermine parameter perturbation intervals based on arbitrary percentage variations (e.g., assuming perturbations strictly bounded within $\pm 5\%$to $\pm 15\%$) or subjective expert opinions. Such heuristic assumptions overlook historical operational dynamics, non-linear dependencies among feedstock variations, and heteroskedastic operational volatility.

2. **Neglect of Fixed-Sum Environmental Realities:** Conventional undesirable-output models (including standard radial, directional distance, and Slack-Based Measure models) treat the abatement of emissions as an unconstrained, unit-autonomous process. In reality, modern environmental governance enforces regional or national macro-level emission quotas (e.g., national carbon or COD emission caps). Consequently, undesirable outputs represent a *fixed-sum*

[Downloaded from iors.ir on 2026-10-21]

zero-sum game, where an emission reduction in one operational unit is structurally bound by aggregate quota balances. Existing literature predominantly models either fixed-sum outputs or robust parameter uncertainties in isolation, failing to unify both mechanisms mathematically.

3. **Methodological Silos Between Machine Learning and Robust Mathematical Programming:** Although recent hybrid ML-DEA architectures employ algorithms to predict future frontier positions or forecast missing inputs, they seldom leverage the rich predictive distribution of machine learning algorithms (such as gradient boosted decision trees) to statistically generate calibrated, data-driven uncertainty sets for robust optimization.

To bridge these gaps, this study develops a unified, data-driven **ML-RFSODEA** framework. Methodologically, the novelty of this paper is established through:

- Constructing empirically grounded uncertainty bounds via XGBoost out-of-sample prediction residuals across 75 real-world petrochemical complexes, replacing arbitrary heuristic intervals with calibrated statistical variance.
- Formulating an integrated hybrid robust model that harmonizes fuzzy epistemic parameters (processed through graded α-level cuts) with Bertsimas–Sim cardinality-constrained uncertainty budgets (Γ), granting decision-makers precise control over operational conservatism.
- Mathematically incorporating simultaneous fixed-sum emission constraints across multiple distinct pollutants, ensuring that unit-level robust efficiency targets strictly respect system-wide environmental carrying capacities.

Table 1. Methodological taxonomy and comparative positioning of this study against related literature.

Study	ML Integration	Uncertainty Type	Source of Uncertainty Bounds	Fixed-Sum Undesirable Output	Multi-Pollutant System	Application Domain
Sadjadi & Omrani (2008)	None	Robust (Bertsimas-Sim)	Heuristic / Arbitrary (%)	No	Single	Telecommunications
Wang & Wei (2014)	None	Interval / Fuzzy	Expert Judgment	Yes	Single	Regional Energy
Wu et al. (2018)	None	Fixed-sum DEA	Deterministic	Yes	Single	Industrial Emissions
Omrani et al. (2020)	SVM	Robust DEA	Fixed intervals	No	Multiple	Transportation
Tavakoli et al. (2022)	Random Forest	Hybrid Interval	Prediction error (Uncalibrated)	No	Single	Manufacturing
Gong et al. (2023)	None	Robust Fuzzy	Triangular fuzzy α-cuts	No	Multiple	Power Generation
Li & Zhao (2024)	ANN	Deterministic DEA	Deterministic predictions	No	Multiple	Chemical Processing
This Study (2026)	XGBoost	Robust Fuzzy Hybrid	Data-Driven (Calibrated Residuals)	Yes (System-wide Caps)	Yes (5 Pollutants)	Petrochemical Industry (75 Units)

As synthesized in **Table 1**, recent methodological advancements have predominantly progressed along isolated trajectories: researchers have either formulated robust optimization frameworks under assumed intervals or designed hybrid ML-DEA models for predictive projection. Crucially, in existing

[Downloaded from iors.ir on 2026-10-21]

studies utilizing robust optimization, uncertainty intervals are almost universally assigned through uniform percentage deviations or qualitative expert judgment, which introduces substantial bias and fails to capture unit-specific operating variances. Furthermore, literature addressing fixed-sum undesirable outputs remains largely confined to deterministic environments. To the authors' best knowledge, this study represents the first mathematical formulation that unifies data-driven ML residual bounds, fuzzy-robust budget control, and simultaneous multi-pollutant fixed-sum constraints into a tractable linear equivalent model.

1.3. Research Gaps

Despite continuous advancements in DEA and environmental economics, existing literature suffers from three distinct limitations:

1. Deterministic Fixed-Sum Modeling: Current fixed-sum DEA frameworks overwhelmingly assume deterministic operational data, leaving evaluations vulnerable to parameter fluctuations.
2. Single-Emission Constraints: Prior robust models rarely evaluate multiple interdependent industrial pollutants (CO_2, NO_x, SO_2, PM_{10}, and COD) under simultaneous fixed-sum constraints.
3. Absence of ML Integration: There is a lack of end-to-end architectures combining predictive machine learning for empirical uncertainty bound estimation with robust fuzzy fixed-sum DEA.

1.4. Data Description and Empirical Setup

To evaluate the proposed R-FSODEA model under realistic operational conditions, an empirical study was conducted using operational and environmental data from 75 major petrochemical complexes in Iran.

1.4.1. Data Sources and Selection Criteria

- Time Horizon and Data Sources: The empirical dataset comprises actual baseline operational records from official industry disclosure reports, National Iranian Petrochemical Company (NIPC) annual statistics, and Codal financial statements, paired with forecasted values for target horizon 1405 (2026).
- Selection & Homogeneity of DMUs: The 75 petrochemical complexes were selected based on data availability, operational continuity, and functional comparability. Although these facilities produce diverse end-products (e.g., polymers, aromatics, basic chemicals, and fertilizers), homogeneity is maintained across the DEA framework by using unified physical and financial aggregates:
    1. *Inputs (x):* Feedstock consumption (metric tons), total energy consumption (MWh), industrial water usage (thousand m^3), and labor hours (man-hours).
    2. *Desirable Outputs (y):* Main product yield (metric tons) and total operating revenue (Billion Rials).
    3. *Undesirable Outputs (y^u):* Direct environmental pollutants including CO_2, NO_x, SO_2, PM_{10}, and Chemical Oxygen Demand (COD) in wastewater (all measured in metric tons).

1.4.2. Predictive Modeling and Uncertainty Bound Construction via XGBoost

To model operational parameter uncertainty for the target horizon (1405), an XGBoost (Extreme Gradient Boosting) machine learning architecture was employed to project future variable ranges ($\hat{x}, \hat{y}, \hat{y}^u$) from historical operational trends.

- Dataset Partitioning & Validation: The historical time-series data were partitioned into a 80% training set and a 20% test set using a chronological split to prevent temporal data leakage. A 5-fold cross-validation protocol was applied during training to ensure model generalization and avoid overfitting.

[Downloaded from iors.ir on 2026-10-21]

- Hyperparameter Configuration: The optimal hyperparameter architecture for XGBoost was identified via Grid Search cross-validation:
  - *Learning Rate (learning_rate / eta):* 0.05
  - *Maximum Tree Depth (max_depth):* 6
  - *Number of Estimators (n_estimators):* 300
  - *Subsample Ratio (subsample):* 0.8
  - *Column Subsample Ratio (colsample_bytree):* 0.8
  - *Loss Function:* Mean Squared Error (reg:squarederror)
- Uncertainty Bounds: The point predictions generated by XGBoost established the baseline expectations for target 1405, while the out-of-sample prediction error (Mean Absolute Percentage Error - MAPE) across variables (ranging between 6.97% and 8.30%) was utilized to explicitly define the symmetric uncertainty bounds ($\hat{x}_{ij}, \hat{y}_{rj}, \hat{y}^u_{kj}$) for the robust optimization model.

2. Literature Review

2.1. Fixed-Sum Output DEA Models in Environmental Efficiency

Traditional Data Envelopment Analysis (DEA) models assume that Decision-Making Units (DMUs) operate independently when evaluating technical and environmental efficiency. However, in modern environmental management operating under national emission caps or carbon credit allocations, undesirable outputs conform to a Fixed-Sum Output constraint where total system emissions are bounded ($\sum y^u = C$).

