Journal of Mathematical Modeling
Vol. -, No. -, 2026, pp. -. Research Article

p-robust network cost efficiency with genetic algorithms and machine learning

Rita Shakouri*

Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Technical and Vocational University (TUV), Tehran, Iran
Email(s): rshakori@tvu.ac.ir

Abstract. Original data envelopment analysis models for expected cost-efficiency evaluation lack robustness in the presence of uncertainty and high-dimensional data. This gap becomes more critical when dealing with big data in the petroleum industry, where selecting relevant variables from large, noisy datasets significantly affects performance results. To address this gap, we propose an uncertainty-integrated, two-stage network data envelopment analysis framework that incorporates artificial intelligence techniques, genetic algorithm and random forest for optimal feature selection. Genetic algorithm simulates natural selection to identify the most relevant variables, reducing dimensionality and enhancing model stability across probabilistic scenarios. In the second stage, Wilcoxon statistical testing and a p-robust approach are applied to ensure consistent and reliable ranking of decision-making units under uncertain conditions. Random forest complements this framework by capturing hidden data patterns, improving accuracy and interpretability. The model is validated using real-world data from ten oilfields, demonstrating substantial improvements over the traditional data envelopment analysis models in feature selection, expected cost-efficiency measurement, and decision robustness. This study offers a practical and intelligent decision-support tool for expected cost efficiency measurement under uncertainty in complex petroleum environments.

Keywords: Two-stage network, data envelopment analysis, genetic algorithm, machine learning, expected cost-efficiency, stochastic p-robust, Wilcoxon test.
AMS Subject Classification 2010: 34A34, 65L05.

1 Introduction

Cost efficiency (CE) evaluates the ability of a decision-making unit (DMU) to produce current outputs at the lowest possible cost. The concept dates back to Farrell [13], who introduced foundational ideas

*Corresponding author
Received: 20 April 2025/ Revised: 29 September 2025/ Accepted: 4 October 2025
DOI: 10.22124/jmm.2025.30184.2705

Table 1: Some studies for evaluating CE in DEA and NDEA approaches

	Authors	Features
1	Farrell [13] (1957), Jahanshahloo and et al. [20] (2007), Toloo [47] (2014), Piran et al. [35] (2021)	Cost-DEA
2	Fukuyama and Matousek [14] (2011), Blagojevich et al. [4] (2020), Kao and Hwang [21] (2008), Lozano [29] (2011), Sarkar [37] (2018), Wanke and Barros [48] (2016), Fakharzadeh Jahromi [10] (2023), Cesaroni [6] (2018), Seyedboveir et al. [40] (2017), Kordrostami and Jahani Sayyad Noveiri [22] (2022)	Cost-NDEA

Table 2: Some studies for evaluating CE under uncertainty approaches

	Authors	Features
1	Al-Khasawneh et al. [1] (2020), Lee and Huang [24] (2017), Lotfi et al. [28] (2020), Jahani and Kordrostami [19] (2022)	Stochastic
2	Ashrafi and Kaleibar [2] (2017), Pourmahmood and Aramy [36] (2024), Bagherzadeh Valami [3] (2010)	Fuzzy
3	Camanho and Dyson [5] (2005), Mousavizadeh et al. [32] (2020), Soleimani-Chamkhorami and Ghobadi [44] (2021), Conceição et al. [8] (2013)	Robust
4	Kuosmanen and Post [23] (2001), Schaffnit et al. [38] (1997), Thompson et al. [45] (1990), Thompson et al. [46] (1996), Lei and Li [25] (2012)	Interval

that underpin data envelopment analysis (DEA). In this framework, measuring CE requires input and output quantities as well as precise input prices at each DMU. However, as Cooper et al. [9] pointed out, the practical utility of CE measurement is limited due to the difficulty in obtaining accurate price data and the potential for short-term fluctuations in those prices. Compared to technical efficiency (TE), CE often offers more practical insights because it directly relates to input costs, which are generally more reliable in real-world settings. Importantly, a DMU may be technically efficient while still being cost-inefficient. Traditional DEA models typically treat DMUs as “black boxes” and do not account for the internal structure of processes. However, in real-world systems such as in energy or production sectors—many DMUs consist of multiple interrelated subsystems. To better model such complexities, the two-stage network DEA (NDEA) models have been introduced. In this structure, outputs of the first stage serve as inputs for the second stage, enabling a more nuanced analysis of efficiency [12]. Table 1 summarizes key studies evaluating CE using DEA and NDEA models.

Another limitation in prior research is the common assumption of data certainty. In practice, data particularly price and cost data are often uncertain due to market volatility, resource constraints, or economic instability. Various methodologies, such as stochastic programming, robust optimization, interval analysis, and fuzzy logic, have been developed to handle this challenge. Initial efforts to manage uncertainty in DEA models often relied on stochastic theories introduced by scholars like Sengupta [39]. Table 2 presents a summary of CE evaluations in DEA/NDEA frameworks under uncertainty.

It is worth noting that none of the research mentioned above has considered the genetic algorithm (GA) with the NDEA models to evaluate the expected cost efficiency. While in NDEA models that aim to assess cost efficiency within complex and multi-stage structures, the genetic algorithm can be employed to identify the optimal combination of inputs, outputs, and associated costs. Despite advances in modeling CE and accounting for uncertainty, few studies have integrated intelligent algorithms—such as GAs or machine learning (ML)—with NDEA models to optimize expected cost efficiency. Yet, GAs have

proven effective in navigating large, nonlinear solution spaces and identifying optimal combinations of inputs, outputs, and associated costs. Introduced by Holland [17], GAs simulate the process of natural selection and are increasingly used in conjunction with DEA models [15, 16, 18, 30]. Their integration with ML can further enhance prediction accuracy and efficiency evaluation, especially in uncertain environments [11, 31, 33, 34]. This study aims to address the gap by developing a robust framework that integrates GA and ML techniques within a two-stage NDEA model. Building upon the centralized NDEA approach by Camanho et al. [5], the model enhances CE evaluation under uncertainty by incorporating a stochastic p-robust optimization method, as proposed by Snyder et al. [43] (interested readers can see [41, 42], and references therein for further studies in considering uncertainty on mathematical programming problems). This technique ensures solution robustness across various probabilistic scenarios by minimizing deviations from the optimal value. An essential component of this framework is intelligent feature selection. The GA is employed to identify the most relevant variables that influence CE, mimicking evolutionary processes to optimize model performance. Then, ML techniques such as RF are applied to reveal patterns in high-dimensional, noisy data—thereby improving the interpretability and reducing model complexity. The proposed model is applied to assess the CE of 10 oilfields in Iran under conditions of data uncertainty. Through this application, the framework demonstrates its ability to produce reliable performance rankings, improve decision-making under uncertainty, and support long-term planning through scalable and interpretable analytics.
The contributions of this study are as follows:

- Enhanced decision-making under uncertainty:
  The proposed two-stage NDEA framework, combined with genetic algorithms and random forests, equips managers with a more robust tool to assess cost efficiency. This enables them to make informed decisions based on reliable, uncertainty-aware performance indicators.

- Improved variable selection and model interpretability:
  By automatically identifying the most relevant input and output variables, the model reduces cognitive and computational overload for managers. This leads to clearer insights into performance drivers and actionable areas for improvement.

- Data-driven strategic planning:
  The integration of AI methods allows managers to extract valuable knowledge from noisy and complex big datasets. This supports more strategic resource allocation and optimized planning across oilfields.

- Reliable ranking and benchmarking:
  The use of Wilcoxon statistical testing and p-robust optimization ensures that performance comparisons among oilfields are not only data-driven but also statistically sound. This facilitates transparent benchmarking and prioritization of operational interventions.