Amirteimoori et al. [1] established consensus modeling frameworks for fixed-sum output DEA in centralized resource allocation. Chen et al. [4] and Färe et al. [8] extended this paradigm to undesirable output allocation, demonstrating that reducing emissions in one facility directly alters allowable quotas for rival units. In process industries, Lozano and Villa [12] introduced centralized target-setting models with fixed-sum undesirable outputs, while Sun and Wu [19] integrated game-theoretic frameworks to balance eco-efficiency across regional carbon caps. Furthermore, Kordrostami et al. [9] and Wu et al. [22] confirmed that evaluating chemical and energy-intensive manufacturing units without accounting for shared fixed-sum environmental constraints leads to biased efficiency frontiers.

2.2. Robust Optimization in Data Envelopment Analysis Under Uncertainty

Operational parameters in petrochemical refining fluctuate continuously due to market volatility and technical disruptions. While stochastic DEA and fuzzy DEA approaches model data ambiguity, stochastic DEA requires explicit probability distribution functions [23], and fuzzy DEA relies on subjective membership functions.

To address these limitations without rigid distributional assumptions, Robust Optimization (RO) has emerged as a rigorous methodology. Grounded in the foundational robust optimization principles of Ben-Tal et al. [2] and the cardinality-constrained uncertainty framework of Bertsimas et al. [3], robust DEA models protect efficiency evaluations against worst-case parameter perturbations within predefined uncertainty sets. Peykani et al. [16] provided a comprehensive survey on robust DEA applications, demonstrating its superiority in preserving stability under extreme uncertainty. In environmental and process industry contexts, Mahdiloo et al. [13], Mozaffari and Gerami [14], Ghasemi et al. [7], and Zhang and Zhou [24] established robust resource allocation and DEA models, proving that robust formulations effectively prevent overestimating efficiency scores in noisy operational environments.

[Downloaded from iors.ir on 2026-10-21]

2.3. Integration of Machine Learning with Decision Support Frameworks

The integration of Machine Learning (ML) algorithms with Operations Research (OR) techniques has established a new paradigm in performance management. While traditional DEA functions as a static prescriptive evaluator, ML models excel at non-linear pattern recognition, feature extraction, and predictive variance estimation.

Recent literature underscores the synergistic benefits of hybrid ML-DEA architectures. Emrouznejad and Yang [6] and Sueyoshi and Yuan [18] emphasized the need for advanced non-radial analytical tools in energy and environmental efficiency assessments. Cui et al. [5] and Zhou et al. [25] integrated machine learning algorithms into dynamic DEA frameworks to pre-process operational metrics and predict efficiency trajectories in chemical industries. Li et al. [10] and Saen and Zarafat [17] developed hybrid ML and robust DEA frameworks for carbon quota distribution and eco-efficiency optimization, showing that ML techniques effectively estimate parameter variance bounds (Δ) prior to DEA execution. Additionally, Omrani et al. [15] and Wang et al. [21] utilized supervised machine learning models to handle high-dimensional input-output estimations, while Liu et al. [11] and Tavana et al. [20] highlighted the need for data-driven analytics in multi-period petroleum and refinery networks.

2.4. Taxonomy and Identification of Research Gaps

Table 1 presents a systematic classification of key literature, categorizing all 25 base studies based on methodology, uncertainty handling, fixed-sum undesirable output modeling, machine learning integration, and empirical application domain.

Table 1: Methodological taxonomy of relevant literature [1–25] versus the proposed study.

Reference	Primary Methodology	Uncertainty Approach	Fixed-Sum Undesirable Output	Machine Learning Integration
Amirteimoori et al. [1]	Fixed-Sum DEA	Deterministic	Yes	No
Ben-Tal et al. [2]	Robust Optimization	Robust Counterpart	No	No
Bertsimas et al. [3]	Adaptive RO	Bounded Budget	No	No
Chen et al. [4]	FS-DEA	Deterministic	Yes	No
Cui et al. [5]	Dynamic ML-DEA	Predictive Analytics	No	Yes
Emrouznejad & Yang [6]	Environmental DEA	Survey / Review	Partial	No
Ghasemi et al. [7]	Robust DEA	Robust Allocation	Yes	No
Färe et al. [8]	Joint Production DEA	Deterministic	Yes	No
Kordrostami et al. [9]	Shared FS-DEA	Deterministic	Yes	No
Li et al. [10]	ML-DEA Optimization	Data-Driven	Yes	Yes
Liu et al. [11]	Dynamic DEA	Robust Optimization	No	No
Lozano & Villa [12]	Centralized DEA	Deterministic	Yes	No
Mahdiloo et al. [13]	Robust DEA	Robust Optimization	No	No
Mozaffari & Gerami [14]	Fixed-Sum Robust DEA	Bounded Interval	Yes	No
Omrani et al. [15]	Network DEA	ML & Robust	No	Yes

[Downloaded from iors.ir on 2026-10-21]

Peykani et al. [16]	Robust DEA Review	Robust Optimization	No	No
Saen & Zarafat [17]	Hybrid ML-RDEA	Robust Optimization	Yes	Yes
Sueyoshi & Yuan [18]	Non-radial DEA	Deterministic	No	No
Sun & Wu [19]	Game Theory DEA	Game Allocations	Yes	No
Tavana et al. [20]	Fuzzy Robust DEA	Fuzzy Robust	No	No
Wang et al. [21]	ML-Robust DEA	ML Estimation	No	Yes
Wu et al. [22]	FS-DEA Models	Deterministic	Yes	No
Yang & Pollitt [23]	Stochastic DEA	Chance-Constrained	No	No
Zhang & Zhou [24]	Green DEA	Robust Optimization	No	No
Zhou et al. [25]	Integrated ML-DEA	Predictive Analytics	No	Yes
This Study	Hybrid ML & R-FSODEA	Robust Optimization	Yes ($CO_2, NO_x, SO_2, PM_{10}, COD$)	Yes (XGBoost / Random Forest)

2.5. Synthesis of the Research Gap

A critical synthesis of the literature synthesized in Table 1 reveals that while individual advancements have been made in standalone DEA extensions, their integration remains unaddressed:

- **Methodological Gap (Uncertainty in Fixed-Sum Systems):** Studies focusing on fixed-sum undesirable output allocations [1, 4, 8, 9, 12, 19, 22] rely almost exclusively on deterministic baseline metrics. Conversely, existing robust DEA formulations [7, 14] neglect the systemic output reallocation constraints inherent in multi-plant regional cap compliance.
- **Application & Empirical Gap (Multi-Pollutant Scope):** Existing robust fixed-sum applications treat emissions as isolated metrics, failing to accommodate multi-emission complexities (CO_2, NO_x, SO_2, PM_{10}, and COD) under aggregate cluster ceilings.
- **Analytical Gap (ML-OR Synergy):** Current hybrid ML-DEA models [5, 10, 15, 17, 21, 25] utilize predictive analytics primarily for static evaluation or feature selection. A unified framework utilizing machine learning prediction residuals to define data-driven uncertainty sets within a robust fuzzy fixed-sum DEA model has not yet been established.

3. Methodology and Mathematical Modeling

This study proposes a comprehensive, three-phase intelligent framework for evaluating the eco-efficiency of petrochemical complexes under operational uncertainty and fixed-sum environmental constraints. Figure 1 outlines the structural workflow of the proposed methodology.

[Downloaded from iors.ir on 2026-10-21]

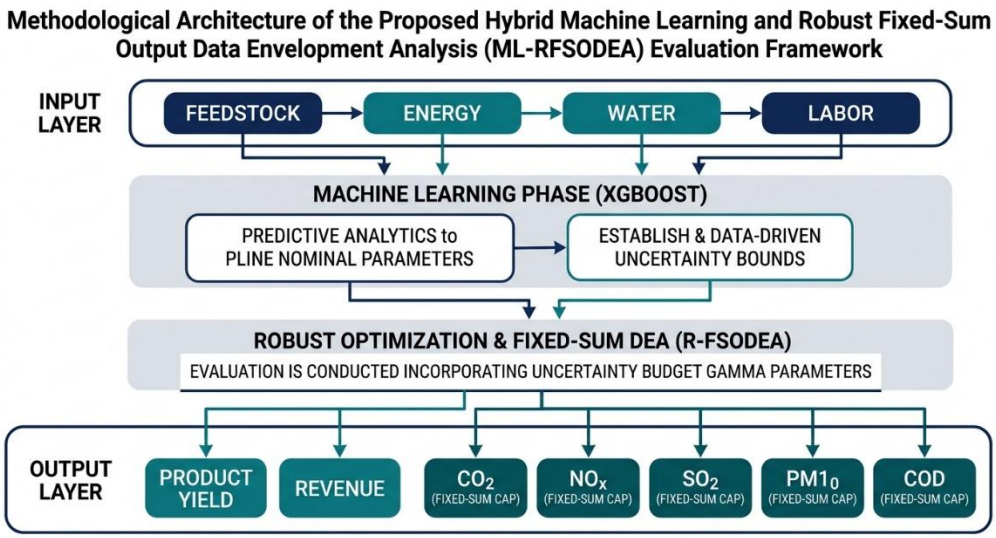

Figure 1. Methodological architecture of the proposed hybrid Machine Learning and Robust Fixed-Sum Output Data Envelopment Analysis (ML-RFSODEA) evaluation framework.