- Scalability and adaptability:
  The modular design of the framework makes it adaptable for other sectors facing similar challenges with uncertainty and big data. Managers can expand its use beyond oilfields to other segments of the energy production and supply chains.

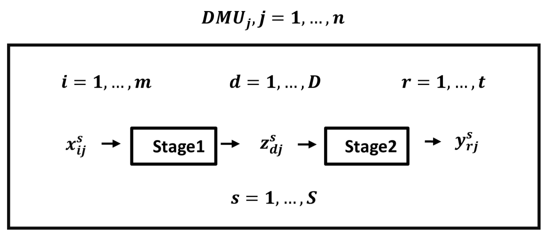

Figure 1: A serial network system

- Risk reduction and operational stability:
  By incorporating probabilistic robustness, the model helps to reduce the risk of over- or underestimating efficiency due to data fluctuations, supporting more stable long-term planning.

The remainder of this paper is structured as follows: Section 2 introduces key preliminaries and definitions related to the centralized two-stage NDEA model. Section 3 discusses the stochastic p-robust approach to centralized NDEA modeling. Section 4 presents the concepts of a hybrid algorithm for feature selection and subsequently a real case study demonstrating the proposed model is presented in Section 5. Results and discussion are provided in Section 6. Finally, conclusions are offered in the last section.

2 Preliminary concepts

2.1 Structure of two-stage NDEA models

Different models have been introduced to gain the overall efficiency of a system with a network structure and sub-units. One of the fundamental models in this scope was offered by Liang et al. [26, 27], who considered a centralized serial network system (Figure 1). Suppose we have n DMUs, and each DMU_j $(j = 1, \dots, n)$ uses m inputs x^s_{ij}, $(i = 1, \dots, m)$ in the first stage to generate D outputs z^s_{dj}, $(d = 1, \dots, D)$ under scenario $s = \{1, \dots, S\}$. Then, these D outputs become the inputs to the second stage and will be dispatched to the second stage as intermediate measures. The outputs from the second stage are y^s_{rj}, $(r = 1, \dots, t)$ under the scenario s. Therefore, x^s_{ij}, z^s_{dj}, and y^s_{rj} display the i^{th} input, the d^{th} intermediate output, and the r^{th} output of DMU_j by the s^{th} scenario. The following model is proposed to evaluate the overall efficiency of the system and its sub-processes for DMU_0:

$$e_0^{c-s} = \max \sum_{r=1}^{t} u_r y^s_{ro} \tag{1a}$$

$$\text{s.t.} \sum_{i=1}^{m} v_i x^s_{io} = 1, \qquad \forall s \in S, \tag{1b}$$

$$\sum_{r=1}^{t} u_r y^s_{rj} - \sum_{d=1}^{D} w^1_d z^s_{dj} \leq 0, \qquad \forall j, \forall s \in S, \tag{1c}$$

$$\sum_{d=1}^{D} w^1_d z^s_{dj} - \sum_{i=1}^{m} v_i x^s_{ij} \leq 0, \qquad \forall j, \forall s \in S, \tag{1d}$$

$$v_i, u_r, w_d^1 \geq 0, \qquad \forall i, r, d. \tag{1e}$$

Based on the radial Constant Returns to Scale (CRS) model of Charnes et al. [7], the input-oriented DEA models for each efficiency score of stage are

$$e_0^{1s} = \max \frac{\sum_{d=1}^{D} w_d^1 z_{dj_0}^s}{\sum_{i=1}^{m} v_i x_{ij_0}^s} \tag{2}$$

and

$$e_0^{2s} = \max \frac{\sum_{r=1}^{t} u_r y_{rj_0}^s}{\sum_{d=1}^{D} w_d^1 z_{dj_0}^s}. \tag{3}$$

Definition 1. If e_0^{1s} and e_0^{2s} are the efficiency scores of the first and second stages under the s^{th} scenario, respectively, then the efficiency score of the overall system is as

$$e_0 = e_0^{1s} \times e_0^{2s} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}^s}{\sum_{i=1}^{m} v_i x_{ij_0}^s}. \tag{4}$$

The overall efficiency score of a DMU is defined based on equalizing the role of the intermediate weights of the model (1). In the sequel, first we briefly explain the CE proposed model of Schaffnit et al. [38], and afterwards present the mathematical formulation of two-stage DEA model based on it.

2.2 The CE evaluation with certain prices

The CE measures the output value of the minimum input cost. Consequently, we expect that input prices will be incorporated into the CE calculations. Farrell [13] provided the CE measure for each DMU in an input-oriented form, assuming that the input prices are known as follows:

$$CE_0^s = \min \frac{\sum_{i=1}^{m} p_{ij_0} x_{ij_0}^{s^*}}{\sum_{i=1}^{m} p_{ij_0} x_{ij_0}^s} \tag{5a}$$

$$\text{s.t. } \sum_{j=1}^{n} \lambda_j x_{ij}^s \leq x_{ij_0}^{s^*}, \qquad \forall i, \forall s \in S, \tag{5b}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj}^s \geq y_{rj_0}^s, \qquad \forall r, \forall s \in S, \tag{5c}$$

$$\lambda_j \geq 0, \qquad \forall j \tag{5d}$$

$$x_{ij_0}^{s^*} \text{ is free.} \tag{5e}$$

In this model, p_{ij_0} is the price of the i^{th} input for DMU$_j$ and $x_{ij_0}^{s^*}$ is the minimum amount of the i^{th} input for DMU$_0$ under evaluation. Alternatively, CE can be determined by incorporating weight constraints using the radial CRS model developed by Charnes et al. [7]. Since only the relative prices of inputs are important for measuring CE, the only restrictions on the weights used in the evaluation are that the value of the input weights must correspond to the respective values of the input prices for each DMU, such that

$$\frac{v_{i^a}}{v_{i^b}} = \frac{\rho_{i^a j_0}}{\rho_{i^b j_0}}, \quad i^a, i^b = 1, \ldots, m. \tag{6}$$

In the equation (6),v_{i^a} and v_{i^b} are the input weights utilized for the CE evaluation with the DEA model, and $\rho_{i^a j_0}$ and $\rho_{i^b j_0}$ are the input prices considered at DMU_0, for any two inputs i^a and i^b applied by the DMU. Therefore, the resulting CE model with certain and known prices at each DMU and also using the standard DEA model with the addition of weight constraints is as follows:

$$CE^s = \max \sum_{r=1}^{t} u_r y^s_{r j_0} \tag{7a}$$

$$\text{s.t.} \sum_{i=1}^{m} v_i x^s_{i j_0} = 1, \tag{7b}$$

$$\sum_{r=1}^{t} u_r y^s_{rj} - \sum_{i=1}^{m} v_i x^s_{ij} \leq 0, \qquad \forall j, \forall s \in S, \tag{7c}$$

$$v_{i^a} - \frac{\rho_{i^a j_0}}{\rho_{i^b j_0}} v_{i^b} = 0, \qquad i^a < i^b\ \forall i, \tag{7d}$$

$$u_r \geq 0, \quad v_i \geq 0, \qquad \forall r, i. \tag{7e}$$

It can be seen that the CE^s value gained from (7) is equal to the cost efficiency measure gained from (5). Now, considering DMU_j as a whole, and by adopting a centralized perspective, based on the method introduced by Schaffnit et al. [42], the centralized model to evaluate CE of DMUs with certain prices is proposed as follows:

$$e_0^{c-s} = \max \sum_{r=1}^{t} u_r y^s_{r j_0} \tag{8a}$$

$$\text{s.t.} \sum_{i=1}^{m} v_i x^s_{i j_0} = 1, \tag{8b}$$

$$\sum_{d=1}^{D} w^1_d z^s_{dj} - \sum_{i=1}^{m} v_i x^s_{ij} \leq 0, \qquad \forall j,\ \forall s \in S, \tag{8c}$$

$$\sum_{r=1}^{t} u_r y^s_{rj} - \sum_{d=1}^{D} w^1_d z^s_{dj} \leq 0, \qquad \forall j,\ \forall s \in S, \tag{8d}$$

$$w^1_{d^c} - \frac{\rho'_{d^c j_0}}{\rho'_{d^t j_0}} w^1_{d^t} = 0, \qquad d^c < d^t, \quad \forall d, \tag{8e}$$

$$v_{i^a} - \frac{\rho_{i^a j_0}}{\rho_{i^b j_0}} v_{i^b} = 0, \qquad i^a < i^b, \quad \forall i, \tag{8f}$$

$$u_r \geq 0, \quad w^1_d \geq 0, \quad v_i \geq 0, \qquad \forall r, d, i. \tag{8g}$$