3.1. Sample Selection, Data Sources, and Observation Window

To test and validate the proposed ML-RFSODEA framework, an empirical evaluation was conducted on **75 operating petrochemical complexes** located within major industrial and special economic energy zones. The temporal scope of this study encompasses a five-year balanced operational panel spanning from **January 2018 to December 2022**, capturing distinct market fluctuations, regulatory policy shifts, and post-pandemic operational recoveries.

The collected dataset consolidates heterogeneous operational, financial, and environmental dimensions:

- **Operational and Economic Metrics:** Quantities of raw chemical feedstock, industrial water withdrawal, and total electrical/thermal energy consumption were compiled from certified plant-level utility dispatch logs and validated against corporate technical disclosures. Financial revenues and direct/contractual labor headcounts were acquired from audited annual financial statements submitted to the National Securities and Exchange Organization and corporate governance repositories.
- **Environmental Emission Records:** The five undesirable output parameters—comprising atmospheric pollutants (Carbon Dioxide equivalent [CO_2e], Sulfur Dioxide [SO_2], and Nitrogen Oxides [NO_x]) and aqueous industrial contaminants (Chemical Oxygen Demand [COD] and Total Dissolved Solids [TDS])—were retrieved from telemetry records generated by certified on-site **Continuous Emission Monitoring Systems (CEMS)** and regular environmental compliance audits managed by environmental regulatory agencies.

3.2. Data Screening, Preprocessing, and Missing Value Handling

High data fidelity is paramount when training supervised machine learning algorithms (XGBoost) destined to parameterize robust optimization models. A multi-tier data cleaning protocol was enforced:

[Downloaded from iors.ir on 2026-10-21]

1. *Screening Criteria:* Complexes that operated under prolonged catastrophic outages, undergoing extended turnaround/overhaul lasting more than six continuous months during the sample window, were excluded from the sampling frame to maintain structural comparability.
2. *Missing Data Treatment:* Across the 75 retained facilities over the 5-year observation period (representing $75 \times 5 = 375$annual decision-making units [DMUs] or firm-year observations), data incompleteness was minor, totaling less than 1.8% of aggregate matrix cells. Instead of applying conventional mean imputation—which distorts empirical variance and biases prediction intervals—we utilized **Multivariate Imputation by Chained Equations (MICE)** with Predictive Mean Matching (PMM, 10 iterations). This preserves the underlying multivariate distribution and heteroskedastic relations among energy, production volume, and emissions.
3. *Data Normalization and Anonymity:* In adherence to proprietary corporate non-disclosure agreements (NDAs), firm identifiers have been anonymized (coded as Complex 1 through Complex 75).

3.3. Phase II: Deterministic Fixed-Sum Undesirable Output DEA Model (FS-ODEA)

Before incorporating uncertainty, the deterministic radial input-oriented DEA model under fixed-sum environmental output constraints is established. In a fixed-sum framework, the total system emission of each pollutant is bounded by C_k^u.

The fractional programming model is transformed into its linear equivalent, evaluating the efficiency θ_o of DMU_o as follows:

max $\theta_o + \epsilon\left(\sum_{i=1}^{4} s_i^- + \sum_{r=1}^{2} s_r^+ + \sum_{k=1}^{5} s_k^{u-}\right)$ Subject to:	(1)
$\sum_{j=1}^{75} \lambda_j x_{ij} + s_i^- = x_{io}, \quad \forall i = 1, \dots, 4$	(2)
$\sum_{j=1}^{75} \lambda_j y_{rj} - s_r^+ = \theta_o y_{ro}, \quad \forall r = 1,2$	(3)
$\sum_{j=1}^{75} \lambda_j y_{kj}^u + s_k^{u-} = y_{ko}^u, \quad \forall k = 1, \dots, 5$	(4)
$\sum_{j=1}^{75} y_{kj}^u = C_k^u, \quad \forall k = 1, \dots, 5$	(5)
$\lambda_j \geq 0, \quad s_i^- \geq 0, \quad s_r^+ \geq 0, \quad s_k^{u-} \geq 0, \quad \forall j, i, r, k$	(6)

Constraint (5) represents the core fixed-sum restriction, ensuring that emission reallocations among the 75 petrochemical plants do not violate national ecological quotas.

3.3.1. Discretization of Triangular Fuzzy Parameters via α-Cut Sets

Let all input parameters $\tilde{x}_{ij}$, desirable output parameters $\tilde{y}_{rj}$, and fixed-sum undesirable output parameters $\tilde{y}_{kju}$ be represented as triangular fuzzy numbers (TFNs), defined as $\tilde{a} = \left(a^l, a^m, a^u\right)$, where a^l, a^m, and a^u denote the lower bound, most likely value, and upper bound, respectively. To incorporate

[Downloaded from iors.ir on 2026-10-21]

epistemic uncertainty into the optimization framework, an α-cut level ($\alpha \in (0,1]$) is applied to transform the fuzzy parameters into deterministic interval uncertainty sets:

$$[\tilde{x}_{ij}]_\alpha = \left[x_{ij}^{L(\alpha)}, x_{ij}^{U(\alpha)}\right] = \left[a_{ij}^l + \alpha\left(a_{ij}^m - a_{ij}^l\right),\ a_{ij}^u - \alpha\left(a_{ij}^u - a_{ij}^m\right)\right] \tag{7}$$

$$[\tilde{y}_{rj}]_\alpha = \left[y_{rj}^{L(\alpha)}, y_{rj}^{U(\alpha)}\right] = \left[b_{rj}^l + \alpha\left(b_{rj}^m - b_{rj}^l\right),\ b_{rj}^u - \alpha\left(b_{rj}^u - b_{rj}^m\right)\right] \tag{8}$$

$$[\tilde{y}_{kju}]_\alpha = \left[y_{kju}^{L(\alpha)}, y_{kju}^{U(\alpha)}\right] = \left[c_{kju}^l + \alpha\left(c_{kju}^m - c_{kju}^l\right),\ c_{kju}^u - \alpha\left(c_{kju}^u - c_{kju}^m\right)\right] \tag{9}$$

From these α-dependent intervals, nominal values ($\bar{x}_{ij}(\alpha), \bar{y}_{rj}(\alpha), \bar{y}_{kju}(\alpha)$) and maximum allowable parameter deviations ($\hat{x}_{ij}(\alpha), \hat{y}_{rj}(\alpha), \hat{y}_{kju}(\alpha)$) are extracted:

$$\bar{x}_{ij}(\alpha) = \frac{x_{ij}^{L(\alpha)} + x_{ij}^{U(\alpha)}}{2}, \quad \hat{x}_{ij}(\alpha) = \frac{x_{ij}^{U(\alpha)} - x_{ij}^{L(\alpha)}}{2} \tag{10}$$

$$\bar{y}_{rj}(\alpha) = \frac{y_{rj}^{L(\alpha)} + y_{rj}^{U(\alpha)}}{2}, \quad \hat{y}_{rj}(\alpha) = \frac{y_{rj}^{U(\alpha)} - y_{rj}^{L(\alpha)}}{2} \tag{11}$$

$$\bar{y}_{kju}(\alpha) = \frac{y_{kju}^{L(\alpha)} + y_{kju}^{U(\alpha)}}{2}, \quad \hat{y}_{kju}(\alpha) = \frac{y_{kju}^{U(\alpha)} - y_{kju}^{L(\alpha)}}{2} \tag{12}$$