In model (8), v_i, w^1_d, and u_r are non-negative decision variables of inputs, intermediate measures, and outputs, respectively. In addition, x^s_{ij}, z^s_{dj}, and y^s_{rj} show uncertain parameters related to the i^{th} input, the d^{th} intermediate measure, and the r^{th} output of DMU_j in terms of the s^{th} scenario, respectively. In this model, v_{i^a} and v_{i^b} are the input weights utilized for the CE evaluation in the first stage; also, $w^1_{d^c}$ and $w^1_{d^t}$ are the weights of intermediate factors; $\rho'_{i^a j_0}$, $\rho'_{i^b j_0}$, $\rho'_{d^c j_0}$ and $\rho'_{d^t j_0}$ are the input prices for any input weights v_{i^a}, v_{i^b}, $w^1_{d^c}$ and $w^1_{d^t}$ in the first and second stages, respectively; a, b, c, and t are applied to prevent repetition of i, and d (for more details see Appendix A). It is noted that the e_0^{c-s} in model (8) is the optimal efficiency score for the centralized model.

3 Stochastic p-robust for two-stage NDEA model

In this section, first the p-robust concept is described, and then the formulation of two-stage NDEA is proposed under uncertainty based on the stochastic p-robust approach. Finally, the process of feature selection combining GA and ML approaches is presented.

3.1 p-robust concept

Let S be a set of scenarios, and P_s be a deterministic maximization problem for scenario index s, so that there is a different problem, that is, P_s for each scenario $s \in S$. Moreover, let $\Phi_s^* > 0$ be the optimal objective score for P_s. In addition, let X be the feasible vector in terms of the weights of inputs and outputs, respectively, and $\Phi_s(X)$ be the objective score of problem P_s in solution X. So, X is called p-robust ($p \geq 0$ is constant) if for all $s \in S$ the below inequality holds: X is called p-robust ($p \geq 0$ is constant) if in all $s \in S$ the below inequality holds:

$$\frac{\Phi_s^* - \Phi_s(X)}{\Phi_s^*} \leq p. \tag{9}$$

The left-hand side of inequality (9) displays the relative regret in the s^{th} scenario, and $p \geq 0$ is a parameter that indicates the robustness level between different scores of each scenario. The relative regret in each scenario is limited by p. Inequality (9) can be shown as below:

$$\Phi_s(X) \geq (1-p)\Phi_s^*. \tag{10}$$

Ultimately, to control the relative regret associated with the scenarios, the p-robust restrictions are combined with the model.

3.2 Stochastic p-robust centralized NDEA model

To tackle uncertain conditions, the deterministic models cannot lead to correct results. In fact, uncertainty can change the final results and DMUs rankings. Hence, the classical NDEA model must be robust to uncertainty. In this case, this paper suggests the stochastic p-robust technique for the CE of the two-stage network in the centralized NDEA model to cope with this issue as follows:

$$f_0^{c-s} = \max \sum_{s=1}^{S} \xi^s \sum_{r=1}^{t} u_r y_{ro}^s \tag{11a}$$

$$\text{s.t.} \sum_{r=1}^{t} u_r y_{ro}^s \geq (1-p) e_0^{c-s^*}, \qquad \forall s \in S, \tag{11b}$$

$$\sum_{i=1}^{m} v_i x_{ij_0}^s = 1, \tag{11c}$$

$$\sum_{d=1}^{D} w_d^1 z_{dj}^s - \sum_{i=1}^{m} v_i x_{ij}^s \leq 0, \qquad \forall j, \forall s \in S, \tag{11d}$$

$$\sum_{r=1}^{t} u_r y_{rj}^s - \sum_{d=1}^{D} w_d^1 z_{dj}^s \leq 0, \qquad \forall j, \forall s \in S. \tag{11e}$$

$$w_{d^c}^1 - \frac{\rho'_{d^c j_0}}{\rho'_{d^t j_0}} w_{d^t}^1 = 0, \qquad d^c < d^t, \quad \forall d, \tag{11f}$$

$$v_{i^a} - \frac{\rho_{i^a j_0}}{\rho_{i^b j_0}} v_{i^b} = 0, \qquad i^a < i^b, \quad \forall i, \tag{11g}$$

$$u_r \geq 0, \quad w_d^1 \geq 0, \quad v_i \geq 0, \qquad \forall r, d, i. \tag{11h}$$

In model (11), the objective function maximizes the expected CE score of DMUs based on the data from each scenario. In this model, the uncertainty in the parameters is defined by discrete scenarios. In the objective function, ξ^s is the probability that scenario s happens (i.e., it is obscure which scenario will occur in the future, in other words, there is no information about the probability of each scenario). The second constraint is the p-robust constraint. This set of restrictions may not allow the scenario efficiency to take a score less than $100p\%$ of the ideal efficiency score achieved in each scenario. The parameter p can flexibly control the relative regret among all scenarios. It is noted that if $p = \infty$, then the p-robust constraints in model (11) become redundant. Usually, p-values should not be less than 0.2. The upper limit can be adjusted through trial and error and may be increased to one. The third to fifth sets of constraints are the original NDEA constraints which must be held for all $s \in S$, and the rest of constraints are the same as model (8).

Theorem 1. *Model (11) is feasible and convex.*

Proof. Assume the DMU under evaluation is DMU_{k_0}. Let $z_{k_0}^s = \max\{z_{d0}^s \mid 1 \leq d \leq D\} > 0$, $x_{k_0}^s = \max\{x_{i0}^s \mid 1 \leq i \leq m\} > 0$, and $y_{k_0}^s = \max\{y_{r0}^s \mid 1 \leq r \leq t\} > 0$. Then setting $(w_1^1, \ldots, w_D^1, v_1, \ldots, v_m) = (0, \ldots, 0, \ldots, 1/x_{k_0}^s, 0, \ldots)$, constraints $\sum_{i=1}^m v_i x_{ij_0} = 1$, and $\sum_{d=1}^D w_d^1 z_{dj}^s - \sum_{i=1}^m v_i x_{ij}^s \leq 0$, imply $\sum_{d=1}^D w_d^1 z_{dj}^s \leq 1$. Since $\sum_{r=1}^t u_r y_{rj}^s \leq \sum_{d=1}^D w_d^1 z_{dj}^s \leq 1$, we get $f_j^{c-s*} \leq \frac{1}{1-p}$. This confirms the feasibility of the model (11).

To prove the convexity, it is sufficient to show the convexity of its feasible region. Let Ω be the feasible region of the model (11), and we have: $(v'_1, \ldots, v'_i, w_1^{1'}, \ldots, w_d^{1'}, u'_1, \ldots, u'_r) \in \Omega$, and $(v''_1, \ldots, v''_i, w_1^{1''}, \ldots, w_d^{1''}, u''_1, \ldots, u''_r) \in \Omega$. Then for each $\beta \in [0,1]$, the following relationships are held:

$$\begin{aligned} &\beta v'_i + (1-\beta) v''_i \geq 0, && \forall i; \\ &\beta w_d^{1'} + (1-\beta) w_d^{1''} \geq 0, && \forall d; \\ &\beta u'_r + (1-\beta) u''_r \geq 0, && \forall r. \end{aligned}$$

Therefore, we can get

$$\begin{aligned} \sum_{i=1}^m (\beta v'_i + (1-\beta) v''_i) x_{ij} &= \beta \sum_{i=1}^m v'_i x_{ij}^s + (1-\beta) \sum_{i=1}^m v''_i x_{ij}^s = \beta + (1-\beta) = 1, \\ \sum_{i=1}^m (\beta w_d^{1'} + (1-\beta) w_d^{1''}) z_{dj}^s &= \beta \sum_{i=1}^m w_d^{1'} z_{dj}^s + (1-\beta) \sum_{i=1}^m w_d^{1''} z_{dj}^s \\ &\leq \sum_{i=1}^m v'_i x_{ij}^s + (1-\beta) \sum_{i=1}^m v''_i x_{ij}^s = \sum_{i=1}^m (\beta v'_i + (1-\beta) v''_i) x_{ij}^s, \end{aligned}$$

$$\sum_{i=1}^{m}(\beta u'_r+(1-\beta)u''_r)y^s_{rj}=\beta\sum_{i=1}^{m}u'_r y^s_{rj}+(1-\beta)\sum_{i=1}^{m}u''_r y^s_{rj}$$
$$\leq \sum_{i=1}^{m} w^{1'}_d z^s_{dj}+(1-\beta)\sum_{i=1}^{m} w^{1''}_d z^s_{dj}=\sum_{i=1}^{m}(\beta w^{1'}_d+(1-\beta)w^{1''}_d)z^s_{dj},$$
$$(\beta v_{i^{a'}}+(1-\beta)v_{i^{a''}})-\frac{\rho_{i^a j_0}}{\rho_{i^b j_0}}(\beta v_{i^{b'}}+(1-\beta)v_{i^{b''}})$$
$$=\beta\left(v_{i^{a'}}-\frac{\rho_{i^a j_0}}{\rho_{i^b j_0}}v_{i^{b'}}\right)+(1-\beta)\left(v_{i^{a''}}-\frac{\rho_{i^a j_0}}{\rho_{i^b j_0}}v_{i^{b''}}\right)=0,$$
$$(\beta w^{1'}_{d^c}+(1-\beta)w^{1''}_{d^c})-\frac{\rho'_{a^c j_0}}{\rho'_{a^t j_0}}(\beta w^{1'}_{d^t}+(1-\beta)w^{1''}_{d^t})$$
$$=\beta\left(w^{1'}_{d^c}-\frac{\rho'_{a^c j_0}}{\rho'_{a^t j_0}}w^{1'}_{d^t}\right)+(1-\beta)\left(w^{1''}_{d^c}-\frac{\rho'_{a^c j_0}}{\rho'_{a^t j_0}}w^{1''}_{d^t}\right)=0.$$

Since

$$\begin{aligned}
&\beta v'_i+(1-\beta)v''_i\in\Omega, && \forall i;\\
&\beta w^{1'}_d+(1-\beta)w^{1''}_d\in\Omega, && \forall d;\\
&\beta u'_r+(1-\beta)u''_r\in\Omega, && \forall r.
\end{aligned}$$

We have

$$\beta(v'_1,\dots,v'_i,w^{1'}_1,\dots,w^{1'}_d)+(1-\beta)(v''_1,\dots,v''_i,w^{1''}_1,\dots,w^{1''}_d)\in\Omega,$$
$$\beta(w^{1'}_1,\dots,w^{1'}_d,u'_1,\dots,u'_r)+(1-\beta)(w^{1''}_1,\dots,w^{1''}_d,u''_1,\dots,u''_r)\in\Omega.$$

Thus, it can be concluded that Ω is a convex set. □

4 A hybrid algorithm for feature selection and weight

Oilfields are inherently complex systems with multiple inputs (e.g., crude oil intake), intermediates (e.g., distillation temperature, process yields), and outputs (e.g., refined petroleum). These variables are often nonlinear and interdependent, posing significant challenges to traditional optimization methods. In contrast, GA provides effective solutions by mimicking natural evolution to find optimal combinations of key variables. Through techniques like the crossover and mutation, GA identifies relevant variables while reducing the problem dimensions, thus enhancing computational efficiency and optimizing oilfield performance within operational and economic constraints. To address this, a hybrid algorithm approach of GA and random forest (RF) is introduced in two phases as below:
Phase 1: Feature selection

- The RF algorithm computes importance scores for all variables.
- Variables with scores below a predefined threshold (α) are eliminated.
- The selected set of impactful variables is denoted as $\mathscr{F}_\zeta$.

Table 3: The new GA-RF algorithm for the proposed NDEA model

NDEA parameters	**Description**
$X = \{x^s_{ij}\}$	Input variables
$Z = \{z^s_{dj}\}$	Intermediate variables
$Y = \{y^s_{rj}\}$	Output variables
α	Feature importance threshold
N_g	Number of generations
N_τ	Population size
Phase 1: Feature selection	
Step 1.1:	$\mathscr{F} = \{f_1, \dots, f_m\}$
Step 1.2:	$w_i = RF(X,Y)$
Step 1.3:	$\mathscr{F}_\zeta = \{f_i \mid w_i \geq \alpha\}$
Phase 2: Genetic algorithm	
Initialize:	$E_0 = \{\varepsilon_1, \dots, \varepsilon_{N_\tau}\}$
For	$g = 1$ to N_g
Overall efficiency:	$f(\varepsilon_i) = f_0^{c-s}$
Subject to constraints:	
constraints:	The set of constraints on model (11)
Genetic operations:	
Selection:	$E'_g = \text{select}(E_g)$
Crossover:	$E''_g = \text{crossover}(E'_g)$
Mutation:	$= \text{mutate}(E''_g)E_{g+1}$
Output:	
Optimal weights:	$E^* = \arg\max_{p \in E_{N_g}} f(\varepsilon)$
Selected features:	$\mathscr{F}_\zeta$
Optimal efficiency:	$f(\varepsilon^*) = f_0^{c-s^*}$

Phase 2: Weight optimization via GA

- GA starts with initial populations representing different weight combinations.
- The fitness of each solution is evaluated using the NDEA model.
- Genetic operations are applied:
  [Selection] Higher efficiency solutions have a greater chance of reproduction.
  [Crossover] Combines parent solutions to generate offspring with potentially better performance.
  [Mutation] Introduces variability to avoid premature convergence and ensure global optimization.
  In this phase, weights for inputs and intermediates are optimized sequentially, satisfying all model constraints.

Final Outcome:
The best set of variable weights is selected to maximize overall expected cost efficiency. The process is summarized in Table 3, which outlines the steps and operations of the GA-RF algorithm.