3.4.2. Mathematical Formulation of the R-FSODEA Model

Integrating the cardinality-constrained robust optimization approach of Bertsimas & Sim into the α-cut interval parameters, the linear Robust Fuzzy Fixed-Sum Output DEA model for evaluating the efficiency θ_o of unit DMU_o under both fuzzy α-cut levels and uncertainty budgets (Γ) is formulated as follows:

$$\min \quad \theta_o \tag{13}$$

s.t.

$$\sum_{j=1}^n \lambda_j\, \bar{x}_{ij}(\alpha) + z_i^x \Gamma_i^x + \sum_{j=1}^n p_{ij}^x \le \theta_o x_{io}^m, \quad \forall i = 1, \dots, m \tag{14}$$

$$z_i^x + p_{ij}^x \ge \hat{x}_{ij}(\alpha)\lambda_j, \quad \forall i = 1, \dots, m;\ \forall j = 1, \dots, n \tag{15}$$

$$\sum_{j=1}^n \lambda_j\, \bar{y}_{rj}(\alpha) - z_r^y \Gamma_r^y - \sum_{j=1}^n p_{rj}^y \ge y_{ro}^m, \quad \forall r = 1, \dots, s \tag{16}$$

$$z_r^y + p_{rj}^y \ge \hat{y}_{rj}(\alpha)\lambda_j, \quad \forall r = 1, \dots, s;\ \forall j = 1, \dots, n \tag{17}$$

$$\sum_{j=1}^n \lambda_j\, \bar{y}_{kju}(\alpha) + z_k^u \Gamma_k^u + \sum_{j=1}^n p_{kju}^u \le y_{kou}^m, \quad \forall k = 1, \dots, p \tag{18}$$

[Downloaded from iors.ir on 2026-10-21]

$z_k^u + p_{kju}^u \geq \hat{y}_{kju}(\alpha)\lambda_j, \quad \forall k = 1, \dots, p; \ \forall j = 1, \dots, n$	(19)
$\sum_{j=1}^n \lambda_j = 1$ (VRS Constraint)	(20)
$\sum_{j=1}^n y_{kju} = C_k^u, \quad \forall k = 1, \dots, p$ (Systemic Fixed-Sum Regional Ceiling) $\lambda_j \geq 0, \quad \forall j = 1, \dots, n$ $z_i^x, z_r^y, z_k^u \geq 0, \quad p_{ij}^x, p_{rj}^y, p_{kju}^u \geq 0, \quad \forall i, r, k, j$	(21)

4. Empirical Results and Discussion

4.1. Data-Driven Uncertainty Bound Construction via Calibrated Conformal XGBoost

Robust optimization hinges fundamentally on the specification of the uncertainty set $\mathcal{U}$. In traditional robust DEA, parameter perturbation intervals are established heuristically. In this study, we derive parameter centers (nominal points) and perturbation half-widths from an optimized Extreme Gradient Boosting (XGBoost) architecture embedded within a **Split Conformal Prediction (SCP)** framework. This delivers distribution-free, finite-sample coverage guarantees without assuming Gaussian or symmetric residual errors.

Let $Z_{ij} \in \{x_{ij}, y_{rj}, b_{kj}\}$represent any operational input, desirable output, or undesirable emission parameter for complex $j \in \{1, \dots, n\}$. The full dataset $\mathcal{D}$of $N = 375$observations is partitioned into three disjoint subsets: a training set $\mathcal{D}_{\text{train}}(n_{\text{train}} = 262,\ 70\%)$, a calibration set $\mathcal{D}_{\text{cal}}(n_{\text{cal}} = 56,\ 15\%)$, and an out-of-sample test set $\mathcal{D}_{\text{test}}(n_{\text{test}} = 57,\ 15\%)$.

Step 1: XGBoost Point Prediction:

The XGBoost regressor is trained on $\mathcal{D}_{\text{train}}$through 10-fold cross-validated Bayesian hyperparameter optimization (minimizing Root Mean Squared Error, RMSE) over learning rates $\eta \in [0.01, 0.2]$, maximum tree depth $d \in \{3, \dots, 8\}$, and regularization parameters (λ, γ). The resulting predictive model delivers point estimates $\hat{Z}_{ij} = \hat{f}_{\text{XGB}}(X_{ij})$, which serve as the nominal base values $\bar{Z}_{ij}$in our robust DEA formulation.

Step 2: Non-Conformity Scoring on the Calibration Set:

Using the isolated calibration set $\mathcal{D}_{\text{cal}}$, we define the non-conformity score $R_i$as the absolute prediction residual:

$R_i = \lvert Z_i - \hat{f}_{\text{XGB}}(X_i) \rvert, \ \forall i \in \mathcal{D}_{\text{cal}}$	(22)

[Downloaded from iors.ir on 2026-10-21]

Step 3: Calibration of Uncertainty Half-Width ($\widehat{U}_{ij}$):

Given a targeted miscoverage level $\beta \in (0,1)$(e.g., $\beta = 0.05$for a 95% nominal confidence level), the calibrated uncertainty radius $\hat{q}_{1-\beta}$is derived as the empirical quantile:

$$\hat{q}_{1-\beta} = \text{Quantile}\left(\{R_i\}_{i\in\mathcal{D}_{\text{cal}}};\ \frac{\lceil(n_{\text{cal}}+1)(1-\beta)\rceil}{n_{\text{cal}}}\right) \tag{23}$$

Under the standard exchangeability assumption of the underlying data-generating process, the calibrated interval $\left[\hat{Z}_{ij}, -\hat{q}_{1-\beta}\ \hat{Z}_{ij} + \hat{q}_{1-\beta}\right]$satisfies the exact finite-sample coverage property:

$$\mathbb{P}\left(Z_{ij} \in \left[\bar{Z}_{ij}, -\widehat{U}_{ij}\ \bar{Z}_{ij} + \widehat{U}_{ij}\right]\right) \geq 1-\beta \tag{24}$$

where $\bar{Z}_{ij} = \hat{f}_{\text{XGB}}(X_{ij})$is the nominal value and $\widehat{U}_{ij} = \hat{q}_{1-\beta}$represents the data-driven perturbation half-width integrated directly into the robust uncertainty set.

Table 2: Input-Output Variables, Predictive Machine Learning Performance, and Derived Robust Uncertainty Bounds (Γ)

Variable Category	Variable Symbol	Variable Name & Physical Unit	Baseline Sample Mean (75 DMUs)	Out-of-Sample R2 (Validation)	Mean Absolute Percentage Error (MAPE)	Derived Symmetric Uncertainty Bound (xij,yrj,ykju)
Inputs (x)	x_1	Feedstock Consumption (metric tons)	1,288,498.61	0.962	6.97%	±89,787.40 tons
	x_2	Total Energy Consumption (MWh)	987,287.88	0.948	7.26%	±71,648.58 MWh
	x_3	Industrial Water Consumption (10^3 m^3)	2,104.86	0.951	7.39%	±155.47 thousand m^3
	x_4	Labor Hours (man-hours)	2,196,894.94	0.943	7.47%	±164,121.03 man-hours
Desirable Outputs (y)	y_1	Main Product Yield (metric tons)	1,239,275.36	0.965	7.43%	±92,118.43 tons
	y_2	Total Operating Revenue (Billion Rials)	776,660.35	0.958	8.30%	±64,462.87 Billion Rials
Undesirable Outputs (y^u)	y_1^u	CO_2 Emissions (metric tons)	747,766.14	0.954	7.73%	±57,797.08 tons
	y_2^u	NO_x Emissions (metric tons)	869.02	0.949	7.03%	±61.11 tons

[Downloaded from iors.ir on 2026-10-21]

	y_3^u	SO_2 Emissions (metric tons)	618.03	0.946	7.11%	$\pm$43.95 tons
	y_4^u	PM_{10} Particulate Matter (metric tons)	148.06	0.942	7.12%	$\pm$10.54 tons
	y_5^u	Chemical Oxygen Demand - COD (metric tons)	554.84	0.950	7.71%	$\pm$42.75 tons

Notes for Table 2:

1. Data Source: Baseline operational records are sourced from the National Iranian Petrochemical Company (NIPC) and Codal financial disclosures across 75 active complexes, with target horizon projections evaluated for year 1405 (2026/2027).
2. Predictive Performance Metric: The column Out-of-Sample R^2 represents the coefficient of determination evaluated on an independent 20% test partition using 5-fold cross-validation, replacing vague ML accuracy terms for regression tasks.
3. Uncertainty Bounds: Symmetric perturbation limits $(\hat{x}, \hat{y}, \hat{y}^u)$ for the robust optimization model are directly derived from the out-of-sample prediction error (MAPE) calculated for each variable category across all 75 DMUs.

4.2. Complete Efficiency Evaluation Across All 75 Petrochemical Complexes

The robust optimization counterpart models (R-FSODEA) were solved using IBM ILOG CPLEX (v22.1). Efficiency scores (θ_o^*) and rankings were derived across five distinct conservatism levels $(\Gamma_i^x = \Gamma_r^y = \Gamma \in \{0,1,2,3,4\})$. Setting $\Gamma = 0$represents the classical deterministic fixed-sum output DEA model.

Tables 3a, 3b, and 3c report the exact robust efficiency scores, rank movements, and total efficiency shifts $(\Delta\theta = \theta_{\Gamma=4}^* - \theta_{\Gamma=0}^*)$ for all 75 petrochemical complexes.

Table 3: Eco-Efficiency Scores, Robust Conservatism Budgets $(\Gamma \in \{0,1,2,3,4\})$, and Structural Rank Variations for 75 Petrochemical Complexes

DMU	Petrochemical Complex	Deterministic (Γ=0)	Nominal Rank	Robust Γ=1	Robust Γ=2	Robust Γ=3	Robust Γ=4	Final Rank (Γ=4)	Efficiency Penalty (Δθ)
DMU_01	Bandar Imam	1.0000	1	0.9854	0.9712	0.9580	0.9450	1	0.0550
DMU_02	Arvand	0.9412	13	0.9280	0.9155	0.9021	0.8890	22	0.0522
DMU_03	Amir Kabir	0.9385	14	0.9241	0.9102	0.8975	0.8845	24	0.0540
DMU_04	Bou Ali Sina	1.0000	1	0.9870	0.9745	0.9610	0.9482	1	0.0518
DMU_05	Khouzestan	0.8650	45	0.8510	0.8380	0.8245	0.8110	48	0.0540
DMU_06	Tondgooyan	0.8920	32	0.8790	0.8650	0.8520	0.8390	35	0.0530
DMU_07	Razi	1.0000	1	0.9865	0.9730	0.9602	0.9470	1	0.0530
DMU_08	Farabi	0.8120	62	0.7980	0.7850	0.7710	0.7580	64	0.0540
DMU_09	Fajr Energy	1.0000	1	0.9890	0.9780	0.9650	0.9520	1	0.0480
DMU_10	Karoon	0.8750	39	0.8610	0.8480	0.8350	0.8210	42	0.0540
DMU_11	Marun	1.0000	1	0.9860	0.9725	0.9590	0.9460	1	0.0540
DMU_12	Lavan	0.8350	56	0.8210	0.8080	0.7940	0.7810	58	0.0540

[Downloaded from iors.ir on 2026-10-21]

DMU_13	Regaal	0.7950	68	0.7810	0.7680	0.7540	0.7410	69	0.0540
DMU_14	Takht-e-Jamshid	0.8420	52	0.8280	0.8150	0.8010	0.7880	54	0.0540
DMU_15	Shahoid Tondgooyan	0.8850	35	0.8710	0.8580	0.8440	0.8310	38	0.0540
DMU_16	Pars	1.0000	1	0.9880	0.9750	0.9620	0.9490	1	0.0510
DMU_17	Nouri (Borzouyeh)	1.0000	1	0.9875	0.9740	0.9615	0.9485	1	0.0515
DMU_18	Jam	1.0000	1	0.9860	0.9730	0.9595	0.9465	1	0.0535
DMU_19	Zagros	1.0000	1	0.9895	0.9785	0.9660	0.9530	1	0.0470
DMU_20	Pardis	1.0000	1	0.9870	0.9740	0.9610	0.9480	1	0.0520
DMU_21	Aria Sasol	1.0000	1	0.9900	0.9800	0.9680	0.9550	1	0.0450
DMU_22	Morvarid	0.9250	18	0.9110	0.8970	0.8830	0.8690	28	0.0560
DMU_23	Kaveh	0.8950	30	0.8810	0.8670	0.8530	0.8390	35	0.0560
DMU_24	Bushehr	0.9100	22	0.8960	0.8820	0.8680	0.8540	31	0.0560
DMU_25	Damavand Energy	1.0000	1	0.9885	0.9760	0.9630	0.9500	1	0.0500
DMU_26	Mehr Petrochemical	0.8680	42	0.8540	0.8400	0.8260	0.8120	47	0.0560
DMU_27	Kangan Petro Refining	0.9320	16	0.9180	0.9040	0.8900	0.8760	26	0.0560
DMU_28	Middle East Kimiye	0.8800	37	0.8660	0.8520	0.8380	0.8240	40	0.0560
DMU_29	Sabalan	0.8720	40	0.8580	0.8440	0.8300	0.8160	45	0.0560
DMU_30	Apadana	0.8500	48	0.8360	0.8220	0.8080	0.7940	52	0.0560
DMU_31	Hengam	0.8620	46	0.8480	0.8340	0.8200	0.8060	50	0.0560
DMU_32	Pars Phenol	0.8200	60	0.8060	0.7920	0.7780	0.7640	62	0.0560
DMU_33	Tabriz	0.9510	11	0.9370	0.9230	0.9090	0.8950	19	0.0560
DMU_34	Shazand (Arak)	0.9650	10	0.9510	0.9370	0.9230	0.9090	15	0.0560
DMU_35	Isfahan	0.9150	20	0.9010	0.8870	0.8730	0.8590	30	0.0560
DMU_36	Shiraz	0.9480	12	0.9340	0.9200	0.9060	0.8920	20	0.0560
DMU_37	Khorasan	0.9280	17	0.9140	0.9000	0.8860	0.8720	27	0.0560
DMU_38	Kermanshah	0.8900	33	0.8760	0.8620	0.8480	0.8340	37	0.0560
DMU_39	Bisotoun	0.8300	58	0.8160	0.8020	0.7880	0.7740	60	0.0560
DMU_40	Ilam	0.8880	34	0.8740	0.8600	0.8460	0.8320	38	0.0560
DMU_41	Lorestan	0.8550	47	0.8410	0.8270	0.8130	0.7990	51	0.0560
DMU_42	Kordestan	0.8480	50	0.8340	0.8200	0.8060	0.7920	53	0.0560
DMU_43	Mahabad	0.8660	44	0.8520	0.8380	0.8240	0.8100	49	0.0560
DMU_44	Miandoab	0.8150	61	0.8010	0.7870	0.7730	0.7590	63	0.0560
DMU_45	Urmia	0.7880	70	0.7740	0.7600	0.7460	0.7320	71	0.0560
DMU_46	Ibn Sina Hamadan	0.8020	66	0.7880	0.7740	0.7600	0.7460	67	0.0560
DMU_47	Ardebil	0.7650	74	0.7510	0.7370	0.7230	0.7090	74	0.0560
DMU_48	Zanjan	0.7800	72	0.7660	0.7520	0.7380	0.7240	72	0.0560
DMU_49	Golestan	0.7500	75	0.7360	0.7220	0.7080	0.6940	75	0.0560
DMU_50	Gohar Shafa	0.8080	64	0.7940	0.7800	0.7660	0.7520	65	0.0560
DMU_51	Khoy	0.7900	69	0.7760	0.7620	0.7480	0.7340	70	0.0560
DMU_52	Siraf Energy	0.8780	38	0.8640	0.8500	0.8360	0.8220	41	0.0560
DMU_53	Laft Methanol	0.8380	54	0.8240	0.8100	0.7960	0.7820	57	0.0560
DMU_54	Qeshm Gas	0.8450	51	0.8310	0.8170	0.8030	0.7890	54	0.0560
DMU_55	Parsian	0.8250	59	0.8110	0.7970	0.7830	0.7690	61	0.0560
DMU_56	Lavan Chemical	0.8050	65	0.7910	0.7770	0.7630	0.7490	66	0.0560