5 A real case study

Oilfields are complex industrial systems that involve numerous input, intermediate, and output variables, all of which significantly impact on their overall performance. These variables include factors such as crude oil intake, distillation column temperature, oilfield process efficiencies, and production levels of various petroleum products. Table 4 shows the set of input, intermediate, and output variables of the oilfield. Given the large number of variables and the nonlinear relationships between them, traditional optimization methods are often ineffective in selecting key variables and optimizing oilfield performance. In this case, GA was utilized as a powerful computational tool that simulates natural evolution to identify the optimal combination of influential variables. By mimicking the natural selection process through cross-over and mutation operations, GA effectively selects the most relevant variables while also reducing the complexity of the problem. This approach not only improves computational efficiency but also optimizes oilfield performance by taking into account operational and economic constraints.

By integrating GA with ML, features that significantly impact the economic, technical, and environmental efficiency of the oilfield can be identified. This process helps not only selecting the most critical features but also eliminating those that are less significant, thereby reducing complexity and enhancing model accuracy. Therefore, the selection criteria for these features are based on the following four parameters:

- Direct importance and impact on oilfield performance,
- Accurate measurability,
- Low correlation between variables,
- Coverage of different performance aspects (technical, economic, and qualitative).

The optimal set of features identified serves as a critical foundation for developing effective policies and initial performance improvement strategies within the oilfield. Subsequently, feature importance is evaluated using RF to determine which features had the greatest impact on overall efficiency. To execute the proposed algorithm, the parameters of the GA need adjustment. The parameter settings are provided in Table 5. Two genetic operators (mutation and single crossover with rates of 3% and 85%, respectively) are utilized.

In this study, after applying the GA-RF method, six variables were selected from a total of 30 that significantly influence the improvement of oilfield performance. These selected variables are: $\mathscr{F}_\zeta = \{x_1, x_3, z_6, z_8, y_1, y_{10}\}$ that are presented in Table 6.

6 Results and discussion

This section discusses the advantages of the proposed model using a dataset of Persian Gulf oilfields. Table 7 presents the results obtained from the normalized dataset under two scenarios for 10 selected oilfields in the Persian Gulf. First, we calculate the ideal efficiency scores of the proposed model in both scenarios using model (11), as offered by the oilfield system analyzers (i.e., s_1 = Pessimistic, s_2 = Optimistic). According to Snyder and Daskin (2006), we consider all scenarios equiprobable, that is, each scenario has a probability of $q^s = 0.5$. The results of this calculation are illustrated in Figure 2. For

Table 4: The initial input, intermediate, and output variables

x_i	**Input variables**
1.	Crude oil consumption (x_1): The volume of crude oil entering the oilfield
2.	Labor costs (x_2): Total cost of direct and indirect workforces
3.	Operational costs (x_3): Expenses on energy (electricity and gas), equipment maintenance, and repairs
4.	Facility area (x_4): The oilfield's infrastructural capacity
5.	Chemicals and catalysts costs (x_5): The consumption of materials required in the refining process
6.	API gravity of crude oil (x_6): Density of oil
7.	Hydrogen flow rate (x_7): Hydro treating and hydrocracking processes
8.	Steam flow rate (x_8): The influence of the distillation tower performance
9.	Gas to oil ratio (x_9): Amount of gas accompanying crude oil
10.	Operating pressure in the distillation tower (x_{10})
z_d	Intermediate variables
1.	Primary oilfield products: Produced gasoline (z_1), diesel, kerosene, and liquefied petroleum gas (LPG)
2.	Product quality (z_2): Sulfur content and API gravity of the products
3.	Process efficiency (z_3): Conversion ratio of crude oil into useful products
4.	Generated waste (z_4): The amount of unusable or substandard output
5.	Refining efficiency rate (z_5): Overall refining efficiency rate
6.	Temperature at each stage of the distillation tower (z_6)
7.	Pressure in the hydrocracking reactor (z_7)
8.	Catalytic cracking reaction rate (z_8)
9.	Residence time in cracking reactors (z_9)
10.	Pollutant separation rate (z_{10})
y_r	**Output variables**
1.	Revenue from product sales (y_1): Total revenue from selling gasoline, diesel, and other products
2.	Profitability ratio (y_2)
3.	Energy efficiency (y_3): The ratio of recovered useful energy to total energy consumption
4.	The percentage of aromatic compounds in gasoline (y_4)
5.	The percentage of conversion of crude oil to final products (y_5)
6.	The amount of industrial waste and effluent (y_6)
7.	The percentage of coke produced (y_7)
8.	Economic efficiency (y_8)
9.	The amount of diesel production (y_9)
10.	The amount of gasoline production (y_{10})

Table 5: Parameter setting

GA parameters	**Value**
Selection mechanism	Elitism
Mutation rate	3%
Crossover rate	85%
Maximum generations	150

instance, the columns of this figure report the ideal efficiency scores according to data collected from the first and second scenarios, respectively. According to the results, DMU_5 and DMU_8 attained the efficiency score of one.

Table 6: Feature reduction and selection with GA-RF

Input variable	**Intermediate variable**	**Output variable**
Crude oil (x_1)	Temperature tower (z_6)	Revenue (y_1)
Operational costs (x_3)	Catalytic cracking (z_8)	Mount of gasoline (y_{10})

Table 7: The normalized data set of two scenarios for 10 oilfields

DMUs	x_1		x_3		y_1		y_{10}		z_6		z_8	
	s_1	s_2	s_1	s_2	s_1	s_2	s_1	s_2	s_1	s_2	s_1	s_2
1	0.513	0.684	0.381	0.508	2.820	3.760	0.170	0.226	0.591	0.788	0.116	0.155
2	1.000	1.333	0.495	0.659	4.557	6.076	0.299	0.399	1.029	1.372	0.164	0.219
3	0.414	0.552	0.251	0.335	3.028	4.037	0.144	0.192	0.276	0.367	0.070	0.093
4	0.344	0.459	0.343	0.457	2.167	2.889	0.133	0.178	0.448	0.598	0.146	0.195
5	0.518	0.691	0.178	0.237	1.906	2.542	0.170	0.226	0.392	0.523	0.005	0.006
6	0.684	0.912	0.172	0.229	2.490	3.320	0.295	0.394	0.416	0.554	0.240	0.320
7	0.689	0.918	0.361	0.482	3.674	4.899	0.214	0.286	0.953	1.270	0.059	0.079
8	0.368	0.491	0.261	0.348	1.952	2.603	0.123	0.164	0.370	0.493	0.088	0.117
9	0.577	0.769	0.172	0.229	2.751	3.668	0.198	0.264	0.886	1.181	0.147	0.196
10	0.320	0.426	0.110	0.146	1.480	1.973	0.106	0.142	0.225	0.301	0.052	0.070

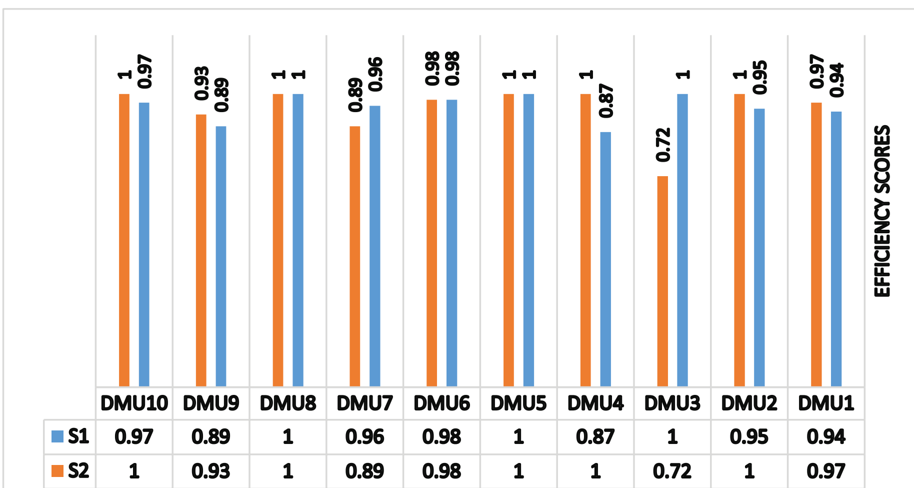

Figure 2: Ideal efficiency scores of model (8) in the two scenarios

Following this, we solved model (11) based on the ideal score results for the efficiency evaluation of DMUs. The results of solving this model with different p-values and equal probability of 0.5 for each scenario are reported in Table 8.