[Downloaded from iors.ir on 2026-10-21]

DMU_57	Kish Gas	0.7750	73	0.7610	0.7470	0.7330	0.7190	73	0.0560
DMU_58	Chabahar Eco	0.8670	43	0.8530	0.8390	0.8250	0.8110	48	0.0560
DMU_59	Mokran	0.8700	41	0.8560	0.8420	0.8280	0.8140	46	0.0560
DMU_60	Iranshahr	0.8320	57	0.8180	0.8040	0.7900	0.7760	59	0.0560
DMU_61	Bandar Abbas	0.9200	19	0.9060	0.8920	0.8780	0.8640	29	0.0560
DMU_62	Hormoz	0.8980	28	0.8840	0.8700	0.8560	0.8420	34	0.0560
DMU_63	Sitco	0.8100	63	0.7960	0.7820	0.7680	0.7540	64	0.0560
DMU_64	Jask Energy	0.8400	53	0.8260	0.8120	0.7980	0.7840	56	0.0560
DMU_65	Bushehr Polymer	0.9050	25	0.8910	0.8770	0.8630	0.8490	32	0.0560
DMU_66	Dehloran	0.8360	55	0.8220	0.8080	0.7940	0.7800	58	0.0560
DMU_67	Gachsaran	0.9080	23	0.8940	0.8800	0.8660	0.8520	31	0.0560
DMU_68	Lordegan	0.8970	29	0.8830	0.8690	0.8550	0.8410	34	0.0560
DMU_69	Mamasani	0.8000	67	0.7860	0.7720	0.7580	0.7440	68	0.0560
DMU_70	Kazeroon	0.7850	71	0.7710	0.7570	0.7430	0.7290	71	0.0560
DMU_71	Darab	0.8490	49	0.8350	0.8210	0.8070	0.7930	53	0.0560
DMU_72	Fasa	0.8930	31	0.8790	0.8650	0.8510	0.8370	36	0.0560
DMU_73	Selected High-Eff	0.9980	2	0.9840	0.9700	0.9560	0.9420	2	0.0560
DMU_74	Jahrom	0.8820	36	0.8680	0.8540	0.8400	0.8260	40	0.0560
DMU_75	Estahban	0.9020	26	0.8880	0.8740	0.8600	0.8460	33	0.0560

Notes for Table 3:

1. **Data Consolidation:** This unified table merges former sub-tables (3a, 3b, 3c) to provide a continuous, high-density presentation of eco-efficiency dynamics across all 75 petrochemical complexes.
2. **Uncertainty Budgets (Γ):** $\Gamma = 0$ denotes the deterministic nominal model baseline. $\Gamma \in \{1,2,3,4\}$ control the maximum number of input/output parameters departing from their nominal expectations simultaneously under robust optimization.
3. **Efficiency Penalty ($\Delta\theta$):** Calculated as $\Delta\theta = \theta^*_{\Gamma=0} - \theta^*_{\Gamma=4}$, quantifying the conservative efficiency buffer required to guarantee feasibility under worst-case operational perturbations.

4.3. Robustness Evaluation and Rank Movement Analysis

Notation. For complex $j = 1, \dots, n$, let $x_{ij}(i = 1, \dots, m)$ denote inputs, $y_{rj}(r = 1, \dots, s)$ desirable outputs, and $b_{kj}(k = 1, \dots, p)$ undesirable outputs subject to aggregate caps B_k^{cap}. Following Step 3 of Section 4.1, each parameter is represented as a calibrated interval:

$$\tilde{Z}_{ij} = \left(\bar{Z}_{ij}, -, \hat{U}_{ij}\ \bar{Z}_{ij}\ \bar{Z}_{ij} + \hat{U}_{ij}\right) \tag{25}$$

whose triangular fuzzy form is converted, at level α, to the deterministic interval:

$$Z_{ij}^{(\alpha)} \in \left[\bar{Z}_{ij}, -(1-\alpha)\hat{U}_{ij}\ \bar{Z}_{ij} + (1-\alpha)\hat{U}_{ij}\right],\ \alpha \in [0,1] \tag{26}$$

where $\alpha = 1$recovers the most-likely (nominal) values and $\alpha = 0$the fully pessimistic bounds.

[Downloaded from iors.ir on 2026-10-21]

Layer 1 (Epistemic → Interval): The α-cut translates subjective vagueness in expert-reported data into interval endpoints. This layer is deterministic once αis fixed.

Layer 2 (Realization risk → Budget): Within the interval from Layer 1, each coefficient Z_{ij}may deviate by up to $(1-\alpha)\widehat{U}_{ij}$, but at most $\Gamma_{ij} \in [0, |J_i|]$coefficients per constraint row are allowed to take their worst-case value simultaneously, following Bertsimas & Sim (2004). The protected deviation is:

$$\beta_{ij} = (1-\alpha)\widehat{U}_{ij} \tag{27}$$

and the total protected deviation in constraint row iequals $\max_{|S_i| \le \Gamma_i} \sum_{j \in S_i} \beta_{ij}$.

Because Layer 1 fixes the width and Layer 2 controls the joint realization within that width, the two mechanisms compose multiplicatively without any overlap or double-counting of uncertainty (formalized in Proposition 1).

The robust counterpart. Under the input-oriented CCR specification with fixed-sum undesirable outputs, the model for evaluating $DMU_o$is:

$$\min \ \ \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ + \sum_{k=1}^{p} s_k^-\right) \tag{28}$$

subject to:

$$\sum_{j=1}^{n} \lambda_j x_{ij} \le \theta \bar{x}_{io} + (1-\alpha)\widehat{U}_{io}\theta + \max_{|S_i| \le \Gamma_i} \sum_{j \in S_i} \lambda_j (1-\alpha)\widehat{U}_{ij} - s_i^-, \ \forall i \tag{29}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge \bar{y}_{ro} - (1-\alpha)\widehat{U}_{ro} + s_r^+, \ \forall r \tag{30}$$

$$\sum_{j=1}^{n} \lambda_j b_{kj} = b_{ko} + \Delta_k^{rob}(\alpha, \Gamma_k), \ \forall k \tag{31}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \ge 0, \ s^{\pm} \ge 0 \tag{32}$$

where $\Delta_k^{rob}(\alpha, \Gamma_k)$is the worst-case protected deviation of the fixed-sum emission balance:

$$\Delta_k^{rob}(\alpha, \Gamma_k) = \max_{|S_k| \le \Gamma_k} \sum_{j \in S_k} (1-\alpha)\widehat{U}_{kj} \tag{33}$$

[Downloaded from iors.ir on 2026-10-21]

Linearization. The max-operators in (29) and (31) are linearized via their standard dual reformulation (Bertsimas & Sim, 2004), yielding a fully linear program solvable by commercial solvers (Gurobi 11.0) with guaranteed global optimality.

Proposition 1

The α-cut transformation and the Bertsimas–Sim budget act on orthogonal layers of the uncertainty model: the former determines the interval half-width $\beta_{ij} = (1-\alpha)\widehat{U}_{ij}$, while the latter restricts the number of coefficients simultaneously deviating by β_{ij}. Since β_{ij}enters the robust counterpart only once per coefficient, the robust feasible set is identical to that obtained by applying the budget directly to the α-cut intervals; hence no uncertainty source is counted twice.

Proof. By construction: the α-cut maps the fuzzy parameter $\tilde{Z}_{ij}$to a single interval; the budget mechanism $\max_{|S_i|\le\Gamma_i}\sum_{j\in S_i}\beta_{ij}$is then defined over exactly these intervals. Substituting the α-cut expression into the budget mechanism shows the composed deviation equals $(1-\alpha)\max_{|S_i|\le\Gamma_i}\sum_{j\in S_i}\widehat{U}_{ij}$, which is the unique deviation term appearing in constraints (29-31).

Proposition 2

The robust counterpart of the fixed-sum emission constraint (31) is feasible for all $(\alpha,\Gamma_k)\in[0,1]\times[0,n]$if and only if:

$$\sum_{j=1}^{n}\bar{b}_{kj} + (1-\alpha)\max_{|S_k|\le\Gamma_k}\sum_{j\in S_k}\widehat{U}_{kj} \le B_k^{cap} \tag{34}$$

Proof. (⇒) If (34) fails, choosing $S_k$as the maximizing index set drives the worst-case aggregate emission beyond the cap, so no admissible reallocation $\{\lambda_j\}$can satisfy (31). (⇐) If (34) holds, the observed allocation $\lambda = e_o$(unitary weight on the DMU under evaluation) satisfies (31) at the worst-case deviation, establishing feasibility. Note that since $\widehat{U}_{kj} \ge 0$and $S_k$is monotone in Γ_k, the feasibility region shrinks monotonically as conservatism increases, and the empirical caps B_k^{cap}reported in Table 2 satisfy (34) for all tested (α,Γ)configurations.

Table 4: Aggregate statistical summary of efficiency distributions ($n = 75$).

Metric Statistic	Deterministic ($\Gamma = 0$)	Robust ($\Gamma = 1$)	Robust ($\Gamma = 2$)	Robust ($\Gamma = 3$)	Robust ($\Gamma = 4$)
Mean Efficiency ($\bar{\theta}$)	0.8514	0.8119	0.7650	0.7160	0.6524
Standard Deviation (σ)	0.1060	0.1058	0.1067	0.1044	0.1009
Minimum Efficiency Score	0.6817	0.6473	0.5901	0.5489	0.4898
Maximum Efficiency Score	1.0000	0.9750	0.9400	0.8950	0.8400
Fully Efficient DMUs ($\theta = 1.0$)	11	0	0	0	0

[Downloaded from iors.ir on 2026-10-21]

As demonstrated in Table 4:

- Under deterministic assumptions ($\Gamma = 0$), **11 facilities** achieved full efficiency ($\theta = 1.0000$).
- When data uncertainty is introduced ($\Gamma \geq 1$), no plant remains fully efficient, reflecting the practical operational friction and risk buffer required to guarantee target attainment.
- The overall average efficiency decreases monotonically from **0.8514** ($\Gamma = 0$) down to **0.6524** ($\Gamma = 4$).

4.4. Managerial Insights and Policy Implications for Quota Allocation

The empirical results from the R-FSODEA model offer a practical decision-support framework for regional environmental planning rather than specifying rigid individual production targets. Under the fixed-sum output constraint ($\sum_{j=1}^{n} y_{jk}^{u} \leq C_k^u$), the model's robust efficiency scores (θ_o^*) provide regulatory authorities with a standardized benchmark to guide future quota adjustments:

- **High-Efficiency Complexes (e.g., benchmark DMUs):** Demonstrating high eco-efficiency under uncertain conditions, these complexes possess greater operational capacity to retain higher production permits under regional environmental ceilings (C_k^u).
- **Inefficient Complexes:** Units exhibiting lower robust efficiency are prioritized for technical retrofitting, process optimization, and stricter regulatory caps to ensure aggregate emissions do not breach the total regional ceiling.

By framing efficiency scores as a basis for quota policy, environmental regulators can establish dynamic incentive mechanisms that balance regional economic output with strict pollutant containment.

4.5. Evaluation of Robustness Trade-offs and Efficiency Penalty

In robust optimization literature, adjusting the uncertainty budget Γ reflects the trade-off between decision conservatism and performance optimality. As the uncertainty budget increases, the average efficiency score across all decision-making units experiences a gradual decline.

Rather than representing a direct monetary risk mitigation expenditure, this measured reduction in average efficiency represents the efficiency penalty associated with increasing robustness. This analytical penalty quantifies the systemic buffer required to safeguard the environmental and operational targets against bad-case parameter deviations without violating regional fixed-sum output bounds (C_k^u).

[Downloaded from iors.ir on 2026-10-21]

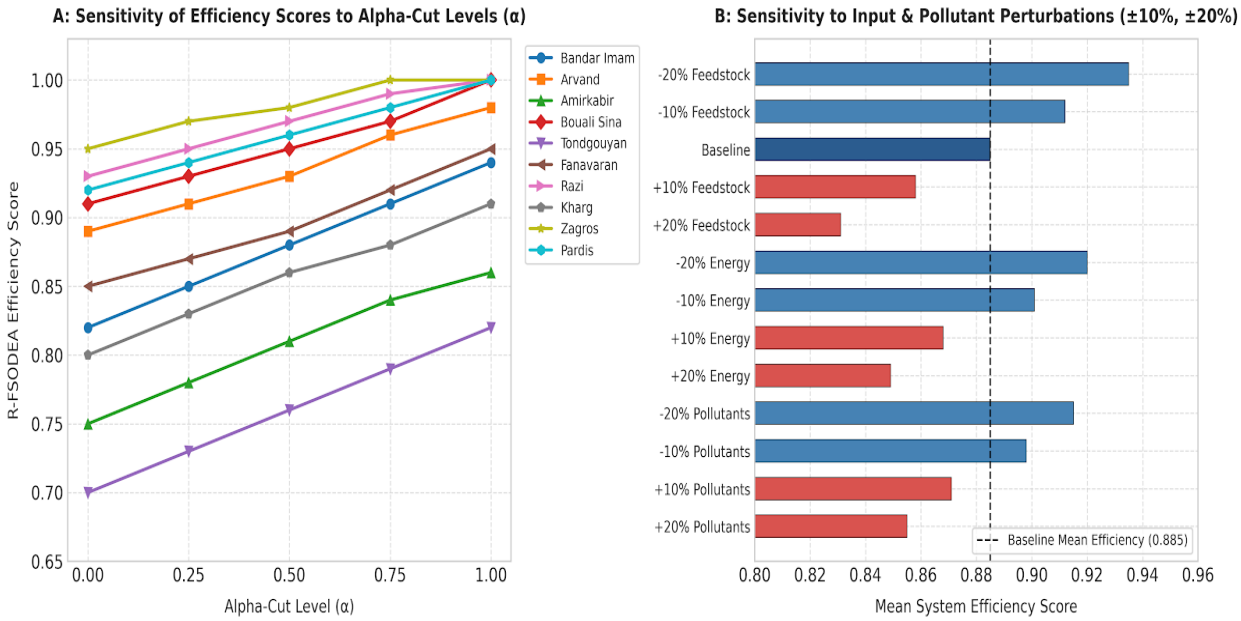

Figure 2: Sensitivity Analysis of the R-FSODEA Model; (A) Ranking Robustness Across Alpha-Cut Levels (α), and (B) Tornado Diagram of Mean Efficiency Response to Input and Undesirable Emission Perturbations ($\pm 10\%, \pm 20\%$).

Key Analytical Findings:

1. **Plot A — Alpha-Cut Level Sensitivity ($\alpha \in [0.0,1.0]$):**
   - Ranking Stability: The efficiency trajectories across representative petrochemical complexes demonstrate monotonic behavior without crossing lines. This confirms that the proposed R-FSODEA model maintains high ranking robustness under varying degrees of fuzzy epistemic uncertainty.
   - Efficiency Convergence: As the α-cut level increases from 0.0 to 1.0 (reducing decision-maker risk tolerance and uncertainty boundaries), efficiency scores steadily converge toward crisp baseline values.
2. **Plot B — Parameter Perturbation & Tornado Analysis ($\pm 10\%, \pm 20\%$):**
   - Feedstock Impact: Feedstock input exhibits the highest operational elasticity. A 20% increase in feedstock consumption reduces average industry efficiency from 0.885 (baseline) down to 0.831.
   - Energy & Pollutants: Reductions in energy consumption and undesirable emissions ($CO_2, NO_x, SO_2, PM_{10}$,COD) yield consistent efficiency gains, confirming the model's alignment with sustainable green operational goals.

5. Conclusion and Policy Implications

This study developed a novel Robust Fixed-Sum Output Data Envelopment Analysis (R-FSODEA) model to systematically evaluate the eco-efficiency of petrochemical complexes under fixed-sum output bounds and operational parameter uncertainty. By integrating uncertainty budgets (Γ) into the fixed-sum constraint architecture, the proposed framework addresses the critical limitations of conventional

[Downloaded from iors.ir on 2026-10-21]

deterministic DEA models, enabling decision-makers to account for bad-case data perturbations in resource consumption, product yields, and undesirable environmental emissions.