Based on these results, model (11) produces infeasible results for most DMUs when small values of p (e.g., $p \leq 0.44$) are applied, which are not reported here. Moreover, in this study for $p \geq 0.46$, the efficiency scores remained constant. Therefore, we did not calculate it for the other p-values higher than 0.46. As the p-values increase, we realize more feasible results. That is, increasing the p-value from 0.44 to 0.46 improves the efficiency score of DMU5, which is also reflected in DMU3. Model (11) maximizes the expected efficiency score across the two scenarios, while p-robust constraints control

Table 8: The results of solving the proposed GA-RF of model (11) with different p-value

DMUs	p-value									
	0.44	0.45	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.53
1	0.887	0.887	0.867	0.868	0.868	0.868	0.868	0.868	0.868	0.868
2	0.798	0.751	0.701	0.700	0.700	0.701	0.701	0.701	0.701	0.701
3	INF	0.674	0.634	0.634	0.634	0.634	0.634	0.634	0.634	0.634
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	INF	INF	0.771	0.771	0.772	0.771	0.772	0.772	0.772	0.772
6	0.895	0.895	0.804	0.804	0.804	0.804	0.804	0.804	0.804	0.804
7	0.663	0.663	0.622	0.622	0.622	0.622	0.622	0.622	0.622	0.622
8	0.531	0.531	0.528	0.528	0.528	0.528	0.528	0.528	0.528	0.528
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.478	0.471	0.460	0.460	0.460	0.460	0.460	0.460	0.460	0.460

Table 9: The results of model (11) and ERR values before and after GA-RF with $p = 0.46$

DMUs	After GA-RF			Before GA-RF		
	Expected Efficiency	Rank	ERR	Expected Efficiency	Rank	ERR
1	0.867	3	0.067	1.000	1	0.000
2	0.701	6	0.209	0.993	9	0.010
3	0.634	7	0.102	1.000	1	0.000
4	1.000	1	0.000	1.000	1	0.000
5	0.771	5	0.210	1.000	1	0.000
6	0.804	4	0.142	1.000	1	0.000
7	0.622	8	0.252	1.000	1	0.007
8	0.528	9	0.310	1.000	1	0.000
9	1.000	1	0.000	1.000	1	0.000
10	0.460	5	0.325	0.987	10	0.040

differences between the model's efficiency scores and the ideal scores. This allows us to rank the DMUs based on the p-values.

To compare model (11) before and after applying GA-RF, we calculate the expected relative regret (ERR) alongside the efficiency scores for each DMU. The ERR is used as a metric to measure decision-making quality and indicates the deviation from the optimal decision. After solving each model and obtaining its efficiency, ERR criterion is computed as $\sum_{s\in S}\xi^s\frac{\left(\Phi_s^*-\Phi_s(X)\right)}{\Phi_s^*}$. Indeed, the ERR value shows the relative difference between the efficiency of model (11) and the ideal efficiency from each scenario. Hence, the small value of this measure shows that the model makes close results to the ideal efficiencies. It should be noted that, here, $p = 0.46$ and $\xi^s = 0.5$. The results are shown in Table 9 and displayed in Figure 3.

Table 9 indicates that 80% of the total DMUs had an efficiency of zero before implementing the GA-RF model. This suggests that the model was overfitting and lacked generalization, potentially failing to capture the real complexities of the problem. Additionally, a review of Table 9 reveals that after applying GA-RF, the ERR values are positive, although small. However, the variations in ERR across

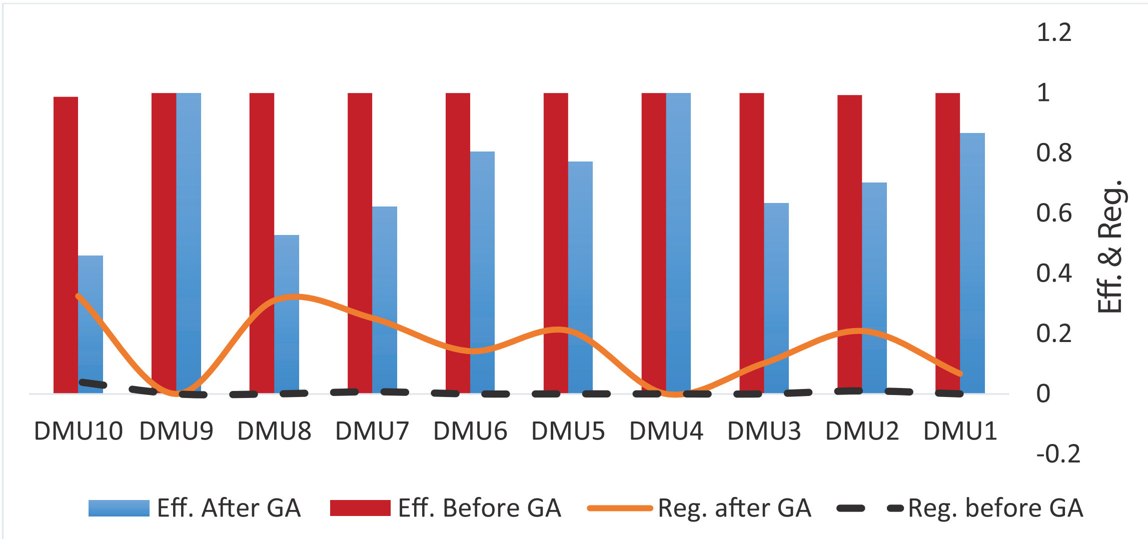

Figure 3: Comparison of ERR and expected efficiency of model (11) before and after GA-RF

Table 10: Wilcoxon rank correlation analysis

Statistical Metric	**value**	**Interpretation**
Wilcoxon Statistics (W)	0.0	The W value indicates that all post-GA efficiency values are less than or equal to pre-GA values.
γ-value[1]	0.0117	Since $\gamma < 0.05$, difference in efficiency before and after GA is statistically significant.
Mean efficiency before GA	0.9987	Before GA, most units had an efficiency of 1.
Mean efficiency after GA	0.8287	After GA, some units had efficiency values below 1, indicating better differentiation in the model.
Test Conclusion	Significant Difference	The GA has caused a meaningful change in efficiency, improving the model's precision.

different data sets are logical and justifiable. Moreover, the model successfully maintains a balance between minimizing regret and ensuring generalization capability, as illustrated in Figure 3. The dashed and smooth lines in Figure 3 also show different rankings for the two methods used. Notably, before applying the GA-RF method, many DMUs, approximately 80% of the total, ranked equally at one. This suggests that the model is unable to accurately differentiate between these units. However, after using the GA-RF method, the model's ability to distinguish between the ranks of the units improved significantly.

Table 10 displays the results of the Wilcoxon rank correlation analysis. By comparing and assessing the significance of the differences between two sets of ranks, we can determine whether the proposed model yields accurate results and if a more efficient model can be selected.

According to Table 10, the Wilcoxon test results demonstrate that using the GA resulted in a decrease in mean efficiency, suggesting improved model accuracy and differentiation in DMU analysis. This confirms the validity of the proposed model.

[1] Significance level

6.1 Managerial implications

The managerial implications of this study are substantial when viewed from a broader perspective. The results obtained from the implementation of model (11) under varying values of the robustness parameter provide critical insights into the model's behavior and performance under uncertainty. Specifically, when small values of p (e.g., $p \leq 0.44$) are applied, the model produces infeasible results for the majority of DMUs that are not reported here. This indicates a failure to ensure robustness at lower confidence levels. In practice, this suggests that excessive risk aversion renders the model overly conservative, leading to impractical efficiency estimates. However, beyond a certain threshold, the efficiency scores stabilize, revealing a point where the model achieves both feasibility and reliability. This sensitivity point is crucial for managerial decision-making, as it provides guidance on setting robustness parameters at moderate levels to balance conservatism and usability particularly relevant in highly volatile industries such as oil and gas. Moreover, the comparison of model (11) before and after integrating the GA-RF algorithm highlights a significant improvement in both generalization and ranking ability. Initially, approximately 80% of DMUs received identical efficiency scores, reflecting poor differentiation and possible overfitting. After incorporating GA-RF for feature selection and modeling, the refined model successfully distinguished between DMUs and minimized the ERR. Although the ERR values remained small, they confirmed that the updated model generated results closer to the ideal solutions across scenarios. From a managerial standpoint, this demonstrates the tangible value of embedding AI-based feature selection methods such as GA-RF into performance evaluation frameworks. Doing so enhances interpretability, sharpens the identification of inefficiencies, and supports more informed resource allocation and targeted performance improvement strategies. Furthermore, the Wilcoxon test results in Table 8 revealed a statistically significant reduction in average efficiency after applying GA-RF. While this outcome may appear counterintuitive, it reflects improved accuracy and discriminative power. A lower mean efficiency does not imply underperformance; rather, it signals the model's heightened ability to uncover hidden inefficiencies and promote strategic differentiation among DMUs. This shift ultimately provides managers with a clearer, more reliable basis for decision-making, avoiding the misleading optimism caused by overfitted models.

7 Conclusions

Modeling oilfield networks and evaluating associated costs under real-world uncertainty is a complex task that requires advanced and robust methodologies. This study introduced a two-stage NDEA framework that integrates GA and RF for optimal feature selection and uncertainty handling. By automatically identifying the most relevant input and output variables, the model not only enhances interpretability and reduces dimensionality but also improves the reliability of performance evaluation across probabilistic scenarios. The application of the Wilcoxon test confirmed statistically significant differences in cost efficiency among oilfields and enabled more credible and data-driven benchmarking of their performance. In addition, the incorporation of a stochastic p-robust method further strengthened the model's ability to support stable and consistent decision-making under fluctuating conditions. The framework also enables managers to extract meaningful insights from high-dimensional and noisy datasets, providing strategic guidance for resource allocation and operational planning. Its adaptability across various sectors and ability to reduce the risk of misjudgment make it a valuable decision-support tool in uncertain, data-rich environments. In particular, the managerial implications of this research are noteworthy. The results

Table 11: Decision variables in p-robust centralized NDEA model

Symbol	Description
$x^{s^*}_{ij_0}$	Minimum amount of i-th input for DMU under evaluation in scenario s
λ_j	Linear combination coefficients in DEA model
u_r	Weight of r-th output
v_i	Weight of i-th input
w^1_d	Weight of d-th intermediate factor in first stage
v_{i^a}	Weight of i^a-th input for CE evaluation
v_{i^b}	Weight of i^b-th input for CE evaluation
$w^1_{d^c}$	Weight of d^c-th intermediate factor in first stage
$w^1_{d^t}$	Weight of d^t-th intermediate factor in first stage

Table 12: Price parameters in p-robust centralized NDEA model

Symbol	Description
p_{ij_0}	Price of i-th input for DMUj
$\rho_{i^a j_0}$	Price of i^a-th input at DMU under evaluation
$\rho_{i^b j_0}$	Price of i^b-th input at DMU under evaluation
$\rho'_{d^c j_0}$	Price of d^c-th intermediate factor at DMU under evaluation
$\rho'_{d^t j_0}$	Price of d^t-th intermediate factor at DMU under evaluation

demonstrate that setting the robustness parameter at ensures feasible and stable efficiency outcomes, which provides practical guidance for managers making decisions under data uncertainty. Furthermore, the integration of GA-RF improves model discriminability and enhances the ranking of decision-making units, reducing overfitting and helping identify inefficiencies more accurately. By minimizing expected regret and supporting generalizability, the model offers a robust framework for cost-effective planning, resource prioritization, and performance improvement across operationally volatile industries such as oil and gas. These insights strengthen the practical value of the model and underscore its relevance as a tool for informed and resilient managerial decision-making.

Appendix A

Variables, parameters and constants of CE models are defined in the following tables:

Table 13: Indices and Sets in p-robust centralized NDEA model

Symbol	Description
i	Index of inputs, $i = 1,\ldots,m$
j	Index of DMUs, $j = 1,\ldots,n$
r d	Index of outputs, $r = 1,\ldots,t$ Index of intermediate factors, $d = 1,\ldots,D$
s	Index of scenarios, $s \in S$
S	Set of scenarios
0	Index of DMU under evaluation
m	Total number of inputs
n	Total number of DMUs
t	Total number of outputs
D	Total number of intermediate factors

Table 14: Sub-indices to Avoid Repetition in p-robust centralized NDEA model

Symbol	Description
i^a, i^b	Sub-indices for inputs to avoid repetition of index i
d^c, d^t	Sub-indices for intermediate factors to avoid repetition of index d
a, b, c, t	Auxiliary indices for distinguishing between different variables

References

[1] J.A. Al-Khasawneh, N. Essaddam, T. Hussain, *Total productivity and cost efficiency dynamics of US merging banks: a nonparametric bootstrapped analysis of the fifth merger wave*, Q. Rev. Econ. Finance. **78** (2020) 199–211.

[2] A. Ashrafi, M.M. Kaleibar, *Cost, revenue and profit efficiency models in generalized fuzzy data envelopment analysis*, Fuzzy Inf. Eng. **9** (2017) 237–246.

[3] H. Bagherzadeh Valami, *Cost efficiency with triangular fuzzy number input prices: an application of DEA*, Chaos Soliton Fract. **42** (2009) 1631–1637.

[4] M. Blagojevich, P. Ralevic, D. Sarac, *An integrated approach to analyzing the cost efficiency of postal networks*, Util. Policy **62** (2020).

[5] A.S. Camanho, R.G. Dyson, *Cost efficiency measurement with price uncertainty: a DEA application to bank branch assessments*, Eur. J. Oper. Res. **61** (2005) 432–446.

[6] G. Cesaroni, *Industry cost efficiency in data envelopment analysis*, Socio. Econ. Plan. Sci. **61** (2018) 37–43.

[7] A. Charnes, W.W. Cooper, E. Rhodes, *Measuring the efficiency of decision making units*, Eur. J. Oper. Res. (1978) 429–44.

[8] M. Conceicao, A.S. Portela, E. Thanassoulis, *Economic efficiency when prices are not fixed: disentangling quantity and price efficiency*, Omega **47** (2014) 36–44.

[9] W.W. Cooper, R.G. Thompson, R.M. Thrall, *Extensions and new developments in DEA*, Ann. Oper. Res. **66** (1996) 3–45.

[10] A.R. Fakharzadeh Jahromi, H. Rostamzadeh, M.H. Mojtabaei, *Designing and evaluating an optimal budget plan for parallel network systems through DEA methodology*, J. Math. Model. **11** (2023) 665–680.