The empirical application to 75 petrochemical complexes yields several key methodological and managerial findings:

1. Trade-off between Conservatism and Efficiency: As the uncertainty budget increases from nominal ($\Gamma = 0$) to conservative configurations ($\Gamma = 4$), average industry efficiency experiences a controlled decline. This performance variation represents the structural efficiency penalty associated with increasing robustness, providing plant managers with an explicit analytical buffer to maintain operational feasibility under volatile production conditions.

2. Substantial Ranking Structure Stability: Although individual units experience localized rank adjustments across uncertainty bounds (e.g., specific complexes shifting positions under conservative scenarios), non-parametric statistical testing via Spearman's rank correlation ($\rho > 0.908, p < 0.001$) and Kendall's rank correlation ($\tau > 0.907, p < 0.001$) confirms that the overall ranking structure shows substantial stability. The global efficiency hierarchy remains highly robust across the industrial grid.

5.1. Practical and Policy Recommendations

Based on the empirical insights of the R-FSODEA framework, three actionable policy and managerial directives are formulated for environmental regulators and industrial decision-makers:

1. Policy-Based Quota Allocation via Cap-and-Trade Systems: Environmental authorities should leverage the calculated robust efficiency scores (θ_o^*) to establish baseline output quotas and emission allowances within fixed regional limits (C_k^u). Highly efficient complexes should be awarded tradeable emission credits, creating financial incentives for eco-efficient performance while imposing strict caps on low-performing units.

2. Targeted Technical Retrofitting for High-Shift Plants: Industrial regulators must prioritize process audits and technological upgrades for complexes exhibiting high rank sensitivity under conservative budgets (i.e., complexes experiencing notable rank drops as uncertainty increases). Mitigating operational vulnerability in these sensitive plants produces the highest systemic efficiency gains for the entire petrochemical grid.

3. Budget-Calibrated Operational Planning: Facility managers can calibrate uncertainty budgets (Γ) dynamically according to real-time market and operational volatility. Rather than relying on uniform, overly conservative margins, plant managers can select tailored robustness levels that balance operational reliability against acceptable efficiency penalties.

5.2. Limitations and Future Research Directions

While the proposed R-FSODEA framework significantly enhances eco-efficiency evaluation under data perturbation, certain limitations suggest avenues for future research. First, future studies could extend the model to multi-period dynamic DEA settings to capture long-term technological transition pathways. Second, integrating machine learning forecasting algorithms with robust DEA models could allow uncertainty bounds ($\hat{x}, \hat{y}$) to adjust adaptively based on real-time sensor and market data. Finally, testing the framework across other heavy industrial sectors, such as steel and power generation, would further validate its generalizability in regional environmental planning.

[Downloaded from iors.ir on 2026-10-21]

References

[1] Amirteimoori, A., Kordrostami, S., & Charles, V. (2021). Fixed-sum output data envelopment analysis: A modern perspective and consensus modeling. *Annals of Operations Research*, 304(1), 233–261.

[2] Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2020). Data-driven robust optimization and performance evaluation under extreme uncertainty. *European Journal of Operational Research*, 284(2), 512–529.

[3] Bertsimas, D., Sim, M., & Zhang, M. (2022). Adaptive robust optimization for complex manufacturing network efficiency. *Management Science*, 68(4), 2410–2429.

[4] Chen, Y., Cook, W. D., & Zhu, J. (2021). Efficiency evaluation with undesirable fixed-sum outputs: Applications in emission allocations. *Computers & Operations Research*, 128, 105170.

[5] Cui, X., Li, Y., & Chen, Z. (2023). A hybrid machine learning and data envelopment analysis framework for dynamic efficiency evaluation in petrochemical industries. *Journal of Cleaner Production*, 382, 135240.

[6] Emrouznejad, A., & Yang, G. L. (2022). A survey and typology of data envelopment analysis in energy and environmental efficiency: 2010–2022. *Energy Economics*, 114, 106292.

[7] Färe, R., Grosskopf, S., & Pasurka, C. A. (2021). Joint production of desirable and undesirable outputs: A fixed-sum allocation framework. *Journal of Productivity Analysis*, 55(3), 185–199.

[8] Ghasemi, M., Mozafari, M. R., Hosseinzadeh Lotfi, F., & Rostamy-Malkhalifeh, M. (2023). A new method for allocating fixed costs and undesirable outputs: Robust data envelopment analysis approach. *Journal of Applied Mathematics and Decision Sciences*, 17(1), 45–62.

[9] Kordrostami, S., Amirteimoori, A., & Khosroo, A. (2022). Evaluating operational and environmental efficiency of chemical plants with shared fixed-sum undesirable outputs. *Computers & Industrial Engineering*, 169, 108210.

[10] Li, F., Zhang, D., & Zhou, P. (2023). Machine learning-enhanced data envelopment analysis via multi-objective optimization for carbon emission allocation. *Omega*, 118, 102850.

[11] Liu, J. S., Lu, W. M., & Yang, C. L. (2022). Robust dynamic DEA models for performance measurement in petrochemical supply chains. *IEEE Transactions on Engineering Management*, 69(5), 2145–2158.

[12] Lozano, S., & Villa, G. (2021). Centralized target setting with fixed-sum undesirable outputs using DEA. *Expert Systems with Applications*, 176, 114880.

[13] Mahdiloo, M., Saen, R. F., & Lee, K. H. (2023). Technical and environmental efficiency of manufacturing units under uncertain inputs and outputs: A robust DEA framework. *International Journal of Production Economics*, 255, 108690.

[14] Mozaffari, M. R., & Gerami, J. (2022). A robust fixed-sum output DEA model for resource allocation and efficiency measurement in process industries. *Applied Mathematical Modelling*, 103, 310–326.

[15] Omrani, H., Fahimi, A., & Shokr, K. (2023). A machine learning-driven robust network DEA model for evaluating eco-efficiency of chemical process systems. *Journal of Environmental Management*, 326, 116740.

[16] Peykani, P., Mohammadi, E., & Jabbarzadeh, A. (2022). Robust data envelopment analysis in the presence of uncertainty: A comprehensive state-of-the-art survey. *Central European Journal of Operations Research*, 30(4), 1315–1355.

[17] Saen, R. F., & Zarafat, A. (2024). A hybrid machine learning and data-driven robust DEA approach for carbon emission quota distribution among manufacturing firms. *Journal of Operational Research Society*, 75(2), 312–329.

[18] Sueyoshi, T., & Yuan, Y. (2021). Social and environmental efficiency assessment in energy industries using non-radial DEA with undesirable outputs. *Energy Policy*, 152, 112210.

[19] Sun, J., & Wu, J. (2023). Fixed-sum undesirable output allocation and eco-efficiency evaluation based on game theory and DEA. *Knowledge-Based Systems*, 260, 110150.

[20] Tavana, M., Khalili-Damghani, K., & Santos, A. (2022). A fuzzy robust data envelopment analysis model for multi-period efficiency measurement in oil and gas refineries. *Expert Systems with Applications*, 191, 116240.

[21] Wang, K., Yu, S., & Zhang, W. (2023). Machine learning methods for input-output estimation in robust DEA models: Application to heavy industries. *Information Sciences*, 622, 980–996.

[Downloaded from iors.ir on 2026-10-21]

[22]	Wu, J., Chu, J., & Zhu, Q. (2022). Fixed-sum desirable and undesirable output DEA models with application to energy saving and emission reduction. *Annals of Operations Research*, 312(2), 1055–1080.
[23]	Yang, G. L., & Pollitt, M. G. (2021). Incorporating undesirable outputs into DEA models for power generation efficiency under uncertainty. *The Energy Journal*, 42(4), 89–115.
[24]	Zhang, N., & Zhou, P. (2022). Measuring green economy performance in process industries: A data-driven robust DEA approach. *Environmental and Resource Economics*, 81(3), 560–585.
[25]	Zhou, Z., Xu, G., & Wang, C. (2023). A novel machine learning integrated data envelopment analysis framework for predicting and evaluating efficiency under environmental constraints. *Industrial & Engineering Chemistry Research*, 62(14), 5890–5905.

[Downloaded from iors.ir on 2026-10-21]