[11] A. Fallahpour, E.U. Olugu, S.N. Musa, D. Khezrimotlagh, K.Y. Wong, *An integrated model for green supplier selection under fuzzy environment: application of data envelopment analysis and genetic programming approach*, Neural Comput. Appl. **27** (2016) 707–725.

[12] R. Fare, S. Grosskopf, *Network DEA*, Socio-Econ. Plan. Sci. **34** (2000) 35–49.

[13] M.J. Farrell, *The measurement of productive efficiency*, J. R. Stat. Soc. **120** (1957) 253–290.

[14] H. Fukuyama, R. Matousek, *Efficiency of Turkish banking: two-stage network system. Variable returns to scale model*, J. Int. Financ. Mark. Inst. Money **21** (2011) 75–91.

[15] M. Gonzalez, J.J. Lopez-Espin, J. Aparicio, D. Gimenez, J.T. Pastor, *Using genetic algorithms for maximizing technical efficiency in data envelopment analysis*, Procedia Comp. Sci. **51** (2015) 374–383.

[16] M. Gonzalez, J.J. Lopez-Espin, J. Aparicio, D. Giménez, *A parallel application of matheuristics in data envelopment analysis*, In International Symposium on Distributed Computing and Artificial Intelligence, Springer, Berlin (2015b) 172–179.

[17] J.H. Holland, *Adapting in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan 1975, re-issued by MIT Press, 1992.

[18] C.M. Hsu, *An integrated portfolio optimization procedure based on data envelopment analysis, artificial bee colony algorithm and genetic programming*, Inter. J. Syst. Sci. **45** (2014) 2645–2664.

[19] M. Jahani Sayyad Noveiri, S. Kordrostami, *Meta-Frontier Stochastic Cost and Revenue Efficiency Analysis: An Application to Bank Branches*, IJITDM **06** (2022) 1879–1903.

[20] G.R. Jahanshahloo, M. Soleimani-damaneh, A. Mostafaee, *A simplified version of the DEA cost efficiency model*, European J Eur. J. Oper. Res. **184** (2008) 814–815.

[21] C. Kao, S.N. Hwang, *Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan*, Eur. J. Oper. Res. **185** (2008) 418–429.

[22] S. Kordrostami, M. Jahani Sayyad Noveiri, *Cost efficiency analysis of heterogeneous network processes*, ECECSR. **3** (2022) 69–86.

[23] T. Kuosmanen, T. Post, *Measuring economic efficiency with incomplete price information: With an application to European commercial banks*, Eur. J. Oper. Res. **134** (2001) 43–58.

[24] C.C. Lee, T.H. Huang, *Cost efficiency and technological gap in Western European banks: a stochastic metafrontier analysis*, Int. Rev. Econ. Finance **48** (2017) 161–178.

[25] F. Lei, H. Li, *Duality and efficiency computations in the cost efficiency model with price uncertainty*, Comput. Oper. Res. **40** (2013) 594–602.

[26] L. Liang, F. Yang, W.D. Cook, J. Zhu, *DEA models for supply chain efficiency evaluation*, Ann. Oper. Res. **145** (2006) 35–49.

[27] L. Liang, W.D. Cook, J. Zhu, *DEA models for two-stage processes: game approach and efficiency decomposition*, Nav. Res. Log. **55** (2008) 643–653.

[28] F.H. Lotfi, A. Amirteimoori, Z. Moghaddas, M. Vaez-ghasemi, *Cost efficiency measurement with price uncertainty: a data envelopment analysis*, Math. Sci. **14** (2020) 387–396.

[29] S. Lozano, *Scale and cost efficiency analysis of networks of processes*, Expert Syst. with Applications **38** (2011) 6612–6617.

[30] R. Martinez-Moreno, J.J. Lopez-Espin, J. Aparicio, J.T. Pastor, *Application of genetic algorithms to determine closest targets in data envelopment analysis. In: Distributed Computing and Artificial Intelligence*, Springer, Berlin (2013) 111–119.

[31] A. Moradi S. Saati, M. Navabakhsh, *Genetic algorithms for optimizing two-stage DEA by considering unequal intermediate weights*, OPSEARCH, Springer, Oper. Res. Soc. India **60** (2023) 1202–1217.

[32] R. Mousavizadeh, M. Navabakhsh, A. Hafezalkotob, *Cost-efficiency measurement for two-stage DEA network using game approach: an application to electrical network in Iran*, Indian Academy of Sciences **45** (2020) 267.

[33] M.R. Mozafari, S. Ostovan, W.P. Fernandes, *A hybrid genetic algorithm-ratio DEA approach for assessing sustainable efficiency in two-echelon supply chains*, Sustainability **12** (2020) 8075.

[34] P.C. Pendharkar, *A hybrid genetic algorithm and DEA approach for multi-criteria fixed cost allocation*, Soft Comput. **22** (2018) 7315–7324.

[35] F.S. Piran, D.P. Lacerda, A.S. Camanho, M.C. Silva, *Internal benchmarking to assess the cost efficiency of a broiler production system combining data envelopment analysis and throughput accounting*, Int. J. Prod. Econ. **238** (2021) 108173.

[36] J. Pourmahmoud, S. h. Arami, *Evaluating cost efficiency of decision-making units in an uncertain environment*, JMM. **3** (2024) 565–582.

[37] S. Sarkar, *Detection of cost leaders using non-central principal component analysis and data envelopment analysis*, Benchmark. Int. J. **26** (2019) 548–566.

[38] C. Schaffnit, D. Rosen, J.C. Paradi, *Best practice analysis of bank branches: an application of DEA in a large Canadian bank*, Eur. J. Oper. Res. **98** (1997) 269–289.

[39] J.K. Sengupta, *Efficiency measurement in stochastic input-output systems*, Int. J. Syst. Sci. **13** (1982) 273–287.

[40] S. Seyedboveir, S. Kordrostami, B. Daneshian, A. Amirteimoori, *Cost Efficiency Measurement in Data Envelopment Analysis with Dynamic Network Structures: A Relational Model*, APJOR, **05** (2017) 1750023.

[41] R. Shakouri, M. Salahi, S. Kordrostami, *Stochastic p-robust approach to two-stage network DEA model, Quant*. Financ. Econ. **3** (2019) 315–346.

[42] R. Shakouri, M. Salahi, *Stochastic P-Robust Approach to a Centralized Two-Stage DEA System with Resource Waste*, Adv. Oper. Res. **2022** (2022) 1133882.

[43] L.V. Snyder, M.S. Daskin, *Stochastic p-robust location problems*, IIE Trans. **38** (2006) 971–985.

[44] K. Soleimani-Chamkhorami, S. Ghobadi, *Cost-efficiency under inter-temporal dependence*, Ann. Oper. Res. **302** (2021) 289–312.

[45] R.G. Thompson, P.S. Dharmapala, D.B. Humphrey, W.M. Taylor, R.M. Thrall, *Computing DEA/AR efficiency and profit ratio measures with an illustrative bank application*, Ann. Oper. Res. **68** (1996) 303–327.

[46] R.G. Thompson, F.D. Singleton, R.M. Thrall, B.A. Smith, *Comparative site evaluations for locating a high energy physics lab in Texas*, Interfaces **16** (1986) 35–49.

[47] M. Toloo, *A cost efficiency approach for strategic vendor selection problem under certain input prices assumption*, Measurement **85** (2016) 175–183.

[48] P. Wanke, C. Barros, *Two-stage DEA: an application to major Brazilian banks*, Expert Syst. Appl. **41** (2014) 2337–2344.